BIOCHIP JOURNAL, Vol. 1, No. 3, 179-187, September 2007

Identification of Disease Specific Protein Interactions between the Gastric Cancer Causing Pathogen, *H. pylori*, and Human Hosts using Protein Network Modeling and Gene Chip Analysis

Wan Kyu Kim[1,2], Kyuwan Kim[2], Eunjung Lee[2], Edward M. Marcotte[1], Hyong-Ha Kim[3] & Jung-Keun Suh[2,4]

[1]Center for Systems and Synthetic Biology, Department of Chemistry and Biochemistry, Institute for Cellular and Molecular Biology, 2500 Speedway, University of Texas, Austin, Texas 78712, USA.
[2]LG Life Science R & D Park/Bio Institute 104-1, Moonji-Dong, Yuseong-Gu, Daejeon 305-380, Korea
[3]Division of Metrology for Quality Life, Korea Research Institute of Standards & Science, P.O. Box 102, Yuseong, Daejeon 305-600, Korea
[4]Korean German Institute of Technology Biotech Research Center, KGIT Mokdong Center, 905-27, Mok-dong, Yangcheon-gu, Seoul 158-050, Korea
Correspondence and requests for materials should be addressed to J.-K. Suh (suhjung@kgit.re.kr)

Accepted 18 July 2007

Abstract

Proteins are traditionally known as the building blocks or functional units that make up the cellular physiology of living organisms. In the post-genomic view of a protein, however, it can function as an element within a protein network and its role can then be evaluated by protein-protein interaction analysis. The role of proteins within such a network can be defined by their cellular function within the functional modules of the network as well as their individual activity. In this study, we used a protein-protein interaction modeling system to identify the functional modules and proteins involved in the pathogenic interaction between the gastric pathogen, *H. pylori*, and humans. We analyzed 1,590 *H. pylori* proteins against 10,257 human entries expressed in human gastric tissues and identified 4,349 potential protein-protein interactions between 159 *H. pylori* proteins and 108 human proteins. We then investigated the association of gastric cancer with the 108 human proteins found to have an interaction with the *H. pylori* proteins using a GeneChip database that we generated. Among the 108 human proteins, 93 (86%) were shown to be associated with gastric cancer, 91 of which were up-regulated and 2 of which were down-regulated by at least 4 fold in gastric cancer tissues. Additionally, 32 of the proteins were found to be gastric cancer-specific, whereas the remaining proteins were found to be associated with several other forms of cancer. Taken together, these results suggest that protein network modeling in conjunction with GeneChip technology can be a useful tool for the analysis of the complex relationship between human pathogens and their hosts.

Keywords: *H. Pylori*, Protein-protein interaction, Protein network modeling, Host-pathogen interaction, Gene Chip microarray

Introduction

Pathogenic bacteria maintain their presence within their host by interfering with the immune system or by stimulating anti-inflammatory responses[1]. To understand and influence these molecular based interactions between pathogens and their hosts, it is imperative that we identify the subset of key components and regulatory interactions that are responsible for significant functional changes within the host[2]. Computational prediction can help explain these complex systems by integrating many components into a simple network model. The human pathogen, *H. pylori*, is a good model for this type of work because there is genomic sequence data[3-4] and protein-protein interaction data[5] available for this pathogen, and functional genomics studies of this organisms relationship to gastric cancer already exist[6].

H. pylori is a predominately extracellular, flagellated motile gram-negative bacteria. New information regarding the pathogenicity and virulence factors of *H. pylori* that has emerged as a result of studying this organism using various biochemical techniques indicates that infection by *H. pylori* involves a complex interaction of both bacterial and host factors[7]. *H. pylori* generate many effects by interacting with its host. Together, bacterial and host factors interact as dangerous liaisons that trigger a cascade of events resulting in host-pathogen disequilibrium[8]. This interaction seems to be very important in the initial stage of gastric diseases, therefore, characterizing these

molecular mechanisms at various stages of infection is essential to prevent the progress of the disease. Although it is difficult to evaluate this complex pathogen-host interaction experimentally due to a lack of technologies and tools available for data analysis, computational modeling may help to overcome those difficulties.

Computational modeling allows mass prediction of the protein network at low cost. Additionally, computational modeling can help isolate critical proteins from numerous candidate proteins without experimental validation, and this information can then be used by biologists to assign priorities to the proteins or domains to be tested, thereby allowing the construction of a large-scale protein interaction network. Further, this information can also be used to predict the functions of unknown proteins[9].

Therefore, we used a computational modeling system to identify functional modules and proteins involved in the pathogenic interaction between the gastric pathogen, *H. pylori*, and humans. We then analyzed the data using the GenChip database, which we generated, to investigate the association of the proteins identified using the model with gastric cancer. Using this approach, 108 human proteins were shown to be interaction partners with proteins from *H. pylori*, and 93 of these proteins (86%) were identified as gastric cancer related proteins. Additionally, 91 of proteins associated with gastric cancer were up-regulated and 2 were down-regulated over 4 fold in gastric cancer tissues. Further, 32 of these proteins were found to be gastric cancer-specific, however the remaining proteins were found to be associated with several other forms of cancer as well. Taken together, these results suggest the usefulness of protein network modeling in conjunction with GeneChip technology as a tool for analyzing the complex relationship between human pathogens and their hosts.

Table 1. Functional classification of the predicted interactions between human and *H. pylori* proteins

H. pylori proteins		Human proteins	
Functional catalog	No.	Functional catalog	No.
Energy production	1	Enzymes	54
Cell cycle control	6	Secreted protein	1
Metabolism	10	Transcription	16
Translation	4	Channel/Transporter	4
Transcription	4	Membrane protein	7
Replication, recombinationand repair	6	Receptor/G protein	11
Cell wall/membrane biogenesis genes	7	Translation	4
Cell motility genes	9	Proteosome component	3
Protein turnover, chaperones genes	16	Uncharacterized	8
Signal transduction mechanisms genes	2		
Intracellular trafficking and secretion	5		
Defense mechanisms genes	2		
General function prediction only genes	31		
Function unknown genes	56		

Results

Analysis of Protein Interaction between Human and *H. pylori* Proteins

We used protein network modeling to predict the protein-protein interaction of the VacA protein (HP0887) because its partners had already been determined experimentally in humans[10-11]. It has been reported that VacA is capable of binding to RACK1 (a receptor for activated C kinase), a GTPase related protein, and VIP54 (an intermediate filament interacting protein).We predicted that VacA could interact with kinases, ras GTPases and GTP-binding proteins, and filament interacting proteins, as well as proteins involved in the cell cycle, apoptosis, and tumor suppression. The results obtained from modeling the VacA protein suggested that the modeling system we used could predict protein interactions when the interactions between the proteins were functionally related (Figure 3). Therefore, we used our system to predict the protein-protein interaction between human and *H. pylori* proteins. We analyzed 1,590 ORF sequences from *H. pylori* against 67,105 human protein entries in the IPI (EMBL-EBI) database and then rank-ordered the protein domains of *H. pylori* for further biological verification

The first prediction produced a set of interactions between human and *H. pylori* proteins that was too large to deal with experimentally, therefore we used our OncologyDB (GeneChip DB) to try to determine which genes were expressed in normal stomach tissues. We collected GeneChip data for normal stomach tissue from over 30 cases, and then extracted genes that had 50% CALL, which gave 10,257 gene entries. We then used this information to remove the interactions with not expressed proteins and obtained 72,183 interacting pairs. We then selected *H. pylori* protein-domains that had a high prediction score and proteins of public interest from this set for further analysis and

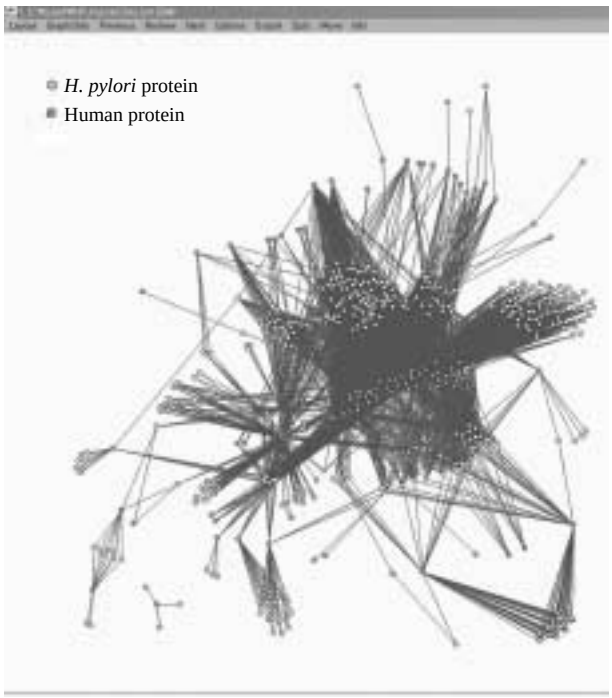

Figure 1. The protein network model generated by the prediction of protein-protein interactions between human (closed circle) and *H. pylori* (open circle) proteins.

found that there was a very dense interaction between 159 *H. pylori* proteins and 108 human proteins (Figure 1). Most of the *H. pylori* proteins identified consisted of predicted ORFs (56 proteins) and ORFs (31 proteins) that were homologous with other genes that had known functions. Interestingly, 16 ORFs related to protein turnover and chaperones were also identified (Table 1), which may imply the importance of bacterial chaperones in pathogenesis against humans. Most of the human proteins identified (54 proteins) were enzymes, including kinases and phosphatases. Additionally, 16 human transcription factors and activators were identified, however only 1 secreted human protein was identified (Table 1).

Modeling Pathogenesis using the Protein Network Models

We then used the protein network information obtained to determine which modules from humans and *H. pylori* are important in pathogenesis. For example, vacuolating cytotoxin (VacA) is believed to be the major virulence factor of *H. pylori*. It is synthesized as a 140-kDa precursor, proteolytically processed, and then secreted by bacterial cells[12-14]. VacA that has been secreted then interacts with the surface of target cells, which leads to internalization and

Table 2. Functional classification of disease specific genes and prioritized genes identified by network modeling of the interactions between human and *H. pylori* proteins in combination with GeneChip analysis using stomach cancer tissues.

93 Human genes		Prioritized genes		
Functional class	No.	Functiona lclass		No.
Enzyme	47	Enzyme	Kinase	2
Secreted protein	1		Phosphatase	2
Transcription factor	14		Protease	1
Channel/ transporter	4		Lipase	1
Membrane protein	4	Secreted protein		1
Receptor/ G protein	8	Transcription factor		6
Translation factor	4	Membrane protein		3
Proteosome component	3	Receptor/G protein		4
Uncharacterized	5			

translocation of the toxin into the host cytosol. Once in the cytosol, the toxin forms intracellular vacuoles that originate from the late endosomal and lysosomal compartments of eukaryotic cells[15-16]. Using the protein network model, we then predicted the following human protein partners network modules for VacA: apoptotic module, cell cycle module, cytoskeletal remodeling module, *ras* GTPase module, and tumor suppressor module (Figure 3). It has already been demonstrated that GTPase activity is required for vacuole formation, therefore it is possible that VacA may activate GTPase, which results in vacuolation. All of the possible interaction modules are summarized in Figure 3.

Identification of Disease Related Genes/proteins

We analyzed the 108 human proteins selected by the protein interaction analysis using the GeneChip database (OncologyDB) to determine if they were associated with disease. We first compared the 108 human proteins identified by the prediction system with a list of genes known to be associated with stomach cancer. Among the 108 proteins detected by the prediction system, 93 (86%) were found to be associated with stomach cancer, 91 of which are up-regulated and 2 of which are down-regulated more than 4 fold in gastric cancer tissues (Figure. 2). Thirty two of these genes are gastric cancer-specific, how-

ever the remaining genes are also known to be associated with several other forms of cancer. Thirty of the 32 gastric cancer-specific genes are up-regulated in gastric cancer tissues and 2 are down regulated. Sixteen of the genes show high incidence in greater than 4 cancer types. Additionally, 93 of the proteins that were identified using the OncologyDB have been functionally classified (Table 2), and most of the genes (~50%) that have been functionally classified are enzymes, including kinases and phophatases. We also found that 14 of the genes identified using the OncologyDB are involved in transcription (transcription factors and activators). The entire list of genes identified by the model is summarized in the Table 3.

Discussion

We used a novel algorithm[10-11] designed to predict protein-protein interactions and protein network modeling to evaluate the relationship between *H. pylori* and human hosts. This bioinformatic method can provide new direction for functional studies of uncharacterized proteins without the need to perform extensive experiments. The results of this study revealed a very dense interaction network (4,350 protein-protein interactions) between 159 *H. pylori* proteins and 108 human proteins. Therefore, we focused on only approximately 100 *H. pylori* ORFs instead of all 1,600 ORFs identified, which saved time and labor as well as reduced the cost of the project.

Gastric cancer, which is recognized as the second most common cancer world-wide, has a high morbidity and poor prognosis[17]. Even though the etiology of gastric cancer is thought to be multifactorial, inflammation seems to be the most important trigger and the role of *Helicobacter pylori* and its association with gastric cancer has been acknowledged[18]. Additionally, our understanding of the relationship between *H. pylori* and gastric cancer has changed dramatically in recent years, and predicting host susceptibility factors and molecular targets forms the basis of our under-

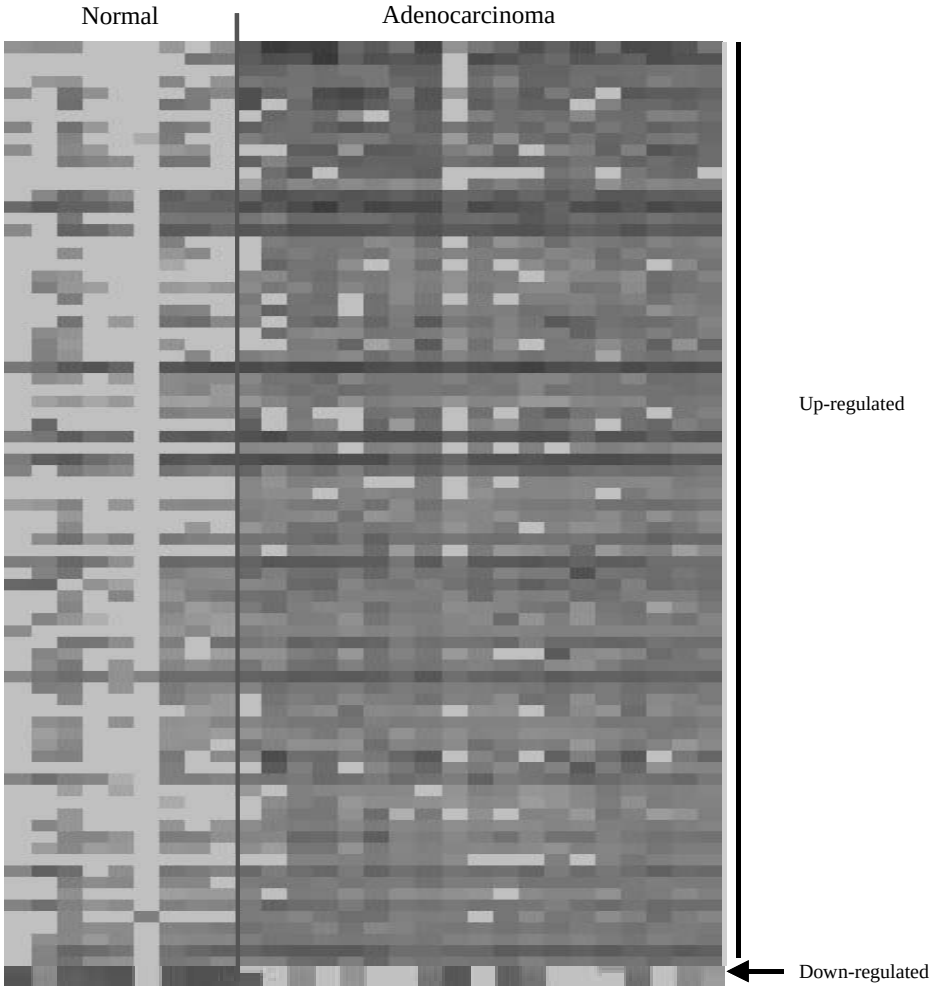

Figure 2. The expression pattern of 93 human genes predicted to interact with *H. pylori* proteins that are differentially expressed (up-regulated or down-regulated in cancer tissue by over 4-fold when compared to normal tissue) in stomach cancer tissues.

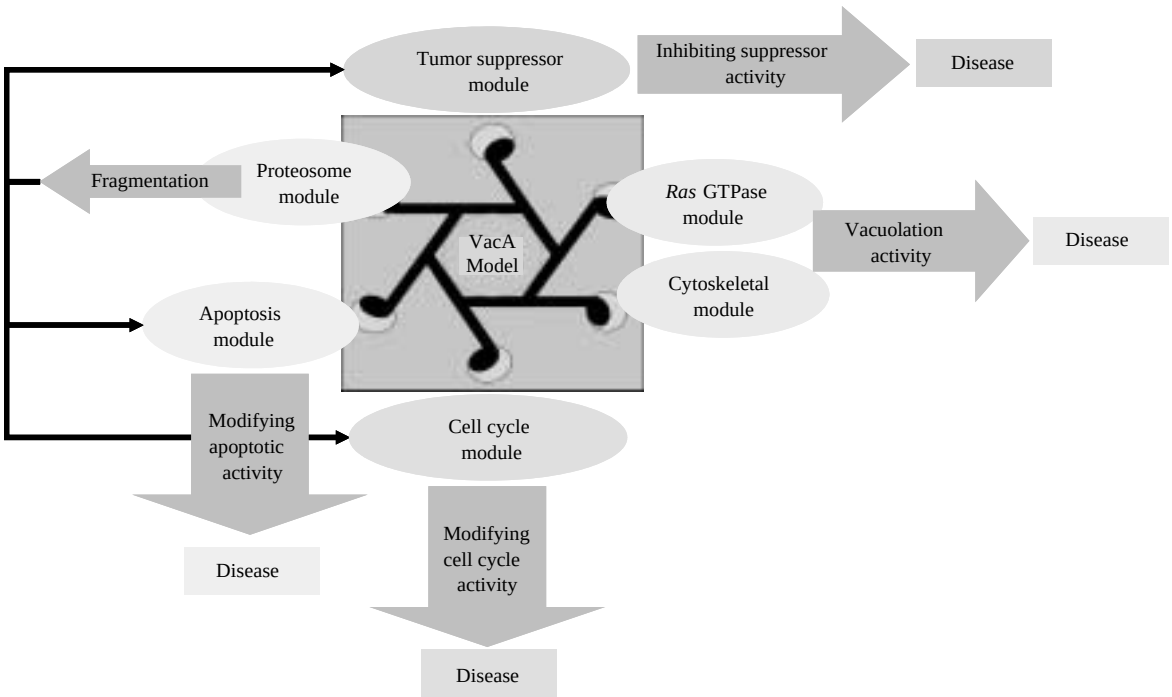

Figure 3. The possible relationships of VacA with pathogenesis inferred from the information provided by the protein network model.

standing of the pathogenesis, prevention and treatment strategies of gastric cancer[8].

Therefore, we analyzed the interactions between human and *H. pylori* proteins and integrated this interaction data with information obtained from a GeneChip database. Generally, the correlation between general genomic data and proteomic data is not high, which makes it difficult to integrate these data sets. However, there was very high correlation between the interaction data obtained in our study and the genomic database. Up to 90% of the proteins identified by the protein interaction analysis were found in cancer-related genes (overexpressed or down regulated in cancer tissues) that were identified through GeneChip analysis, therefore, by integrating the GeneChip data and the protein interaction data we were able to easily identify the disease related genes or proteins.

We identified 93 stomach cancer related genes by integrating the GeneChip data with the protein interaction data. Of these 93 genes, 24 were already known to be associated with cancer, however, the remaining 69 genes need to be evaluated to determine if they are also associated with cancer. If so, these genes may be novel targets for the diagnosis and treatment of cancer.

It is difficult to elucidate the mechanisms underlying diseases by analyzing genes one at a time, however, even though most projects investigating disease mechanisms begin by analyzing the entire genome, they ultimately analyze genes individually. We analyzed proteins both at the individual and at the genomic level to attempt to understand the disease or pathogenesis mechanisms in terms of functional modules, which are more directly related to the cellular phenotype. This approach provided more comprehensive disease and pathogenesis models, as indicated by our model for VacA, which shows that VacA may control host cellular activities related to membrane trafficking, cell growth and development, proteosmal degradation, and the cell cycle, as well as RNA processing. Our model also explains several known activities of VacA, including the activities controlling membrane trafficking, the cell cycle, and cell growth and development[19-20]. Additionally, our model indicated that RNA processing is associated with VacA, which has not been suggested before; therefore this association should be evaluated in a future study.

Materials and Methods

Protein-protein Interaction Prediction

To predict the protein-protein interaction between

Table 3. The interaction of 93 human proteins with *H. pylori* proteins that are differentially expressed (up-regulated (+) or down-regulated (−) in cancer tissue by over 4-fold when compared to those in normal tissue) in stomach cancer tissues.

Unigene ID	Symbol	Fold change	UniGene description
Hs.90744	PSMD11	+4.85	proteasome (prosome, macropain) 26S subunit, non-ATPase, 11
Hs.39871	KIAA0727	+4.62	KIAA0727 protein
Hs.239189	GLS	+4.76	glutaminase
Hs.197298	NS1-BP	+4.27	NS1-binding protein
Hs.154036	TSSC3	+4.21	tumor suppressing subtransferable candidate 3
Hs.11538	ARPC1B	+4.42	actin related protein 2/3 complex, subunit 1A (41 kD)
Hs.256583	ILF3	+4.93	interleukin enhancer binding factor 3, 90 kD
Hs.166096	ELF3	+5.41	E74-like factor 3 (ets domain transcription factor, epithelial-specific)
Hs.31334	TOM	+4.63	putative mitochondrial outer membrane protein import receptor
Hs.286229	PLXNC1	+4.56	plexin C1
Hs.24879	PPAP2C	+4.49	phosphatidic acid phosphatase type 2C
Hs.6900	RNF13	+5.31	ring finger protein 13
Hs.20166	PSCA	−6.02	prostate stem cell antigen
Hs.52644	SKAP-HOM	+4.63	SKAP55 homologue
Hs.118397	AEBP1	+5.68	AE-binding protein 1
Hs.211568	EIF4G1	+6.45	eukaryotic translation initiation factor 4 gamma, 1
Hs.155218	E1B-AP5	+8.02	E1B-55 kDa-associated protein 5
Hs.211568	EIF4G1	+6.45	eukaryotic translation initiation factor 4 gamma, 1
Hs.129673	EIF4A1	+7.55	eukaryotic translation initiation factor 4A, isoform 1
Hs.184669	ZNF144	+4.33	zinc finger protein 144 (Mel-18)
Hs.84728	KLF5	+6.34	Kruppel-like factor 5 (intestinal)
Hs.99910	PFKP	+4.88	phosphofructokinase, platelet
Hs.1390	PSMB2	+4.81	proteasome (prosome, macropain) subunit, beta type, 2
Hs.24763	RANBP1	+6.21	RAN binding protein 1
Hs.75243	BRD2	+4.30	biphenyl hydrolase-like (serine hydrolase; breast epithelial mucin-associated antigen)
Hs.182429	P5	+7.45	protein disulfide isomerase-related protein
Hs.88556	HDAC1	+4.42	histone deacetylase 1
Hs.155206	STK25	+4.01	serine/threonine kinase 25 (Ste20, yeast homolog)
Hs.74566	DPYSL3	+5.51	dihydropyrimidinase-like 3
Hs.74619	PSMD2	+6.66	proteasome (prosome, macropain) 26S subunit, non-ATPase, 2
Hs.118174	TTC3	+4.48	tetratricopeptide repeat domain 3
Hs.9930	SERPINH1,2	+12.56	serine (or cysteine) proteinase inhibitor, clade H (heat shock protein 47), member 2
Hs.155462	MCM6	+4.91	minichromosome maintenance deficient (mis5, S. pombe) 6
Hs.173902	PPP2R1A	−5.96	protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), alpha isoform
Hs.821	ZFP92	+5.43	zinc finger protein homologous to Zfp92 in mouse
Hs.181165	EEF1A1	+5.61	eukaryotic translation elongation factor 1 alpha 1
Hs.89414	CXCR4	+5.17	chemokine (C-X-C motif), receptor 4 (fusin)
Hs.74578	DDX9	+6.85	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 9
Hs.75693	PRCP	+5.11	prolylcarboxypeptidase (angiotensinase C)
Hs.795	H2AFO	+6.73	H2A histone family, member O
Hs.79353	TFDP1	+4.03	transcription factor Dp-1
Hs.179735	ARHC	+4.84	ras homolog gene family, member C
Hs.25797	SF3B4	+4.98	splicing factor 3b, subunit 4, 49kD
Hs.74111	RALY	+4.27	RNA-binding protein (autoantigenic)
Hs.180446	KPNB1	+5.27	karyopherin (importin) beta 1
Hs.77348	HPGD	+4.63	hydroxyprostaglandin dehydrogenase 15-(NAD)
Hs.78996	PCNA	+5.42	proliferating cell nuclear antigen
Hs.198951	JUNB	+4.72	jun B proto-oncogene
Hs.149846	ITGB5	+4.60	integrin, beta 5
Hs.1624	EFNA1	+4.91	ephrin-A1
Hs.1686	GNA11	+7.41	guanine nucleotide binding protein (G protein), alpha 11 (Gq class)
Hs.82932	CCND1	+5.08	cyclin D1 (PRAD1: parathyroid adenomatosis 1)
Hs.118442	CCNC	+4.23	cyclin C
Hs.150403	DDC	+5.71	dopa decarboxylase (aromatic L-amino acid decarboxylase)
Hs.169840	TTK	+4.60	TTK protein kinase
Hs.174070	E2-EPF	+5.01	ubiquitin carrier protein

Table 3. Continued.

Unigene ID	Symbol	Fold change	UniGene description
Hs.76084	LMNB2	+8.42	lamin B2
Hs.21486	STAT1	+4.82	signal transducer and activator of transcription 1, 91 kD
Hs.180832	RARS	+5.55	arginyl-tRNA synthetase
Hs.177559	IFNGR2	+5.35	interferon gamma receptor 2 (interferon gamma transducer 1)
Hs.256583	ILF3	+4.93	interleukin enhancer binding factor 3, 90 kD
Hs.82251	MYO1C	+5.26	myosin IC
Hs.197540	HIF1A	+8.02	hypoxia-inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)
Hs.159557	KPNA2	+4.63	karyopherin alpha 2 (RAG cohort 1, importin alpha 1)
Hs.220689	G3BP	+4.90	Ras-GTPase-activating protein SH3-domain-binding protein
Hs.239737	CTBP1	+4.49	C-terminal binding protein 1
Hs.83429	TNFSF10	+4.29	tumor necrosis factor (ligand) superfamily, member 10
Hs.81182	HNMT	+4.28	histamine N-methyltransferase
Hs.180398	LPP	+4.63	LIM domain-containing preferred translocation partner in lipoma
Hs.88556	HDAC1	+4.42	histone deacetylase 1
Hs.79706	PLEC1	+4.93	plectin 1, intermediate filament binding protein, 500 kD
Hs.168075	KPNB2	+7.07	karyopherin (importin) beta 2
Hs.54089	BARD1	+4.52	BRCA1 associated RING domain 1
Hs.8546	NOTCH3	+6.06	Notch (Drosophila) homolog 3
Hs.81800	CSPG2	+6.42	chondroitin sulfate proteoglycan 2 (versican)
Hs.155191	VIL2	+4.30	villin 2 (ezrin)
Hs.149846	ITGB5	+4.6)	integrin, beta 5
Hs.750	FBN1	+4.03	fibrillin 1 (Marfan syndrome)
Hs.31638	RSN	+4.91	restin (Reed-Steinberg cell-expressed intermediate filament-associated protein)
Hs.2903	PPP4C	+4.19	protein phosphatase 4 (formerly X), catalytic subunit
Hs.16003	RBBP4	+4.63	retinoblastoma-binding protein 4
Hs.85226	LIPA	+4.79	lipase A, lysosomal acid, cholesterol esterase (Wolman disease)
Hs.75424	ID1	+4.39	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
Hs.74576	GDI1	+9.29	GDP dissociation inhibitor 1
Hs.249495	HNRPA1	+5.04	heterogeneous nuclear ribonucleoprotein A1
Hs.152151	PKP4	+9.07	plakophilin 4
Hs.83942	CTSK	+4.12	cathepsin K (pycnodysostosis)
Hs.77348	HPGD	+4.63	hydroxyprostaglandin dehydrogenase 15-(NAD)
Hs.10029	CTSC	+5.10	cathepsin C
Hs.78943	BLMH	+14.69	bleomycin hydrolase
Hs.85201	CLECSF2	+4.27	C-type (calcium dependent, carbohydrate-recognition domain) lectin, superfamily member 2
Hs.100555	DDX18	+4.06	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 18 (Myc-regulated)
Hs.75626	CD58	+4.30	CD58 antigen, (lymphocyte function-associated antigen 3)

human and *H. pylori* proteins, 1,590 genes from *H. pylori* 26695 were used in conjunction with human genes collected from GeneChip data for normal stomach tissue. Human genes were selected by evaluating more than 30 cases, and then extracting the genes that had 50% CALL, which resulted in 10,257 gene entries. We then categorized all of the domains of the genes collected for *H. pylori* and humans using InterPro (http://www.ebi.ac.uk/ interpro/), which provides a large non-redundant and integrated source of domains or functional sites classified by Pfam, PRINTS, PROSITE, ProDom, SMART, and TIGRFAM. To search for InterPro domains from protein sequences, InterProScan v3.1 was downloaded from EBI (European Bioinformatics Institute, ftp://ftp.ebi.ac.uk/pub/databases/interpro/iprscan/) and installed locally. Interpro matches of all SWISS-PROT/TrEMBL proteins (protein2ipr.dat) were also obtained from the same site. Repeated domains with the same InterPro IDs are often found in a single protein, therefore these domains were counted once for the number of domain species. The InterPro domains in the parent-child relationship were also merged into one domain species. In addition to the InterPro domains, the profiles of Prosite release can also be found using InterProScan. Although some of these patterns are weak and occur frequently, they can be added to the domain catalog without producing many false positive PIDs because the threshold will also increase according to the frequency of the domain involved. All of the protein-protein interactions were predicted based on the domain information using PreDIN[11] and PreSPI[10] algo-

rithms. To generate the network model of protein interactions between humans and *H. pylori*, we selected *H. pylori* protein-domains that had high prediction scores, which gave a protein network that had ~4,350 protein interactions.

GeneChip Microarray Analysis

The Oncology DB contains expression profiles and associated clinical data from 17 normal stomach tissue samples. The expression data was generated using Affymetrix, Inc.'s GeneChip microarray technology, which currently contains 60,000 human gene fragments.

Sample categories having many missing data were not analyzed. Normal tissue samples were further selected by performing a "Principle Component Analysis (PCA)", and individual variations within the normal tissue samples were analyzed using the "Fold Change Analysis" function of the GX2000 program. Briefly, on a given organ, the fold change of a probe set was calculated by comparing a normal sample with other normal samples obtained from the same organ as a set. Probe sets with more than 2-, 3-, and 4-fold changes were counted and presented as percentages of entire probe sets (63,175). In normal samples, 7.69% and 5.18% of the entire probe sets showed more than 3- and 4-fold individual variations, respectively. Therefore, we used a 4-fold change as the threshold for stomach cancer analysis. Known genes in the "Gene List" were prioritized by reviewing information available from OMIM, GeneCard, PubMed, and MedMiner. To facilitate the process, we have built an in-house website called "FindKnownGenes". UniGene ID, Gene Symbol, UniGene Description, GeneCard, OMIM, and MedMiner data are displayed after inputting the Gene Symbol, Affymetrix ID, or GenBank Accession ID.

Acknowledgements

This study was partly supported by a grant from the Daedeok Innopolis Commercialization of Research Results Project (M12008-D07010160-200), the Ministry of Science and Technology, Korea, and by the Basic Research Project funded by the Government Contribution, Republic of Korea, and KOSEF through the National Core Research Center for Nanomedical Technology (R15-2004-024-00000-0).

References

1. Mills, K.H. Regulatory T cells: Friend or foe in immunity to infection? *Nat. Rev. Immunol.* **4**, 841-855 (2004).
2. Thakari, T., Pilione, M., Kirimanjeswara, G., Harvill, E.T. & Albert, R. Modeling systems-level regulation of host immune responses. *PLoS Comp. Biol.* **3**, 1022-1039 (2007).
3. Tomb, J.F. *et al*. The complete genome sequence of the gastric pathogen *Helicobacter pylori. Nature* **388**, 539-547 (1997).
4. Alm, R.A. *et al*. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori. *Nature* **397**, 176-180 (1999).
5. Rain, J.C. *et al*. The protein-protein interaction map of Helicobacter pylori. *Nature* **409**, 211-215 (2001).
6. Scott, D.R., Marcus, E. A., Wen, Y., Oh, J. & Sachs, G. Gene expression *in vivo* shows that *Helicobacter pylori* colonizes an acidic niche on the gastric surface. *Proc. Natl. Acad. Sci*. **104**, 7235-7240 (2007).
7. Hardin, R.J. & Wright, R.A. Helicobacter pylori: review and update. *Hospital Physician* 22-31 (May 2002).
8. Peter, S. & Beglinger, C. *Helicobacter pylori* and gastric cancer: the causal relationship. *Digestion* **75**, 25-35 (2007).
9. Sprinzak, E. & Margalit, H. Correlated sequence-signatures as markers of protein-protein interaction. *J. Mol. Biol*. **311**, 681-692. (2000).
10. Han, D.S., Kim, H.S., Jang, W.H., Lee, S.D. & Suh, J.K. PreSPI: a domain combination based prediction system for protein-protein interaction. *Nucleic Acids Res*. **32**, 6312-6320 (2004).
11. Kim, W., Park J. & Suh, J.K. Large Scale Statistical Prediction of Protein-Protein Interaction by Potentially Interacting Domain (PID) pair. *Genome Informatics* **13**, 42-50 (2002).
12. Schmitt, W. & Hass, R. Genetic analysis of the *Helicobacter pylori* vacuolating cytotoxin: structural similarities with IgA protease type of exported protein. *Mol. Microbiol*. **12**, 307-319 (1994).
13. Phadnis, S.H., Ilver, D., Janzon, L., Normark, S. & Westblom, T.U. Pathological significance and molecular characterization of the vacuolating toxin gene of *Helicobacter pylori. Infect. Immun*. **62**, 1557-1565 (1994).
14. Nguyen, V.Q., Caprioli, R.M. & Cover, T.L. Carboxy-terminal proteolytic processing of *Helicobacter pylori* vacuolating toxin. *Infect. Immun*. **69**, 543-546 (2001).
15. Telford, J.L. *et al*. Gene structure of the *Helicobacter pylori* cytotoxin and evidence of its key role in gastric disease. *J. Exp. Med.* **179**, 1653-1658 (1994).
16. Molinari, M. *et al*. Vacuoles induced by *Helicobacter pylori* toxin contain both late endosomal and lysosomal markers. *J. Biol. Chem*. **272**, 25339-25344 (1997).

17. Parkin, D.M., Bray, F., Ferlay, J. & Pisani, P. Global cancer statistics. *CA Cancer J. Clin*. **55**, 74-108 (2002).
18. Peek, R.M.Jr. & Blaser, M.J. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nat. Rev. Cancer* **2**, 28-37 (2002).
19. Cover, T.L. & Blanke, S.R. *Helicobacter pylori* VacA, a paradigm for toxin multifunctionality. *Nat. Rev. Microbiol*. **3**, 320-332 (2005).
20. Hennig, E.E., Godlewski, M.M., Butruk, E. & Ostrowski, J. *Helicobacter pylori* VacA cytotoxin interacts with fibronectin and alters HeLa cell adhesion and cytoskeletal organization *in vitro*. *FEMS Immunol. Med. Microbiol*. **44**,143-150 (2005).