Analysis of Gene Expression Profile Triggered by Signal Peptide of Eosinophil Cationic Protein

ECPsp triggers gene expression

Yu-Shu Liu, Chung-Hsaio Chao, Hao-Teng Chang*
Graduate Institute of Molecular Systems Biomedicine, College of Medicine, China Medical University, Taichung, Taiwan
*htchang@mail.cmu.edu.tw

Yong Wang
Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

Margaret Dah-Tsyr Chang
Institute of Molecular and Cell Biology & Department of Medical Science, National Tsing Hua University, Taichung, Taiwan

Tun-Wen Pai
Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung, Taiwan

***Abstract*—The signal peptide of eosinophil cationic protein (ECPsp) is known to play an important role in translocating ECP to extracellular space. However, we previously discovered that ECPsp has a novel function of inhibiting microbial growth and regulating the gene expression of tumor growth factor-alpha (TGF-α) and epidermal growth factor receptor (EGFR) in mammalian cells. In the present study, we first generated a DNA microarray dataset, which showed that ECPsp up-regulated inflammatory molecules including cytokines, chemokines, interferon-induced molecules, and Toll-like receptors. We then generated a function linkage network by integrating the microarray dataset with the KEGG pathway database, and discovered that STAT1, an important factor regulating cytokine expression and release, served as a hub to connect the pathways of cytokine stimulation (TGF-α and EGFR) and inflammatory responses. Furthermore, integrating the ECPsp interactome dataset with the functional linkage network elucidated that STAT1 served as a hub to connect 3 functional clusters, including cell proliferation and survival, protein translational regulation, and inflammatory responses. Our approach involving experimental and computational systems biology provided predicted pathways and potential regulation for further characterization of the novel function of ECPsp under inflammatory conditions.**

Keywords-eosinophil cationic protein, signal peptide, inflammation, functional linkage network, gene regulation

I. INTRODUCTION

Human eosinophil cationic protein (ECP) is an important molecule related to some diseases such as asthma and inflammation. It is secreted by activated eosinophils, and serves as one of the major components of eosinophil granule proteins [1]. Structurally, ECP contains 3 α-helices, 5 β-strands, and 8 loops [2], with a molecular mass of approximately 16–22 kDa due to varying extents of post-translational modification [3, 4]. In addition, ECP, also called ribonuclease 3 (RNase3), belongs to the human RNaseA superfamily, with a very low ribonucleolytic activity but a high cytotoxic activity [5]. The members of the RNaseA superfamily have been defined by their similar sequences and structures, and possess ribonucleolytic activity for the regulation of gene expression at the mRNA level [6]. To date, 13 human protein sequences have been found to be similar to bovine pancreatic RNaseA [7-10]. The sequence of ECP contains 19 arginines and 2 lysines, which leads to a high isoelectric point (pI = 10.8) [11, 12]. ECP also has a great cytotoxic activity, for its highly positive charge facilitates interaction between ECP and negatively-charged molecules such as heparin sulfate and cell membrane lipids [11, 13]. Through interaction with the cell or organism surface, ECP translocates into cells by endocytosis and causes cell damage such as airway inflammation during asthma and intestinal mucosa damage in Crohn's disease [14-16]. ECP has been defined to response against pathogen infection, and to remove the invading microorganisms by the innate immune systems. It also possesses cytotoxicity against parasites, bacteria, viruses, helminth, brugia, and mammalian cells [15, 17-19]. Therefore, ECP has been categorized as an antimicrobial peptide (AMP). The mechanism of ECP-triggered cell damage is suggested by the fact that ECP destabilizes the cell membrane lipids *via* the processes of pore forming, permeability change and membrane leakage [5, 12, 20].

On the other hand, ECP possesses a signal peptide with 27 amino acids cleaved by signal peptidase (SP) in the endoplasmic reticulum (ER). The signal peptide of ECP (ECPsp), like other secretory proteins, contains a short, positively charged amino-terminal region, a central hydrophobic region, and a polar carboxy-terminal region. Previous studies have indicated that ECPsp possesses

This study was supported by National Science Council, Taiwan (NSC-98-2311-B-039-003-MY3, NSC-99-2627-B-039-002, and NSC-99-3112-B-007-002).

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

cytotoxicity to inhibit the growth of lower organisms such as *Escherichia coli* (*E. coli*) and *Pichia pastoris* (*P. pastoris*), but not in mammalian cells, because ECPsp can be cleaved into ECPsp-1-17 and ECPsp-18-27 fragments by signal peptide peptidase (SPP) in mammalian cells [21]. In addition, ECPsp-1-17 enhances the expression of tumor growth factor-alpha (TGF-α) and epidermal growth factor receptor (EGFR) in A431 and HL-60 cell lines [22].

Interestingly, ECPsp has been suggested to be a novel peptide possessing dual biological functions, protein secretion and regulation of gene expression. In this study, we use a systems biology approach to further characterize the biological function of ECPsp. In particular, we investigate the ECPsp-triggered gene expression using DNA microarray and the genome wide interactome of ECPsp. By combining these two datasets, we explore the integrated network from STRING 8.3 and KEGG to provide potential pathways and functional clusters to understand ECPsp-induced gene expression profiling.

II. Methods and Materials

A. Cells and Cell Culture

HL-60 clone 15, a human promyelocytic leukemia cell line (FIRDI, Taiwan), was maintained in IMDM (HyClone, USA) containing 20% heat-inactivated fetal bovine serum (FBS, Invitrogen, USA). Beas-2B, a human bronchial epithelial cell line (ATCC CRL-9609), was cultured in RPMI-1640 (Invitrogen, USA) supplemented with 10% FBS. All cells were cultured at 37°C in an incubator under 5% CO_2 and 95% air.

B. RNA Extraction and DNA Microarray

Plasmids expressing enhanced green fluorescence protein (eGFP) and ECPsp-eGFP were transfected into Beas-2B cells using TurboFect™ (Thermo, USA) according to the manufacturer's instructions. After 48 hr, total RNA of Beas-2B cells was isolated using TRIzol reagent (Invitrogen, USA). Two micrograms of total RNA and 0.5 μg oligo-dT primer (EPICENTRE® Biotechnologies) were mixed in DEPC water and heated at 65°C for 2 min. The reverse transcription mixture contained Moloney Murine Leukemia Virus (M-MLV) reaction buffer (EPICENTRE® Biotechnologies), 2 mM dNTP (PROTECH, Taiwan), 10 mM DTT (EPICENTRE® Biotechnologies), 2 units of M-MLV reverse transcriptase (EPICENTRE® Biotechnologies) and DEPC water. cDNA was produced at 37°C for 1 h and subsequently 5 min at 85°C to inactivate the reverse transcriptase activity.

The Human Whole Genome OneArray® v5 (HOA, Phalanx Biotech Group, Taiwan) has in total 30,275 DNA oligonucleotide probes, and each probe has 60 nucleotides in the sense strand. Of these probes, 29,187 probes correspond to the annotated genes in the RefSeq v38 and Ensembl v56 databases. Furthermore, 1,088 control probes are also included for monitoring the sample quality and the hybridization process. The signal intensity of each spot was loaded into the Rosetta Resolver System® (Rosetta Biosoftware) to process the data analysis. The error model of the Rosetta Resolver System® removed both systematic and random errors from the data. We filtered out spots for which the flag was less than 0. Spots that passed the criteria were normalized by the 50% media scaling normalization method. The technical repeat data were tested by Pearson correlation coefficient calculation to check the reproducibility ($R > 0.95$). Normalized spot intensities were transformed to gene expression $\log_2$ ratios between the control and treatment groups. Spots with $p < 0.05$ and $\log_2 \geq 1$ were tested for pathway analysis.

C. Pull Down Assay and Liquid Chromatography Mass Spectrometry

HL-60 cells were dissolved in PBS containing protease inhibitor cocktail (Amresco, USA) and homogenized by sonication. After centrifugation at 14,000 rpm at 4°C, 2 mg of the cell lysate was transferred to a 1.5-ml microcentrifuge tube separately, and 100 μl streptavidin beads resuspended in cold-PBS were added to clean the bead-binding proteins at 4°C on an inverted shaker. After 4 h, the beads were removed and the pre-cleaned lysate was mixed with 50 μl fresh streptavidin beads that had been incubated with or without 30 μg biotinylated ECPsp-1-17 peptide for 4 h. After further incubation at 4°C overnight, the supernatant was discarded and the beads were washed using wash buffer (30 mM Tris-HCl pH = 7.4, 150 mM NaCl and 0.1% NP-40) 8 times and subsequently using cold PBS 5 times to remove the unbound proteins. The precipitated proteins were eluted by 50 μl of elution buffer (8 M Guanidine-HCl pH = 1.5) at room temperature and neutralized by adding 5 μl of 1 M Tris-HCl (pH = 7.5) immediately. After desalting, the proteins were transferred to a new tube and digested by trypsin for subsequent LC-MS/MS analyses.

D. Pathway Enrichment Analysis

According to the originally generated gene expression profiles, the proteins upregulated more than 2-fold ($\log_2 \geq 1$) were selected and assembled as a gene expression subdataset. We constructed the protein functional linkage network from the STRING 8.3 database [23]. The proteins in the gene expression subdataset were then put into the STRING network to further analyze the functional relationships among those molecules, including TGF-α and EGFR, both induced by ECPsp. Finally, the ECPsp interactome identified by pull down analyses was defined as the ECPsp interactome signature and combined with the network.

To further elucidate the molecular pathways related to the ECPsp interactome signature, we conducted pathway enrichment analysis. This revealed significant enrichment of several pathways in the signature. The pathway database used here was MolSigDB (https://www.broadinstitute.org/gsea/msigdb/), a collection of curated gene sets from online pathways including KEGG, publications in PubMed, and knowledge of domain experts.

We enumerated and ranked all annotated gene sets in MolSigDB for which there was a significant overlap of genes with our ECPsp interactome signature. The basic assumption was that function-related gene sets tend to share more common genes than expected by chance. We developed two scores, a *p*-value score and an enrichment score, to assess the significance of gene overlap for a pair of gene sets. These two scores were further combined as a single gene overlap score to define pathway enrichment.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

To determine whether the gene overlap was statistically significant, we fixed the total number of genes in humans (N), the number of genes in the gene set from MolSigDB (N_1), and the number of genes in our expression signature (N_2). We then treated the number of overlap genes as a random variable X. Under the null hypothesis that the two gene sets were independent, X followed a hypergeometric distribution:

$$P(X=i)=\frac{\binom{N_1}{i}\binom{N-N_1}{N_2-i}}{\binom{N}{N_2}}$$

From here, we were then able to calculate a p-value score, which was defined as the probability that the gene overlap would assume a value greater than or equal to the observed value, m, by chance:

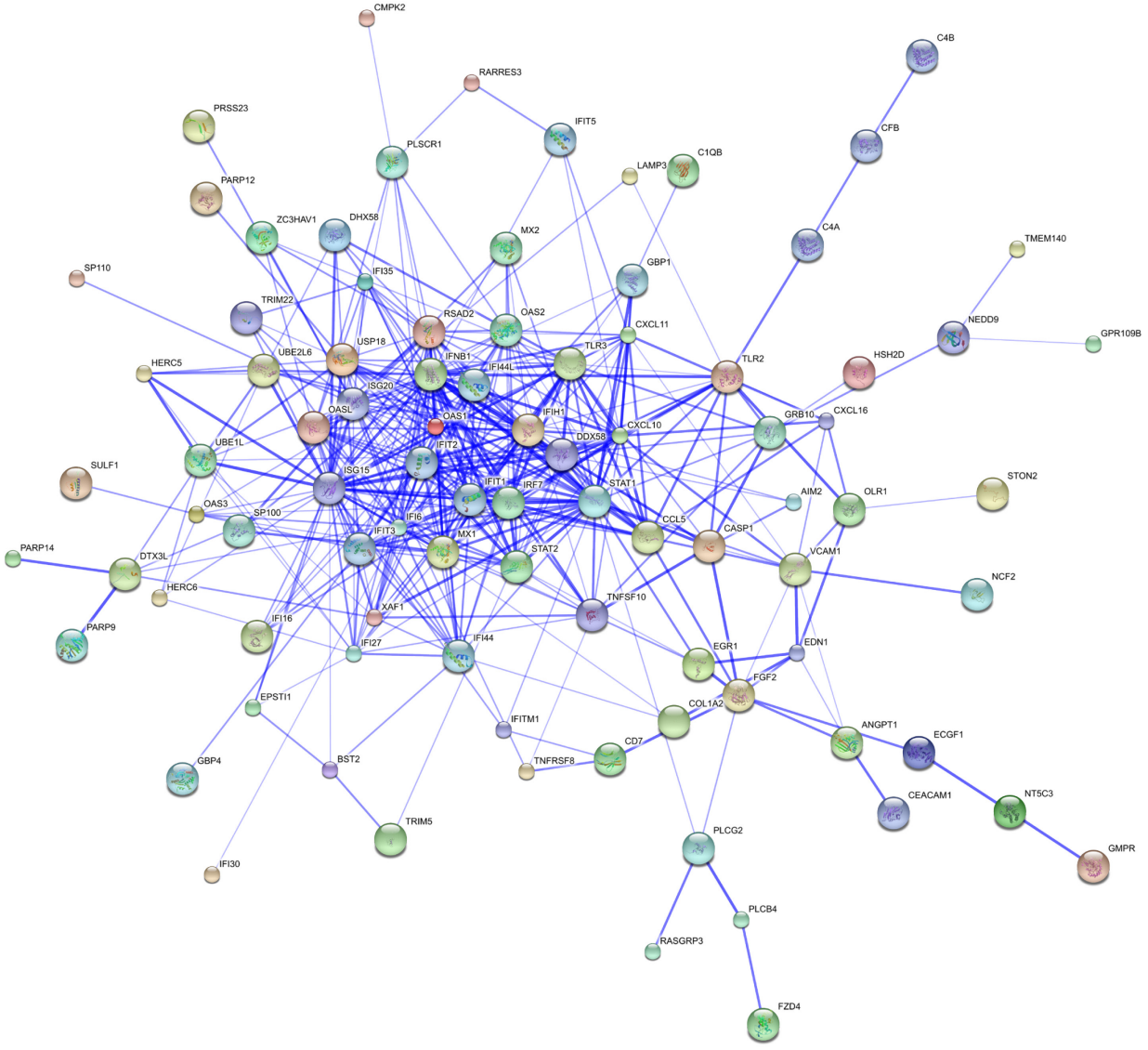

Figure 1. Functional linkage networks of ECPsp-induced genes. Here, the nodes in the networks are the ECPsp-induced genes that have a greater than 2-fold expression as compared with the control. The edges in the network are obtained from STRING by applying the confidence threshold.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

$$P(X \geq m) = 1 - \sum_{i=0}^{m-1} P(X = i)$$

The gene overlap was statistically significant if the *p*-value score was smaller than a chosen cutoff. As many pathways were tested for our single gene expression signature, multiple testing corrections were applied by calculating the false discovery rate (FDR).

The above procedure gave us all gene sets for which the gene overlap with our expression signature was statistically significant. To further quantify the extent of the overlap, we also calculated an enrichment score (*FE*), defined as the ratio of the observed overlap versus the expected overlap by chance:

$$FE = \frac{Nm}{N_1 N_2}$$

A score larger than 1 indicated that there is more gene overlap than expected by chance.

We selected all gene sets for which the both *p*-value and enrichment scores were significant (*p*-value < 0.05, enrichment score > 2). These ranked pathways or gene sets assisted us to characterize the function of the ECPsp interactome signature, and further the function of ECPsp.

III. RESULTS

A. *Inflammatory molecules are upregulated by ECPsp, and are central in the functional linkage network*

To understand the gene expression profile triggered by ECPsp, the transcriptome was analyzed using DNA microarray conducted by Phalanx Biotech in Taiwan. Each total RNA derived from the cells transfected with control plasmids or plasmids containing ECPsp was duplicated, and to confirm and balance the experimental variations, two technical repeats were carried out. According to the gene expression data, 83 genes were upregulated by more than 2-fold as compared with the control. The 83 genes were casted into the functional linkage network obtained from the STRING 8.3 database and the confidence sub-network was generated by choosing a strict cutoff (Fig. 1). In this network, a "focused" area is clearly displayed, in which nodes are related to inflammatory responses including cytokines, chemokines, and interferon-induced/related molecules, and inflammatory receptors such as Toll-like receptors. The focused area revealed that ECPsp might induce gene expression of inflammatory molecules and trigger inflammatory pathways. This discovery is crucial in studying the biological function of ECPsp. ECPsp has been recognized as a conventional signal peptide for the secretion of mature ECP *via* the endoplasmic reticulum (ER) and Golgi apparatus. In 2007, we discovered a novel function of ECPsp, which induced TGF-α and EGFR overexpression, using the differential display technique and real-time PCR [22]. Another study also indicated that ECPsp possesses cytotoxicity towards bacteria and yeasts, but not mammalian cells [21]. These findings demonstrated that ECPsp not only executes its conventional function, protein secretion, but also regulates gene expression in a cell.

B. *Combination of ECPsp-induced genes and the TGF-α/EGFR pathway*

Our previous study indicated that ECPsp induced TGF-α and EGFR expression at the mRNA and protein levels [22]. TGF-α is one of the ligands stimulating EGFR phosphorylation and triggering the following signaling pathways. KEGG also revealed that the STAT pathway was triggered by TGF-α *via*

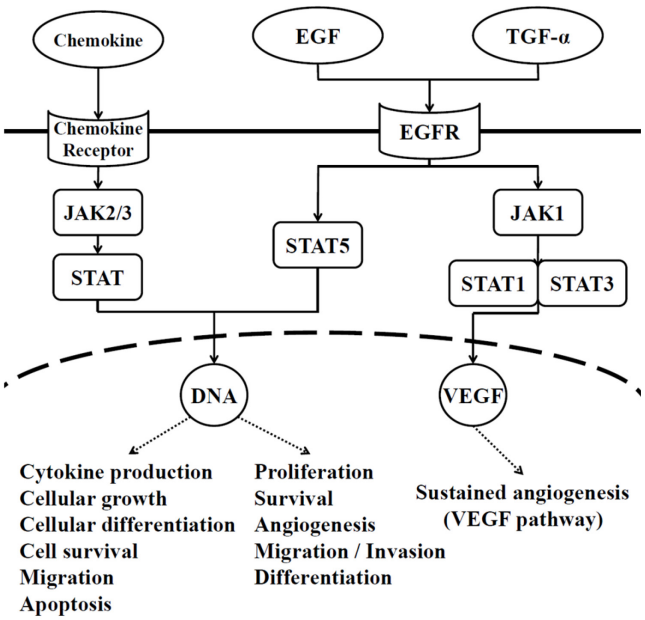

Figure 2. Schematic graph of the STAT pathway modified from KEGG.

EGFR, resulting in DNA repair, cell proliferation, and cell survival. In addition, STATs were regulated by chemokines and chemokine receptors, and served as an upstream molecule driving the gene expression of cytokines (Fig. 2). Thus, incorporation of TGF-α, EGFR, and the ECPsp-induced genes may enable discovery of new regulatory networks initiated by ECPsp. In order to reduce the complexity, the focused area in Fig. 1 was extracted and combined with TGF-α and EGFR. After incorporation, the network showed that the TGF-α /EGFR pathway connected with inflammatory molecules *via* STAT1, an important factor regulating cytokine expression and release (Fig. 3) [24]. Therefore, ECPsp may initiate alternative inflammatory responses in cells simultaneously when ECP is expressed at the same time.

C. *STAT1 serves as a hub in connecting TGF-α pathways and cytokine/chemokine production.*

To obtain a high confidence network, we further studied the molecules that are induced more than 4 folds by ECPsp, and the underlying functional linkage network is shown in Fig. 4. The functional linkage network combined with the ECPsp interactome dataset, TGF-α, EGFR and STAT1/2. According to Fig. 4, STAT1 still serves as a hub to connect three functional clusters. One cluster contains interferon-inducing

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

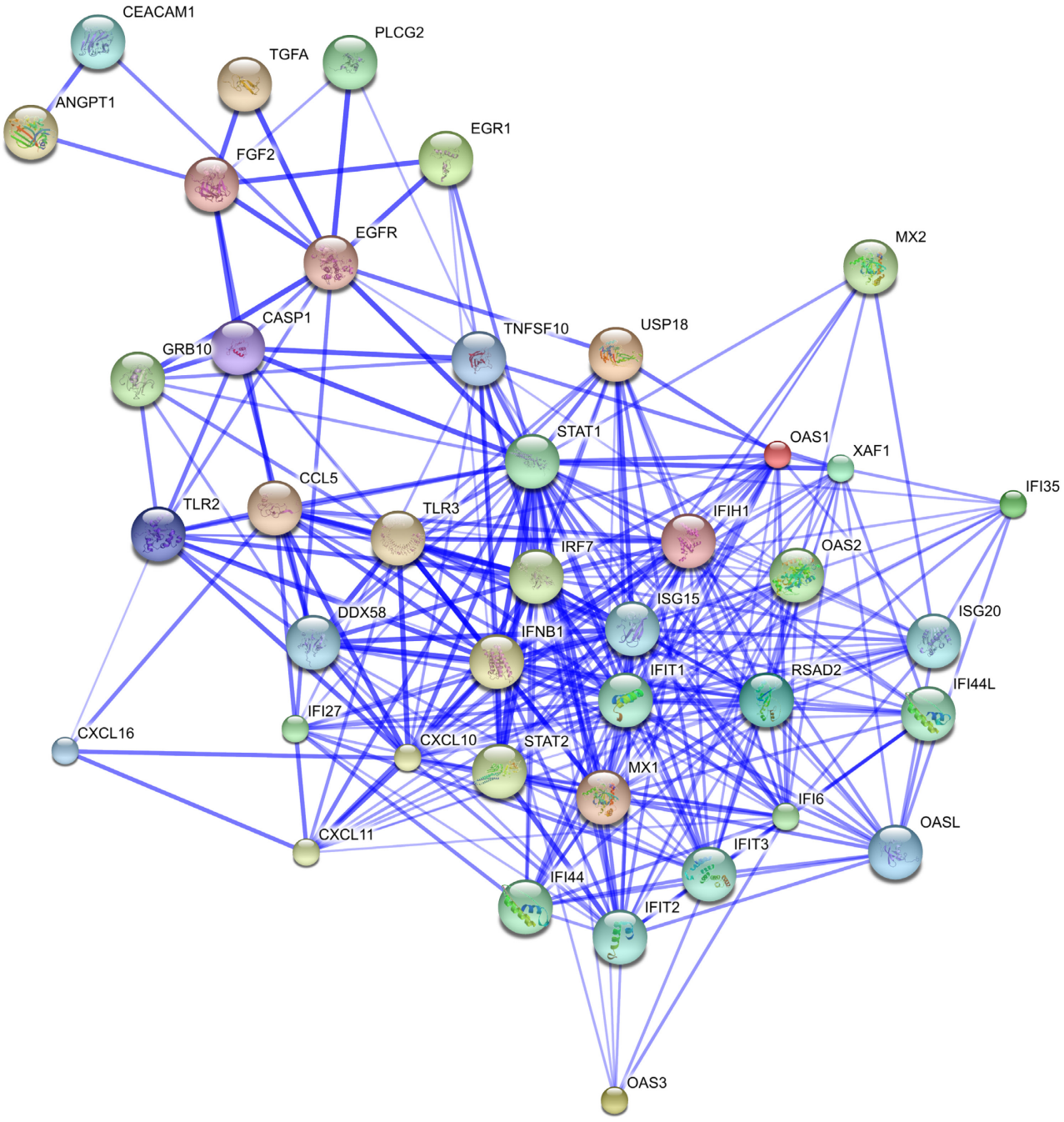

Figure 3. Confidence network of combination of the focused network, and TGF-α and EGFR partners. To connect the gene expression profile with the experimental discoveries, the molecules of the focused network in Fig. 1 were selected and incorporated with TGF-α, EGFR, and related partners.

factors, cytokine, and chemokines including IFI44L, IFIH1, CCL5, and CXCL11. Another cluster is of ribosomal proteins, which may control the protein translational regulation. The other cluster is of molecules related to cell growth and proliferation, such as TGF-α, EGFR, ANGPT1, and CBL. This network provides important clues for researches to further investigate novel functions of ECPsp.

Figure 4. Genes with a greater than 4-fold expression were selected (dark blue circles), and STAT1/2 (light blue circles), TGF-α, and EGFR (red circles), and the proteins identified by pull down assay (gold yellow circles) were incorporated.

TABLE I. LIST OF THE ENRICHED PATHWAYS BY OVERLAPPING WITH THE ECPSP INTERACTOME SIGNATURE

Description of the pathway	FE Score	FDR Score	NumOverlap / Num gene	*p*-value
REACTOME_POST_CHAPERONIN_TUBULIN_FOLDING_PATHWAY	193.79	0.002	3 / 16	4.27E-07
REACTOME_FORMATION_OF_TUBULIN_FOLDING_INTERMEDIATES_BY_CCT_TRIC	167.36	0.002	3 / 16	6.77E-07
REACTOME_PREFOLDIN_MEDIATED_TRANSFER_OF_SUBSTRATE_TO_CCT_TRIC	131.50	0.003	3 / 16	1.44E-06
REACTOME_CHAPERONIN_MEDIATED_PROTEIN_FOLDING	73.64	0.012	3 / 16	8.50E-06
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION	62.41	0.015	3 / 16	1.40E-05
RADMACHER_AML_PROGNOSIS	46.61	0.031	3 / 16	3.38E-05
KEGG_GAP_JUNCTION	40.91	0.039	3 / 16	4.99E-05

On the other hand, pathway enrichment analysis was performed using the ECPsp interactome signature. Table I shows that the genes in these 7 pathways are significantly enriched in the ECPsp interactome dataset, which indicates that these 7 pathways are involved in the gene expression triggered by ECPsp.

IV. DISCUSSION AND CONCLUSION

Under asthmatic conditions, ECP would be released by activated eosinophils, and then damaged bronchial epithelial cells. At the same time, ECPsp are cleaved by human SPP, and the cleaved ECPsp triggers overexpression of TGF-α and EGFR. This study provided the information that ECPsp

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

indeed regulates gene expression in cells, and the upregulated genes are classified as inflammation-related partners. It is relevant that under inflammatory conditions, eosinophils may release cytokines to stimulate cell growth and survival, and chemokines attract macrophages to clean the pathogens or the damaged cell debris after ECP is released to bronchia. In the present study, we discovered that ECPsp upregulated the inflammatory molecules, and this regulation might be driven by the TGF-α and EGFR pathway *via* STAT1. This finding provides clues for researchers to study the roles of ECP and ECPsp under inflammatory conditions. Therefore, we hypothesize that ECPsp may serve as the secondary effect to induce inflammatory gene expression and trigger following immune processes, including the recruitment of macrophages, cleaning damaged cells and repairing the bronchial epithelial cells, based on the integration of experimental findings, DNA microarray, and proteomics analyses.

REFERENCES

[1] I. Olsson and P. Venge, "Cationic proteins of human granulocytes. II. Separation of the cationic proteins of the granules of leukemic myeloid cells," *Blood,* vol. 44, pp. 235-46, Aug 1974.

[2] G. Mallorqui-Fernandez*, et al.*, "Three-dimensional crystal structure of human eosinophil cationic protein (RNase 3) at 1.75 A resolution," *J Mol Biol,* vol. 300, pp. 1297-307, Jul 28 2000.

[3] C. G. Peterson*, et al.*, "Purification and characterization of eosinophil cationic protein from normal human eosinophils," *Eur J Haematol,* vol. 40, pp. 415-23, May 1988.

[4] H. F. Rosenberg, "The eosinophil ribonucleases," *Cell Mol Life Sci,* vol. 54, pp. 795-803, Aug 1998.

[5] H. F. Rosenberg, "Recombinant human eosinophil cationic protein. Ribonuclease activity is not essential for cytotoxicity," *J Biol Chem,* vol. 270, pp. 7876-81, Apr 7 1995.

[6] A. A. Makarov and O. N. Ilinskaya, "Cytotoxic ribonucleases: molecular weapons and their targets," *FEBS Lett,* vol. 540, pp. 15-20, Apr 10 2003.

[7] H. F. Rosenberg, "RNase A ribonucleases and host defense: an evolving story," *J Leukoc Biol,* vol. 83, pp. 1079-87, May 2008.

[8] G. Z. Cheng*, et al.*, "Human ribonuclease 9, a member of ribonuclease A superfamily, specifically expressed in epididymis, is a novel sperm-binding protein," *Asian J Androl,* vol. 11, pp. 240-51, Mar 2009.

[9] H. F. Rosenberg*, et al.*, "Human eosinophil cationic protein. Molecular cloning of a cytotoxin and helminthotoxin with ribonuclease activity," *J Exp Med,* vol. 170, pp. 163-76, Jul 1 1989.

[10] S. Cho*, et al.*, "The ribonuclease A superfamily of mammals and birds: identifying new members and tracing evolutionary histories," *Genomics,* vol. 85, pp. 208-20, Feb 2005.

[11] E. Carreras*, et al.*, "Surface-exposed amino acids of eosinophil cationic protein play a critical role in the inhibition of mammalian cell proliferation," *Mol Cell Biochem,* vol. 272, pp. 1-7, Apr 2005.

[12] E. Carreras*, et al.*, "Both aromatic and cationic residues contribute to the membrane-lytic and bactericidal activity of eosinophil cationic protein," *Biochemistry,* vol. 42, pp. 6636-44, Jun 10 2003.

[13] T. C. Fan*, et al.*, "A heparan sulfate-facilitated and raft-dependent macropinocytosis of eosinophil cationic protein," *Traffic,* vol. 8, pp. 1778-95, Dec 2007.

[14] G. J. Gleich, "Mechanisms of eosinophil-associated inflammation," *J Allergy Clin Immunol,* vol. 105, pp. 651-63, Apr 2000.

[15] S. Winterkamp*, et al.*, "Secretion and tissue content of eosinophil cationic protein in Crohn's disease," *J Clin Gastroenterol,* vol. 30, pp. 170-5, Mar 2000.

[16] S. Pegorier*, et al.*, "Eosinophil-derived cationic proteins activate the synthesis of remodeling factors by airway epithelial cells," *J Immunol,* vol. 177, pp. 4861-9, Oct 1 2006.

[17] J. B. Domachowske*, et al.*, "Eosinophil cationic protein/RNase 3 is another RNase A-family ribonuclease with direct antiviral activity," *Nucleic Acids Res,* vol. 26, pp. 3358-63, Jul 15 1998.

[18] R. I. Lehrer*, et al.*, "Antibacterial properties of eosinophil major basic protein and eosinophil cationic protein," *J Immunol,* vol. 142, pp. 4428-34, Jun 15 1989.

[19] H. A. Molina*, et al.*, "Toxic effects produced or mediated by human eosinophil granule components on Trypanosoma cruzi," *Am J Trop Med Hyg,* vol. 38, pp. 327-34, Mar 1988.

[20] J. D. Young*, et al.*, "Mechanism of membrane damage mediated by human eosinophil cationic protein," *Nature,* vol. 321, pp. 613-6, Jun 5-11 1986.

[21] C. M. Wu and M. D. Chang, "Signal peptide of eosinophil cationic protein is toxic to cells lacking signal peptide peptidase," *Biochem Biophys Res Commun,* vol. 322, pp. 585-92, Sep 17 2004.

[22] H. T. Chang*, et al.*, "Signal peptide of eosinophil cationic protein upregulates transforming growth factor-alpha expression in human cells," *J Cell Biochem,* vol. 100, pp. 1266-75, Apr 1 2007.

[23] C. von Mering*, et al.*, "STRING 7--recent developments in the integration and prediction of protein interactions," *Nucleic Acids Res,* vol. 35, pp. D358-62, Jan 2007.

[24] L. Canaff*, et al.*, "The proinflammatory cytokine, interleukin-6, up-regulates calcium-sensing receptor gene transcription via Stat1/3 and Sp1/3," *J Biol Chem,* vol. 283, pp. 13586-600, May 16 2008.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE