

Metabolic Pathways Associated with Kimchi, a Traditional Korean Food, Based on *In Silico* Modeling of Published Data

Ga Hee Shin¹, Byeong-Chul Kang¹, Dai Ja Jang^{2*}

¹Data Science Center, Insilicogen, Inc., Yongin 16954, Korea,

²Processing Technology Research Group, Korea Food Research Institute, Seongnam 13539, Korea

Kimchi is a traditional Korean food prepared by fermenting vegetables, such as Chinese cabbage and radishes, which are seasoned with various ingredients, including red pepper powder, garlic, ginger, green onion, fermented seafood (*Jeotgal*), and salt. The various unique microorganisms and bioactive components in kimchi show antioxidant activity and have been associated with an enhanced immune response, as well as anti-cancer and anti-diabetic effects. Red pepper inhibits decay due to microorganisms and prevents food from spoiling. The vast amount of biological information generated by academic and industrial research groups is reflected in a rapidly growing body of scientific literature and expanding data resources. However, the genome, biological pathway, and related disease data are insufficient to explain the health benefits of kimchi because of the varied and heterogeneous data types. Therefore, we have constructed an appropriate semantic data model based on an integrated food knowledge database and analyzed the functional and biological processes associated with kimchi *in silico*. This complex semantic network of several entities and connections was generalized to answer complex questions, and we demonstrated how specific disease pathways are related to kimchi consumption.

Keywords: computer simulation, metabolism, semantic modeling, traditional Korean food

Introduction

Kimchi is a salty, fermented preparation of cabbage that is widely consumed in the traditional Korean diet. Certain types of kimchi are prepared by mixing salted cabbage with spicy ingredients, such as red pepper powder, garlic, and ginger, while others are prepared without red chili pepper or are soaked in a savory liquid. Kimchi is prepared by mixing cabbage well with other vegetables, fish seasoning, and salt, and packing it into jars. The various unique microorganisms [1] and bioactive components present in kimchi have shown antioxidant activity and have been associated with an enhanced immune response, as well as anti-cancer and anti-diabetic effects. For example, components produced during the fermentation process show anti-cancer effects via the inhibition of the *Ras* oncogene signaling pathway [2], and anti-atherosclerotic effects are seen via a reduction in low-density lipoprotein-cholesterol oxidation [3-6]. During

the fermentation process, bacteria in the environment, predominantly from the genera *Lactobacilli*, *Leuconostoc*, and *Weissella*, produce byproducts that contribute to the pungent flavor of kimchi and include beneficial bioactive compounds [7, 8].

In particular, the red chili pepper, the component of kimchi that makes it spicy, is a potent antioxidant due to its high vitamin C levels and capsaicin content. Furthermore, red pepper inhibits the decay of microorganisms and prevents food from spoiling. The physiological effects of kimchi thus originate from bioactive compounds present in its ingredients, and the functions of these are amplified by the fermentation process. Therefore, the health-promoting, probiotic, and functional properties of kimchi can be optimized by manipulating the types and amounts of ingredients, and by using appropriate probiotic starters and kimchi preparation methods such as fermentation. Optimally produced kimchi can be a very healthy food [9].

Received June 8, 2016; Revised November 23, 2016; Accepted December 27, 2016

*Corresponding author: Tel: +82-31-780-9172, Fax: +82-31-780-9170, E-mail: djjang@kfri.re.kr

Copyright © 2016 by the Korea Genome Organization

© It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>).

However, the available genome, biological pathway, and related disease data are still insufficient to explain the health benefits of kimchi. An integrated database of these data is not available, because it is difficult to control for the heterogeneous data types. Furthermore, biological processes are interrelated on many levels and their regulation presents complexity. These interactions must be understood in detail to successfully manage risks in the development of biological products. Evidence-based information on kimchi should be provided by both public and proprietary information sources, including public databases, experimental results from individual studies, text-mining analyses, and knowledge management systems. The model integrating this data would represent the accumulated knowledge in this domain that could be browsed and mined interactively. In this study, we have constructed the proper semantic data model and extracted integrated information on kimchi's nutritional content, and its effects on physiology and disease, based on analyses of the functional and biological characteristics of kimchi in the scientific literature.

Methods

Semantic modeling

We used a semantic tool for knowledge base management, BioXM (Biomax Information AG, Munchen, Germany), which was developed for object-oriented semantic integration. In this approach, semantically identical objects and the associations between them are identified and mapped based on data and descriptive meta-information [10]. This primarily involves the mapping of biological objects and descriptive data from literature to common objects like genes, diseases, or chemical. Associations between entities are mapped as relations (e.g., chemical X is associated with disease A) and object relation information is contextually structured (e.g., gene B is regulated negatively/positively by chemical X). Using objects as nodes and relations as edges, a “semantic network” can be generated that provides information on the connections between participating objects. Various elementary concepts, such as elements, relations, and annotations, are defined in Fig. 1 and the study workflow is illustrated in Fig. 2. We modified this complex semantic

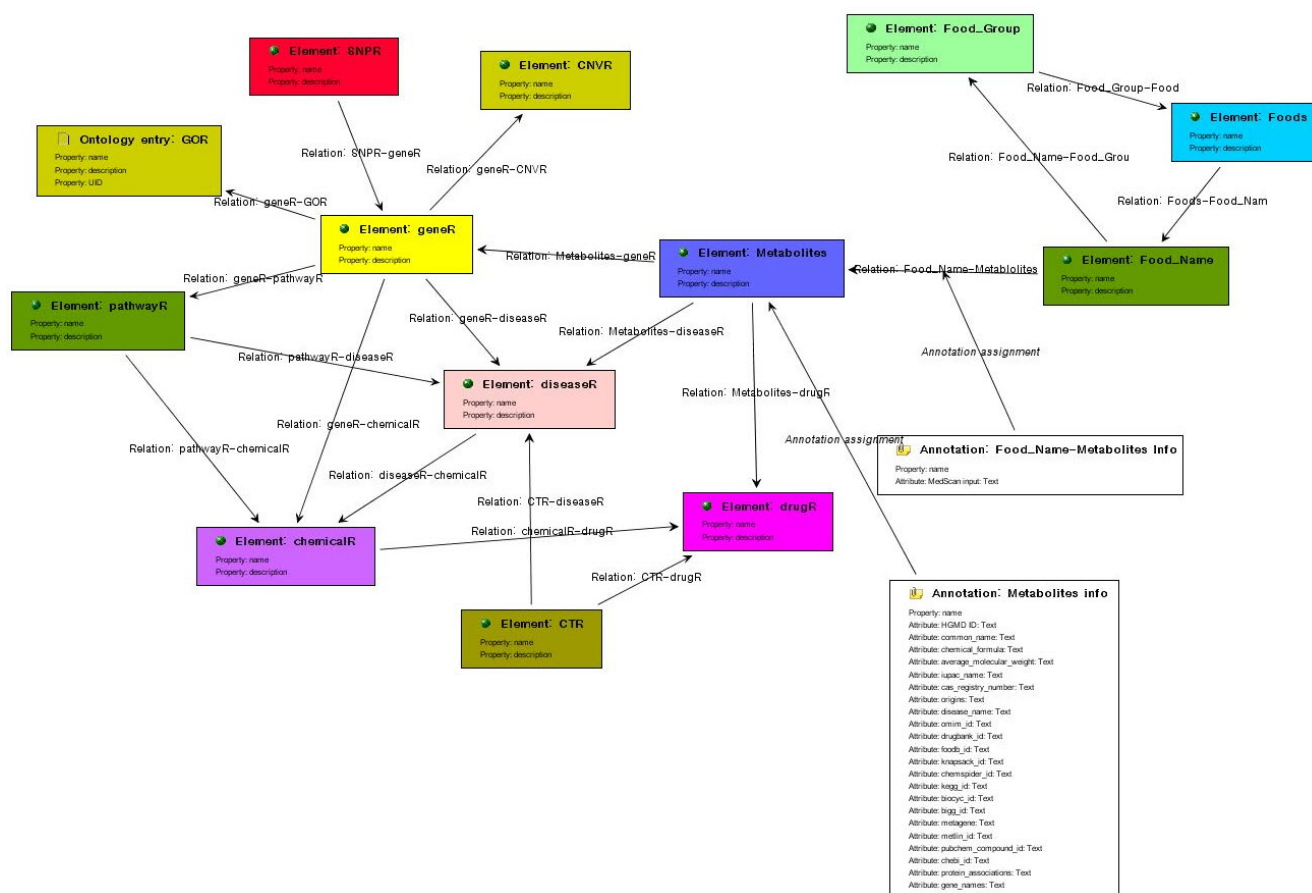


Fig. 1. Korean Food Standard Reference semantic modeling. Semantic data modeling for integration of the biological relationships among various semantic elements, such as foods, nutrients, genes, chemicals, diseases, and pathways, derived from public databases.

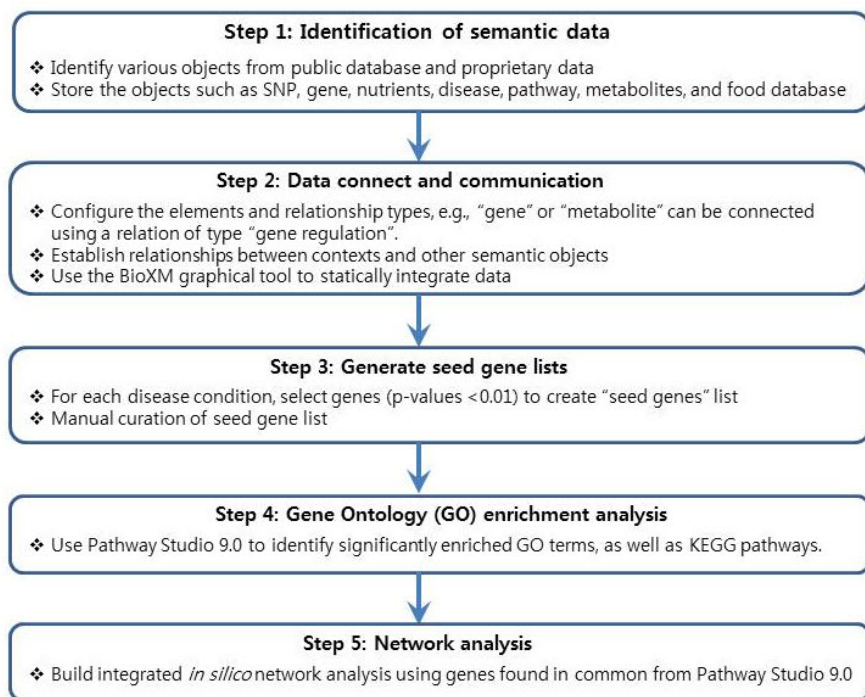


Fig. 2. The study workflow. SNP, single nucleotide polymorphism; KEGG, Kyoto Encyclopedia of Genes and Genomes.

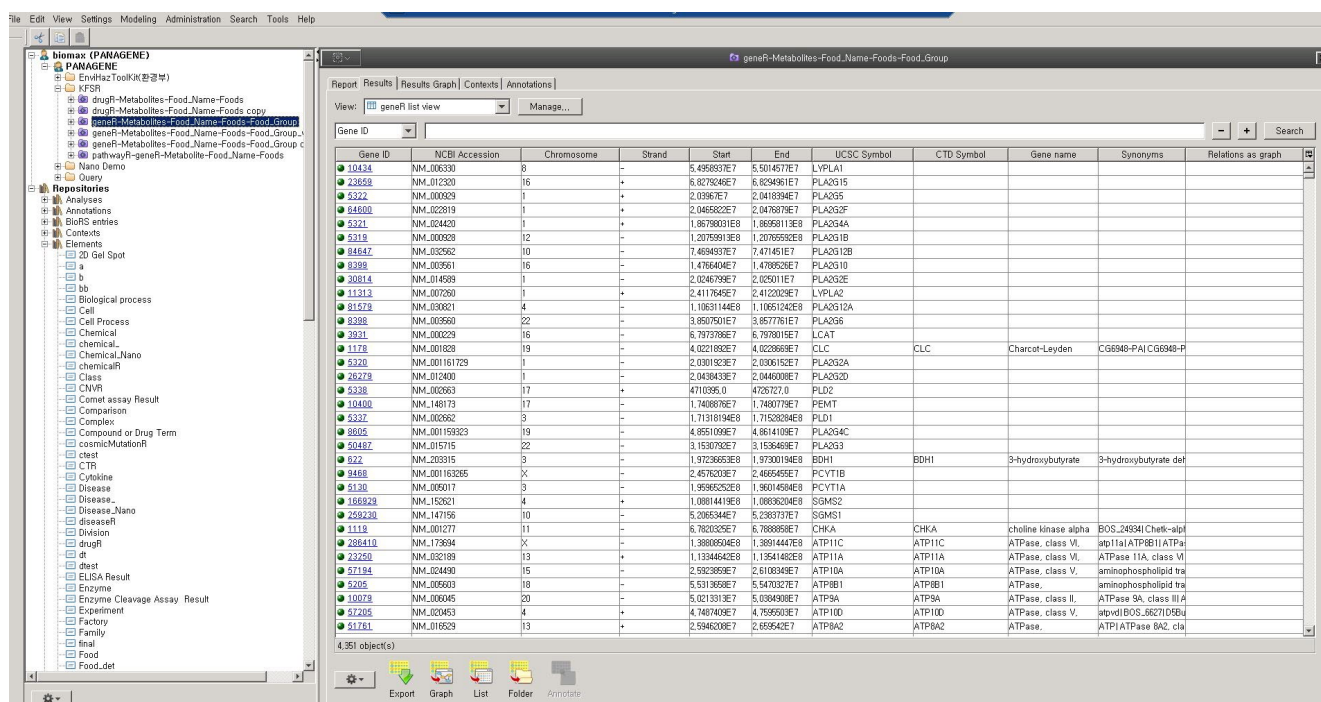


Fig. 3. Kimchi-related genes determined by semantic modeling of data in the Korean Food Standard Reference (KFSR). A total of 4,351 genes were extracted from the KFSR. The blue box indicates a query on how genes are extracted by semantic modeling.

network to detect connections, extract patterns, and answer complex questions, as is shown in Fig. 3. To access the data easily, we used the BioXM tool, which provides visual browsing of the knowledge network (Fig. 3), as well as data

mapping and query wizards (Fig. 4) that automatically adapt to the changes of a given data model. This system enabled us to create knowledge networks with flexible workflows for handling data and information types.

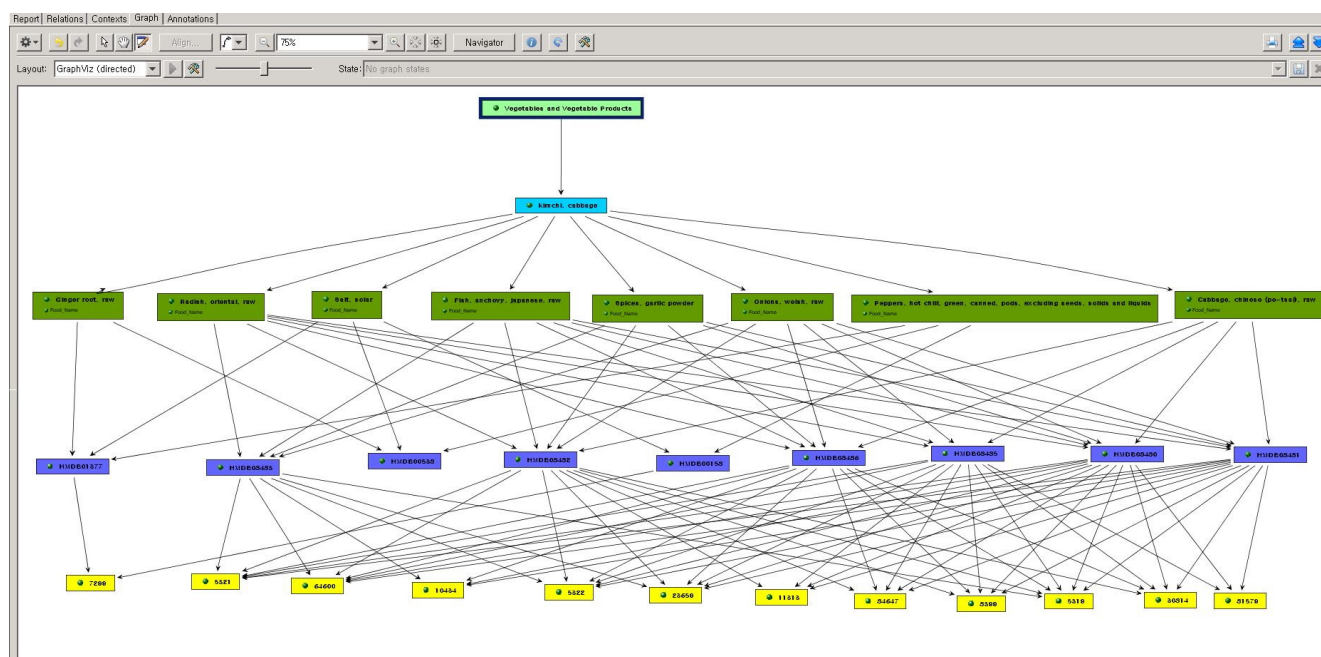


Fig. 4. *In silico* metabolic pathway of kimchi. Visualization of the knowledge network retrieved by querying genes and metabolites that are related to kimchi.

Table 1. Constructed database states

Semantic object	Description	Example
Element	Represents a basic unit of a knowledge model	The "Foods" element type can be used to create the "Kimchi" element; the "Metabolites" element type can be used to create the "Capsaicin" metabolite term element.
Relation	Describes a relation between semantic objects	"Foods-Metabolites" relation; "Foods" are associated with "Metabolites" relation
Annotation	Extends the properties of a semantic object by a set of attributes	Gene Ontology ID, KEGG ID, HMGD ID, Disease Name, Drug ID, Omin ID Pathway, Chemical ID
Ontology	Classifies semantic objects according to a defined hierarchical nomenclature of concepts	Gene Ontology to classify biological functions
Context	Represents sets of semantic objects	Metabolic pathways, protein complexes, or disease

Data collection by text-mining and data analysis

Pathway Studio software 9.0 (Elsevier, Atlanta, GA, USA) was used to collect information on biological processes and diseases related to kimchi, based on information in the Korean Food Standard Reference (KFSR). Pathway Studio software allows the automatic extraction of regulatory and physical interactions from MEDLINE abstracts using a natural-language-processing technology called MedScan [11]. Interactions extracted by MedScan, which contain a formalized set of relationships, are imported into the Pathway Studio database and analyzed further using data-mining tools for knowledge inference [12]. Since MedScan keeps the reference of the original article with information on the extracted interaction, it also allows a

quick review of extracted facts and identification of relevant publications. Thus, it facilitated our pathway reconstruction by selecting the appropriate interactions. All interactions found were manually curated, and only validated interactions were included in the pathways.

Results

Semantic data modeling

Elements and relations were maintained and linked by relation and object. This semantic model enabled elements to be mapped to known elements. Relations between biological data are used to efficiently extract interactions of interest. The semantic modeling in this study was configured in the generic BioXM knowledge management en-

vironment to create a knowledge base for this translational systems biology approach. We constructed semantic networks for the discovery of a network biomarker using BioXM software (Biomax Information AG), which is a customizable knowledge base management tool for scientific data. The graphical data model is shown in Fig. 1. As shown in Table 1, we created elements such as “Foods,” “Foods Group,” “Food Name,” “Metabolites,” and nine other elements to describe the available biological knowledge. Information on the interconnection of elements is captured in relations such as “is related to” for “Foods” or “regulates” for a metabolite that regulates expression of a gene. This network consisted of 13 elements and 20 relations. The semantic network provided information on connections between included objects instantly. Therefore, this integrated semantic data model was used to create new knowledge networks with flexible workflows.

Data configuration

This semantic model enabled us to integrate existing public databases and experimental data derived from the literature. To populate the knowledge base with data from public databases, we manually generated mappings in import-templates. A graphical wizard for import-template generation provided a selection of possible import options and mappings.

Construction of the KFSR

The United States Department of Agriculture (USDA) food and nutrient databases provide the basic infrastructure for food and nutrition research, nutrition monitoring, policy, and dietary practice. The databases date back to 1892 and they are unique, as the only databases available in the public domain that perform these functions. The Nutrient Data Laboratory develops and maintains the USDA Standard Reference and related data products [13]. The Standard Reference is the major source of food composition data in the United States. It provides the foundation for most food composition databases used in food policy, research, dietary practice, and nutrition monitoring. The USDA databases were used to construct the KFSR. This product is the basis of numerous other applications that address nutrients in foods. The database contains information on agricultural commodities, as well as data on formulated and processed foods. Currently, the database contains values for up to 140 different food components, including nitrogen (protein), total fat, carbohydrates, and moisture, as well as individual fatty acids, amino acids, vitamins, and minerals in >7,500 foods. We have added five traditional Korean foods, i.e., kimchi, bulgogi, bibimbap, doenjang, and gochujang. To construct a useful KFSR, we integrated 11 public databases as shown in Table 2, with a total of 26,848,859 records.

Table 2. Public data integrated into the BioXM knowledge base for the Korean Foods Standard Reference (KFSR)

Entity	Record No.	Entity data source	Relation	Record No.
Foods	5	Korean foods	Food_Group-Foods	5
Food_Name	7,914	USDA (http://www.usda.gov)	Foods-Food_Name	31
Food_Group	27	USDA (http://www.usda.gov)	Food_Name-Food_Group	7,909
Metabolites	40,286	HMDB (http://www.hmdb.ca)	Food_Name-Metabolites	12,697
geneR	23,252	UCSC (http://www.genome.ucsc.edu)	Metabolites-drugR	1,575
CNVR	20,052	Cancer genome (http://cancergenome.nih.gov)	Metabolites-diseaseR	1,019
diseaseR	9,648	CTD (http://ctdbase.org)	Metabolites-geneR	327,973
chemicalR	144,435	CTD (http://ctdbase.org)	geneR-CNVR	21,591
pathwayR	362	CTD (http://ctdbase.org)	geneR-chemicalR	308,405
CTR (clinical trials)	1,273	CT (http://clinicaltrials.gov)	geneR-pathwayR	60,057
drugR	6,712	DB (http://drugbank.ca)	chemicalR-drugR	1,702
GOR (Gene Ontology)	38,092	GO (http://www.geneontology.org)	CTR-drugR	1,419
SNPR	14,901,097	1000 genomes (http://www.1000genomes.org)	CTR-diseaseR	1,210
			pathwayR-chemicalR	43,139
			diseaseR-chemicalR	842,368
			geneR-SNPR	15,240,996
			geneR-diseaseR	9,780,690
Total	15,193,155		Total	26,848,859

USDA, United States Department of Agriculture; HMDB, Human Metabolome Database; UCSC, University of California, Santa Cruz; CNVR, copy number variation region; CTD, comparative toxicogenomics database; SNPR, single nucleotide polymorphism relation.

Data analysis of kimchi from the knowledge base

We found a total 4,351 genes strongly associated with kimchi's secondary metabolites using the KFSR, as shown in Fig. 3. First, 1,237 kimchi metabolites were extracted from >7,500 foods in KFSR and 4,351 genes were associated with 1,237 kimchi metabolites. Through the semantically integrated datasets, we discovered genes that were associated with kimchi metabolites. This network is shown in Fig. 4. We used the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway to classify genes by metabolic function. The genes in Table 3 are classified into six categories, and 309 of these genes were found to be involved in lipid metabolism. At first, we initiated a cardiovascular disease network analysis related to kimchi using 30 genes. Of these, 27 were directly connected by interactions such as regulation, promoter binding, direct regulation, protein modification, and miRNA regulation. Genes like *ADCY5*, *ADCY1*, *ADCY2*, *GNAS*, *CTNNT1*, *EDNI*, and *RAC1* were the most highly connected nodes with node degrees of ≥ 10 ($p < 0.001$) (Fig. 5A). Adenyl cyclases (ADCYs) are critical regulators of metabolic and cardiovascular function, and *ADCY5* is increased in hearts with pressure-overloaded left ventricular hypertrophy [14, 15]. In addition, guanine nucleotide-binding protein Gs (*GNAS*) overexpression results in increased sensitivity to apoptotic stimulation and cardiomyopathy [16], as shown in Fig. 4A. In addition, *GNAS* leads to the activation of β -catenin which is a multifunctional protein coded by the *CNNB1* gene. Although there has long been speculation on the association between kimchi and heart disease, Lee *et al.* [16] demonstrated that kimchi may not adversely affect blood pressure and cardiac function.

Kimchi is effective in enhancing immune function [3, 18, 19]; however, the scientific evidence to support this claim has been insufficient. In this work, we performed an immune disease network analysis related to kimchi, based on public information and using 39 selected genes. Twenty-three genes were directly connected to each other by interactions such as regulation, promoter binding, direct regulation, protein modification, and miRNA regulation. In particular, *PTPRC*, *LCK*, *JAK3*, *ZAP70*, and *VEGFA* were the most highly connected nodes with node degrees of ≥ 7 ($p < 0.001$) (Fig. 5B). Specifically, *PTPRC* and *JAK3* were found to regulate innate immune responses [20-22]. This analytic information database, the KFSR, may suggest ways to analyze and more effectively use existing scientific data.

Discussion

For a long time, many nutritionists have believed, based

on metabolomics data, that well-fermented kimchi has numerous anti-biotic activities [23, 24]. Recently, this has led to a significant increase in kimchi export. Kimchi has gradually gained popularity, even among foreigners, and it has become a globally recognized food. However, the genome, biological pathway, and related disease data are still

Table 3. Functional classifications of genes extracted by Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway map IDs

KEGG pathway map	No. of genes
1. Metabolism	
1.1 Carbohydrate metabolism	283
1.2 Energy metabolism	109
1.3 Lipid metabolism	309
1.4 Nucleotide metabolism	154
1.5 Amino acid metabolism	229
1.6 Metabolism of other amino acids	99
1.7 Glycan biosynthesis and metabolism	104
1.8 Metabolism of cofactors and vitamins	170
1.9 Metabolism of terpenoids and polyketides	16
1.10 Biosynthesis of other secondary metabolites	11
1.11 Xenobiotic biodegradation and metabolism	102
2. Genetic information processing	
2.1 Transcription	23
2.2 Translation	95
2.3 Folding, sorting, and degradation	108
2.4 Replication and repair	34
3. Environmental information processing	
3.1 Membrane transport	16
3.2 Signal transduction	429
3.3 Signaling molecules and interactions	127
4. Cellular Process	
4.1 Transport and catabolism	221
4.2 Cell motility	89
4.3 Cell growth and death	93
4.4 Cell communication	168
5. Organismal systems	
5.1 Immune system	220
5.2 Endocrine system	210
5.3 Circulatory system	117
5.4 Digestive system	207
5.5 Excretory system	89
5.6 Nervous system	291
5.7 Sensory system	41
5.8 Development	95
5.9 Environmental adaptation	84
6. Human disease	
6.1 Cancers	225
6.2 Immune diseases	39
6.3 Neurodegenerative diseases	154
6.4 Substance dependence	136
6.6 Cardiovascular diseases	30
6.7 Endocrine and metabolic diseases	25
6.8 Infectious diseases	298

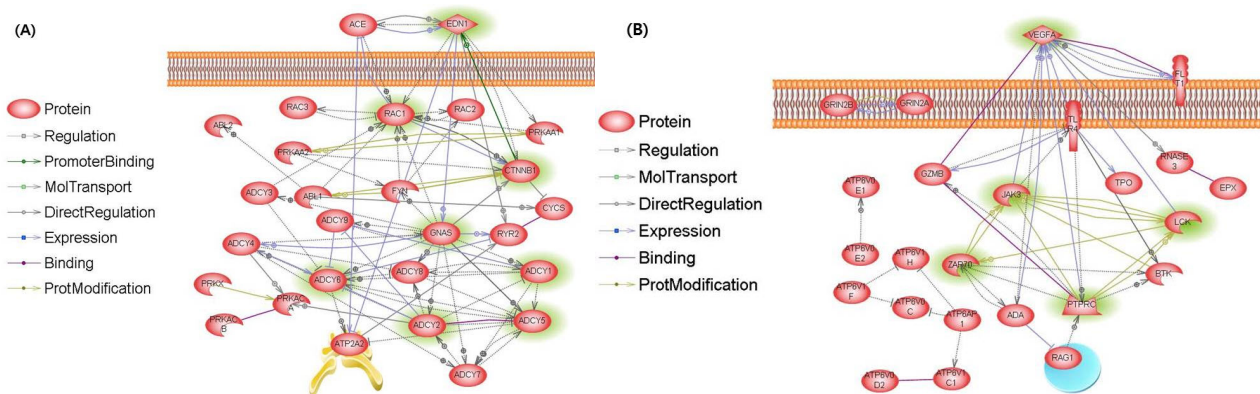


Fig. 5. The relationship of kimchi to a disease network. (A) Cardiovascular disease. (B) Immune disease. The genes associated with kimchi nutrients are highlighted in green.

insufficient to explain the health benefits of kimchi. Therefore, we investigated, *in silico*, a kimchi network based on public information using network-based systems biology tools, including BioXM and Pathway Studio. The *in silico* network data used in our study were based on published literature. With a long-term aim to search for the metabolic pathways affected by kimchi and kimchi's effects on various diseases, we first re-analyzed the data for better comparability. Then, a number of specific biomolecular networks for kimchi were built in various ways. Genes analyzed in Pathway Studio were shown to be key players in various cellular pathways and mechanisms, such as inflammation, cell proliferation, and immunity, which could be altered by underlying causes, such as disease. Moreover, these biological processes were also related to obesity. This interpretation supports the view that there are common, underlying markers for targeted biological processes. Our results should be considered only the initial effort in understanding the interrelated characteristics of the highly complex metabolic pathway of kimchi. Experimental studies will also be needed to verify this *in silico* kimchi network.

Acknowledgments

This work was supported by a grant from the Korea Food Research Institute.

References

- Lee JS, Heo GY, Lee JW, Oh YJ, Park JA, Park YH, *et al.* Analysis of kimchi microflora using denaturing gradient gel electrophoresis. *Int J Food Microbiol* 2005;102:143-150.
- Park KY, Cho EJ, Rhee SH, Jung KO, Yi SJ, Jhun BH. Kimchi and an active component, beta-sitosterol, reduce oncogenic H-Ras(v12)-induced DNA synthesis. *J Med Food* 2003;6:151-156.
- Park KY. The nutritional evaluation, and antimutagenic and anticancer effects of kimchi. *J Korean Soc Food Sci Nutr* 1995;24:169-182.
- Islam MS, Choi H. Antidiabetic effect of Korean traditional Baechu (Chinese cabbage) kimchi in a type 2 diabetes model of rats. *J Med Food* 2009;12:292-297.
- Kim HJ, Lee JS, Chung HY, Song SH, Suh H, Noh JS, *et al.* 3-(4'-hydroxyl-3',5'-dimethoxyphenyl)propionic acid, an active principle of kimchi, inhibits development of atherosclerosis in rabbits. *J Agric Food Chem* 2007;55:10486-10492.
- Kim EK, An SY, Lee MS, Kim TH, Lee HK, Hwang WS, *et al.* Fermented kimchi reduces body weight and improves metabolic parameters in overweight and obese patients. *Nutr Res* 2011;31:436-443.
- Cheigh HS, Park KY. Biochemical, microbiological, and nutritional aspects of kimchi (Korean fermented vegetable products). *Crit Rev Food Sci Nutr* 1994;34:175-203.
- Kwon EA, Kim M. Microbial evaluation of commercially packed kimchi products. *Food Sci Biotechnol* 2007;16:615-620.
- Lee H, Yoon H, Ji Y, Kim H, Park H, Lee J, *et al.* Functional properties of *Lactobacillus* strains isolated from kimchi. *Int J Food Microbiol* 2011;145:155-161.
- Bornhövd C, Buchmann AP. A prototype for metadata-based integration of internet sources. In: Proceedings of the 11th International Conference on Advanced Information Systems Engineering; 1999 Jun 14-18; Heidelberg, Germany. Berlin: Springer, 1990.
- Novichkova S, Egorov S, Daraselia N. MedScan, a natural language processing engine for MEDLINE abstracts. *Bioinformatics* 2003;19:1699-1706.
- Yuryev A, Mulyukov Z, Kotelnikova E, Maslov S, Egorov S, Nikitin A, *et al.* Automatic pathway building in biological association networks. *BMC Bioinformatics* 2006;7:171.
- David H, Linda L, Pamela P. USDA National Nutrient Database for Standard Reference, Release 24 [Internet]. Beltsville: Nutrient Data Laboratory, USDA National Nutrient Database for Standard Reference, 2016 [cited 2016 Dec 1]. Available from: www.ars.usda.gov/nutrientdata.
- Patel TB, Du Z, Pierre S, Cartin L, Scholich K. Molecular biological approaches to unravel adenylyl cyclase signaling and

- function. *Gene* 2001;269:13-25.
15. Hu CL, Chandra R, Ge H, Pain J, Yan L, Babu G, *et al.* Adenylyl cyclase type 5 protein expression during cardiac development and stress. *Am J Physiol Heart Circ Physiol* 2009;297:H1776-H1782.
 16. Lee SM, Cho Y, Chung HK, Shin DH, Ha WK, Lee SC, *et al.* Effects of kimchi supplementation on blood pressure and cardiac hypertrophy with varying sodium content in spontaneously hypertensive rats. *Nutr Res Pract* 2012;6:315-321.
 17. Frey UH, Adamzik M, Kottenberg-Assenmacher E, Jakob H, Manthey I, Broecker-Preuss M, *et al.* A novel functional haplotype in the human GNAS gene alters Galphas expression, responsiveness to beta-adrenoceptor stimulation, and peri-operative cardiac performance. *Eur Heart J* 2009;30:1402-1410.
 18. Kim KH, Kim SH, Park KY. Effects of kimchi extracts on production of nitric oxide by activated macrophages, transforming growth factor β 1 of tumor cells and interleukin-6 in splenocytes. *J Food Sci Nutr* 2001;6:126-132.
 19. Kim MJ, Kwon MJ, Song YO, Lee EK, Youn HJ, Song YS. The effects of kimchi on hematological and immunological parameters *in vivo* and *in vitro*. *J Korean Soc Food Sci Nutr* 1997;26:1208-1214.
 20. Stark GR, Darnell JE Jr. The JAK-STAT pathway at twenty. *Immunity* 2012;36:503-514.
 21. Hermiston ML, Xu Z, Weiss A. CD45: a critical regulator of signaling thresholds in immune cells. *Annu Rev Immunol* 2003;21:107-137.
 22. Vyshkina T, Leist TP, Shugart YY, Kalman B. CD45 (PTPRC) as a candidate gene in multiple sclerosis. *Mult Scler* 2004;10:614-617.
 23. Jung JY, Lee SH, Kim JM, Park MS, Bae JW, Hahn Y, *et al.* Metagenomic analysis of kimchi, a traditional Korean fermented food. *Appl Environ Microbiol* 2011;77:2264-2274.
 24. Park KY, Jeong JK, Lee YE, Daily JW 3rd. Health benefits of kimchi (Korean fermented vegetables) as a probiotic food. *J Med Food* 2014;17:6-20.