Research Article

Identification of Novel Thyroid Cancer-Related Genes and Chemicals Using Shortest Path Algorithm

Yang Jiang,[1] Peiwei Zhang,[2] Li-Peng Li,[1] Yi-Chun He,[1] Ru-jian Gao,[1] and Yu-Fei Gao[1]

[1]*Department of Surgery, China-Japan Union Hospital of Jilin University, Changchun 130033, China*
[2]*The Key Laboratory of Stem Cell Biology, Institute of Health Sciences, Shanghai Institutes for Biological Sciences,*
 Chinese Academy of Sciences, Shanghai 200031, China

Correspondence should be addressed to Yu-Fei Gao; gaoyufei1975@sina.cn

Thyroid cancer is a typical endocrine malignancy. In the past three decades, the continued growth of its incidence has made it urgent to design effective treatments to treat this disease. To this end, it is necessary to uncover the mechanism underlying this disease. Identification of thyroid cancer-related genes and chemicals is helpful to understand the mechanism of thyroid cancer. In this study, we generalized some previous methods to discover both disease genes and chemicals. The method was based on shortest path algorithm and applied to discover novel thyroid cancer-related genes and chemicals. The analysis of the final obtained genes and chemicals suggests that some of them are crucial to the formation and development of thyroid cancer. It is indicated that the proposed method is effective for the discovery of novel disease genes and chemicals.

1. Introduction

Thyroid cancer (TC) is a typical endocrine malignancy. During the past three decades, its incidence has been nearly tripled in the whole world, such as the United States and other developed countries [1]. Thus, it has been a formidable and urgent task to uncover the mechanism behind it, thereby efficiently improving the medical treatment. Research has been focused on the findings of possible driving genes of this disease, especially those genes with high frequent mutations, over-expressions, or fusions for a long time. Until recent years, this research process just started to accelerate.

With the advent of advanced technology including the next-generation sequencing technologies, findings of genetic and epigenetic alterations are speeding up [2]. In other words, the gradual accumulation of somatic mutations and chromosomal rearrangements that are related to many crucial tumor initiation and development genes has been found [3]. For example, high prevalence of mutations and gene fusions in effectors of the PI3K-AKT and MAPK pathway occurred in most patients with TC, suggesting its important contributions to tumor initiation and development. Meanwhile,

dysregulation of hundreds of gene expressions, such as DPP4, MET, LGALS3, and TIMP1, have been common events in this disease [4]. This achievement towards the uncovering of mechanism behind TC is inspiring. However, despite the unprecedented rate of discovery of novel mutations and gene fusions in TC, evidence towards the tumor genesis of TC is still not convincing because of the still large search space.

In addition to the influence of our genomes, it is evident that cancer is also influenced by environmental chemicals from our daily lives. This is partly because environmental exposures can cause DNA mutations and change epigenetic mechanisms [5]. For example, we might contact fluoride and arsenic in drinking water, and toxic gases from burning of fuel and industrial emissions. Current studies show that outdoor air pollution and second-hand smoke often contain chemicals, such as arsenic and polycyclic aromatic hydrocarbons, which further increase risks of numerous cancers [6]. Exposure to toxic level of arsenic can significantly increase DNA methylation of p16 and p53 promoter regions [7] and change miRNA expression [8]. However, many chemicals' effects towards cancer have not been researched and illustrated. Considering the important influences of chemicals

towards cancer, we are also interested in searching for novel chemicals related to TC.

We realized that with the simple results from experiments, it would be difficult to meet up our expectation on the detection of novel genes and chemicals related to TC due to the time- and money-consuming process. Thus, more effective and rapid alternative methods must be used to assist the searching process of genes and chemicals related to TC. Considering the efficiency of computational approach, it might be a potential way, which can be used to complete this arduous searching task in a more effective and time-saving way. Until now, several computational methods have been developed in the field of biological network analysis and other related areas, such as construction and analysis of gene regulation, gene coexpression or other biological networks [9–14], and drug designs [15–21]. Recently, some computation methods were proposed to identify new candidate disease genes based on the knowledge of the known disease genes [22–25]. These methods only considered the disease genes. However, it is easy to improve their methods to identify both genes and chemicals that were related to certain disease. In this study, we generalized their methods by constructing a weighted graph containing the information of protein-protein interactions, chemical-chemical interactions, and chemical-protein interactions and applied this method to study TC. Similar to the methods in [22–25], according to known TC-related genes that were collected from TSGene Database [26], UniPort [27], and NCI (National Cancer Institute) [28] and known TC-related chemicals retrieved from CTD (Comparative Toxicogenomics Database) [29], some new candidate genes and chemicals were discovered by our method. The analysis results of these new candidate genes and chemicals showed that some of them are crucial to the formation and development of TC. We hope that this method could contribute to uncovering the mechanism of TC.

2. Materials and Methods

2.1. Materials. The TC-related genes were collected from three sources: 209 TC-related genes were achieved from UniProt (http://www.uniprot.org/) [27] after we input "human thyroid cancer reviewed" as the keywords; 16 genes were chosen in the catalogue of thyroid cancer from TSGene database (http://bioinfo.mc.vanderbilt.edu/TSGene/search.cgi) [26]; 251 TC-related genes were retrieved from NCI (https://gforge.nci.nih.gov/, released April 2009) [28]. After integrating the above 476 genes, we finally obtained 444 different TC-related genes, which were provided in Online Supporting Information (see S1 in Supplementary Material available online at http://dx.doi.org/10.1155/2015/964795).

The TC-related chemicals were retrieved from CTD (http://ctdbase.org/) [29], which included the interactions between chemicals and genes and their associations with diseases that were manually curated from 110,142 articles (http://ctdbase.org/about/dataStatus.go, accessed 2014 August). Only the 44 chemicals that were markers of TC, were therapeutic to TC, or had known mechanism in TC
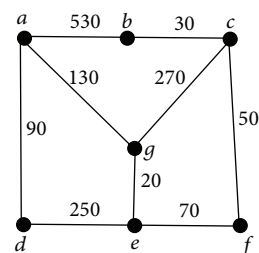
Protein or chemical	Protein or chemical	Combined_score
a	b	470
b	c	970
d	e	750
e	f	930
e	g	980
a	g	870
c	f	950
c	g	730
a	d	910

Figure 1: An example to display the construction of the weighted graph, where a, b, and c represent chemicals and d, e, f, and g represent proteins.

were analyzed. The pubchem IDs of these chemicals were also provided in Online Supporting Information S1.

2.2. A Weighted Graph Constructed from Interactions of Chemicals and Proteins. The core idea of our method is to construct a hybrid weighted graph containing the information of proteins, chemicals, and their associations. This idea has been applied to our previous study on assigning chemicals and enzymes to metabolic pathway [30]. To do that, we employed the information of protein-protein interactions, chemical-chemical interactions, and chemical-protein interactions.

The information concerning protein-protein interaction was retrieved from STRING (Search Tool for the Retrieval of Interacting Genes/Proteins, version 9.1, http://string.embl.de/) [31], a large scale database containing direct (physical) and indirect (functional) interactions of proteins, which are derived from genomic context, high-throughput experiments, (conserved) coexpression, or previous knowledge (refer to http://string.embl.de/). Some computational models have been built based on these information [32–35]. Each obtained interaction contains two proteins and one score, which measures the strength of the interaction, that is, the likelihood of the interaction's occurrence. For latter formulation, let us denote the score of the interaction between proteins p_1 and p_2 by $S_{pp}(p_1, p_2)$. In particular, if proteins p_1 and p_2 cannot comprise an interaction according to the current data in STRING, $S_{pp}(p_1, p_2)$ was set to zero.

The information concerning chemical-chemical interaction and chemical-protein interaction was retrieved from STITCH (Search tool for interactions of chemicals, version 4.0, http://stitch.embl.de/) [36], a sister project of STRING which provides the known and predicted interactions of chemicals and proteins. These interactions are confirmed by evidence derived from experiments, databases, and the
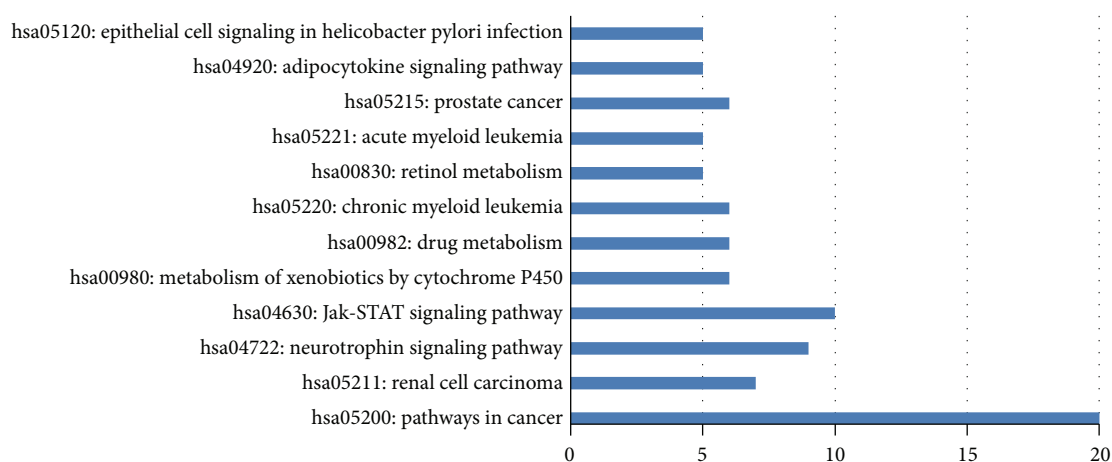
FIGURE 2: The top twelve KEGG pathways that were enriched by 169 significant candidate genes.
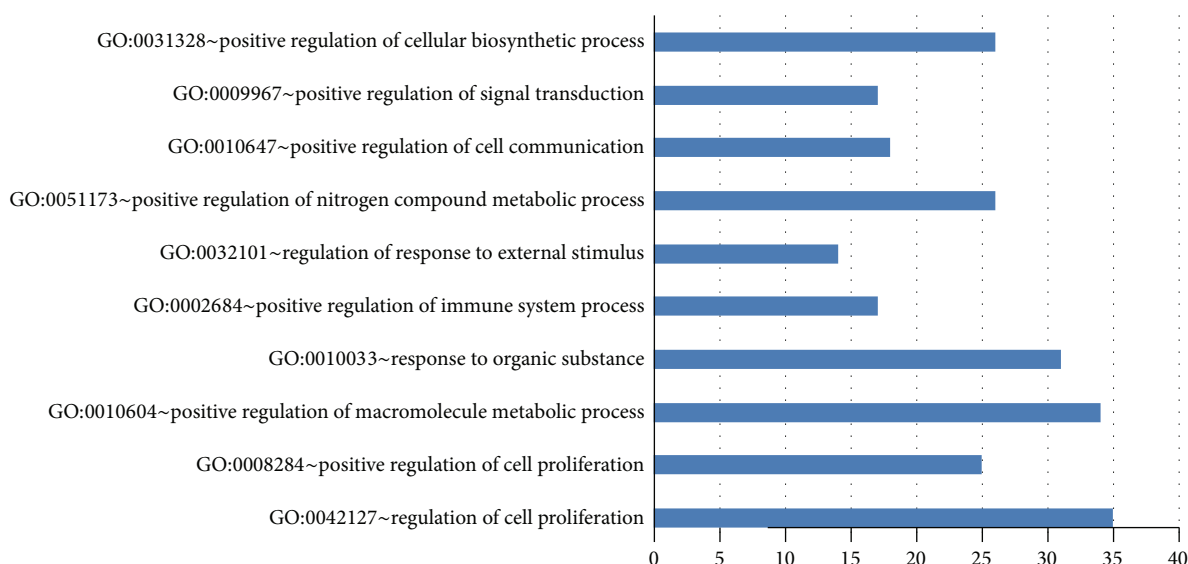

FIGURE 3: The top ten GO terms that were enriched by 169 significant candidate genes.

literature (refer to http://stitch.embl.de/). Each obtained chemical-chemical interaction contains two chemicals and five scores: "Similarity," "Experiment," "Database," "Textmining," and "Combined_score," which measure the strength of the interaction from different aspects, such as their structures, activities, reactions, cooccurrence in literature, and integration of the above information. To widely indicate the interaction between chemicals, we used the last score, that is, "Combined_score," to measure the strength of the interaction. For two chemicals c_1 and c_2, the "Combined_score" of the interaction between them was denoted by $S_{cc}(c_1, c_2)$. Similarly, $S_{cc}(c_1, c_2)$ was set to zero if c_1 and c_2 do not occur as an interaction in STITCH. Each obtained chemical-protein interaction contains one chemical, one protein, and five scores. With the similar argument, we used the "Combined_score" to indicate the strength of the interaction between one chemical and one protein. Let $S_{cp}(c, p)$ denote the "Combined_score" of the interaction between chemical c

and protein p. Also, $S_{cp}(c, p) = 0$ if c and p cannot comprise a chemical-protein interaction. It is necessary to point out that all chemicals in the retrieved chemical-chemical and chemical-protein interactions must have records in KEGG (Kyoto Encyclopedia of Genes and Genomes) [37] because the number of chemicals in STITCH is too large and most of chemicals have few associations with human tissues.

Based on the information concerning protein-protein interactions, chemical-chemical interactions, and chemical-protein interactions, a weighted graph $G = (V, E)$ was constructed as follows: V contained all proteins and chemicals occurring in the above three kinds of information and E consisted of all pairs of nodes such that the corresponding proteins or chemicals can comprise an interaction. It is easy to know that each edge in G represented an interaction. On the other hand, as mentioned in the above paragraph, each interaction was assigned a score to indicate its strength; that is, different interactions may have different strength. To note

this fact in G and use the shortest path algorithm to search for new candidate genes and chemicals, each edge was labeled a weight as follows. Since the range of the interaction scores is between 1 and 999, the weight of an edge e with endpoints n_1 and n_2 was defined by

$$
w(e)
$$
$$
= \begin{cases} 1000 - S_{pp}(p_1, p_2) \\ \quad \text{If } n_1 \text{ and } n_2 \text{ represented proteins } p_1 \text{ and } p_2 \\ 1000 - S_{cp}(c, p) \\ \quad \text{If } n_1 \text{ and } n_2 \text{ represented chemical } c \text{ and protein } p \\ 1000 - S_{cc}(c_1, c_2) \\ \quad \text{If } n_1 \text{ and } n_2 \text{ represented chemicals } c_1 \text{ and } c_2. \end{cases}
$$
$$(1)$$

To clearly display the procedures of construction of the graph, a small example is shown in Figure 1. In the example, there are three chemicals a, b, and c and four proteins d, e, f, and g. The interactions, including their "Combined_score," between them are listed in the table of Figure 1 and the constructed graph based on these interactions is shown at the top of Figure 1.

2.3. Method for Discovery of New Candidate Genes and Chemicals. The following method for finding new candidate TC-related genes and chemicals was almost same as that in our previous study [25]. The only difference was that the input of the current method contained both genes and chemicals, while the method in [25] only considered genes. Readers can refer to our previous study [25] for the detailed procedures of the method and its principle. For the integrity of this study, the brief description of the method was as follows: (I) search all shortest paths connecting any pair of TC-related genes and chemicals using Dijkstra's algorithm [38]; (II) for each node (gene or chemical) in G, count its betweenness that was defined as the number of paths containing it as an inner node; (III) select the nodes (genes or chemicals) with betweenness larger than zero as candidate genes and chemicals; (IV) produce 1,000 sets by randomly selecting nodes (genes or chemicals) from G; the numbers of genes and chemicals in each set were the same as those in known TC-related gene and chemical set; (V) for each set, search all shortest paths connecting any pair of genes or chemicals in G; (VI) count the betweenness of candidate genes and chemicals on each randomly produced sets; (VII) for each candidate gene and chemical, compare its betweenness on known TC-related gene and chemical set and those on randomly produced sets, thereby calculating its permutation FDR that was defined as "the number of randomly produced sets on which the betweenness was larger than that on the known TC-related gene and chemical set"/1000.

3. Results and Discussions

3.1. Candidate Genes and Chemicals. Of the 444 TC-related genes and 44 TC-related chemicals, we searched the shortest paths in G such that the endpoints of them were TC-related genes or TC-related chemicals. Accordingly, the betweenness of each gene and chemical in G was computed, obtaining 636 candidate genes and 174 candidate chemicals whose betweenness was larger than zero; that is, these genes and chemicals occurred in at least one shortest path as inner nodes. These genes and chemicals are listed in Online Supporting Information S2, in which their betweenness is also provided.

To exclude false discoveries, the permutation test was executed by constructing 1,000 randomly selected gene and chemical sets for calculating the permutation FDR of each candidate gene and chemical, which is also provided in Online Supporting Information S2. Then, we selected 0.05 as a threshold to exclude false discoveries, obtaining 169 candidate genes and 49 candidate chemicals with permutation FDRs smaller than 0.05. The information of these genes and chemicals is available in Online Supporting Information S3. For convenience, we termed these genes and chemicals as significant candidate genes and significant candidate chemicals, respectively. The following discussion was based on these significant candidate genes and significant candidate chemicals.

3.2. Gene Enrichment Analysis. DAVID [39] is a powerful tool that could be used to make integrative and systematic of large gene lists. Thus, it was used in this study to analyze the 169 significant candidate genes. The analysis results included two parts: KEGG pathway enrichments (Online Supporting Information S4) and gene ontology (GO) enrichments (Online Supporting Information S5). GO enrichments include three parts: biological process enrichment (BP enrichment), cellular component enrichment (CC enrichment), and molecular function enrichment (MF enrichment). Since our method was mainly based on protein-protein interactions, BP enrichment analysis was more convincing, while other two results were not very reasonable. Thus, we only gave the discussion based on the BP enrichment.

For the KEGG pathway enrichment analysis results, 169 candidate genes are enriched in 19 KEGG pathways (see Online Supporting Information S4). Among these 19 KEGG pathways, twelve of them were with P value (modified Fisher exact P value) less than 0.05. Figure 2 shows these twelve KEGG pathways and the number of enriched genes among the significant candidate genes ("count"). Hsa05200 (pathways in cancer, "count" = 20) is the most significant pathway, which enriched 20 significant candidate genes, such as FGFR2, FGF6, DVL3, EPAS1, and PPARG. Since all these genes enriched in this pathway were reported related to cancer formation and development, it further revealed the validity of our method. Hsa05211 (renal cell carcinoma, "count" = 7) is the second significant pathway with 7 genes related to renal cell carcinoma. Hsa04722 (neurotrophin signaling pathway, "count" = 7) is the third significant pathway, enriching 7 genes, such as KRAS, PLCG1, and NTF3. Among

them, NTF3 in neurotrophin signaling pathway has been reported with the association to cancer [40]. Other pathways, such as hsa05221 (acute myeloid leukemia, "count" = 5) and hsa05215 (prostate cancer, "count" = 6), also revealed that the significant candidate genes are associated with cancer.

For the BP enrichment analysis, results are shown in Online Supporting Information S5. Ranked by P value, top ten BP GO terms are depicted in Figure 3. The mainly enriched GO terms are associated with cell proliferation. For example, genes in GO:0042127 (regulation of cell proliferation, "count" = 35) and GO:0008284 (positive regulation of cell proliferation, "count" = 25) are all reported related to cell proliferation. Also, GO:0010604 (positive regulation of macromolecule metabolic process, "count" = 34) and GO:0051173 (positive regulation of nitrogen compound metabolic process, "count" = 26) are associated with metabolic process. Since proliferative signaling and activating metastasis are two hallmarks of cancer [41], it is convincing that the result of BP enrichment analysis further supports the validity of our method.

Thus, this enrichment analysis further proved the importance of genes discovered by our method. We hope that it could be used to gain better understandings of the mechanism of TC.

3.3. Analysis of Some Significant Candidate Genes.

Among 169 significant candidate genes, we selected some important genes to elucidate their potential values to be TC-related genes. Since they have been reported to be associated with the tumorigenesis or development of other types of cancers, we thought it might lend credence to our method and make our findings more convincing.

The gene CYP2B6 (cytochrome P450, family 2, subfamily B, polypeptide 6) mainly encodes enzymes which are involved in many reactions, specifically in anticancer drug metabolism. A report based on one Japanese population showed that polymorphism of CYP2B6 is significantly associated with prostate cancer risk [42]. Also, decreased expression of CYP2B6 is shown in prostate cancer, and it has been recognized as growth inhibitory [43].

FURIN, also known as PACE, encodes furin protein. High expression of furin has been detected in different cancer types, such as ovarian cancer [44] and head and neck cancer cells [45]. And the inhibition of its expression can help decrease the tumorigenesis of cancers [46]. Also, furin overexpression can promote cell invasion in human hepatoma cell lines, which plays a role in the development of hepatocellular carcinoma [47]. Moreover, the gene may involve in the activity of Notch, and the Notch pathway is important during the medullary thyroid cancer (MTC) [48].

MERTK (c-mer proto-oncogene tyrosine kinase) is a protein-coding gene, which belongs to the MER/AXL/TYRO3 receptor kinase family and encodes cell-surface transmembrane receptors that contain regulated kinase activity [49]. Research has found that MERTK is overexpressed in a variety of cancers, such as prostate cancer, non-small-cell lung cancer, and breast cancer [50]. Also, its overexpression can result in the activation of oncogenic

signaling pathways and drive cell transformation in cancer cells [51].

OAS2 ($2'$-$5'$-oligoadenylate synthetase 2, 69/71 kDa) is involved in immune response of viral infection, because it activates RNase L as a result of the elimination of viruses. In a recent study of cervical cancer, researchers found that genes related to antiviral response were increasingly expressed, including OAS2 which is directly involved in viral RNA degradation [52].

PPARG (peroxisome proliferator-activated receptor gamma) is a member of PPAR subfamily of nuclear receptors, which plays a crucial role in the regulation of gene transcription and adipocyte differentiation. Currently, the activation of PPARG has been recognized as one key step in colorectal cancer progression [53], and its deacetylation can determine lipid synthesis and growth in breast tumor [54].

To summarize, even though these 169 significant candidate genes have not been found associated with TC until now, a wealth of evidence has proved their relations to other types of cancer. Therefore, previous researches have validated the reliability of our method and the importance of our findings. We hope our method will be helpful to search novel TC-related genes and be further promoted to the exploration of other biological questions.

3.4. Analysis of Some Candidate Chemicals.

Besides the significant candidate genes, we also discovered 49 significant candidate chemicals that are deemed to be related to thyroid cancer development. Most of them (29 out of 49) can be supported by published literatures. Here, we only gave detailed discussions for three of them. All of these 29 chemicals are briefly discussed in Online Supporting Information S6.

Chloride ion (CID000000312) is a common ion in human cells, which plays a crucial role in cell invasion due to its ability to change the osmotic balance between the inner- and extra-cellular space [55]. The reason behind invading cancer cells that can pass though extracellular matrix is partly because it has the ability to reduce its volume. Several major chloride channels on the cell membrane are responsible for this invasive behavior of cancer cells. Research has found that inhibition of the sodium-potassium-chloride cotransporter isoform-1 (NKCC1) can decrease cell invasion by 50% [56].

Hydrogen cyanide (HCN, CID000000768) is the product of various tobaccos, existing in the smoke as a colorless gas. In the study of gastroesophageal cancer based on selected ion flow tube mass spectrometry (SIFT-MS), hydrogen cyanide is significantly different between cancer and healthy groups [57]. Hydrogen cyanide is also recognized to have cardiovascular and respiratory toxicity, which might be a potential factor to cause lung cancer [58].

Aniline (CID000006115) consists of a phenyl group attached to an amino group, and it is the precursor of industrial chemicals. It is reported that the incidence of bladder cancer is clearly related to exposure to aniline [59]. Potential reasons might be due to an increase in iron overload in the spleen and upregulation of TNF-α, IL-1, and IL-6. Also, the expression of cyclin dependent kinases (CDKs) is upregulated by aniline [60].

4. Conclusion

During the fight with thyroid cancer, discovery of its related genes and chemicals and uncovering the mechanism behind it are important to today's research and future's drug design for designing effective treatments. Only with the assistance of experiment methods would be an onerous and low efficient way. In this study, we sufficiently used known resource, such as protein-protein interactions, chemical-chemical interactions, chemical-protein interactions, and known thyroid cancer-related genes and chemicals, to search new candidate thyroid cancer-related genes and chemicals by the shortest path algorithm. The proposed method generalized our previous method that can only discover disease genes. Further analysis of the selected genes and chemicals implies that some of them have direct or indirect relationship with the formation and development of thyroid cancer, thereby suggesting the effectiveness of our method. We hope that our method and the findings could shed new light on the mechanism research of thyroid cancer.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Authors' Contribution

Yang Jiang and Peiwei Zhang contributed equally to this work.

Acknowledgments

References

[1] Y. E. Nikiforov and M. N. Nikiforova, "Molecular genetics and diagnosis of thyroid cancer," *Nature Reviews Endocrinology*, vol. 7, no. 10, pp. 569–580, 2011.

[2] M. A. Ginos, G. P. Page, B. S. Michalowicz et al., "Identification of a gene expression signature associated with recurrent disease in squamous cell carcinoma of the head and neck," *Cancer Research*, vol. 64, no. 1, pp. 55–63, 2004.

[3] A. Mathur, W. Moses, R. Rahbari et al., "Higher rate of BRAF mutation in papillary thyroid cancer over time: a single-institution study," *Cancer*, vol. 117, no. 19, pp. 4390–4395, 2011.

[4] S. Chevillard, N. Ugolin, P. Vielh et al., "Gene expression profiling of differentiated thyroid neoplasms: diagnostic and clinical implications," *Clinical Cancer Research*, vol. 10, no. 19, pp. 6586–6597, 2004.

[5] A. Baccarelli and V. Bollati, "Epigenetics and environmental chemicals," *Current Opinion in Pediatrics*, vol. 21, no. 2, pp. 243–251, 2009.

[6] A. Prüss-Ustün, C. Vickers, P. Haefliger, and R. Bertollini, "Knowns and unknowns on burden of disease due to chemicals: a systematic review," *Environmental Health: A Global Access Science Source*, vol. 10, no. 1, article 9, 2011.

[7] S. Chanda, U. B. Dasgupta, D. GuhaMazumder et al., "DNA hypermethylation of promoter of gene p53 and p16 in arsenic-exposed people with and without malignancy," *Toxicological Sciences*, vol. 89, no. 2, pp. 431–437, 2006.

[8] C. J. Marsit, K. Eddy, and K. T. Kelsey, "MicroRNA responses to cellular stress," *Cancer Research*, vol. 66, no. 22, pp. 10843–10848, 2006.

[9] J. M. Stuart, E. Segal, D. Koller, and S. K. Kim, "A gene-coexpression network for global discovery of conserved genetic modules," *Science*, vol. 302, no. 5643, pp. 249–255, 2003.

[10] L. Chen, B.-Q. Li, and K.-Y. Feng, "Predicting biological functions of protein complexes using graphic and functional features," *Current Bioinformatics*, vol. 8, no. 5, pp. 545–551, 2013.

[11] D. Warde-Farley, S. L. Donaldson, O. Comes et al., "The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function," *Nucleic Acids Research*, vol. 38, no. 2, Article ID gkq537, pp. W214–W220, 2010.

[12] L. Chen, W.-M. Zeng, Y.-D. Cai, and T. Huang, "Prediction of metabolic pathway using graph property, chemical functional group and chemical structural set," *Current Bioinformatics*, vol. 8, no. 2, pp. 200–207, 2013.

[13] D. Marbach, J. C. Costello, R. Küffner et al., "Wisdom of crowds for robust gene network inference," *Nature Methods*, vol. 9, no. 8, pp. 796–804, 2012.

[14] T. Huang, L. Chen, Y.-D. Cai, and K.-C. Chou, "Classification and analysis of regulatory pathways using graph property, biochemical and physicochemical property, and functional property," *PLoS ONE*, vol. 6, no. 9, Article ID e25297, 2011.

[15] L. Chen, W.-M. Zeng, Y.-D. Cai, K.-Y. Feng, and K.-C. Chou, "Predicting anatomical therapeutic chemical (ATC) classification of drugs by integrating chemical-chemical interactions and similarities," *PLoS ONE*, vol. 7, no. 4, Article ID e35254, 2012.

[16] Y. C. Wang, N. Deng, S. Chen, and Y. Wang, "Computational study of drugs by integrating omics data with kernel methods," *Molecular Informatics*, vol. 32, no. 11-12, pp. 930–941, 2013.

[17] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kanehisa, "Prediction of drug-target interaction networks from the integration of chemical and genomic spaces," *Bioinformatics*, vol. 24, no. 13, pp. i232–i240, 2008.

[18] L. Chen, J. Lu, T. Huang et al., "Finding candidate drugs for hepatitis C based on chemical-chemical and chemical-protein interactions," *PLoS ONE*, vol. 9, no. 9, Article ID e107767, 2014.

[19] F. Napolitano, Y. Zhao, V. M. Moreira et al., "Drug repositioning: a machine-learning approach through data integration," *Journal of Cheminformatics*, vol. 5, article 30, 2013.

[20] L. Chen, J. Lu, N. Zhang, T. Huang, and Y.-D. Cai, "A hybrid method for prediction and repositioning of drug anatomical therapeutic chemical classes," *Molecular BioSystems*, vol. 10, no. 4, pp. 868–877, 2014.

[21] L. Wu, N. Ai, Y. Liu, Y. Wang, and X. Fan, "Relating anatomical therapeutic indications by the ensemble similarity of drug sets," *Journal of Chemical Information and Modeling*, vol. 53, no. 8, pp. 2154–2160, 2013.

[22] B.-Q. Li, T. Huang, L. Liu, Y.-D. Cai, and K.-C. Chou, "Identification of colorectal cancer related genes with mRMR and shortest path in protein-protein interaction network," *PLoS ONE*, vol. 7, no. 4, Article ID e33393, 2012.

[23] M. Jiang, Y. Chen, Y. Zhang et al., "Identification of hepatocellular carcinoma related genes with k-th shortest paths in a protein-protein interaction network," *Molecular BioSystems*, vol. 9, no. 11, pp. 2720–2728, 2013.

[24] J. Zhang, M. Jiang, F. Yuan et al., "Identification of age-related macular degeneration related genes by applying shortest path algorithm in protein-protein interaction network," *BioMed Research International*, vol. 2013, Article ID 523415, 8 pages, 2013.

[25] Y.-F. Gao, Y. Shu, L. Yang et al., "A graphic method for identification of novel glioma related genes," *BioMed Research International*, vol. 2014, Article ID 891945, 8 pages, 2014.

[26] M. Zhao, J. Sun, and Z. Zhao, "TSGene: a web resource for tumor suppressor genes," *Nucleic Acids Research*, vol. 41, no. 1, pp. D970–D976, 2013.

[27] UniProt Consortium, "Update on activities at the Universal Protein Resource (UniProt) in 2013," *Nucleic Acids Research*, vol. 41, pp. D43–D47, 2012.

[28] S. McNeil, A. Budhu, N. Grantees et al., *Imaging*, National Cancer Institute, 2013.

[29] A. P. Davis, C. G. Murphy, R. Johnson et al., "The comparative toxicogenomics database: update 2013," *Nucleic Acids Research*, vol. 41, no. 1, pp. D1104–D1114, 2013.

[30] Y. F. Gao, L. Chen, Y. D. Cai, K. Y. Feng, T. Huang, and Y. Jiang, "Predicting metabolic pathways of small molecules and enzymes based on interaction information of chemicals and proteins," *PLoS ONE*, vol. 7, no. 9, Article ID e45944, 2012.

[31] L. J. Jensen, M. Kuhn, M. Stark et al., "STRING 8—a global view on proteins and their functional interactions in 630 organisms," *Nucleic Acids Research*, vol. 37, no. 1, pp. D412–D416, 2009.

[32] L. Hu, T. Huang, X.-J. Liu, and Y.-D. Cai, "Predicting protein phenotypes based on protein-protein interaction network," *PLoS ONE*, vol. 6, no. 3, Article ID e17668, 2011.

[33] L.-L. Hu, T. Huang, Y.-D. Cai, and K.-C. Chou, "Prediction of body fluids where proteins are secreted into based on protein interaction network," *PLoS ONE*, vol. 6, no. 7, Article ID e22989, 2011.

[34] P. Gao, Q.-P. Wang, L. Chen, and T. Huang, "Prediction of human genes' regulatory functions based on proteinprotein interaction network," *Protein and Peptide Letters*, vol. 19, no. 9, pp. 910–916, 2012.

[35] L. L. Hu, T. Huang, X. Shi, W.-C. Lu, Y.-D. Cai, and K.-C. Chou, "Predicting functions of proteins in mouse based on weighted protein-protein interaction network and protein hybrid properties," *PLoS ONE*, vol. 6, no. 1, Article ID e14556, 2011.

[36] M. Kuhn, C. von Mering, M. Campillos, L. J. Jensen, and P. Bork, "STITCH: interaction networks of chemicals and proteins," *Nucleic Acids Research*, vol. 36, no. 1, pp. D684–D688, 2008.

[37] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.

[38] T. H. Gormen, C. E. Leiserson, R. L. Rivest, and C. Stein, Eds., *Introduction to Algorithms*, MIT Press, Cambridge, Mass, USA, 1990.

[39] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists," *Nucleic Acids Research*, vol. 37, no. 1, pp. 1–13, 2009.

[40] A. A. Bapat, G. Hostetter, D. D. von Hoff, and H. Han, "Perineural invasion and associated pain in pancreatic cancer," *Nature Reviews Cancer*, vol. 11, no. 10, pp. 695–707, 2011.

[41] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: the next generation," *Cell*, vol. 144, no. 5, pp. 646–674, 2011.

[42] T. Kurosaki, M. Suzuki, Y. Enomoto et al., "Polymorphism of cytochrome P450 2B6 and prostate cancer risk: a significant association in a Japanese population," *International Journal of Urology*, vol. 16, no. 4, pp. 364–368, 2009.

[43] J. Kumagai, T. Fujimura, S. Takahashi et al., "Cytochrome P450 2B6 is a growth-inhibitory and prognostic factor for prostate cancer," *The Prostate*, vol. 67, no. 10, pp. 1029–1037, 2007.

[44] R. E. Page, A. J. P. Klein-Szanto, S. Litwin et al., "Increased expression of the pro-protein convertase furin predicts decreased survival in ovarian cancer," *Cellular Oncology*, vol. 29, no. 4, pp. 289–299, 2007.

[45] D. E. Bassi, H. Mahloogi, R. L. de Cicco, and A. Klein-Szanto, "Increased furin activity enhances the malignant phenotype of human head and neck cancer cells," *The American Journal of Pathology*, vol. 162, no. 2, pp. 439–447, 2003.

[46] D. E. Bassi, R. L. de Cicco, H. Mahloogi, S. Zucker, G. Thomas, and A. J. P. Klein-Szanto, "Furin inhibition results in absent or decreased invasiveness and tumorigenicity of human cancer cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 18, pp. 10326–10331, 2001.

[47] R.-N. Chen, Y.-H. Huang, Y.-C. Lin et al., "Thyroid hormone promotes cell invasion through activation of furin expression in human hepatoma cell lines," *Endocrinology*, vol. 149, no. 8, pp. 3817–3831, 2008.

[48] M. Cook, X.-M. Yu, and H. Chen, "Notch in the development of thyroid C-cells and the treatment of medullary thyroid cancer," *American Journal of Translational Research*, vol. 2, no. 1, pp. 119–125, 2010.

[49] A. Verma, S. L. Warner, H. Vankayalapati, D. J. Bearss, and S. Sharma, "Targeting Axl and Mer kinases in cancer," *Molecular Cancer Therapeutics*, vol. 10, no. 10, pp. 1763–1773, 2011.

[50] C. T. Cummings, D. DeRyckere, H. S. Earp, and D. K. Graham, "Molecular pathways: MERTK signaling in cancer," *Clinical Cancer Research*, vol. 19, no. 19, pp. 5275–5280, 2013.

[51] K. Q. Nguyen, W. I. Tsou, D. A. Calarese et al., "Overexpression of MERTK receptor tyrosine kinase in epithelial cancer cells drives efferocytosis in a gain-of-function capacity," *The Journal of Biological Chemistry*, vol. 289, no. 37, pp. 25737–25749, 2014.

[52] K. L. Mine, N. Shulzhenko, A. Yambartsev et al., "Gene network reconstruction reveals cell cycle and antiviral genes as major drivers of cervical cancer," *Nature Communications*, vol. 4, article 1806, 2013.

[53] L. Sabatino, A. Fucci, M. Pancione et al., "UHRF1 coordinates peroxisome proliferator activated receptor gamma (PPARG) epigenetic silencing and mediates colorectal cancer progression," *Oncogene*, vol. 31, no. 49, pp. 5061–5072, 2012.

[54] R. Pestell, L. Tian, C. Wang et al., "Abstract P2-06-02: Pparg deacetylation by SIRT1 determines breast tumor lipid synthesis and growth," *Cancer Research*, vol. 73, p. P2-06-02, 2014.

[55] O. Veiseh, F. M. Kievit, R. G. Ellenbogen, and M. Zhang, "Cancer cell invasion: treatment and monitoring opportunities in nanomedicine," *Advanced Drug Delivery Reviews*, vol. 63, no. 8, pp. 582–596, 2011.

[56] B. R. Haas and H. Sontheimer, "Inhibition of the sodium-potassium-chloride cotransporter isoform-1 reduces glioma invasion," *Cancer Research*, vol. 70, no. 13, pp. 5597–5606, 2010.

[57] S. Kumar, J. Huang, J. R. Cushnir, P. Španěl, D. Smith, and G. B. Hanna, "Selected ion flow tube-MS analysis of headspace vapor from gastric content for the diagnosis of gastro-esophageal cancer," *Analytical Chemistry*, vol. 84, no. 21, pp. 9550–9557, 2012.

[58] P. J. Branton, K. G. McAdam, D. B. Winter, C. Liu, M. G. Duke, and C. J. Proctor, "Reduction of aldehydes and hydrogen cyanide yields in mainstream cigarette smoke using an amine functionalised ion exchange resin," *Chemistry Central Journal*, vol. 5, no. 1, article 15, 2011.

[59] T. Carreón, M. Hein, K. Hanley, S. Viet, and A. Ruder, "0094 Bladder cancer incidence among workers exposed to o-toluidine, aniline and nitrobenzene at a rubber chemical manufacturing plant," *Occupational & Environmental Medicine*, vol. 71, pp. A9–A10, 2014.

[60] S. Kannan, R. Fielder, J. Tristan, E. Longoria, and A. Castillon, "Molecular mechanism of aniline induced bladder cancer," *The FASEB Journal*, vol. 27, pp. 793–791, 2013.