Genomics & Informatics Vol. 7(2) 122-130, June 2009

Integrative Analysis of Microarray Data with Gene Ontology to Select Perturbed Molecular Functions using Gene Ontology Functional Code

Changsik Kim, Jiwon Choi and Sukjoon Yoon*

Department of Biological Sciences, Sookmyung Women's University, Seoul 140-742, Korea

Abstract

A systems biology approach for the identification of perturbed molecular functions is required to understand the complex progressive disease such as breast cancer. In this study, we analyze the microarray data with Gene Ontology terms of molecular functions to select perturbed molecular functional modules in breast cancer tissues based on the definition of Gene ontology Functional Code. The Gene Ontology is three structured vocabularies describing genes and its products in terms of their associated biological processes, cellular components and molecular functions. The Gene Ontology is hierarchically classified as a directed acyclic graph. However, it is difficult to visualize Gene Ontology as a directed tree since a Gene Ontology term may have more than one parent by providing multiple paths from the root. Therefore, we applied the definition of Gene Ontology codes by defining one or more GO code(s) to each GO term to visualize the hierarchical classification of GO terms as a network. The selected molecular functions could be considered as perturbed molecular functional modules that putatively contributes to the progression of disease. We evaluated the method by analyzing microarray dataset of breast cancer tissues; i.e., normal and invasive breast cancer tissues. Based on the integration approach, we selected several interesting perturbed molecular functions that are implicated in the progression of breast cancers. Moreover, these selected molecular functions include several known breast cancer-related genes. It is concluded from this study that the present strategy is capable of selecting perturbed molecular functions that putatively play roles in the progression of diseases and provides an improved interpretability of GO terms based on the definition of Gene Ontology codes.

*Corresponding author: E-mail yoonsj@sookmyung.ac.kr
Tel +82-2-710-9415, Fax +82-2-2077-7322
Accepted 25 May 2009

Keywords: breast cancer, gene ontology code, gene ontology functional code, microarray, molecular functions

Introduction

A major challenge in molecular biology is the identification of genes or molecular functions involved in diseases and other biological process. Over the past decades, DNA microarrays have been widely used for this problem due to their ability of parallel monitoring of the genome-wide transcriptional profiling. The conventional analysis of microarray data starts with normalization, followed by the test statistics to compare expression levels in different phenotypes for each gene, yielding thousands of differentially expressed genes (Dhanasekaran *et al*., 2001). However, this strategy is limited due to the fact that differentially expressed genes can be analyzed one-at-a-time. In addition, most genes are known to function in concert rather than alone, and their products interact with each other and with DNA by forming dynamic topological interactomes. Therefore, a systems biology approach that can identify perturbed molecular functions with differentially expressed genes would accelerate the understanding the basic molecular mechanism of certain phenotypes or diseases.

The Gene Ontology (The Gene Ontology Consortium, 2000) is a database of structured controlled ontology that describe gene product in terms of their biological processes, cellular components, and molecular functions. GO is a hierarchically structure forming an acyclic digraph with top-down directions, which provides an efficient navigation for the structure of the ontology. However, the same GO term may occur in different lines of ontology structure. It means that GO may not be represented as directed tree since a GO term may have more than one parent providing multiple paths from the root. To construct an ordered GO tree for the purpose of global visualization, GO terms need to be distinguished from one another if they are occurred in different locations on the hierarchical classification of gene ontology. Based on the biological point of view, Lee *et al*. (2004) justified that what is more important is not a GO term itself, but which path the GO terms takes from the root in the gene ontology. It means that each location of a GO term could be considered distinct if a distinct path leads to it from the root of gene ontology. Therefore, we de-

fined Gene ontology Functional Code (GFC) to each GO term based on the definition of GO codes by Lee *et al*. (2004). Note that a unique GO code is assigned to a GO term in each location of the classification of gene ontology. It therefore means that the visualization of gene ontology based on GO codes will provide more interpretability of gene ontology than the one based on GO terms.

Many progressive diseases such as cancer are caused by combinatorial effect of multiple genes rather than a single gene. In this study, we integrated the microarray data with GO terms to find the perturbed molecular functions in breast cancer tissues using GO codes in molecular function. To this end, we identify several perturbed molecular functions that are implicated to play roles in the progression of breast cancer.

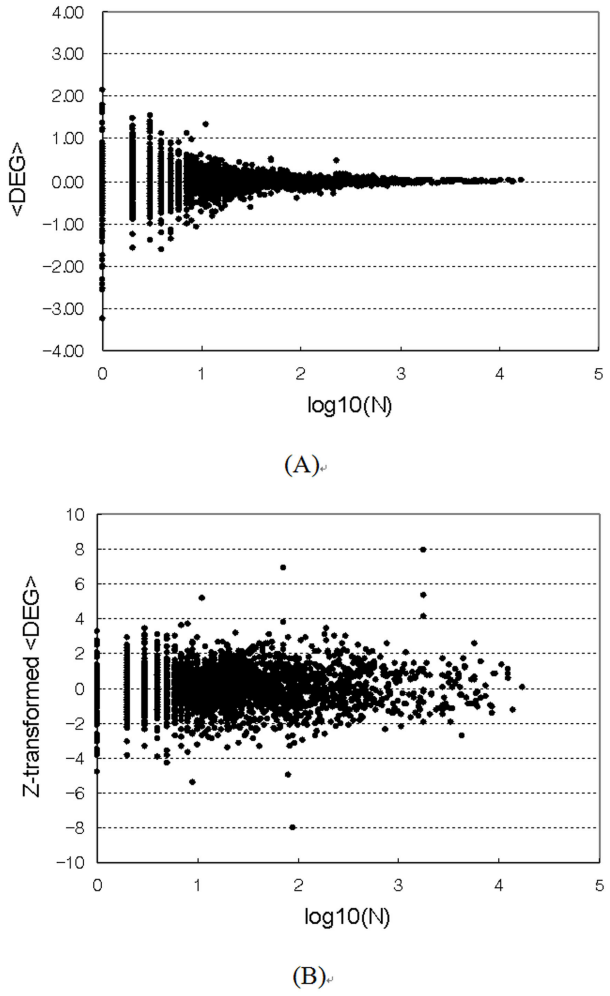

Fig. 1. (A) The distribution of average of log2-fold change of probes for all GFC codes according to the size of all GFC codes. (B) The distribution of "z-transformed" average of log2-fold change of probes for all GFC codes according to the size of probes for all GFC codes. Note the N represents the number of probes belong to each GFC code, *(DEG)* represents the average of log-fold change of probes for each GFC code, and *Z-transformed (DEG)* represents the average of log-fold change of probes with z-transformation for each GFC code.

Methods

The hierarchical classification of gene ontology is basically represented as a directed acyclic graph (DAG) with three categories, namely molecular function, biological process and cellular component (Ashburner *et al*., 2000). Based on this hierarchical classification, GO terms can be represented as node by an acyclic digraph. However, GO DAG is difficult to be visualized as a directed tree since a GO term may have more than one parent by providing multiple paths from the root. Thus, we have adopted the definition of GO code (Lee *et al*., 2004; Fig. 1), in which GO terms can be efficiently handled in a tree structure by defining one or more GO code(s) to each GO term, to visualize the hierarchical classification of GO terms as a network. That is, a GO code is assigned to a GO term in each location of the classification of gene ontology. A GO term is transformed into a GO code $\alpha_1\alpha_2\alpha_3\cdots\alpha_H$ using the unique path from the root of the classification of GO to each GO term, in which H is the length of a longest path from the root to each GO term in the classification of GO. The GO codes can be represented as nodes in the ordered tree of GO category. Each node is located on the level N of GO tree for $N=1,2,\cdots H$. See Lee *et al*.'s (2004) study for more detail of GO code. In this study, we have focused

Table 1. The distribution of GFC classes and average number of entries per each GFC class corresponding to the level of GFC. Note that the numbers inside the parenthesis represent the number of GFC classes with at least 1 probe is assigned to its GFC classes and the number of average entries per each GFC classes with at least 1 probes are assigned to its GFC classes, respectively

GFC level	Number of GFC classes	Number of average entries
1	1	30,130 (30,130)
2	19 (18)	4,225 (4,460)
3	927 (244)	179 (682)
4	676 (418)	298 (483)
5	1,537 (746)	134 (275)
6	4,507 (1,555)	71 (205)
7	2,259 (1,119)	39 (79)
8	2,160 (1,131)	28 (54)
9	1,970 (1,012)	20 (39)

on the GO code for the GO term of molecular function, which is defined as Gene ontology Functional Code (GFC). Note that for the probe set of *Affymetrix GeneChip Human Genome U133 Plus 2.0* Array chip, the distribution of GFC classes and average number of entries per each GFC class corresponding to the level of GFC is listed in Table 1.

To visualize the perturbed molecular functions in breast cancer tissues using GFC, we analyzed the microarray dataset from Turashvilli *et al.* (2007), which consists of 2 types of breast cancer tissues; i.e., invasive lobular and ductal carcinomas. This dataset includes a total of 30 samples that consist of 10 samples from normal ductal cells, 10 samples from normal lobular cells, 5 samples from invasive ductal carcinoma cells, and 5 samples from invasive lobular carcinoma cells, which were microdissected from cryosections of 10 mastectomy specimens from postmenopausal patients.

Before selecting the perturbed molecular functional modules, we applied the two steps of data processing. First, the log2 transformation of fold change between normal and disease states was applied for each sample array. Second, the average of log2-fold change was calculated for the probe set of each GFC. The distribution of average of log2-fold change for each GFC depending on the size of GFCs is shown in Fig. 1A. It is observed that the GFCs with relatively large size tend to have significantly smaller average values of log2-fold change than the ones with small size. It means that it is difficult to discriminate the perturbed molecular functional modules with relatively large size using the average value of log2-fold change. To overcome this difficulty, the average values of log2-fold change were z-transformed depending on the size of GFCs as follows. First, the GFCs were divided into 152 subsets, in which each subset consists of GFCs with same or similar sizes. Second, the average value of log2-fold change in each subset was standardized such that its mean and standard deviation are 0 and 1, respectively. After z-transformation, it

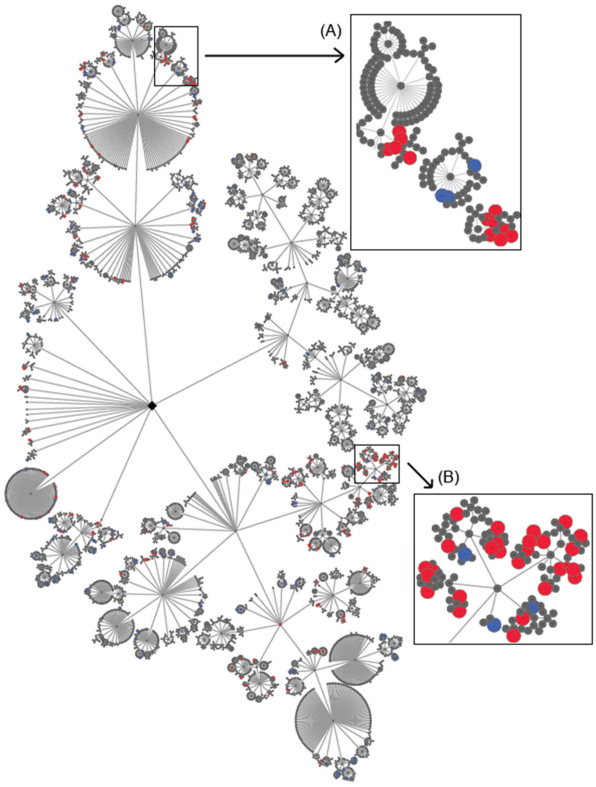

Fig. 2. The global network of hierarchical classes of Gene ontology Functional Code (GFC). Note that red nodes represent the GFC code with up-regulated probes and blue nodes the GFC code with down-regulated probes.

Table 2. Dysregulated GFC codes with Z-transformed (DEG) ≥ 2.0 or ≤ -2.0 in Fig. 2 (A)

GFC code	GFC description	N	(DEG)
1-3-45-19-1-1-0-0-0-0-0-0-0-0-0	C-C chemokine binding	36	2.409314
1-3-45-19-1-1-1-0-0-0-0-0-0-0-0	C-C chemokine receptor activity	36	2.409314
1-3-45-19-1-3-0-0-0-0-0-0-0-0-0	C-X3-C chemokine binding	34	2.206857
1-3-45-19-1-3-1-0-0-0-0-0-0-0-0	C-X3-C chemokine receptor activity	34	2.206857
1-3-45-19-1-4-1-0-0-0-0-0-0-0-0	C-C chemokine receptor activity	24	3.143907
1-3-45-33-11-0-0-0-0-0-0-0-0-0-0	Interleukin-6 binding	29	−2.05751
1-3-45-33-14-2-0-0-0-0-0-0-0-0-0	Nerve growth factor binding	1	−2.79655
1-3-45-33-15-0-0-0-0-0-0-0-0-0-0	Platelet-derived growth factor binding	5	−2.1969
1-3-45-62-4-1-1-0-0-0-0-0-0-0-0	IgA receptor activity	2	2.332873
1-3-45-62-4-3-1-0-0-0-0-0-0-0-0	IgE receptor activity	2	2.332873
1-3-45-62-4-4-1-0-0-0-0-0-0-0-0	IgG receptor activity	2	2.332873
1-3-45-62-4-5-0-0-0-0-0-0-0-0-0	IgM binding	2	2.332873
1-3-45-62-4-5-1-0-0-0-0-0-0-0-0	IgM receptor activity	2	2.332873
1-3-45-62-4-7-0-0-0-0-0-0-0-0-0	Polymeric immunoglobulin binding	2	2.332873
1-3-45-62-4-7-1-0-0-0-0-0-0-0-0	Polymeric immunoglobulin receptor activity	2	2.332873

is therefore more viable to discriminate the perturbed GFCs of which size are relatively large (Fig. 1B).

Results and Discussion

The global network of hierarchical molecular functional classes of GO was constructed using GFC in Fig. 2. The "z-transformed" average of log2-fold change for each GFC was integrated with the network to visualize the perturbed molecular functional modules in breast cancer tissues. Note that the colored (e.g., red or blue) nodes represent the perturbed molecular functional classes in Fig. 2, and the sub-network with enriched color nodes could be chosen for further analysis of per-

Table 3. Dysregulated GFC codes with Z-transformed (DEG) ≥ 2.0 or ≤ -2.0 in Fig. 2 (B)

GFC code	GFC description	N	(DEG)
1-4-7-15-2-64-0-0-0-0-0-0-0-0-0	Phospholipase A2 activity	45	2.326891
1-4-7-15-4-0-0-0-0-0-0-0-0-0-0	Lipase activity	191	3.401221
1-4-7-15-4-4-0-0-0-0-0-0-0-0-0	Phospholipase activity	162	2.113644
1-4-7-15-4-4-5-0-0-0-0-0-0-0-0	Phospholipase D activity	20	−2.17466
1-4-7-15-5-1-2-1-0-0-0-0-0-0-0	Endodeoxyribonuclease activity, producing 3'-phosphomonoesters	6	2.951635
1-4-7-15-5-1-2-1-1-0-0-0-0-0-0	3'-flap endonuclease activity	3	2.792583
1-4-7-15-5-1-2-2-1-0-0-0-0-0-0	5'-flap endonuclease activity	3	2.792583
1-4-7-15-5-1-2-4-0-0-0-0-0-0-0	Flap endonuclease activity	3	2.792583
1-4-7-15-5-1-3-1-10-0-0-0-0-0-0	Single-stranded DNA specific exodeoxyribonuclease activity	8	2.140097
1-4-7-15-5-1-3-1-2-0-0-0-0-0-0	5'-3' exodeoxyribonuclease activity	5	2.82153
1-4-7-15-5-2-1-1-0-0-0-0-0-0-0	Endodeoxyribonuclease activity, producing 3'-phosphomonoesters	6	2.951635
1-4-7-15-5-2-1-1-1-0-0-0-0-0-0	3'-flap endonuclease activity	3	2.792583
1-4-7-15-5-2-1-2-1-0-0-0-0-0-0	5'-flap endonuclease activity	3	2.792583
1-4-7-15-5-2-1-4-0-0-0-0-0-0-0	Flap endonuclease activity	3	2.792583
1-4-7-15-5-2-2-1-1-0-0-0-0-0-0	3'-flap endonuclease activity	3	2.792583
1-4-7-15-5-2-2-2-0-0-0-0-0-0-0	Endoribonuclease activity, producing 3'-phosphomonoesters	32	−2.06783
1-4-7-15-5-2-3-1-1-0-0-0-0-0-0	5'-flap endonuclease activity	3	2.792583
1-4-7-15-5-3-1-1-0-0-0-0-0-0-0	3'-5'-exodeoxyribonuclease activity	9	2.642723
1-4-7-15-5-3-1-1-2-0-0-0-0-0-0	Single-stranded DNA specific 3'-5' exodeoxyribonuclease activity	2	2.86147
1-4-7-15-5-3-1-3-0-0-0-0-0-0-0	Exoribonuclease II activity	2	2.86147
1-4-7-15-5-3-2-0-0-0-0-0-0-0-0	5'-3' exonuclease activity	5	2.82153
1-4-7-15-5-3-2-1-0-0-0-0-0-0-0	5'-3' exodeoxyribonuclease activity	3	2.792583
1-4-7-15-5-3-2-1-2-0-0-0-0-0-0	Double-stranded DNA specific 5'-3' exodeoxyribonuclease activity	3	2.792583
1-4-7-15-5-3-2-1-3-0-0-0-0-0-0	Single-stranded DNA specific 5'-3' exodeoxyribonuclease activity	1	2.590043
1-4-7-15-5-3-3-1-10-0-0-0-0-0-0	Single-stranded DNA specific exodeoxyribonuclease activity	8	2.140097
1-4-7-15-5-3-3-1-2-0-0-0-0-0-0	5'-3' exodeoxyribonuclease activity	5	2.82153
1-4-7-15-5-3-5-1-10-0-0-0-0-0-0	Single-stranded DNA specific exodeoxyribonuclease activity	8	2.140097
1-4-7-15-5-3-5-1-2-0-0-0-0-0-0	5'-3' exodeoxyribonuclease activity	5	2.82153
1-4-7-15-5-3-5-2-2-0-0-0-0-0-0	5'-3' exoribonuclease activity	5	2.82153
1-4-7-15-5-3-7-2-2-0-0-0-0-0-0	5'-3' exoribonuclease activity	5	2.82153
1-4-7-15-5-4-2-0-0-0-0-0-0-0-0	Oligoribonucleotidase activity	39	−2.29416
1-4-7-15-5-5-10-0-0-0-0-0-0-0-0	tRNA-specific ribonuclease activity	26	−2.21839
1-4-7-15-5-5-3-0-0-0-0-0-0-0-0	Exoribonuclease activity	109	2.701695
1-4-7-15-5-5-3-2-2-0-0-0-0-0-0	5'-3' exoribonuclease activity	5	2.82153
1-4-7-15-6-0-0-0-0-0-0-0-0-0-0	Phosphoric ester hydrolase activity	1,203	2.025344
1-4-7-15-6-2-50-5-0-0-0-0-0-0-0	Protein tyrosine/threonine phosphatase activity	4	−2.25336
1-4-7-15-6-3-16-0-0-0-0-0-0-0-0	Phospholipase C activity	73	6.864734
1-4-7-15-6-3-4-3-0-0-0-0-0-0-0	3',5'-cyclic-nucleotide phosphodiesterase activity	72	2.20463

turbed molecular functional modules. As an example, we chose the two sub-networks with relatively enriched perturbed GFCs in Fig. 2 (A) and (B) for more detailed analysis.

The perturbed molecular functional modules in Fig. 2 (A) and (B) are listed in Table 2 and 3, and the differentially expressed genes for those GFC in Table 1 and 3 are listed in Table 4 and 5, respectively. Table 2 includes 15 perturbed molecular functional modules that are mainly encompassed by three major molecular functional activities: 1) G-protein coupled peptide receptor activity, 2) immunoglobulin binding activity, and 3) interleukin binding activity. Table 3 includes 38 perturbed molecular functional modules that are mainly encompassed by peptide receptor activity. There are significant indications that these molecular functions are associated with human cancers as follows.

The G-protein coupled peptide receptor activity encompasses the perturbed molecular functional modules such as C-C chemokine binding/receptor activity and C-3X-C chemokine binding/receptor activity (Table 2) including differentially expressed genes such as CX3CR1, DARC, CXCR4, and CCRL2 (Table 4). Recent review study by Dorsam *et al.* (2007) addressed that G-protein-coupled receptors (GPCRs) have crucial roles in tumor growth and metastasis and malignant cells often seize control of the normal physiological functions of GPCRs to survive, proliferate autonomously, evade the immune system, increase their blood supply, invade their surrounding tissues and disseminate to other tissues. Chemokines are leukocyte chemoattractants that are divided structurally into four subgroups, based on the composition of cysteine residues in their amino-terminal portion. It has been shown that target organs for breast cancer metastases are enriched in the chemokine CXCL-12, inducing the specific migration of breast tumor cells that express the relevant receptor for this chemokine, CXCR4 (Muller *et al.*, 2001).

The immunoglobulin binding activity encompasses the several perturbed molecular functional modules such as IgA/E/G/M receptor activity and polymeric immunoglobulin binding/receptor activity (Table 2) including differentially expressed gene such as FCGR1B (Table 4). Qiu *et al.* (2003) demonstrated that human cancers of breast epithelial cancer lines produce immunoglobulins in both cytoplasmic and secreted form. Their study indicates that prevalent expression of immuglobulin in human carcinomas and its growth-promoting functions may have important implications in growth regulation of human cancers. Monocyte activation in cancer patients could be an indication of anticancer activity. However, Goodale *et al.* (2009) showed that recruitment of macrophase can promote tumor growth and angiogenesis by analyzing the collected peripheral blood using flow cytometry for monocyte activation. They demonstrated that metastatic breast cancer patients have a higher monocyte FCGR1B (CD64) index relative to normal donors and localized breast cancer patients.

The interleukin binding activity encompasses three perturbed molecular functional modules such as interleukin-6 binding, nerve growth factor binding, and platelet-derived growth factor binding (Table 2) including differentially expressed genes such as IL15RA, IL17RB, and PDGFRA (Table 4). Interleukin-17 (IL-17) is the core member of a group of cytokines (i.e., the IL-17 family), which was identified as a transcript from a rodent T-cell hybridoma by Rouvier *et al.* (1993). IL-17 induces the production of many other cytokines (e.g., IL-6, IL-1beta, TGF-beta, TNF-beta), chemokines (e.g., IL-8) and prostaglandins (e.g., PGE_2) from many cell types (fibroblast,

Table 4. Dysregulated genes (fold change ≥ 2.0 or ≤ -2.0) with GFC codes in Table 2

GFC code	GFC description	Dysregulated genes
1-3-45-19-1-1-0-0-0-0-0-0-0-0-0	C-C chemokine binding	CX3CR1, DARC, CXCR4, CCRL2
1-3-45-19-1-1-1-0-0-0-0-0-0-0-0	C-C chemokine receptor activity	CX3CR1, DARC, CXCR4, CCRL2
1-3-45-19-1-3-0-0-0-0-0-0-0-0-0	C-X3-C chemokine binding	CX3CR1, DARC, CXCR4, CCRL2
1-3-45-19-1-3-1-0-0-0-0-0-0-0-0	C-X3-C chemokine receptor activity	CX3CR1, DARC, CXCR4, CCRL2
1-3-45-19-1-4-1-0-0-0-0-0-0-0-0	C-C chemokine receptor activity	CXCR4, CCRL2
1-3-45-33-11-0-0-0-0-0-0-0-0-0-0	Interleukin-6 binding	IL15RA, IL17RB
1-3-45-33-14-2-0-0-0-0-0-0-0-0-0	Nerve growth factor binding	
1-3-45-33-15-0-0-0-0-0-0-0-0-0-0	Platelet-derived growth factor binding	PDGFRA
1-3-45-62-4-1-1-0-0-0-0-0-0-0-0	IgA receptor activity	FCGR1B
1-3-45-62-4-3-1-0-0-0-0-0-0-0-0	IgE receptor activity	FCGR1B
1-3-45-62-4-4-1-0-0-0-0-0-0-0-0	IgG receptor activity	FCGR1B
1-3-45-62-4-5-0-0-0-0-0-0-0-0-0	IgM binding	FCGR1B
1-3-45-62-4-5-1-0-0-0-0-0-0-0-0	IgM receptor activity	FCGR1B
1-3-45-62-4-7-0-0-0-0-0-0-0-0-0	Polymeric immunoglobulin binding	FCGR1B
1-3-45-62-4-7-1-0-0-0-0-0-0-0-0	Polymeric immunoglobulin receptor activity	FCGR1B

endothelial cells, epithelial cells, keratinocytes and macrophases). It is well known that TGF-beta plays a role as an antiproliferative factor in normal epithelial cells and at early stages of onogenesis (Hill *et al*., 2009). The IL-17 family has been linked to many immune/autoimmune related diseases including anti-tumor immunity (Aggarwal and Gurney, 2002). It should be noted that this study suggests molecular functional modules for breast cancer are down-regulated in Table 2. The platelet-derived growth factors (PDGFA-A and -B) are important factors regulating cell proliferation, cellular differentiation, cell growth, development and many diseases including cancer. PDGF and its receptor are important targets in cancer therapy based on angiogenesis (Sennino *et al*., 2009).

The peptide receptor activity encompasses 38 perturbed molecular functional modules (Table 3) including several differentially expressed genes (Table 5). There

Table 5. Dysregulated genes (fold change ≥ 2.0 or ≤ -2.0) with GFC codes in Table 3

GFC code	GFC description	Dysregulated genes
1-4-7-15-2-64-0-0-0-0-0-0-0-0-0	Phospholipase A2 activity	
1-4-7-15-4-0-0-0-0-0-0-0-0-0-0	LIpase activity	EDNRA, PLCE1, BDKRB2, PLCB1, PLCL2, PLCH1, PLD1, GPLD1, CES1, BC005884, ESD
1-4-7-15-4-4-0-0-0-0-0-0-0-0-0	Phospholipase activity	CES1, BC005884, ESD, EDNRA, PLCE1, BDKRB2, PLCB1, PLCL2, PLCH1, PLD1, GPLD1
1-4-7-15-4-4-5-0-0-0-0-0-0-0-0	Phospholipase D activity	PLD1, GPLD1
1-4-7-15-5-1-2-1-0-0-0-0-0-0-0	Endodeoxyribonuclease activity, producing 3'-phosphomonoesters	EXO1, FEN1
1-4-7-15-5-1-2-1-1-0-0-0-0-0-0	3'-flap endonuclease activity	EXO1
1-4-7-15-5-1-2-2-1-0-0-0-0-0-0	5'-flap endonuclease activity	EXO1
1-4-7-15-5-1-2-4-0-0-0-0-0-0-0	Flap endonuclease activity	EXO1
1-4-7-15-5-1-3-1-10-0-0-0-0-0-0	Single-stranded DNA specific exodeoxyribonuclease activity	ISG20
1-4-7-15-5-1-3-1-2-0-0-0-0-0-0	5'-3' exodeoxyribonuclease activity	FEN1
1-4-7-15-5-2-1-1-0-0-0-0-0-0-0	Endodeoxyribonuclease activity, producing 3'-phosphomonoesters	EXO1, FEN1
1-4-7-15-5-2-1-1-1-0-0-0-0-0-0	3'-flap endonuclease activity	EXO1
1-4-7-15-5-2-1-2-1-0-0-0-0-0-0	5'-flap endonuclease activity	EXO1
1-4-7-15-5-2-1-4-0-0-0-0-0-0-0	Flap endonuclease activity	EXO1
1-4-7-15-5-2-2-1-1-0-0-0-0-0-0	3'-flap endonuclease activity	EXO1
1-4-7-15-5-2-2-2-0-0-0-0-0-0-0	Endoribonuclease activity, producing 3'-phosphomonoesters	ANG, QSER1
1-4-7-15-5-2-3-1-1-0-0-0-0-0-0	5'-flap endonuclease activity	EXO1
1-4-7-15-5-3-1-1-0-0-0-0-0-0-0	3'-5'-exodeoxyribonuclease activity	FEN1, ISG20
1-4-7-15-5-3-1-1-2-0-0-0-0-0-0	Single-stranded DNA specific 3'-5' exodeoxyribonuclease activity	ISG20
1-4-7-15-5-3-1-3-0-0-0-0-0-0-0	Exoribonuclease II activity	ISG20
1-4-7-15-5-3-2-0-0-0-0-0-0-0-0	5'-3' exonuclease activity	FEN1, EXO1
1-4-7-15-5-3-2-1-0-0-0-0-0-0-0	5'-3' exodeoxyribonuclease activity	FEN1
1-4-7-15-5-3-2-1-2-0-0-0-0-0-0	Double-stranded DNA specific 5'-3' exodeoxyribonuclease activity	EXO1
1-4-7-15-5-3-2-1-3-0-0-0-0-0-0	Single-stranded DNA specific 5'-3' exodeoxyribonuclease activity	
1-4-7-15-5-3-3-1-10-0-0-0-0-0-0	Single-stranded DNA specific exodeoxyribonuclease activity	ISG20
1-4-7-15-5-3-3-1-2-0-0-0-0-0-0	5'-3' exodeoxyribonuclease activity	FEN1
1-4-7-15-5-3-5-1-10-0-0-0-0-0-0	Single-stranded DNA specific exodeoxyribonuclease activity	ISG20
1-4-7-15-5-3-5-1-2-0-0-0-0-0-0	5'-3' exodeoxyribonuclease activity	FEN1
1-4-7-15-5-3-5-2-2-0-0-0-0-0-0	5'-3' exoribonuclease activity	EXO1, FEN1
1-4-7-15-5-3-7-2-2-0-0-0-0-0-0	5'-3' exoribonuclease activity	EXO1, FEN1
1-4-7-15-5-4-2-0-0-0-0-0-0-0-0	Oligoribonucleotidase activity	ANG, AZGP1, ZKSCAN1, DIS3L2
1-4-7-15-5-5-10-0-0-0-0-0-0-0-0	tRNA-specific ribonuclease activity	QSER1
1-4-7-15-5-5-3-0-0-0-0-0-0-0-0	Exoribonuclease activity	CNOT6L, DCUN1D3, XRN1, AU148213, EXOD1, EXOSC3, EXO1, FEN1, ISG20, REXO2, FLJ20433, USP52

Table 5. Continued

GFC code	GFC description	Dysregulated genes
1-4-7-15-5-5-3-2-2-0-0-0-0-0-0	5'-3' exoribonuclease activity	EXO1, FEN1
1-4-7-15-6-0-0-0-0-0-0-0-0-0-0	Phosphoric ester hydrolase activity	CTDSPL2, CTDSPL, PON2, PFKFB1, PFKFB2, IMPA2, INPP5B, NT5C1B, NT5E, DLG7, CTDP1, CILP, PPAP2B, PPAP2C, DLG1, LCK, PPM1A, PPM1H, RGS12, TIAM1, BAG4, MS4A1, DCLK1, BRAF, STYX, CLDN4, EGFR, F3, FZD1, NGFR, TLR3, LILRB5, IL18RAP, KLRD1, DARC, CD69, LEPR, FCGR2C, IGHG3, BF664114, RAC1, PLXNB1, INTS6, IGHG1, TLR8, FBN1, DUSP19, PTPLA, BC029442, PTPRS, PTPRB, DUSP16, PTPRK, PTPN3, DUSP4, PTPRZ1, CDC14A, PTPN21, PTPRC, EYA4, DUSP6, CDKN3, EPM2A, MTMR3, CDC14B, C20orf57, SGPP1, PDE8B, W73272, PDE4D, AI638433, PDE1C, PDE1A, GDPD1, CNOT6L, USP52, EXO1, ISG20, FEN1, DCUN1D3, EXOSC3, XRN1, AU148213, REXO2, EXOD1, EDNRA, PLCE1, BDKRB2, PLCB1, PLCL2, PLCH1, PLD1, GPLD1
1-4-7-15-6-2-50-5-0-0-0-0-0-0-0	Protein tyrosine/threonine phosphatase activity	DUSP4, PTPRZ1
1-4-7-15-6-3-16-0-0-0-0-0-0-0-0	Phospholipase C activity	EDNRA, PLCE1, BDKRB2, PLCB1, PLCL2, PLCH1
1-4-7-15-6-3-4-3-0-0-0-0-0-0-0	3',5'-cyclic-nucleotide phosphodiesterase activity	PDE8B, W73272, PDE4D, AI638433, PDE1C, PDE1A

are significant amount of *in vitro* observation that peptide receptors are expressed in large quantities in certain tumors. Reubi (2003) summarized and critically evaluated the *in vitro* data on peptide and peptide receptor expression in human cancers, which could be considered to be the molecular basis for peptide receptor targeting of tumors. Moreover, there are considerable amount of genes in Table 5 that are implicated to play roles in the progression of breast cancer or other cancers, including BDKRB2 (Greco *et al*., 2005), PLCB1 (Naor, 2009), PLD1 (Eisen and Brown, 2002), GPLD1 (Derevianko *et al*., 1996; Williams *et al*., 2001), EXO1 (Naderi *et al*., 2007), ANG (Campo *et al*., 2005), ISG20/HEM45 (Pentecost, 1998), AZGP1/ZAG (Hassan *et al*., 2009), ZKSCAN1 (Pennanen *et al*., 2009), REXO2 (Flanagan *et al*., 2009), CTDSPL (Murabito *et al*., 2007), NT5E (Zhou *et al*., 2007), DLG1 (Fuja *et al*., 2004), PPM1A (Lin *et al*., 2006), TIAM1 (Lane *et al*., 2008), BAG4 (Yang *et al*., 2006), BRAF (Hollestelle, *et al*., 2007), CLDN4 (Kulka *et al*., 2009), F3/TF (Amirkhosravi *et al*., 1998), FZD1 (Benhaj *et al*., 2006), NGFR (Reis-Filho *et al*., 2006), LEFR (Han *et al*., 2008), RAC1 (Han *et al*., 2008), PLXNB1 (Rody *et al*., 2007), IGHG1 (Kabbage *et al*., 2008), IGHG3 (Bin Amer *et al*., 2008), TLR3 (Salaun *et al*., 2006), FBN1 (Chen *et al*., 2007), DUSP4 (Venter *et al*., 2005), PTPN3 (Wang *et al*., 2004), CDKN3/KAP (Lee *et al*., 2000), and etc.

Based on the integration of GFC codes and microarray data of breast cancer tissues, we present several perturbed molecular functions, which are implicated in the progression of breast cancer. It was also found that these selected molecular functions include several known breast cancer related genes. Moreover, it is believed that our strategy provides efficient visual interpretability of microarray data integrated with GFC to understand the perturbed molecular functions in complex progressive diseases such as breast cancer. Therefore, the selected perturbed molecular functions in this study could be further investigated to understand the basic molecular functions, which discriminate different breast cancer types.

Acknowledgements

This research was supported by Sookmyung Women's University Research Grants 1-0803-0198.

References

Aggarwal, S., and Gurney, A.L. (2002). IL-17: prototype member of an emerging cytokine family. *J. Leukoc. Biol.* 71, 1-8.

Amirkhosravi, A., Meyer, T., Warnes, G., Amaya, M., Malik, Z., Biggerstaff, J.P., Siddiqui, F.A., Sherman, P., and Francis, J.L. (1998). Pentoxifylline Inhibits Hypoxia-induced Upregulation of Tumor Cell Tissue Factor and Vascular Endothelial Growth Factor. *Thromb. Haemost.* 4, 598-602.

Ashburner, M., Ball, C.A., Blake, J.A., *et al*. (2000). Gene

Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25-29.
Sennino, B., Kuhnert, F., Tabruyn, S.P., Mancuso, M.R., Hu-Lowe, D.D., Kuo, C.J., and McDonald, D.M. (2009). Cellular Source and Amount of Vascular Endothelial Growth Factor and Platelet-Derived Growth Factor in Tumors Determine Response to Angiogenesis Inhibitors. *Cancer Res.* 69, 4527.
Benhaj, K., Akcali, K.C., and Ozturk, M. (2006). Redundant expression of canonical Wnt ligands in human breast cancer cell lines. *Oncol. Rep.* 15, 701-707.
Bin Amer, S.M., Maqbool, Z., Nirmal, M.S., Qattan, A.T., Hussain, S.S., Jeprel, H.A., Tulbah, A.M., Malik, O.A., and Al-Tweigeri, T.A. (2008). Gene expression profiling in women with breast cancer in a Saudi population. *Saudi Med. J.* 29, 507-513.
Campo, L., Turley, H., Han, C., Pezzella, F., Gatter, K.C., Harris, A.L., and Fox, S.B. (2005). Angiogenin is up-regulated in the nucleus and cytoplasm in human primary breast carcinoma and is associated with markers of hypoxia but not survival. *J. Pathol.* 205, 585-591.
Chen, W., Salto-Tellez, M., Palanisamy, N., *et al.* (2007). Targets of genome copy number reduction in primary breast cancers identified by integrative genomics. Genes Chromo. *Cancer* 46, 288-301.
Danasekaran, S.M., Barrette, T.R., Ghosh, D., *et al.* (2001). Delineation of prognostic biomarkers in prostate cancer metastasis. *Nature* 412, 822-826.
Derevianko, A., Graeber, T., D'Amico, R., and Simms, H.H. (1996). Altered oxygen tension modulates cytokine-induced signal transduction in polymorphonuclear leukocytes: regulation of the GPLD pathway. *Shock* 2, 97-105.
Eisen, S.F., and Brown, H.A. (2002). Selective estrogen receptor (ER) modulators differentially regulate phospholipase D catalytic activity in ER-Negative breast cancer cells. *Mol. Pharmacol.* 62, 911-920.
Flanagan, J.M., Munoz-Alegre, M., Henderson, S., *et al.* (2009). Gene-body hypermethylation of ATM in peripheral blood DNA of bilateral breast cancer patients. *Hum. Mol. Genet.* 18, 1332-1342
Fuja, T.J., Lin, F., Osann, K.E., and Bryant, P.J. (2004). Somatic mutations and altered expression of the candidate tumor suppressors CSNK1 epsilon, DLG1, and EDD/hHYD in mammary ductal carcinoma. *Cancer Res.* 64, 942-951.
Goodale, D., Phay, C., Brown, W., Gray-Statchuk, L., Furlong, P., Lock, M., Chin-Yee, I., Keeney, M., and Allan, A.L. (2009). Flow cytometric assessment of monocyte activation markers and circulating endothelial cells in patients with localized or metastatic breast cancer. *Cytometry B Clin. Cytom.* 76, 107-117.
Greco, S., Elia, M.G., Muscella, A., Romano, S., Storelli, C., and Marsigliante, S. (2005). Bradykinin stimulates cell proliferation through an extracellular-regulated kinase 1 and 2-dependent mechanism in breast cancer cells in primary culture. *J. Endocrinol.* 186, 291-301.
Han, C.Z., Du, L.L., Jing, J.X., Zhao, X.W., Tian, F.G., Shi, J., Tian, B.G., Liu, X.Y., and Zhang, L.J. (2008). Associations among lipids, leptin, and leptin receptor gene Gin223Arg polymorphisms and breast cancer in China. *Biol. Trace. Elem. Res.* 126, 38-48.
Han, G., Fan, B., Zhang, Y., Zhou, X., Wang, Y., Dong, H., Wei, Y., Sun, S., Hu, M., Zhang, J., and Wei, L. (2008). Positive regulation of migration and invasion by vasodilator-stimulated phosphoprotein via Rac1 pathway in human breast cancer cells. *Oncol. Rep.* 20, 929-938.
Hassan, M.I., Waheed, A., Yadav, S., Singh, T.P., and Ahmad, F. (2009). Prolactin inducible protein in cancer, fertility and immunoregulation: structure, function and its clinical implications. *Cell. Mol. Life Sci.* 66, 447-459.
Hollestelle, A., Elstrodt, F., Nagel, J.H., Kallemeijn, W.W., and Schutte, M. (2007). Phosphatidylinositol-3-OH kinase or RAS pathway mutations in human breast cancer cell lines. *Mol. Cancer Res.* 5, 195-201.
Reubi, J.C. (2003). Peptide receptors as molecular targets for cancer diagnosis and therapy. *Endocr. Rev.* 24, 389-427.
Hill, J.J., Tremblay, T.L., Cantin, C., O'Connor-McCourt, M., Kelly, J.F. and Lenferink, A.E. (2009). Glycoproteomic analysis of two mouse mammary cell lines during transforming growth factor (TGF)-β induced epithelial to mesenchymal transition. *Proteome Sci.* 7, 2.
Kabbage, M., Chahed, K., Hamrita, B., Guillier, C.L., Trimeche, M., Remadi, S., Hoebeke, J., and Chouchane, L. (2008). Protein alterations in infiltrating ductal carcinomas of the breast as detected by nonequilibrium pH gradient electrophoresis and mass spectrometry. *J. Biomed. Biotechnol.* 2008, 564127.
Kulka, J., Szász, A.M., Németh, Z., Madaras, L., Schaff, Z., Molnár, I.A., and Tokés, A.M. (2009). Expression of Tight Junction Protein Claudin-4 in Basal-Like Breast Carcinomas. *Pathol. Oncol. Res.* 15, 59-64.
Lane, J., Martin, T.A., Mansel, R.E., and Jiang, W.G. (2008). The expression and prognostic value of the guanine nucleotide exchange factors (GEFs) Trio, Vav1 and TIAM-1 in human breast cancer. *Int. Semin. Surg. Oncol.* 5, 23.
Lee, S.G., Hur, J.U., and Kim, Y.S. (2004). A graph-theoretic modeling on GO space for biological interpretation of gene clusters. *Bioinformatics* 20, 381-388.
Lee, S.W., Reimer, C.L., Fang, L., Iruela-Arispe, M.L., and Aaronson, S.A. (2000). Overexpression of kinase-associated phosphatase (KAP) in breast and prostate cancer and inhibition of the transformed phenotype by antisense KAP expression. *Mol. Cell Biol.* 20, 1723-1732.
Lin, X., Duan, X., Liang, Y.Y., Su, Y., *et al.* (2006). PPM1A Functions as a Smad Phosphatase to Terminate TGFβ Signaling. *Cell* 125, 915-928.
Muller, A., Homey, B., Soto, H., Ge, N., Catron, D., Buchanan, M.E., McClanahan, T., Murphy, E., Yuan, W., Wagner, S.N., Barrera, J.L., Mohar, A., Verástegul, E., and Zlotnik, A. (2001). Involvement of chemokine receptors in breast cancer metastasis. *Nature* 410, 50-56.
Murabito, J.M., Rosenberg, C.L., Finger, D., Kreger, B.E., Levy, D., Splansky, G.L., Antman, K., and Hwang, S.J. (2007). A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. *BMC Med. Genet.* 8(Suppl 1), S6
Naderi, A., Teschendorff, A.E., Barbosa-Morais, N.L., *et al.*

(2007). A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene* 26, 1507-1516

Naor, Z. (2009). Signaling by G-protein-coupled receptor (GPCR): Studies on the GnRH receptor. Front. *Neuroendocrinol.* 30, 10-29.

Pennanen, P.T., Sarvilinna, N.S., Purmonen, S.R., and Ylikomi, T.J. (2009). Changes in protein tyrosine phosphatase type IVA member 1 and zinc finger protein 36 C3H type-like 1 expression demonstrate altered estrogen and progestin effect in medroxyprogesterone acetate-resistant and estrogen-independent breast cancer cell models. *Steroids* 74, 404-409.

Pentecost, B.T. (1998). Expression and estrogen regulation of the HEM45 mRNA in human tumor lines and in the rat uterus. *J. Steroid Biochem. Mol. Biol.* 64, 25-33.

Reis-Filho, J.S., Steele, D., Di Palma, S., Jones, R.L., Savage, K., James, M., Milanezi, F., Schmitt, F.C., and Ashworth, A. (2006). Distribution and significance of nerve growth factor receptor (NGFR/p75NTR) in normal, benign and malignant breast tissue. *Mod. Pathol.* 19, 307-319.

Dorsam, R.T., and Gutkind, J.S. (2007). G-protein-coupled receptors and cancer. *Nat. Rev. Cancer* 7, 79-94.

Rody, A., Holtrich, U., Gaetje, R., Gehrmann, M., Engels, K., von Minckwitz, G., Loibl, S., Diallo-Danebrock, R., Ruckhäberle, E., Metzler, D., Ahr, A., Solbach, C., Karn, T., and Kaufmann, M. (2007). Poor outcome in estrogen receptor-positive breast cancers predicted by loss of plexin B1. *Clin. Cancer Res.* 13, 1115-1122.

Rouvier, E., Luciani, M.F., Mattéi, M.G., Denizot, F., and Golstein, P. (1993). CTLA-8, cloned from an activated T cell, bearing AU-rich messenger RNA instability sequences, and homologous to a herpesvirus saimiri gene. *J. Immunol.* 150, 5445-5456.

Salaun, B., Coste, I., Rissoan, M.C., Lebecque, S.J., and Renno, T. (2006). TLR3 can directly trigger apoptosis in human cancer cells. *J. Immunol.* 176, 4894-4901.

Turashvili, G.., Bouchal, J., Baumforth, K., Wei, W., *et al.* (2007). Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microsection and microarray analysis. *BMC Cancer* 7, 55.

Venter, D.J., Ramus, S.J., Hammet, F.M., de Silva, M., Hutchins, A.M., Petrovic, V., Price, G., and Armes, J.E. (2005). Complex CGH alterations on chromosome arm 8p at candidate tumor suppressor gene loci in breast cancer cell lines. *Cancer Genet. Cytogenet.* 160, 134-140.

Wang, Z., Shen, D., Parsons, D.W., *et al.* (2004). Mutational analysis of the tyrosine phosphatome in colorectal cancers. *Science* 304, 1164-1166.

Williams, K.J., Cowen, R.L., and Stratford, I.J. (2001). Hypoxia and oxidative stress in breast cancer: Tumour hypoxia - herapeutic considerations. *Breast Cancer Res.* 3, 328-331.

Qiu, X., Zhu, X., Zhang, L., Mao, Y., *et al.* (2003). Human epithelial cancers secrete immunoglobulin g with unidentified specificity to promote growth and survival of tumor cells. *Cancer Res.* 63, 6488-6495.

Yang, Z.Q., Streicher, K.L., Ray, M.E., Abrams, J., and Ethier, S.P. (2006). Multiple interacting oncogenes on the 8p11-p12 amplicon in human breast cancer. *Cancer Res.* 66, 11632-11643.

Zhou, P., Zhi, X., Zhou, T., Chen, S., Li, X., Wang, L., Yin, L., Shao, Z., and Ou, Z. (2007). Overexpression of Ecto-5'-Nucleotidase (CD73) promotes T-47D human breast cancer cells invasion and adhesion to extracellular matrix. *Cancer Biol. Ther.* 6, 426-431.