Mol Cell Toxicol (2016) 12:139-148
DOI 10.1007/s13273-016-0018-x

ORIGINAL PAPER

Systematic identification of novel biomarker signatures associated with acquired erlotinib resistance in cancer cells

Young Seok Lee[1], Jin Ki Kim[1], Tae Hwan Park[2], Young Rae Kim[1], Ho Sung Myeong[1], Kang Kwon[3], Young Tae Ro[1], Yun Hee Noh[1] & Sung Young Kim[1]

Received: 10 February 2016 / Accepted: 16 March 2016

Abstract Acquired erlotinib resistance (AER) during cancer treatment remains a major clinical challenge that results in the recurrence and metastasis of cancers. Therefore, we sought to identify differentially expressed genes (DEGs) by performing a meta-analysis of AER-related microarray datasets and discover biomarkers by conducting a systemic *in-silico* analysis. Using the RankProd algorithm, we identified 775 DEGs (536 up-regulated and 239 down-regulated). Functional enrichment analyses of the total DEGs suggested that "cell adhesion" and "cytokine-cytokine receptor interactions" may be closely associated with AER process. Some DEGs shared target sites of the potential microRNA including miR-21, miR-200b/c, miR-429 and miR-9. Target sites of FOXJ1, NFAT, FOXO4, and JUN were also significantly enriched. From the protein-protein interaction network, we clustered four functional modules by *p*-value and node density and found hub genes with many interacting neighbors. Finally, we identified seven candidate hub DEGs (*TIMP3*, *SPARC*, *ITGA1*, *CCNA1*, *SOX2*, *KRT14*, and *PTPRZ1*) for AER development.

Keywords Meta-analysis, Microarray, Differentially expressed genes (DEGs), Acquired drug resistance, Erlotinib

Acquired drug resistance is a major challenge for molecular targeted cancer therapies, as it results in chemotherapeutic failure (e.g., cancer recurrence or metastasis) in most patients who initially respond to anti-cancer drugs[1,2].

Erlotinib, an oral low-molecular weight quinazoline derivative, is a first-generation EGFR-tyrosine kinase inhibitor (TKI) that was approved by the US Food and Drug Administration (FDA) in 2004 as a second-line treatment for advanced non-small cell lung cancer (NSCLC) patients who harbor EGFR-activating mutations (e.g., an in-frame deletion in exon 19 or a point mutation in exon 21)[3]. As an antagonist of the tyrosine kinase activity of EGFR, erlotinib selectively and reversibly binds to the ATP binding site of the intracellular tyrosine kinase domain and prevents auto-phosphorylation of the domain, thereby inhibiting subsequent downstream signaling pathways such as the Ras/Raf/MEK/extracellular signal-regulated kinase (ERK), phosphatidylinositol-3-kinase (PI3K)/AKT/mammalian target of rapamycin (mTOR), or Janus kinase (JAK)/signal transducer and activator of transcription (STAT) pathways[4].

However, as observed with other chemotherapeutic drugs, most patients with EGFR-mutant cancers develop acquired resistance to erlotinib shortly after an initial positive response. Several possible mechanisms underlying acquired gefitinib resistance (AGR) in solid cancers have been suggested: (1) variation in the target oncogene, such as alternative expression of tyrosine kinase isoforms, secondary mutations in the tyrosine kinase domain (e.g., T790 mutation, altered trafficking of EGFR, ERBB3 activation, or HER2 amplification);

Electronic supplementary material The online version of this article (doi: 10.1007/s13273-016-0018-x) contains supplementary material, which is available to authorized users.

[1]Department of Biochemistry, School of Medicine, Konkuk University, Seoul, Korea
[2]Department of Plastic and Reconstructive Surgery, College of Medicine, Yonsei University, Seoul, Korea
[3]School of Korean Medicine, Pusan National University, Busan, Korea
Correspondence and requests for materials should be addressed to S. Y. Kim (✉palelamp@kku.ac.kr)

Table 1. Characteristics of individual studies selected from GEO of NCBI for meta-analysis.

Dataset	Sample		Drug	Cancer cell	Platform
	Erl-S	Erl-R			
GSE62061	12	12	Erlotinib	Head and neck cancer cell (Cal-27, SSC-25, FaDu, SQ20B)	Illumina HumanHT-12 V4.0 expression beadchip
GSE49135	3	3	Erlotinib	Head and neck cancer cell (HN5)	Illumina HumanHT-12 V4.0 expression beadchip
GSE38310	3	6	Erlotinib	Lung cancer (HCC827, ER3, T15-2)	Illumina HumanHT-12 V3.0 expression beadchip

GEO gene expression omnibus, GSE gene expression series, Erl-S erlotinib-sensitive, Erl-R erlotinib-resistant

(2) bypass signaling pathway against inhibition of oncogene addiction, such as compensatory activation of downstream signaling pathways and redundant activation of other survival pathways (e.g., MET amplification, activation of MAPK1 and NFκB signaling, AXL kinase expression, and PTEN loss); and (3) histologic transformation (e.g., epithelial-to-mesenchymal transformation (EMT) and small-cell transformation)[4-9].

Despite large-scale clinical efforts to treat cancer patients who eventually develop acquired erlotinib resistance (AER), the mechanisms and genetic factors responsible for AER have not yet been discovered. Cancer research must include insightful investigations to elucidate AER's complex etiology; the recently developed high-throughput microarray technology allows a detailed observation of gene expression under a variety of conditions[10,11]. In three previous microarray studies on AER cancer cell lines, numerous differentially expressed genes (DEGs) were identified as candidate AER biomarkers[12,13]. However, the lists of candidate genes were largely inconsistent between studies due to small sample sizes, low sample qualities, or differences in the laboratory protocols and platforms used. In order to minimize the uncertainty resulting from these variables, we identified DEGs that were consistently identified in all pair-wise samples by performing a meta-analysis of multiple microarray datasets. We also conducted a systemic analysis of the identified DEGs in a protein-protein interaction (PPI) network to recognize their topological positions at the protein level. To our knowledge, this is the first meta-analysis of multiple gene expression profiles from AER-related microarray datasets.

Selection of microarray datasets for meta-analysis related to AER

We extracted 38 GEO samples (GSMs) from three microarray datasets, which met our criteria for meta-analysis. All three GEO series (GSEs) were microarray expression profiles of only the cancer cell lines that acquire erlotinib-resistance by the step-wise increasing treatment of erlotinib. The microarray results of three GSEs were achieved by using two cancer cell lines such as head and neck cancer cells (GSE62061 and 49135) and lung cancer cells (GSE 38310), on "Illumina" gene chip as GEO platforms (GPLs) (Table 1).

Identification of up- and down-regulated DEGs by meta-analysis

From meta-analysis of microarray datasets based on rank product algorithm, we identified total 775 DEGs (536 up- and 239 down-regulated) across above-mentioned three microarray datasets under the significance threshold of $P < 0.05$, which was depended on the estimated percentage of false-positives. While 416 "gain" DEGs were uniquely identified only in the meta-analysis, 6904 "lost" DEGs were identified in any individual analysis but not in the meta-analysis (Figure 1A). The 20 most significantly up- or down-regulated DEGs, with $P < 1.0E-10$, are shown in Table 2.

Among the up-regulated DEGs, genes with the largest mean $\log_2$FC were *BMPER* (BMP binding endothelial regulation), followed by *DFNB59* (deafness, autosomal recessive 59) and *PSG5* (pregnancy specific beta-1-glycoprotein 5). The down-regulated DEGs with the largest mean $\log_2$FC were determined by the descending order as follows: *TCN1* (transcobalamin I), *AKR1B10* (aldo-keto reductase family 1, member B10), and *PTPRZ1* (protein tyrosine phosphatase, receptor-type, Z polypeptide 1).

In addition, a subset of top 25 up- and down-regulated DEGs across the three microarray datasets was visualized by heat maps that exhibit differential expression of individual datasets (Figure 2).

Functional and pathway enrichment analysis of the total DEGs

The most enriched GO hierarchy-biological process terms were listed in the following descending order: "cell adhesion", "immune response", and "regulation

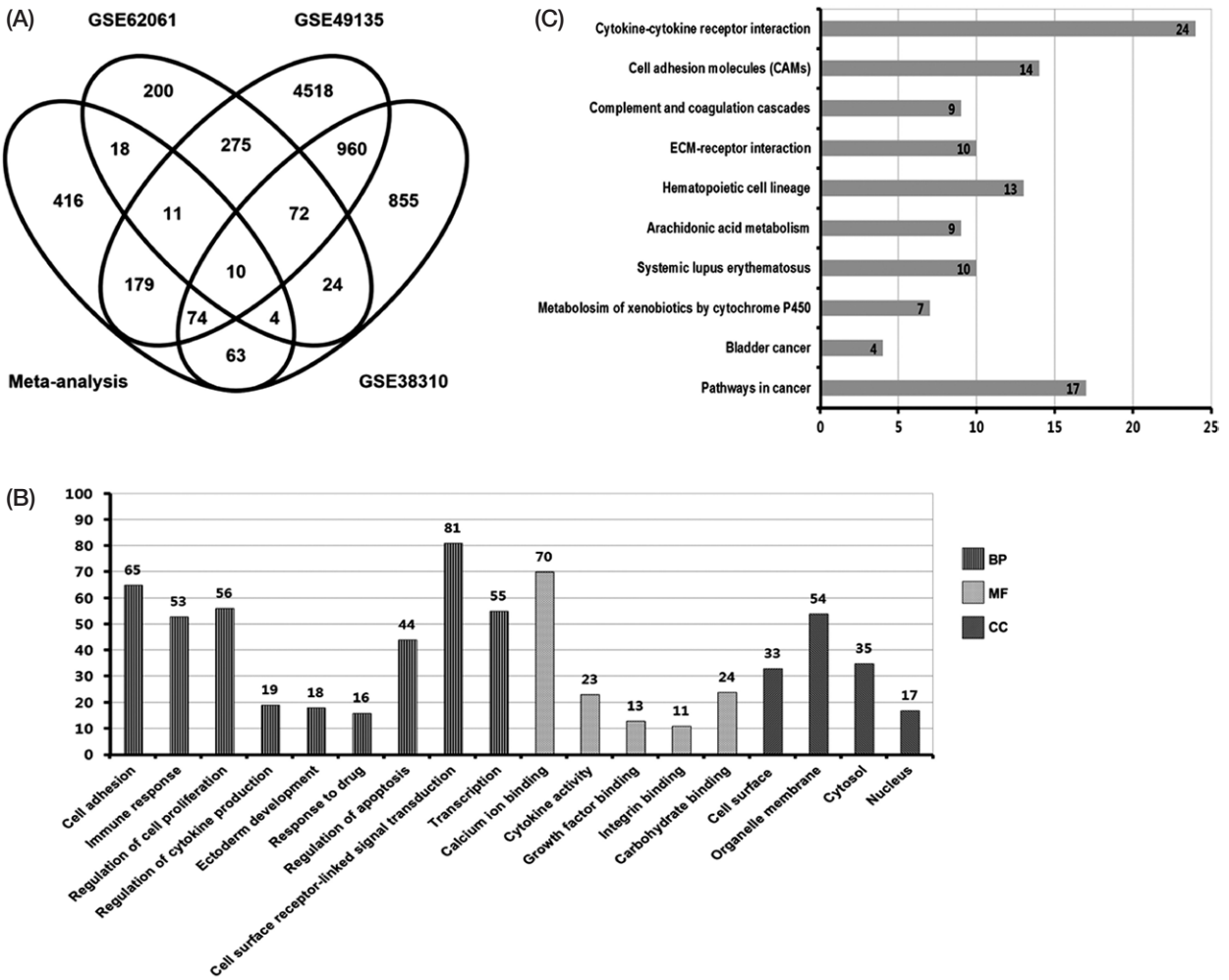

Figure 1. Differential gene expression profiles of this meta-analysis. (A) Venn diagram showing the distribution of DEGs identified from the meta-analysis of 3 microarray datasets and the individual analyses of each dataset. GO function (B) and KEGG pathway (C) terms of total DEGs were functionally enriched by DAVID online resources under a significance threshold of p-value < 0.05. BP, Biological process; MF, Molecular function; CC, Cellular component.

of cell proliferation" (Figure 1B). The most overrepresented GO terms in molecular function and cellular component were "calcium ion binding" and "plasma membrane", respectively.

The most enriched KEGG pathway terms were as follows (in order): "cytokine-cytokine receptor interaction", "cell adhesion molecules (CAMs)", and "complement and coagulation cascades" (Figure 1C).

Gene regulation network analysis of the top 20 up- and down-regulated DEGs

To identify the network regulating gene expression of the top 20 up- and down-regulated DEGs that might directly affect AER, we examined potential microRNAs and transcription factors that target the DEGs depending on their upstream DNA sequence (Table 3). The target gene sites of the following microRNAs were significantly enriched in the DEGs: miR-21, miR-200b/c, miR-429 and miR-9. The target gene sites of transcription factors such as FOXJ1, NFAT, FOXO4, and JUN were also enriched by some DEGs.

PPI network analysis of the top 20 up- and down-regulated DEGs

To understand biological interaction of the identified DEGs at protein level, we constructed a PPI network (include 208 nodes and 984 edges) of proteins encoded by the top 20 up- and down-regulated DEGs from all three huge databases (Figure 3).

The PPI network was further divided into four sub-

Table 2. The top 20 most significantly up- or down-regulated genes in the DEGs identified by meta-analysis.

Entrez ID	Symbol	$-Log_2FC$	*P*-value	Gene name
Up-regulated genes				
168667	*BMPER*	6.95270		BMP binding endothelial regulator
494513	*DFNB59*	6.58465		Deafness, autosomal recessive 59
5673	*PSG5*	6.48217		Pregnancy specific beta-1-glycoprotein 5
23461	*ABCA5*	6.12518		ATP-binding cassette, sub-family A (ABC1), member 5
3672	*ITGA1*	6.04478		Integrin, alpha 1
3730	*KAL1*	5.66942		Kallmann syndrome 1 sequence
8995	*TNFSF18*	5.29755		Tumor necrosis factor (ligand) superfamily, member 18
9586	*CREB5*	5.29616		cAMP responsive element binding protein 5
1544	*CYP1A2*	5.28874		Cytochrome P450, family 1, subfamily A, polypeptide 2
84913	*ATOH8*	5.27516		Atonal homolog 8 (Drosophila)
283392	*LOC283392*	5.25663	< 1.0E-05	Hypothetical LOC283392
3823	*KLRC3*	5.19938		Killer cell lectin-like receptor subfamily C, member 3
3425	*IDUA*	5.13094		Iduronidase, alpha-L-
113263	*GLCCI1*	5.08415		Glucocorticoid induced transcript 1
6678	*SPARC*	5.01613		Secreted protein, acidic, cysteine-rich (osteonectin)
81704	*DOCK8*	5.00027		Dedicator of cytokinesis 8
5675	*PSG6*	4.69312		Pregnancy specific beta-1-glycoprotein 6
55790	*CSGALNACT1*	4.56498		Chondroitin sulfate N-acetylgalactosaminyltransferase 1
7078	*TIMP3*	4.46576		TIMP metallopeptidase inhibitor 3
83660	*TLN2*	4.36085		Talin 2
Down-regulated genes				
6947	*TCN1*	−6.45155		Transcobalamin I (vitamin B12 binding protein, R binder family)
57016	*AKR1B10*	−5.93070		Aldo-keto reductase family 1, member B10 (aldose reductase)
5803	*PTPRZ1*	−3.61576		Protein tyrosine phosphatase, receptor-type, Z polypeptide 1
9635	*CLCA2*	−3.29541		Chloride channel accessory 2
4680	*CEACAM6*	−3.28471	< 1.0E-05	Carcinoembryonic antigen-related cell adhesion molecule 6
338382	*RAB7B*	−3.13399		RAB7B, member RAS oncogene family
1272	*CNTN1*	−3.02910		Contactin 1
8900	*CCNA1*	−2.28873		Cyclin A1
771	*CA12*	−2.77362	0.00071	Carbonic anhydrase XII
793	*CALB1*	−4.46685	0.00076	Calbindin 1, 28 kDa
3861	*KRT14*	−1.52601	0.00083	Keratin 14
221393	*GPR115*	−5.26884	0.00090	G protein-coupled receptor 115
6657	*SOX2*	−2.02380	0.00100	SRY (sex determining region Y)-box 2
3854	*KRT6B*	−1.41676	0.00111	Keratin 6B
3848	*KRT1*	−5.23841	0.00117	Keratin 1
56341	*PRMT8*	−5.74658	0.00125	Protein arginine methyltransferase 8
3860	*KRT13*	−2.82175	0.00125	Keratin 13
1244	*ABCC2*	−2.81959	0.00130	ATP-binding cassette, sub-family C (CFTR/MRP), member 2
3850	*KRT3*	−5.14066	0.00133	Keratin 3
6511	*SLC1A6*	−4.31755	0.00136	Solute carrier family 1 (aspartate/glutamate transporter), member 6

$Log_2FC = log_2$ (class1/class2), FC fold change, class1 erlotinib-sensitive, class2 erlotinib-resistant

networks depending on the *p*-value and node density on CluterONE Cytoscape plugin (Figure 4) and proteins of each sub-network were functionally enriched by GO function (biological function) and KEGG pathway (data not shown). The hub DEGs with degree (defined as number of interactions) greater than 15 in the PPI network were shown in Table 4.

Discussion

Erlotinib is one of the most extensively used molecular-targeted cancer therapies for various solid cancers, but its clinical effectiveness is severely reduced by acquired resistance during treatment. The development of AER is multifactorial, and the precise mechanisms and genetic factors responsible are not yet fully understood. In this respect, we identified 775 genes that were consistently differentially expressed, as determined by a meta-analysis of three independent microarray datasets. Among these, we identified 184 genes as novel DEGs in our meta-analysis that were not identified in the individual studies, suggesting a high possibility of these genes to be novel biomarkers for

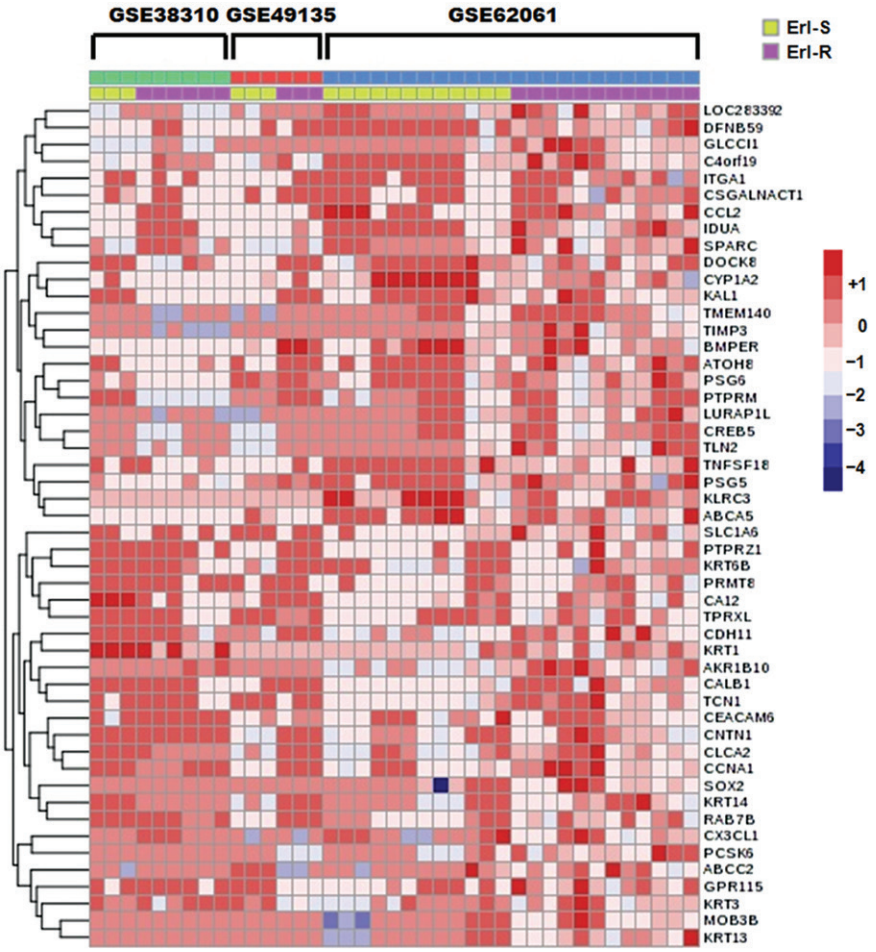

Figure 2. Heat-map representation of expression profiles for the top 25 up- and down-regulated DEGs across three microarray datasets. In heat-map representation, clustering of selected genes was performed by hierarchical clustering algorithm using average linkage method and euclidean distance measure.

Table 3. The potential regulatory elements that target the top 20 up- and down-regulated DEGs.

Regulatory elements	Target sequence	Genes	*P*-Value
microRNA			
miR-21	ATAAGCT	*SOX2, TIMP3, GLCCI1*	1.45E-04
miR-200b/c, miR-429	CAGTATT	*SOX2, CREB5, PTPRZ1, TLN2*	7.32E-04
miR-9	TAGCTTT	*SOX2, TIMP3, CREB5*	1.15E-03
Transcription factor			
FOXJ1	NNNTGTTTATNTR	*SOX2, CREB5, ATOH8*	7.44E-04
NFAT	NWGGAAANWN	*CREB5, KRT14, PTPRZ1*	1.62E-03
FOXO4	TTGTTT	*SOX2, CREB5, ITGA1, CALB1, ATOH8, KRT14, TLN2*	1.83E-03
JUN	TGANTCA	*CREB5, SPARC, KRT14, TLN2, KRT13*	2.77E-03

AER development.

Primary analysis of gene expression patterns, which depend on *p*-values and the $\log_2$FC of DEGs, revealed that several of the top 50 up- or down-regulated genes are found in a variety of tumors and cancers. In particular, most of the top 20 up- and down-regulated genes are involved in the development of various cancers such as breast cancer, colorectal cancer, gastric cancer, ovarian cancer, and lung cancer. In addition, some DEGs were identified in drug-resistant cancer cells:

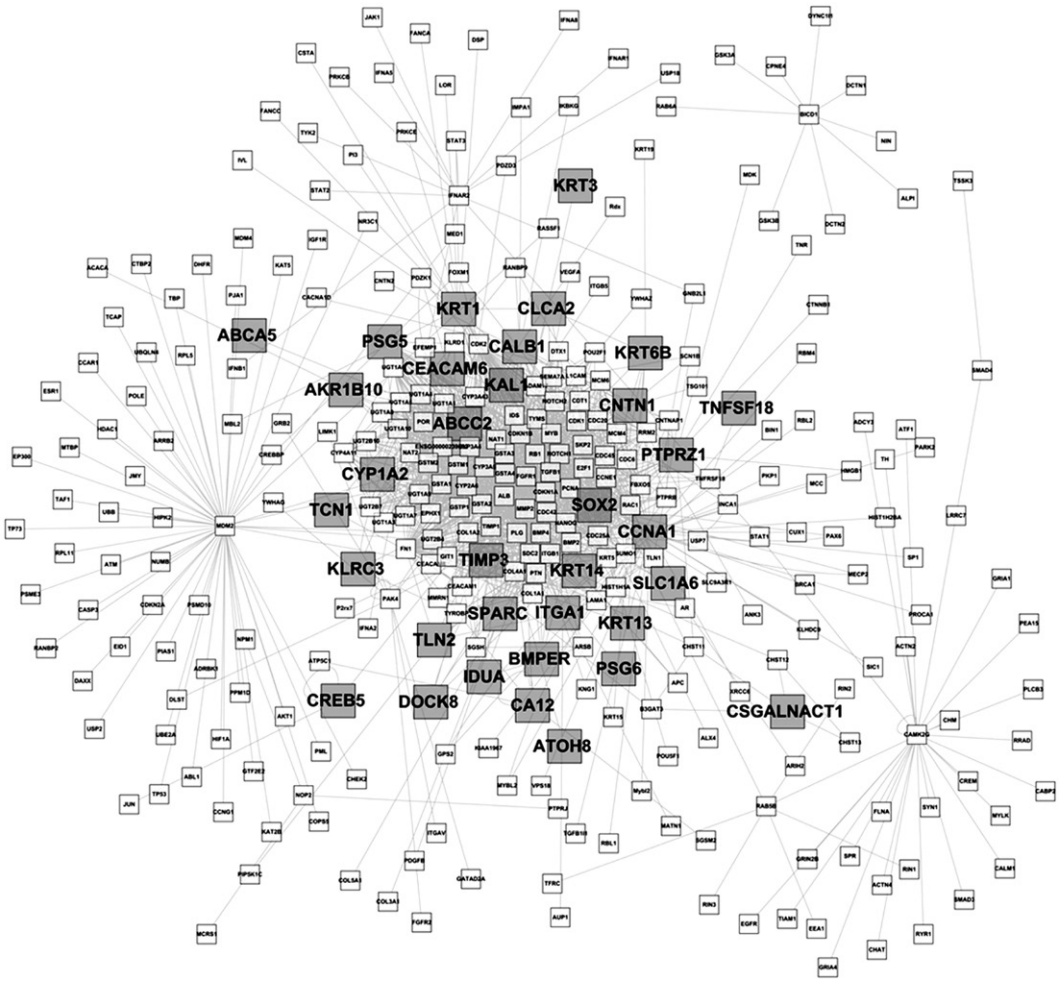

Figure 3. Protein-protein interaction network of the top 20 up- and down-regulated DEGs. We constructed the PPI network of proteins encoded by the top 20 up- and down-regulated DEGs under three huge databases, such as BioGRID, STRING, and HPRD online sources. The color of node signifies proteins that are encoded by the DEG. The color of nodes - Grey; DEG-encoding proteins.

ABCA5 (resistance to multidrug), *ITGA1* (oxaliplatin), *CYP1A2* (trastuzumab), *SPARC* (5-FU), *TIMP3* (tamoxifen), *AKR1B10* (doxorubicin), *CEACAM6* (tamoxifen), *CCNA1* (arsenic trioxide), *SOX2* (gefitinib), *KRT1* (cisplatin) and *ABCC2* (multidrug).

For the full DEG list, functional enrichment analysis by the GO hierarchy and KEGG pathway showed that a large proportion of the genes was concerned with the following cellular processes of oncogenesis and tumor development: epithelial development (ectoderm development, GO0007398), apoptosis (regulation of cell proliferation, GO0042127; regulation of apoptosis, GO0042981), immune response (immune response, GO0006955; regulation of cytokine production, GO0001817; cytokine-cytokine receptor interaction, hsa04060), transcription (transcription, GO0006350), drug metabolism (response to drug, GO0042493; metabolism of xenobiotics by cytochrome P450, hsa00980), signal transduction (cell surface receptor linked signal transduction, GO0007166; Toll-like receptor signaling pathway, hsa04620), and EMT (cell adhesion, GO0007155; cell adhesion molecules, hsa04514; ECM-receptor interaction, hsa04512).

Gene regulation network analysis of the top 20 up- and down-regulated DEGs showed that some genes significantly shared target sites of potential transcription factors and microRNA, which may participate in cellular process of AER. For example, it was reported that miR-200b/c and miR21 confer acquired resistance to erlotinib by modulating activation of EMT and migration in NSCLC cell[14-16]. In the case of transcription factor, Takeuchi K *et al*. revealed that the activation of c-Jun N-terminal kinase (JNK) is critical for apoptosis induced by EGFR-TKIs in NSCLC cell[17].

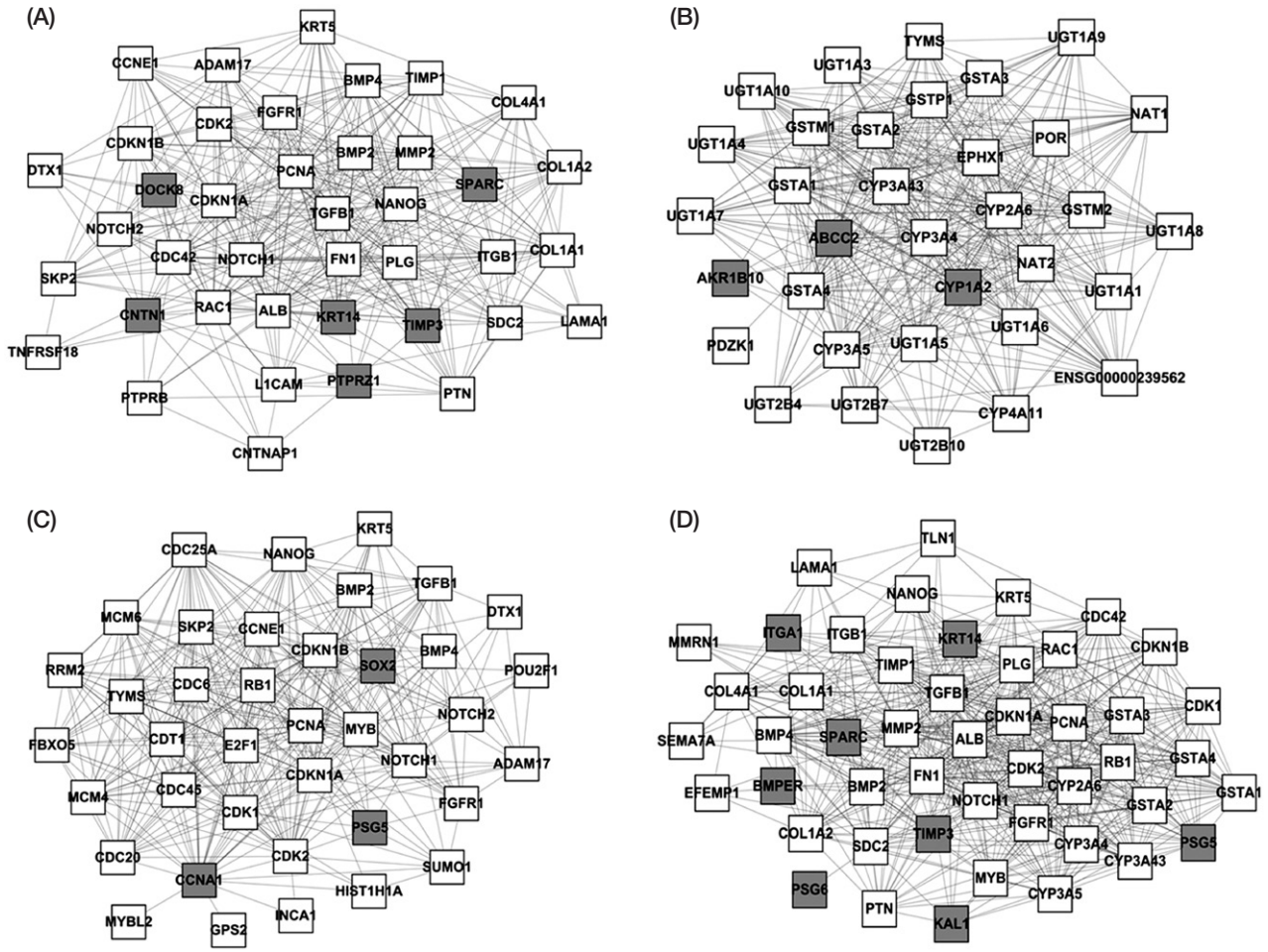

Figure 4. Functional modules in the protein-protein interaction network. From PPI networks of proteins encoded by the top 20 up- and down-regulated DEGs, we clustered four functional modules, using ClusterONE Cytoscape plugin: Module 1 (A), Module 2 (B), Module 3 (C), and Module 4 (D).

To evaluate fully the biological significance of AER-related DEGs in complex diseases such as cancer, comprehensive knowledge of the topological position of each DEG within the PPI network is just as valuable as the fold change and *p*-values of the individual genes. In addition, hub nodes are known to have a large effect on network organization within organic systems and play important functions in the maintenance of the system. Accordingly, we identified functional modules and hub proteins in a PPI network of the top 20 up- and down-regulated genes, by combining several large databases of the known network. The four functional modules, which comprised the most significant DEG-encoding proteins and other known proteins in the network, were significantly enriched by GO and KEGG pathway terms associated with the typical biological processes of AER, including anti-apoptosis and indefinite cell proliferation (modules 1, 3, and 4), cancer signaling pathway (modules 1, 3, and 4), or deregulated drug metabolism (modules 2 and 4). Finally, we identified 7 candidate hub genes that constitute function-

Table 4. The hub genes that degree greater than 15 in PPI network.

Gene	Regulation type	Degree
CCNA1	Down	59
CYP1A2	Up	33
TIMP3	Up	28
SOX2	Down	26
KRT14	Down	24
SPARC	Up	24
CALB1	Down	23
PTPRZ1	Down	21
KRT1	Down	19
ITGA1	Up	17

al modules in the PPI network of the reliable DEGs, including three up-regulated genes (*ITGA1*, *SPARC*, and *TIMP3*) and four down-regulated genes (*CCNA1*, *KRT14*, *SOX2*, and *PTPRZ1*).

Integrins are αβ heterodimeric cell adhesion receptors for cell-cell and cell-extracellular matrix interactions, and play a direct role in the progression, angiogenesis, and metastasis of cancers. It was reported that upregulation of the αvβ3, α5β1, αvβ5, and α6β4 integrins was found in many types of cancer. ITGA1 (the alpha 1 subunit of integrin receptors) is also known to activate the Ras/Raf/MEK/ERK signaling pathway during cellular proliferation and migration of mammary carcinoma cells and is significantly upregulated in colorectal cancer[18,19].

SPARC (secreted protein acidic and rich in cysteine) is a multifaceted protein engaged in different biological processes including bone mineralization, cell proliferation and migration, morphogenesis, differentiation, and angiogenesis[20]. Overexpression of SPARC promotes bone metastasis and EMT in highly metastatic tumors including glioblastomas, melanoma, breast cancer, and prostate cancer and is closely associated with poor patient survival.

The TIMP (tissue inhibitor of metalloproteinase) family consists of four homologous proteins (TIMP1-4) that are involved in multifunctional processes such as cellular proliferation, pro-MMP activation, migration, invasion, and apoptosis. TIMP3 has been reported as a tumor suppressor that inhibits inflammation, tumor growth, and angiogenesis in some cancers[21].

Interestingly, CCNA1 (cyclin A1), which exhibited the highest degree in the PPI network, was identified as a significantly down-regulated gene in the DEG list. CCNA1 is an important cell cycle regulator and belongs to the highly conserved cyclin family. It is characterized by a dramatic periodicity in protein abundance throughout the cell cycle and functions as an activating subunit in the enzymatic complex of cyclin-dependent kinases. Several studies recently demonstrated that expression of CCNA1 was down-regulated and that it might be an important tumor suppressor in head and neck, cervical, and nasopharyngeal cancers[22,23].

SOX2, a member of the sex-determining region Y-box family, is a master transcription factor that is essential for embryonic development, including stem cell fate determination and differentiation. Recent studies revealed that SOX2 was down-regulated in gastric carcinomas, and its exogenous expression suppressed cellular proliferation by inhibiting cell-cycle arrest and apoptosis in gastric epithelial cell lines[24].

Keratins are the intermediate filament (IF)-forming proteins present in epithelial cells, and form the largest protein family, comprising 54 gene products in humans. They are expressed in epithelial cell type or differentiation state-specific manner, play many crucial roles in epithelial cells, and serve as prognostic marker able to determine the origin of epithelial tumors[25].

PTPRZ1 is a member of the receptor-type protein tyrosine phosphatase family. Receptor-type protein tyrosine phosphatases (RPTPs) are involved in regulating cell signaling pathways in cooperation with tyrosine kinases to control cell proliferation, differentiation, adhesion, and migration, which are closely relevant to the pathogenesis of human diseases such as diabetes, autoimmune diseases, and cancer[26].

In conclusion, we identified 775 DEGs that are likely to be involved in the cellular process of AER, by performing a meta-analysis of three microarray datasets for cancers with AER. Also, the systemic PPI network analysis of the significant DEGs provided insight into some possible processes underlying AER. This topological information of multiple gene expression profiles may help in understanding the complex nature of AER development and provide a novel gene expression signature that facilitates future chemotherapy research.

Materials & Methods

Selection of microarray datasets qualified for meta-analysis

According to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines published in 2009, we thoroughly investigated the adequacy of microarray datasets retrieved on Gene Expression Omnibus (GEO) database of National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/geo/) and ArrayExpress database of European Molecular Biology Laboratory- European Bioinformatics Institute (EMBL-EBI) (http://www.ebi.ac.uk/arrayexpress/). We selected the microarray datasets in the meta-analysis if it contained: (1) gene expression profiles of cancer cell lines that acquire drug-resistant or drug-derivative characteristic by step-wise increasing doses of erlotinib; (2) enough and high-quality microarray datasets suitable for meta-analysis; (3) datasets that were established on definite platforms (e.g., microarray chip of Affymetrix or Illumina).

Meta-analysis of microarray datasets with different platforms

We performed meta-analysis of multiple gene expression profiles across microarray datasets gleaned from different platforms, by means of rank product algorithm (RankProd package in R, http://www.r-project.

org/) implemented in the INMEX online program (http://inmex.ca/INMEX/)[27-29]. Before meta-analysis of the datasets, intensity values for gene expression were processed by log2-transformation and quantile normalization to ensure that their mean and unit variance was zero, and all probe IDs from each dataset were annotated as Entrez IDs for data consistency (limma package in R).

Enrichment analysis by GO hierarchy and KEGG pathway

To interpret biological implications of the DEGs in acquired erlotinib-resistant cancer cells, we carried out functional enrichment analysis of Gene Ontology (GO) hierarchy (biological process, molecular function, and cellular component) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, using the DAVID (Database for Annotation, Visualization, and Integrated Discovery) online program under a significance threshold of $P < 0.05$ (http://david.abcc.ncifcrf.gov/).

Gene regulatory network analysis

For prediction of a gene regulation network containing the identified DEGs, we performed enrichment analysis of potential transcription factors and microRNAs based on a comparison of upstream DNA sequences with database of gene annotation sets retrieved from MSigDB (http://www.broadinstitute.org/gsea/msigdb/index.jsp/).

The hypergeometric algorithm and Benjamini-Hochberg adjustment were used for statistical method and multiple test correction of the network analysis, respectively[30].

Protein-protein interaction network analysis

To construct protein-protein interaction (PPI) network of proteins encoded by the DEGs (top 20 up- and down-regulated), we mapped the gene list into the immense database of already-known networks and screened significant protein-protein interactions under Biological General Repository for Interaction Datasets (BioGRID) (http://thebiogrid.org/), STRING (http://string.embl.de), Human Protein Reference Database (HPRD) (http://hprd.org/) online sources[31,32]. The PPI network was screened on a genome-wide scale using Cytoscape software. In the network, distinct modules were further identified by using the Cytoscape plugin, ClusterONE (http://apps.cytoscape.org/apps/clusterone)[33]. The over-represented biological terms of genes that form distinct modules were enriched by DAVID program.

In this study, nodes with degrees larger than 15 were considered as hub proteins depending on degree of interaction.

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

Conflict of Interest The authors declare that they have no conflict of interest.

Author's contributions Y.S.L and S.Y.K conceived and designed this study. Y.S.L and S.Y.K participated in all of the research processes and drafted the manuscript. J.K.K and T.H.P performed a statistical analysis to process the data. Y.R.K and H.S.M assisted in most of the in-silico analyses. K.K, Y.T.R, and Y.H.N reviewed and revised the manuscript to complete this study. All authors received the data, discussed the results, and approved the final manuscript.

References

1. Szakacs, G., Paterson, J. K., Ludwig, J. A., Booth-Genthe, C. & Gottesman, M. M. Targeting multidrug resistance in cancer. *Nat Rev Drug Discov* **5**:219-234 (2006).
2. Chen, K. G. & Sikic, B. I. Molecular pathways: regulation and therapeutic implications of multidrug resistance. *Clin Cancer Res* **18**:1863-1869 (2012).
3. Bareschino, M. A. *et al*. Erlotinib in cancer treatment. *Ann Oncol* **18 Suppl 6**, vi35-41 (2007).
4. Xu, Y., Liu, H., Chen, J. & Zhou, Q. Acquired resistance of lung adenocarcinoma to EGFR-tyrosine kinase inhibitors gefitinib and erlotinib. *Cancer Biol Ther* **9**: 572-582 (2010).
5. Hrustanovic, G., Lee, B. J. & Bivona, T. G. Mechanisms of resistance to EGFR targeted therapies. *Cancer Biol Ther* **14**:304-314 (2013).
6. Beloueche-Babari, M. *et al*. Acquired resistance to EGFR tyrosine kinase inhibitors alters the metabolism of human head and neck squamous carcinoma cells and xenograft tumours. *Br J Cancer* **112 Suppl**, 1206-1214 (2015).
7. Nguyen, K. S., Kobayashi, S. & Costa, D. B. Acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors in non-small-cell lung cancers dependent on the epidermal growth factor receptor pathway. *Clin Lung Cancer* **10**:281-289 (2009).
8. Nurwidya, F. *et al*. Acquired resistance of non-small cell lung cancer to epidermal growth factor receptor tyrosine kinase inhibitors. *Respir Investig* **52**:82-91 (2014).
9. Lin, Y., Wang, X. & Jin, H. EGFR-TKI resistance in NSCLC patients: mechanisms and strategies. *Am J Cancer Res* **4**:411-435 (2014).
10. Ramasamy, A., Mondry, A., Holmes, C. C. & Altman, D. G. Key issues in conducting a meta-analysis of

gene expression microarray datasets. *PLoS Med* **5**:e184 (2008).
11. Zhao, Y. *et al*. Identification of potential therapeutic target genes, key miRNAs and mechanisms in acute myeloid leukemia based on bioinformatics analysis. *Med Oncol* **32**:572 (2015).
12. Giles, K. M. *et al*. Axl mediates acquired resistance of head and neck cancer cells to the epidermal growth factor receptor inhibitor erlotinib. *Mol Cancer Ther* **12**:2541-2558 (2013).
13. Zhang, Z. *et al*. Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. *Nat Genet* **44**:852-860 (2012).
14. Ahmad, A. *et al*. Inhibition of Hedgehog signaling sensitizes NSCLC cells to standard therapies through modulation of EMT-regulating miRNAs. *J Hematol Oncol* **6**:77 (2013).
15. Cufi, S. *et al*. Silibinin suppresses EMT-driven erlotinib resistance by reversing the high miR-21/low miR-200c signature in vivo. *Sci Rep* **3**:2459 (2013).
16. Bryant, J. L. *et al*. A microRNA gene expression signature predicts response to erlotinib in epithelial cancer cell lines and targets EMT. *Br J Cancer* **106**:148-156 (2012).
17. Takeuchi, K., Shin-ya, T., Nishio, K. & Ito, F. Mitogen-activated protein kinase phosphatase-1 modulated JNK activation is critical for apoptosis induced by inhibitor of epidermal growth factor receptor-tyrosine kinase. *FEBS J* **276**:1255-1265 (2009).
18. Boudjadi, S., Carrier, J. C. & Beaulieu, J. F. Integrin alpha1 subunit is up-regulated in colorectal cancer. *Biomark Res* **1**:16 (2013).
19. Kanda, R. *et al*. Erlotinib resistance in lung cancer cells mediated by integrin beta1/Src/Akt-driven bypass signaling. *Cancer Res* **73**:6243-6253 (2013).
20. Kaleagasioglu, F. & Berger, M. R. SIBLINGs and SPARC families: their emerging roles in pancreatic cancer. *World J Gastroenterol* **20**:14747-14759 (2014).
21. Anania, M. C. *et al*. TIMP3 regulates migration, invasion and in vivo tumorigenicity of thyroid tumor cells. *Oncogene* **30**:3011-3023 (2011).
22. Kitkumthorn, N. *et al*. Cyclin A1 promoter hypermethylation in human papillomavirus-associated cervical cancer. *BMC Cancer* **6**:55 (2006).
23. Rivera, A., Mavila, A., Bayless, K. J., Davis, G. E. & Maxwell, S. A. Cyclin A1 is a p53-induced gene that mediates apoptosis, G2/M arrest, and mitotic catastrophe in renal, ovarian, and lung carcinoma cells. *Cell Mol Life Sci* **63**:1425-1439 (2006).
24. Otsubo, T., Akiyama, Y., Yanagihara, K. & Yuasa, Y. SOX2 is frequently downregulated in gastric cancers and inhibits cell growth through cell-cycle arrest and apoptosis. *Br J Cancer* **98**:824-831 (2008).
25. Karantza, V. Keratins in health and cancer: more than mere epithelial cell markers. *Oncogene* **30**:127-138 (2011).
26. Xu, Y. & Fisher, G. J. Receptor type protein tyrosine phosphatases (RPTPs) - roles in signal transduction and human disease. *J Cell Commun Signal* **6**:125-138 (2012).
27. Xia, J. *et al*. INMEX-a web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Res* **41**:W63-70 (2013).
28. Song, G. G. *et al*. Meta-analysis of differentially expressed genes in primary Sjogren's syndrome by using microarray. *Hum Immunol* **75**:98-104 (2014).
29. Fang, F., Pan, J., Xu, L., Li, G. & Wang, J. Identification of potential transcriptomic markers in developing ankylosing spondylitis: a meta-analysis of gene expression profiles. *Biomed Res Int* **2015**:826316 (2015).
30. Liberzon, A. *et al*. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**:1739-1740 (2011).
31. Quan, Z. *et al*. Protein-protein interaction network and mechanism analysis in ischemic stroke. *Mol Med Rep* **11**:29-36 (2015).
32. Rakshit, H., Rathi, N. & Roy, D. Construction and analysis of the protein-protein interaction networks based on gene expression profiles of Parkinson's disease. *PLoS One* **9**:e103047 (2014).
33. Nepusz, T., Yu, H. & Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods* **9**:471-472 (2012).