Inferring Novel Tumor Suppressor Genes with a Protein-Protein Interaction Network and Network Diffusion Algorithms

Lei Chen,[1,4,5] Yu-Hang Zhang,[1,5] Zhenghua Zhang,[2,5] Tao Huang,[1] and Yu-Dong Cai[3]

[1]Institute of Health Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, People's Republic of China; [2]Department of Clinical Oncology, Jing'an District Centre Hospital of Shanghai (Huashan Hospital Fudan University Jing'An Branch), Shanghai 200040, People's Republic of China; [3]School of Life Sciences, Shanghai University, Shanghai 200444, People's Republic of China; [4]College of Information Engineering, Shanghai Maritime University, Shanghai 201306, People's Republic of China

Extensive studies on tumor suppressor genes (TSGs) are helpful to understand the pathogenesis of cancer and design effective treatments. However, identifying TSGs using traditional experiments is quite difficult and time consuming. Developing computational methods to identify possible TSGs is an alternative way. In this study, we proposed two computational methods that integrated two network diffusion algorithms, including Laplacian heat diffusion (LHD) and random walk with restart (RWR), to search possible genes in the whole network. These two computational methods were LHD-based and RWR-based methods. To increase the reliability of the putative genes, three strict screening tests followed to filter genes obtained by these two algorithms. After comparing the putative genes obtained by the two methods, we designated twelve genes (e.g., MAP3K10, RND1, and OTX2) as common genes, 29 genes (e.g., RFC2 and GUCY2F) as genes that were identified only by the LHD-based method, and 128 genes (e.g., SNAI2 and FGF4) as genes that were inferred only by the RWR-based method. Some obtained genes can be confirmed as novel TSGs according to recent publications, suggesting the utility of our two proposed methods. In addition, the reported genes in this study were quite different from those reported in a previous one.

INTRODUCTION

Tumor, also named as neoplasm, generally refers to a malignant tissue with malignant proliferative capacity, which usually but not always forms a solid mass with invasive and metastatic tendency as a space-occupying systematic disease.[1] According to World Health Organization (WHO) statistics, more than 8.8 million people around the world directly die from cancer, taking nearly one-sixth of all global deaths, indicating that cancer may be one of the most significant threats for human health nowadays. Every year, direct economic loss on cancer prevention, diagnosis, and treatment has been validated to reach up to 1.16 trillion dollars.[2] After a century of research, the main pathogeneses of cancers have been partially revealed, confirming both genetic background and environmental factors, inducing genomic abnormality that may initiate and promote the tumorigenesis.[3] However, the detailed genes that contribute to such malignant processes in different tumor subtypes have not been fully discovered.[3]

Genes that directly contribute to tumor-associated malignant processes can be simply divided into two groups: oncogenes and tumor suppressor genes (TSGs), with reverse biological functions in normal conditions.[4] Generally, oncogenes turn out to be genes that have the potential to induce tumor, while tumor suppressors turn out to be genes that protect cells from malignant alterations.[4,5] During the initiation and progression of tumor, genetic variants may motivate the activation of oncogenes and the silencing of tumor suppressors, further inducing the abnormal proliferation and invasion of malignant cells.[5] Therefore, oncogenes and tumor suppressors may be equally significant for tumorigenesis. During the century of research on tumor-associated genes, various studies have paid attention to two such different groups of genes, revealing that two such groups of genes may play different biological roles during tumorigenesis. Two-hit hypothesis emerges to be one of the mainstream theories explaining the genetic contribution on tumor initiation.[6,7] Based on two-hit hypothesis, both or at least one of two such alleles encoding specific proteins may be affected before the phenotype appears.[6] The accumulation of variants in both alleles turns out to be a specific biological characteristic of TSG, but not oncogene, because of the recessive status of most mutant TSGs, indicating that only one mutant allele is not functional enough to induce tumorigenesis.

The exceptions to the two-hit roles have been further identified, perfecting the existing regulatory and pathological mechanisms of tumor

suppressors. Two groups of TSGs have been confirmed to be the exceptions of the two-hit rule: (1) genes that containing dominant-negative mutations, and (2) genes exhibiting haploinsufficiency, both of which allow the expression and function of mutants in one single allele.[8] TP53, which encodes p53 protein, is one of the famous exceptions.[9] Considering that a mutant TP53 allele may prevent the function of the non-mutant one, therefore, only one mutant allele of TP53 may directly contribute to tumorigenesis.[8] Apart from genes that contain dominant native mutations like TP53, another group of TSGs that escapes the two-hit rule turns out to exhibit haploinsufficiency.[10] Usually caused by loss-of-function mutations, haploinsufficiency emerges to be a common genetic mechanism explaining the abnormal phenotype induced by only one functional copy of a certain gene with the other allele as dysfunctional.[11,12] Recessive genes, like most TSGs, can perform their potential biological functions when they exhibit haploinsufficiency.[12] PTCH in medulloblastoma and NF1 in neurofibroma are two typical examples of functional recessive TSGs induced by haploinsufficiency.[13,14] Although such functional TSGs that escape the two-hit theory are generally functional and common TSGs, they only account for a small part of all TSGs, remaining the most of which obey the two-hit hypothesis.

Considering that most TSGs obey the two-hit hypothesis, we find it harder to identify TSGs than oncogenes while comparing the tumor and normal samples (normal samples may also have dysfunctional mutations in only one allele of the TSGs).[6,7] It is hard and time consuming to identify TSGs based on current experimental conditions. In recent years, developing computational methods, which were constructed based on known data, is an alternative way to investigate cancer-related problems. These methods can always yield good performance on known data and may give useful clues for extending current knowledge. For example, several computational models have been built for the identification of cancer-related non-coding RNAs (ncRNAs),[15–20] which can further give help to recognize potential cancer-related ncRNAs. For TSGs, with the development of high-throughput sequencing and the accumulation of experimentally identified TSGs, identifying potential TSGs based on computational methods and current databases, which are faster and cheaper compared with experimental ones, is possible. Recently, Chen et al. proposed a computational method to identify potential TSGs based on known ones.[21] Their method applied the shortest path (SP) algorithm to a protein network and searched for SPs connecting any two known TSGs. Genes lying on the obtained paths were extracted as candidate TSGs, which were further filtered by a permutation test. However, this method cannot fully use the network because the SP algorithm can only identify genes with limited distance to known genes. This study employed two network diffusion algorithms, which can make full use of the whole network and design stricter screening tests to extract more reliable potential TSGs.

In this study, we proposed two computational methods based on two network diffusion algorithms, including Laplacian heat diffusion (LHD) algorithm and random walk with restart (RWR) algorithm. By executing these two algorithms on a protein-protein interaction (PPI) network using known TSGs as seed nodes, several potential

TSGs were accessed. Then, three screening tests, including permutation, association, and function tests, followed to control false-positive genes and extract important ones. As a result, the LHD-based method, using the LHD algorithm as the network diffusion algorithm, yielded 41 putative genes, and the RWR-based method, using the RWR algorithm as the network diffusion algorithm, produced 140 putative genes. By comparing the putative genes yielded by the two methods, only a few genes were identified by both methods, indicating that both methods have their respective prediction advantages and the combined application of two such methods may further improve the prediction effects. In addition, our putative genes were quite different from those reported in a previous study,[21] suggesting that putative genes obtained in this study can be essential supplements for the study of TSG. Finally, based on recent publications, several putative genes yielded by either the LHD-based or RWR-based method can be confirmed to be novel TSGs, validating the reliability of the results.

RESULTS

In this study, we built two computational methods to infer novel TSGs: one was based on the LHD algorithm (LHD-based method), and the other one was based on the RWR algorithm (RWR-based method). These two methods integrated three screening tests. The procedures of these two methods are illustrated in Figure 1. The detailed description of these two methods can be found in the Materials and Methods. This section provides the detailed results yielded by them and makes some comparisons.

Putative Genes Yielded by the LHD-Based Method

The LHD-based method applied the LHD algorithm to the PPI network G using Ensembl IDs of TSGs as seed nodes, yielding large numbers of LHD genes. Then, the obtained LHD genes were evaluated by three tests (permutation test, association test, and function test), thereby extracting important putative genes.

By applying the LHD algorithm to the PPI network G using Ensembl IDs of TSGs as seed nodes, each node in G was assigned with a heat value. We set the threshold 10^{-4} to the heat value, obtaining 2,874 LHD genes. Next, these genes were evaluated by the permutation test with 500 randomly produced sets, yielding a p value for each of them. Moreover, 443 genes with p values less than 0.05 were selected. Then, these 443 genes were measured by maximum interaction scores (MISs) in the association test. Because 900 is set as the cutoff of the highest confidence in STRING, it was set as the threshold of MIS. Finally, for each of 85 genes that remained, we calculated their maximum function scores (MFSs) in the function test. By setting 0.9 as the threshold of MFS, we accessed 41 putative genes. All measurements mentioned above for each LHD gene are listed in Table S1, and the number of candidate genes passing each stage in the LHD-based method is listed in Table 1. The obtained 41 putative genes can be found in Table S2.

Putative Genes Yielded by the RWR-Based Method

Similar to the LHD-based method, the RWR-based method first adopted the RWR algorithm to search candidate genes in the PPI network G using validated TSGs as seed nodes. A great deal of
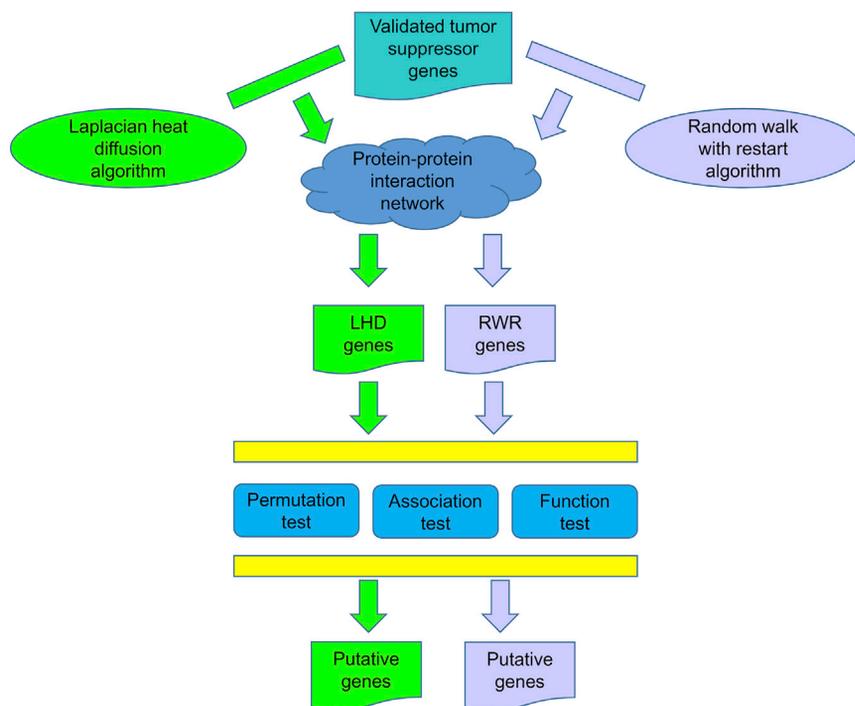
The LHD-based method first applied the LHD algorithm on a PPI network using validated TSGs as seed nodes, producing a large number of LHD genes. Then, these genes were filtered by three screening tests. The RWR-based method followed similar procedures. The only difference was the application of the RWR algorithm on the PPI network rather than the LHD algorithm.

RWR genes can be selected, which were further filtered via permutation, association, and function tests. The remaining genes were termed as the putative genes of this method.

The RWR algorithm was executed on the PPI network G by setting validated TSGs as seed nodes. Each node (gene) received a probability. By setting 10^{-5} as the threshold of probability, we obtained 5,889 RWR genes. Then, these genes were also filtered by a permutation test (with 1,000 randomly produced sets), an association test, and a function test, yielding a p value, an MIS, and an MFS for each RWR gene, which can be found in Table S3. By setting 0.05 as the threshold of p value, 900 as the threshold of MIS, and 0.98 as the threshold of MFS, the number of RWR genes gradually decreased (see Table 1 for the number of RWR genes passing each stage in the method). Finally, 140 putative genes were accessed, which are provided in Table S4.

Comparison of Putative Genes Yielded by the Different Methods
The LHD-based and RWR-based methods yielded 41 and 140 putative genes, respectively. For convenience, we denoted two sets consisting of these genes as G_{LHD} and G_{RWR}, respectively. After taking the union operation of these two gene sets, we obtained 169 putative genes in which 12 were identified by both of the methods, 29 were identified only by the LHD-based method, and 128 were yielded only by the RWR-based method. The distribution of 169 putative genes is illustrated in Figure 2A. To elaborate the difference of G_{LHD} and G_{RWR}, the Jaccard coefficient of these two sets, which is defined as the proportion of members in their intersection among those in their union, i.e., $|G_{LHD} \cap G_{RWR}|/|G_{LHD} \cup G_{RWR}|$, was calcu-

lated, resulting in 7.101%. If these two methods can be proven to be effective, which is addressed in the Discussion, the putative genes obtained by one method can be essential supplements for those obtained by another method. The simultaneous usage of these two methods can provide more potential TSGs.

To further elaborate the reliability of the 169 putative genes, the subnetwork, consisting of the linkages between them and validated TSGs, was extracted from the PPI network G and is shown in Figure 3A. Clearly, the associations between putative genes and validated genes are quite strong. In addition, we extracted the most important linkages (with highest confidence) from this subnetwork to construct another subnetwork, as shown in Figure 3B, from which we can observe that several putative genes have several strong associated TSGs, implying that they can be novel TSGs with high probabilities. In the Discussion, extensive analyses of several important putative genes are given.

In Chen et al.'s study,[21] 205 putative TSGs were reported, which comprised gene set G_{SP}. Among these genes, one gene (EPHA7) was in both G_{LHD} and G_{RWR}, one gene was only in G_{LHD}, and 11 genes were only in G_{RWR} (see Figure 2B). As shown in Figure 2B, 156 genes identified by either the LHD-based or RWR-based method were not in G_{SP}, that is, they were not reported in Chen et al.'s study. If we can prove the utility of our two methods, which is elaborated upon in the Discussion, putative genes reported in this study can help us improve the comprehension of TSGs. To quantify the difference between putative genes in our study and Chen et al.'s study, we further calculated the Jaccard coefficients of G_{SP} and G_{LHD}, G_{SP} and G_{RWR}, and G_{SP} and $G_{LHD} \cup G_{RWR}$; they were 0.820%, 3.604%, and 3.601%, respectively. All of these indicated that our reported genes can be essential supplements for those in Chen et al.'s study. In addition, we list putative genes in at least two sets of G_{SP}, G_{LHD}, and G_{RWR} in Table 2, i.e., these genes were identified by at least two different methods, implying they can be novel TSGs with high probabilities.

DISCUSSION
As mentioned in the previous section, comparing to putative genes reported in Chen et al.'s study,[21] one specific gene (EPHA7) has been identified by all three computational methods, implying its

Table 1. Number of Candidate TSGs in Different Stages of LHD-Based and RWR-Based Methods

|	Network Diffusion Algorithm	Permutation Test	Association Test	Function Test
LHD-based method	2,874	443	85	41
RWR-based method	5,889	1,364	980	140

specific biological function as a potential TSG. According to recent publications, EPHA7 has been confirmed to act as a tumor suppressor in multiple tumor subtypes, including follicular lymphoma,[22] small cell lung cancer,[23] gastric cancer,[24] renal cell carcinoma,[25] prostate cancer,[26] and osteosarcoma,[27] validating its effective anti-tumorigenesis contributions. Take follicular lymphoma as an instance for detailed demonstration. In 2011, a specific study[22] on follicular lymphoma confirmed that the knockdown of EPHA7 may directly induce tumorigenesis in a mouse model, implying the specific tumor suppressor role of such gene.

We identified two potential TSG sets via the LHD-based and RWR-based methods. As shown in Figure 2A, we clustered all putative genes into three groups: (1) putative genes identified by both of the two methods, (2) those identified only by the LHD-based method, and (3) those identified only by the RWR-based method. According to recent publications, some genes in these three groups can be confirmed to definitely contribute to tumorigenesis as functional TSGs, indicating that the two proposed methods are quite efficient and accurate to predict novel TSGs, whether used alone or combined. The detailed analysis is given below.

Putative Genes Identified by Both of the Two Methods

Twelve genes were predicted by the two methods to be putative TSGs. Here, we analyzed five of them, listed in Table 3.

MAP3K10 has been predicted to be a specific TSG. Based on recent publications, MAP3K10 has been widely reported to participate in auto-phosphorylation and subsequent activation via the JUN N-terminal pathway as a homodimer.[28,29] A recent study on integrative genomic analyses in embryonic stem cell (ESC) and NT2 cell lines (normal cells) revealed that MAP3K10 induced DYRK2 phosphorylation combined with SUFU inhibition. This revelation may directly affect the normal biological function of the stem cell-signaling network interacting with TP53, a famous TSG contributing to the downstream proliferation and cell adhesion associated in biological processes.[30] Although no direct evidence confirmed that such gene acts as a tumor suppressor gene during tumorigenesis, such evidence in ESC lines confirmed that our predicted gene may cooperate with TP53, a famous tumor suppressor, implying its potential tumor suppressor functions. Therefore, it is reasonable to forecast such gene as a potential tumor suppressor.
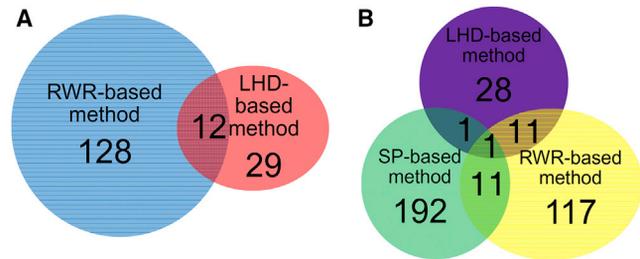

Figure 2. Venn Diagrams to Illustrate Putative Gene Sets Yielded by the Different Methods

(A) A Venn diagram to illustrate the distribution of 169 putative genes that were identified by either the LHD-based method or the RWR-based method. The red and blue circles represent the gene sets consisting of the putative genes yielded by the LHD-based method and the RWR-based method, respectively. 12 genes were identified by both of the two methods. (B) A Venn diagram to illustrate three putative gene sets yielded by three methods. The purple circle represents the gene set yielded by the LHD-based method, the yellow circle represents the gene set yielded by the RWR-based method, and the green circle represents the gene set reported in a previous study (SP-based method).

Apart from MAP3K10, the gene RND1 has also been predicted to be a potential TSG. RND1 lacks intrinsic GTPase activity and controls the rearrangement of actin cytoskeleton and the Rac-dependent neuritic process formation independent of GDP binding in normal cells.[31,32] Furthermore, Rnd subfamily members, including RND1, RND2, and RHOH, may participate in p53-mediated regulatory pathways resistant to certain endogenous and exogenous stimuli like tumorigenesis in human osteosarcoma cell lines, U2OS and SAOS2, implying that RND1 may definitely contribute to the identification and elimination of tumor cells.[33] Such gene may also directly control the balance between cell survival and malignant transformation, validating that RND1 itself may be enough to act as a functional tumor suppressor.[34]

OTX2, as another predicted TSG, participates in the early specification of neuroectoderm in the developmental processes of brain based on a mouse model.[35] As for the potential anti-tumor functions of such gene, recent publications confirmed that OTX2 plays a specific role during tumorigenesis and acts as a core factor in c-*myc*, CRX, and phosphorylated RB pathways.[36] Although OTX2 promotes the malignant transformation of tumor cells as a traditional oncogene, a recent study on retinal detachment confirmed that OTX2 may contribute to the regulation of the p53-signaling pathway and can also act as a potential TSG in multiple cell lines and animal models, like childhood brain tumor medulloblastoma cell line D425MED and patient-derived mouse xenograft models.[37]

Genes like RET and GLI2 are both potential TSGs that were identified by both of the two methods. These two genes have been identified as quite effective tumor suppressors in common cell lines like mouse NIH 3T3 cell lines. RET, as a member of the cadherin superfamily, participates in the regulation of cell proliferation, migration, and differentiation.[38,39] As for its contribution to tumor suppression, at least in colorectal cancer and medullary thyroid carcinoma, RET acts as a
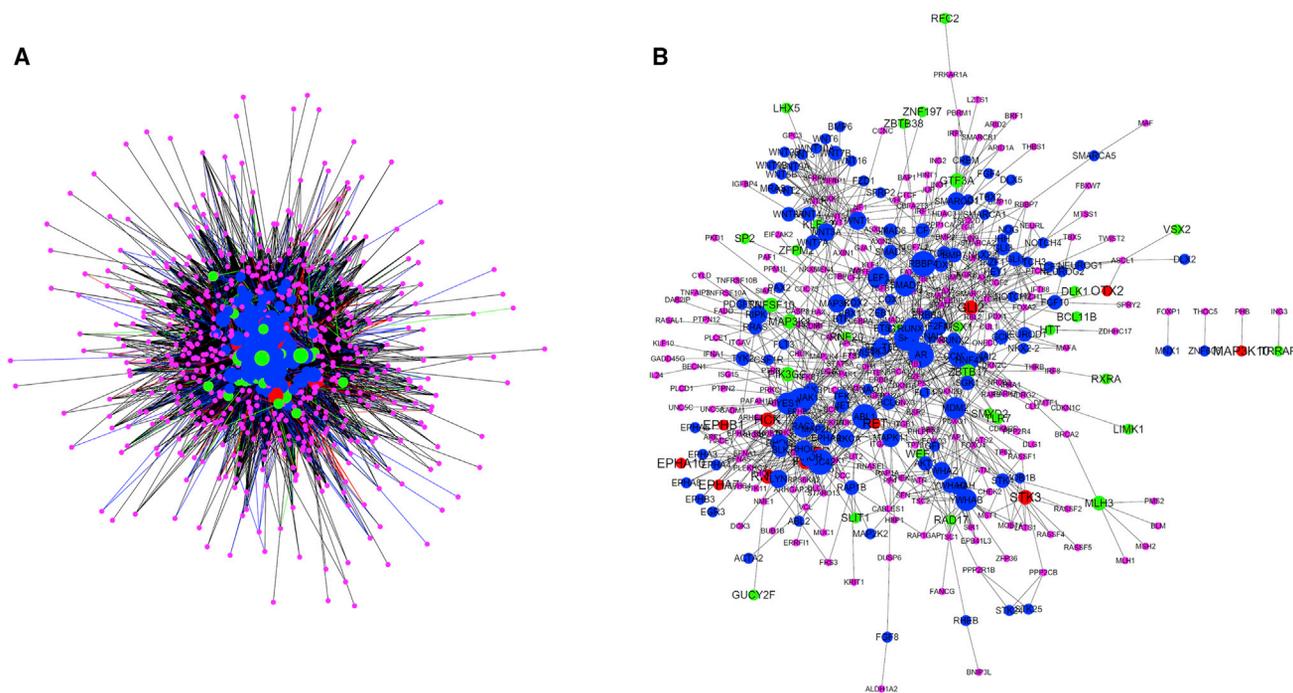
Figure 3. The Subnetwork of PPI Network Containing the Linkages between Putative Genes and Validated TSGs

Pink nodes represent validated TSGs, while green, blue, and red nodes represent putative genes yielded by the LHD-based method, the RWR-based method, and both methods, respectively. The sizes of nodes in green, blue, and red represent their degrees. (A) The subnetwork containing all linkages between putative genes and validated TSGs. Edges in black, blue, green, and red represent PPIs with low, medium, high, and highest confidence, respectively. (B) The subnetwork only containing linkages between putative genes and validated TSGs with highest confidence.

specific TSG in respective cell lines and mouse models.[40] As for GLI2, such gene is a development-associated gene contributing to the formation of lung, trachea, and esophagus, which is further validated in gli3-null and gli3Delta699 mouse models.[41]

Putative Genes Identified Only by the LHD-Based Method

29 genes were identified only by the LHD-based method. We analyzed seven of them, listed in Table 4, to validate the effectiveness of the LHD-based method.

RFC2 was predicted to be a potential TSG in this study. As a member of the activator 1 small subunits family, such gene contributes to the assembly of PCNA and polymerase delta on the DNA template, promoting cell survival.[42] Although various data have been reported to confirm the relationship between RFC2 and tumorigenesis by cell lines or mouse models, only some reports validated the potential tumor suppression function of such gene.[43,44] A specific study on renal cell carcinoma confirmed that, during the initiation and progression of such tumor, RFC2 has a specific rearrangement, partially losing the normal biological function of such gene and indicating that RFC2 may be a potential TSG at least in cell lines.[44]

GUCY2F was also predicted to be a potential TSG. Encoding a guanylyl cyclase predominantly expressed in the retina, such gene contributes to re-synthesis of cGMP required for recovery of the dark

state after phototransduction.[45] GUCY2F may also participate in the development of various tumor subtypes, including Japanese colorectal cancer and myeloma, as a negative regulator, validated by mouse models and related cell lines.[46]

The following two genes, namely, GSC and TNFSF10, were also predicted to be functional potential TSGs. GSC, as a homeodomain-encoding gene, is expressed and involved in a gastrulation process.[47] In concert with another functional gene NKX3.1, GSC contributes to resistance against malignant transformation and cell proliferation.[48] Similarly, TNFSF10, as a member of the tumor necrosis factor superfamily, mainly contributes to the TNF-related apoptosis in spleen, lung, and prostate.[49,50] TNFSF10, as a p53 target gene, contributes to the regulation of p53-dependent cell death, and, during tumorigenesis, it acts as a resistant regulator of abnormal cell proliferation.[51] Therefore, TNFSF10 may definitely be a potential TSG.

LHX5 is another functional gene, which was also predicted as a potential TSG. Involving the control of differentiation and development, such gene contributes to the regulation of neuronal differentiation and migration during development of the CNS.[52] As for its specific contribution to tumorigenesis, this gene may indicate better prognosis of patients with breast cancer by inhibiting the undifferentiated processes of malignant tumor cells, implying its potential contribution as a tumor suppressor.[53]

Table 2. 24 Putative Genes Identified by at Least Two Methods

Ensembl ID	Gene Symbol	LHD-Based Method	RWR-Based Method	SP-Based Method[a]
ENSP00000020945	SNAI2	×[b]	√[c]	√
ENSP00000222330	GSK3A	×	√	√
ENSP00000228682	GLI1	×	√	√
ENSP00000233948	WNT6	×	√	√
ENSP00000253055	MAP3K10	√	√	×
ENSP00000254480	SMARCC1	×	√	√
ENSP00000262158	SMAD7	×	√	√
ENSP00000287934	FZD1	×	√	√
ENSP00000293549	WNT1	×	√	√
ENSP00000308461	RND1	√	√	×
ENSP00000341032	WNT7B	×	√	√
ENSP00000343819	OTX2	√	√	×
ENSP00000347942	RET	√	√	×
ENSP00000354586	GLI2	√	√	×
ENSP00000358309	EPHA7	√	√	√
ENSP00000361892	STK4	×	√	√
ENSP00000362139	EPHA10	√	√	×
ENSP00000363115	FGR	√	√	×
ENSP00000364895	ZBTB17	√	×	√
ENSP00000365012	HCK	√	√	×
ENSP00000368686	E2F4	×	√	√
ENSP00000370912	TEC	√	√	×
ENSP00000381097	EPHB1	√	√	×
ENSP00000390500	STK3	√	√	×

×, the putative gene cannot be identified by the method; √, the putative gene can be identified by the method.
[a]The computational method proposed in Chen et al.'s study.[21]

Genes VSX2 and SLIT1 were also predicted as potential TSGs. VSX2, as a specific regulator in specification and morphogenesis, controls the cell fate specification and differentiation in the developing retina.[54] In retinal tissues, such gene may inhibit the malignant transformation of normal cells to malignant ones.[55] SLIT1 contributes to the histogenesis of the CNS physically or pathologically.[56]

Putative Genes Identified Only by the RWR-Based Method
The RWR-based method identified 128 putative genes. Five of them, which are listed in Table 5, are analyzed here.

SNAI2 has been confirmed to be a transcriptional repressor that modulates both activator-dependent and basal transcription, involving neural crest cell generation and migration.[57] This gene is silenced in prostate cancer and serves as a direct transcriptional regulator of breast tumor cell metastasis targeting BIM (BCL2-Like 11).[58] Although there is a lack of direct evidence, such gene may be a functional potential TSG.

FGF4 is a member of the fibroblast growth factor (FGF) family, participating in cell growth, morphogenesis, and invasion.[59] Interacting with the tumor microenvironment, protein FGF4 encoding promotes the proliferation of tumor cells, indicating that the potential anti-tumor function of such gene still remains to be revealed.[60]

GSK3A is a functional regulator of glycogen synthesis in skeletal muscle.[61] As for the tumor suppressor functions, GSK3A participates in the APC/beta-catenin/Tcf. pathway and modulates drug resistance and chemotherapy-induced necroptosis.[62] Considering that the APC/beta-catenin/Tcf.-signaling pathway is involved in various tumor suppression biological processes, we regard GSK3A as a potential TSG.[63]

WNT16, as another putative gene, participates in the regulation of p53 activity and the phosphoinositide 3-kinase/AKT pathway, validating its potential contribution to tumor suppression processes.[64]

Similarly, DLX5 is a member of a homeobox transcription factor gene, expressed in brain and skeleton.[65] Interacting with TP63, such gene contributes to the negative regulation of tumorigenesis, validating its potential tumor suppression function.[66,67]

As analyzed above, several putative genes inferred by either the LHD-based or the RWR-based method can be confirmed to be novel TSGs by recent publications, indicating the usefulness of these two methods. For the rest of the putative genes that were not analyzed in this study, we believe that many of them can be validated. We leave them to readers for further study.

MATERIALS AND METHODS
Validated TSGs
The validated human TSGs were retrieved from a previous study,[21] in which 716 human TSGs were collected from the TSGene database (https://bioinfo.uth.edu/TSGene/).[68] In this study, these genes were employed to infer novel TSGs with some advanced network diffusion algorithms. Because the diffusion algorithms would be executed on a PPI network, the above 716 genes were mapped into the Ensembl IDs of proteins encoded by them. Also, IDs that did not occur in the PPI network were discarded. Finally, 631 Ensembl IDs were obtained, which are provided in Table S5.

PPI Network
Interacting proteins always share similar functions.[21,69–76] The 631 proteins encoded by TSGs share some common functions because they can protect cells from one step on the path to cancer. Interacting proteins may also have these types of functions but with lower probabilities. This can be done for proteins with longer distances to those encoded by TSGs. Thus, inferring novel TSGs using PPIs based on validated ones is feasible.

In this study, we employed the PPIs reported in STRING (https://string-db.org/, version 10.0),[77] a well-known public database collecting known and predicted PPIs from various sources, such as genomic

Table 3. Important Putative Genes Yielded by Both LHD-Based and RWR-Based Methods

Ensembl ID	Gene Symbol	LHD-Based Method		RWR-Based Method			
		Heat	p Value	Probability	p Value	MIS	MFS
ENSP00000253055	MAP3K10	1.8567E−04	0.036	2.7819E−05	0.029	925	0.9954
ENSP00000308461	RND1	1.2137E−04	0.040	5.2943E−05	<0.001	982	0.9942
ENSP00000343819	OTX2	4.1770E−04	0.028	3.7524E−05	0.024	984	0.9855
ENSP00000347942	RET	1.3647E−04	0.048	1.1216E−04	<0.001	984	0.9867
ENSP00000354586	GLI2	2.1655E−04	0.032	5.4405E−05	<0.001	999	0.9847

MIS, maximum interaction score; MFS, maximum function score.

context predictions, high-throughput lab experiments, (conserved) co-expression, automated textmining, and previous knowledge in databases. To access human PPIs, a file named "9606.protein. links.v10.txt.gz" was downloaded from the download page of the STRING website. This file contained 4,274,001 interactions, each of which consisted of two proteins, represented by Ensembl IDs, and one score with a range between 150 and 999. A high score indicated that the corresponding interaction had a high probability to occur. For later formulation, we denoted the score of the interaction between proteins, p_1 and p_2, as $I(p_1,p_2)$. Ensembl IDs at 19,247 were involved in the above interactions.

Based on the aforementioned PPIs, a large PPI network can be constructed. Ensembl IDs at 19,247 were defined as nodes in the network, and two nodes were connected by an edge if and only if they could comprise a PPI. In addition, the interaction score was assigned to the edge as its weight. As a result, we constructed a PPI network with 19,247 nodes and 4,274,001 edges. For formulation, we denoted this PPI network as G.

Searching Candidate Genes Using the LHD Algorithm

Heat diffusion is a type of network diffusion algorithm that always transmits heat on some seed nodes to other nodes in a network following some rules. The heat assigned to a node indicates its obscure associations to seed nodes. One study[78] has suggested that heat diffusion has good performance to identify disease genes. In this study, one type of heat diffusion algorithm, LHD algorithm, was employed as a basic searching algorithm to infer novel TSGs.

For the PPI network G constructed in the previous section, let A be its adjacent matrix, from which a column-wise normalized matrix A' was built as follows:

$$A'[i,j] = \frac{A[i,j]}{\sum_{k=1}^{n} A[k,j]}. \qquad \text{(Equation 1)}$$

From 631 Ensembl IDs of TSGs, an initial heat distribution, H_0, was formulated as a column vector, which contained 19,247 components. Each component indicated the heat on one node in the PPI network G. The components of the Ensembl IDs of TSGs were set to 1/631, others were set to 0. Then, this vector was updated in the follow manner:

$$H_t[i] = H_0[i]\exp(-\lambda_i t), \qquad \text{(Equation 2)}$$

where H_t represented the heat distribution at time t, and λ_i was the i-th eigenvalue of matrix A'. When the later heat distribution vector and the former one were quite similar, the updating procedures stopped. As a result, each node was assigned with a heat value. Clearly, a node with a high heat value was more important than that with a low heat value. By setting a threshold to heat value, we extracted nodes that were assigned to heat values larger than the threshold and mapped obtained nodes to corresponding genes. Certainly, the TSGs mentioned above in Validated TSGs should not be included. For convenience, genes yielded by the LHD algorithm were called LHD genes.

In this study, we downloaded the program of LHD algorithm from https://cran.r-project.org/web/packages/diffusr/index.html, and we executed it with default parameters on the PPI network G, obtaining several LHD genes.

Searching Candidate Genes Using the RWR Algorithm

The RWR algorithm is a classic ranking algorithm.[79] This algorithm is also a type of network diffusion algorithm that has been successfully applied to identify novel genes, proteins, or chemicals in different networks.[71,72,80,81] The RWR algorithm simulates a walker starting from one seed node or a set of seed nodes and walking randomly on the network. Similar to the LHD algorithm, an initial probability distribution vector, H_0, was constructed, which was the same as that in the LHD algorithm, and was repeatedly updated. Let H_{t+1} denote the probability vector after executing $(t + 1)$-th step, which can be updated from H_t as follows:

$$H_{t+1} = (1 - r)A'^T H_t + rH_0, \qquad \text{(Equation 3)}$$

where A' was the same as that defined in Equation 1, and r was set to 0.8 as suggested in some previous studies.[71,72,80] When H_{t+1} and H_t were quite similar, measured by the L_1 norm of their difference less than 10^{-6}, the updating procedures stopped. H_{t+1} was output as

Table 4. Important Putative Genes Yielded Only by the LHD-Based Method

Ensembl ID	Gene Symbol	LHD-Based Method		RWR-Based Method		MIS	MFS
		Heat	p Value	Probability	p Value		
ENSP00000055077	RFC2	3.1496E−04	0.036	1.8028E−05	0.316	999	0.9146
ENSP00000218006	GUCY2F	2.1541E−04	0.044	1.8699E−05	0.544	904	0.9930
ENSP00000238558	GSC	9.4478E−04	0.030	1.1429E−05	0.170	977	0.9595
ENSP00000241261	TNFSF10	4.9214E−04	0.010	3.9617E−05	0.005	999	0.9460
ENSP00000261731	LHX5	6.8836E−04	0.018	1.4245E−05	0.082	914	0.9436
ENSP00000261980	VSX2	1.0335E−03	0.016	1.6524E−05	0.055	910	0.9491
ENSP00000266058	SLIT1	3.3237E−04	0.012	4.0380E−05	<0.001	959	0.9608

MIS, maximum interaction score; MFS, maximum function score.

the results of the RWR algorithm. Accordingly, each node was assigned to a probability, indicating its associations with seed nodes. We clearly selected nodes with high probabilities, and their corresponding genes were extracted as candidate TSGs. Similarly, a threshold was set to the probability. The obtained genes were termed as RWR genes.

Here we used the program developed by Li and Patra[82] to quickly implement the RWR algorithm.

Screening Tests
Based on validated TSGs, the LHD and RWR algorithms can separately produce some candidate TSGs, namely, LHD genes and RWR genes, respectively. However, the utility of these two algorithms highly relied on the structure of the PPI network G. Some nodes (genes) may have special positions in G, suggesting that they were more likely to be selected by the LHD or RWR algorithm. However, these nodes had little and even no associations with the biological processes of protecting cells from malignant alterations, that is, they were false-positive genes. Thus, controlling this type of candidate TSG is necessary. In addition, to increase the reliability of the obtained genes, we should select the most important candidate TSGs. Therefore, three screening tests were proposed as follows: (1) permutation test, (2) association test, and (3) function test. The first test was to control false-positive genes, and the other two tests helped us to extract important genes. Their descriptions are given below.

Permutation Test
The idea of a permutation test is to evaluate the heat or probability values, produced by the LHD or RWR algorithm, respectively, of candidate TSGs under several gene sets and to compare them with actual ones. In detail, m Ensembl ID sets with size 631 were randomly produced, denoted as $E_1, E_2, \ldots, E_m$. Each set, such as E_i, executed the LHD algorithm or RWR algorithm on the PPI network G with members in this set as seed nodes. Then, each LHD gene or RWR gene (g) received a heat or a probability value. After all 1,000 sets had been tested, each g received one actual heat or probability and 1,000 heat values or probabilities, based on which a measurement, p value, can be calculated as follows:

$$p - value(g) = \frac{\Omega}{m}, \qquad \text{(Equation 4)}$$

where Ω represented the number of randomly produced sets on which g received higher heat values or probabilities than its actual one. Clearly, a candidate TSG with a high p value was less likely to be an actual TSG because it has strong associations with several randomly produced sets. Considering the fact that 0.05 is always deemed as the important cutoff for the significance level of the test, we selected LHD or RWR genes with p values less than 0.05.

Association Test
With regard to the candidate TSGs passing the permutation test, some have strong associations with validated TSGs and should be selected. As mentioned above in PPI Network, interacting proteins

Table 5. Important Putative Genes Yielded Only by the RWR-Based Method

Ensembl ID	Gene Symbol	RWR-Based Method		LHD-Based Method		MIS	MFS
		Probability	p Value	Heat	p Value		
ENSP00000020945	SNAI2	5.3473E−05	0.002	3.9529E−05	−	998	0.9825
ENSP00000168712	FGF4	3.8685E−05	<0.001	4.1323E−05	−	936	0.9836
ENSP00000222330	GSK3A	6.2092E−05	0.010	1.0955E−04	0.076	999	0.9921
ENSP00000222462	WNT16	3.6751E−05	<0.001	4.0873E−05	−	919	0.9854
ENSP00000222598	DLX5	2.9495E−05	0.028	4.7228E−05	−	936	0.9853

MIS, maximum interaction score; MFS, maximum function score; −, the corresponding gene received a heat lower than the threshold of heat in the LHD-based method, i.e., it was not selected as LHD genes. Thus, the p value was not available for this gene.

always share similar functions.[21,69–76] Thus, for each candidate gene g, we linked them to validated TSGs and extracted the MIS as a measurement, named MIS, which was formulated as follows:

$$MIS(g) = \max\left\{I(g, g') \mid g' \text{ is a validated TSG}\right\}. \quad \text{(Equation 5)}$$

Clearly, candidate genes assigned with high MISs were more likely to be actual TSGs. By setting a threshold, we can select important candidate TSGs.

Function Test

The validated TSGs must be highly related to some biological processes. To date, gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways are two popular types of information to describe biological processes. By having an insight into the relationships between TSG (candidate TSGs) and GO terms (KEGG pathways), we can further extract important candidate TSGs that exhibit similar relationships with GO terms and KEGG pathways to at least one validated TSG. To quantify the relationship between one gene and GO terms or KEGG pathways, the enrichment theory[83–86] was employed in this study. For one gene g and one GO term or KEGG pathway T, their relationship defined by enrichment theory was calculated by the following:

$$S(g, T) = -\log_{10}\left(\sum_{k=m}^{n} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}\right), \quad \text{(Equation 6)}$$

where N is the total number of proteins in humans, M is the number of proteins annotated to T, n is the number of proteins in H that consist of g and its direct neighbors in the PPI network reported in STRING, and m is the number of proteins that are in H and annotated to T. Our in-house program using R function phyper was developed to calculate this value. Accordingly, each gene can be encoded into a vector by collecting all outcomes of this function. For gene g, the obtained vector was formulated as $E(g)$. Accordingly, the relationship between two genes on their GO terms and KEGG pathways can be measured by the following:

$$Q(g, g') = \frac{E(g) \cdot E(g')}{\|E(g)\| \cdot \|E(g')\|}. \quad \text{(Equation 7)}$$

Similar to MIS, another measurement, MFS, was calculated for each gene g, which was defined as follows:

$$MFS(g) = \max\left\{Q(g, g') \mid g' \text{ is a validated TSG}\right\}. \quad \text{(Equation 8)}$$

Clearly, we should select candidate TSGs with high MFSs. By setting a threshold to MFS, important candidate TSGs can be extracted.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five tables and can be found with this article online at https://doi.org/10.1016/j.omtm.2018.06.007.

REFERENCES

1. Yari, K., Payandeh, M., and Rahimi, Z. (2016). Association of the hypermethylation status of PTEN tumor suppressor gene with the risk of breast cancer among Kurdish population from Western Iran. Tumour Biol. 37, 8145–8152.

2. McGuire, S. (2016). World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. Adv. Nutr. 7, 418–419.

3. Perera, F.P. (1997). Environment and cancer: who are susceptible? Science 278, 1068–1073.

4. Lobry, C., Oh, P., Mansour, M.R., Look, A.T., and Aifantis, I. (2014). Notch signaling: switching an oncogene to a tumor suppressor. Blood 123, 2451–2459.

5. Kavianpour, M., Ahmadzadeh, A., Shahrabi, S., and Saki, N. (2016). Significance of oncogenes and tumor suppressor genes in AML prognosis. Tumour Biol. 37, 10041–10052.

6. Hino, O., and Kobayashi, T. (2017). Mourning Dr. Alfred G. Knudson: the two-hit hypothesis, tumor suppressor genes, and the tuberous sclerosis complex. Cancer Sci. 108, 5–11.

7. Paige, A.J. (2003). Redefining tumour suppressor genes: exceptions to the two-hit hypothesis. Cell. Mol. Life Sci. 60, 2147–2163.

8. Stoczynska-Fidelus, E., Szybka, M., Piaskowski, S., Bienkowski, M., Hulas-Bigoszewska, K., Banaszczyk, M., Zawlik, I., Jesionek-Kupnicka, D., Kordek, R., Liberski, P.P., and Rieske, P. (2011). Limited importance of the dominant-negative effect of TP53 missense mutations. BMC Cancer 11, 243.

9. Werness, B.A., Parvatiyar, P., Ramus, S.J., Whittemore, A.S., Garlinghouse-Jones, K., Oakley-Girvan, I., DiCioccio, R.A., Wiest, J., Tsukada, Y., Ponder, B.A., and Piver, M.S. (2000). Ovarian carcinoma in situ with germline BRCA1 mutation and loss of heterozygosity at BRCA1 and TP53. J. Natl. Cancer Inst. 92, 1088–1091.

10. Specchia, G., Albano, F., Anelli, L., Storlazzi, C.T., Zagaria, A., Liso, A., Pannunzio, A., Pastore, D., Mestice, A., Greco, G., et al. (2004). Derivative chromosome 9 deletions in chronic myeloid leukemia are associated with loss of tumor suppressor genes. Leuk. Lymphoma 45, 689–694.

11. Berger, A.H., and Pandolfi, P.P. (2011). Haplo-insufficiency: a driving force in cancer. J. Pathol. 223, 137–146.

12. Fero, M.L., Randel, E., Gurley, K.E., Roberts, J.M., and Kemp, C.J. (1998). The murine gene p27Kip1 is haplo-insufficient for tumour suppression. Nature 396, 177–180.

13. Crawford, J.R., Rood, B.R., Rossi, C.T., and Vezina, G. (2009). Medulloblastoma associated with novel PTCH mutation as primary manifestation of Gorlin syndrome. Neurology 72, 1618.

14. Hirbe, A.C., Dahiya, S., Miller, C.A., Li, T., Fulton, R.S., Zhang, X., McDonald, S., DeSchryver, K., Duncavage, E.J., Walrath, J., et al. (2015). Whole Exome Sequencing Reveals the Order of Genetic Changes during Malignant Transformation and Metastasis in a Single Patient with NF1-plexiform Neurofibroma. Clin. Cancer Res. 21, 4201–4211.

15. Chen, X., Yan, C.C., Zhang, X., and You, Z.H. (2017). Long non-coding RNAs and complex diseases: from experimental results to computational models. Brief. Bioinform. 18, 558–576.

16. Chen, X., and Yan, G.Y. (2013). Novel human lncRNA-disease association inference based on lncRNA expression profiles. Bioinformatics 29, 2617–2624.

17. Chen, X., and Huang, L. (2017). LRSSLMDA: Laplacian Regularized Sparse Subspace Learning for MiRNA-Disease Association prediction. PLoS Comput. Biol. 13, e1005912.

18. Chen, X., Xie, D., Zhao, Q., and You, Z.H. (2017). MicroRNAs and complex diseases: from experimental results to computational models. Brief. Bioinform. , Published online October 17, 2017. https://doi.org/10.1093/bib/bbx130.

19. You, Z.H., Huang, Z.A., Zhu, Z., Yan, G.Y., Li, Z.W., Wen, Z., and Chen, X. (2017). PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. PLoS Comput. Biol. 13, e1005455.

20. Chen, X., Huang, L., Xie, D., and Zhao, Q. (2018). EGBMMDA: Extreme Gradient Boosting Machine for MiRNA-Disease Association prediction. Cell Death Dis. 9, 3.

21. Chen, L., Yang, J., Huang, T., Kong, X., Lu, L., and Cai, Y.-D. (2016). Mining for novel tumor suppressor genes using a shortest path approach. J. Biomol. Struct. Dyn. 34, 664–675.

22. Oricchio, E., Nanjangud, G., Wolfe, A.L., Schatz, J.H., Mavrakis, K.J., Jiang, M., Liu, X., Bruno, J., Heguy, A., Olshen, A.B., et al. (2011). The Eph-receptor A7 is a soluble tumor suppressor for follicular lymphoma. Cell 147, 554–564.

23. Peifer, M., Fernández-Cuesta, L., Sos, M.L., George, J., Seidel, D., Kasper, L.H., Plenker, D., Leenders, F., Sun, R., Zander, T., et al. (2012). Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat. Genet. 44, 1104–1110.

24. Sepulveda, J.L., Gutierrez-Pajares, J.L., Luna, A., Yao, Y., Tobias, J.W., Thomas, S., Woo, Y., Giorgi, F., Komissarova, E.V., Califano, A., et al. (2016). High-definition CpG methylation of novel genes in gastric carcinogenesis identified by next-generation sequencing. Mod. Pathol. 29, 182–193.

25. Foster, R.E., Abdulrahman, M., Morris, M.R., Prigmore, E., Gribble, S., Ng, B., Gentle, D., Ready, S., Weston, P.M., Wiesener, M.S., et al. (2007). Characterization of a 3;6 translocation associated with renal cell carcinoma. Genes Chromosomes Cancer 46, 311–317.

26. Li, S., Wu, Z., Ma, P., Xu, Y., Chen, Y., Wang, H., He, P., Kang, Z., Yin, L., Zhao, Y., et al. (2017). Ligand-dependent EphA7 signaling inhibits prostate tumor growth and progression. Cell Death Dis. 8, e3122.

27. Wu, X., Yan, L., Liu, Y., Xian, W., Wang, L., and Ding, X. (2017). MicroRNA-448 suppresses osteosarcoma cell proliferation and invasion through targeting EPHA7. PLoS ONE 12, e0175553.

28. Sapkota, G.P. (2013). The TGFβ-induced phosphorylation and activation of p38 mitogen-activated protein kinase is mediated by MAP3K4 and MAP3K10 but not TAK1. Open Biol. 3, 130067.

29. Miller, A.L., Webb, M.S., Copik, A.J., Wang, Y., Johnson, B.H., Kumar, R., and Thompson, E.B. (2005). p38 Mitogen-activated protein kinase (MAPK) is a key mediator in glucocorticoid-induced apoptosis of lymphoid cells: correlation between p38 MAPK activation and site-specific phosphorylation of the human glucocorticoid receptor at serine 211. Mol. Endocrinol. 19, 1569–1583.

30. Katoh, Y., and Katoh, M. (2008). Integrative genomic analyses on GLI2: mechanism of Hedgehog priming through basal GLI2 expression, and interaction map of stem cell signaling network with P53. Int. J. Oncol. 33, 881–886.

31. Nobes, C.D., Lauritzen, I., Mattei, M.G., Paris, S., Hall, A., and Chardin, P. (1998). A new member of the Rho family, Rnd1, promotes disassembly of actin filament structures and loss of cell adhesion. J. Cell Biol. 141, 187–197.

32. Madigan, J.P., Bodemann, B.O., Brady, D.C., Dewar, B.J., Keller, P.J., Leitges, M., Philips, M.R., Ridley, A.J., Der, C.J., and Cox, A.D. (2009). Regulation of Rnd3 localization and function by protein kinase C alpha-mediated phosphorylation. Biochem. J. 424, 153–161.

33. Ongusaha, P.P., Kim, H.G., Boswell, S.A., Ridley, A.J., Der, C.J., Dotto, G.P., Kim, Y.B., Aaronson, S.A., and Lee, S.W. (2016). RhoE Is a Pro-Survival p53 Target Gene that Inhibits ROCK I-Mediated Apoptosis in Response to Genotoxic Stress. Curr. Biol. 26, 2221–2222.

34. Riou, P., Villalonga, P., and Ridley, A.J. (2010). Rnd proteins: multifunctional regulators of the cytoskeleton and cell cycle progression. BioEssays 32, 986–992.

35. Kurokawa, D., Ohmura, T., Sakurai, Y., Inoue, K., Suda, Y., and Aizawa, S. (2014). Otx2 expression in anterior neuroectoderm and forebrain/midbrain is directed by more than six enhancers. Dev. Biol. 387, 203–213.

36. Li, J., Di, C., Jing, J., Di, Q., Nakhla, J., and Adamson, D.C. (2015). OTX2 is a therapeutic target for retinoblastoma and may function as a common factor between C-MYC, CRX, and phosphorylated RB pathways. Int. J. Oncol. 47, 1703–1710.

37. Gestri, G., Carl, M., Appolloni, I., Wilson, S.W., Barsacchi, G., and Andreazzoli, M. (2005). Six3 functions in anterior neural plate specification by promoting cell proliferation and inhibiting BMP4 expression. Development 132, 2401–2413.

38. Mulligan, L.M. (2014). RET revisited: expanding the oncogenic portfolio. Nat. Rev. Cancer 14, 173–186.

39. Moccia, M., Liu, Q., Guida, T., Federico, G., Brescia, A., Zhao, Z., Choi, H.G., Deng, X., Tan, L., Wang, J., et al. (2015). Identification of Novel Small Molecule Inhibitors of Oncogenic RET Kinase. PLoS ONE 10, e0128364.

40. Macià, A., Gallel, P., Vaquero, M., Gou-Fabregas, M., Santacana, M., Maliszewska, A., Robledo, M., Gardiner, J.R., Basson, M.A., Matias-Guiu, X., and Encinas, M. (2012). Sprouty1 is a candidate tumor-suppressor gene in medullary thyroid carcinoma. Oncogene 31, 3961–3972.

41. Rutter, M., Wang, J., Huang, Z., Kuliszewski, M., and Post, M. (2010). Gli2 influences proliferation in the developing lung through regulation of cyclin expression. Am. J. Respir. Cell Mol. Biol. 42, 615–625.

42. Yao, Z., Hu, K., Huang, H., Xu, S., Wang, Q., Zhang, P., Yang, P., and Liu, B. (2015). shRNA-mediated silencing of the RFC3 gene suppresses hepatocellular carcinoma cell proliferation. Int. J. Mol. Med. 36, 1393–1399.

43. Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S.J., and Qin, J. (2000). BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. Genes Dev. 14, 927–939.

44. Strefford, J.C., Stasevich, I., Lane, T.M., Lu, Y.J., Oliver, T., and Young, B.D. (2005). A combination of molecular cytogenetic analyses reveals complex genetic alterations in conventional renal cell carcinoma. Cancer Genet. Cytogenet. 159, 1–9.

45. Peshenko, I.V., Olshevskaya, E.V., Savchenko, A.B., Karan, S., Palczewski, K., Baehr, W., and Dizhoor, A.M. (2011). Enzymatic properties and regulation of the native isozymes of retinal membrane guanylyl cyclase (RetGC) from mouse photoreceptors. Biochemistry 50, 5590–5600.

46. Shao, R.X., Kato, N., Lin, L.J., Muroyama, R., Moriyama, M., Ikenoue, T., Watabe, H., Otsuka, M., Guleng, B., Ohta, M., et al. (2007). Absence of tyrosine kinase mutations in Japanese colorectal cancer patients. Oncogene 26, 2133–2135.

47. Xue, T.C., Ge, N.L., Zhang, L., Cui, J.F., Chen, R.X., You, Y., Ye, S.L., and Ren, Z.G. (2014). Goosecoid promotes the metastasis of hepatocellular carcinoma by modulating the epithelial-mesenchymal transition. PLoS ONE 9, e109695.

48. Song, L.N., Silva, J., Koller, A., Rosenthal, A., Chen, E.I., and Gelmann, E.P. (2015). The Tumor Suppressor NKX3.1 Is Targeted for Degradation by DYRK1B Kinase. Mol. Cancer Res. 13, 913–922.

49. Seol, J.Y., Mihich, E., and Berleth, E.S. (2015). TNF Apoptosis Protection Fraction (TAPF) prevents apoptosis induced by TNF, but not by Fas or TRAIL, via NF-κB-induced increase in cFLIP. Cytokine 75, 321–329.

50. Dai, X., Zhang, J., Arfuso, F., Chinnathambi, A., Zayed, M.E., Alharbi, S.A., Kumar, A.P., Ahn, K.S., and Sethi, G. (2015). Targeting TNF-related apoptosis-inducing ligand (TRAIL) receptor by natural products as a potential therapeutic approach for cancer therapy. Exp. Biol. Med. (Maywood) 240, 760–773.

51. Kim, M., Kim, T., Johnson, R.L., and Lim, D.S. (2015). Transcriptional co-repressor function of the hippo pathway transducers YAP and TAZ. Cell Rep. 11, 270–282.

52. Miquelajáuregui, A., Varela-Echavarría, A., Ceci, M.L., García-Moreno, F., Ricaño, I., Hoang, K., Frade-Pérez, D., Portera-Cailliau, C., Tamariz, E., De Carlos, J.A., et al. (2010). LIM-homeobox gene Lhx5 is required for normal development of Cajal-Retzius cells. J. Neurosci. 30, 10551–10562.

53. Kamalakaran, S., Varadan, V., Giercksky Russnes, H.E., Levy, D., Kendall, J., Janevski, A., Riggs, M., Banerjee, N., Synnestvedt, M., Schlichting, E., et al. (2011). DNA methylation patterns in luminal breast cancers differ from non-luminal subtypes

and can identify relapse risk independent of other clinical variables. Mol. Oncol. *5*, 77–92.

54. Dyer, M.A. (2003). Regulation of proliferation, cell fate specification and differentiation by the homeodomain proteins Prox1, Six3, and Chx10 in the developing retina. Cell Cycle *2*, 350–357.

55. Li, S., Mo, Z., Yang, X., Price, S.M., Shen, M.M., and Xiang, M. (2004). Foxn4 controls the genesis of amacrine and horizontal cells by retinal progenitors. Neuron *43*, 795–807.

56. Sasaki, T., Komatsu, Y., Watakabe, A., Sawada, K., and Yamamori, T. (2010). Prefrontal-enriched SLIT1 expression in Old World monkey cortex established during the postnatal development. Cereb. Cortex *20*, 2496–2510.

57. Noisa, P., Lund, C., Kanduri, K., Lund, R., Lähdesmäki, H., Lahesmaa, R., Lundin, K., Chokechuwattanalert, H., Otonkoski, T., Tuuri, T., and Raivio, T. (2014). Notch signaling regulates the differentiation of neural crest from human pluripotent stem cells. J. Cell Sci. *127*, 2083–2094.

58. Merino, D., Best, S.A., Asselin-Labat, M.L., Vaillant, F., Pal, B., Dickins, R.A., Anderson, R.L., Strasser, A., Bouillet, P., Lindeman, G.J., and Visvader, J.E. (2015). Pro-apoptotic Bim suppresses breast tumor cell metastasis and is a target gene of SNAI2. Oncogene *34*, 3926–3934.

59. Ogawa, H., Takyu, R., Morimoto, H., Toei, S., Sakon, H., Goto, S., Moriya, S., and Kono, T. (2016). Cell proliferation potency is independent of FGF4 signaling in trophoblast stem cells derived from androgenetic embryos. J. Reprod. Dev. *62*, 51–58.

60. Katoh, M. (2016). FGFR inhibitors: Effects on cancer cells, tumor microenvironment and whole-body homeostasis (Review). Int. J. Mol. Med. *38*, 3–15.

61. Fedrigo, O., Warner, L.R., Pfefferle, A.D., Babbitt, C.C., Cruz-Gordillo, P., and Wray, G.A. (2010). A pipeline to determine RT-QPCR control genes for evolutionary studies: application to primate gene expression across multiple tissues. PLoS ONE *5*, e12545.

62. Grassilli, E., Ianzano, L., Bonomo, S., Missaglia, C., Cerrito, M.G., Giovannoni, R., Masiero, L., and Lavitrano, M. (2014). GSK3A is redundant with GSK3B in modulating drug resistance and chemotherapy-induced necroptosis. PLoS ONE *9*, e100947.

63. Korinek, V., Barker, N., Morin, P.J., van Wichen, D., de Weger, R., Kinzler, K.W., Vogelstein, B., and Clevers, H. (1997). Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science *275*, 1784–1787.

64. Binet, R., Ythier, D., Robles, A.I., Collado, M., Larrieu, D., Fonti, C., Brambilla, E., Brambilla, C., Serrano, M., Harris, C.C., and Pedeux, R. (2009). WNT16B is a new marker of cellular senescence that regulates p53 activity and the phosphoinositide 3-kinase/AKT pathway. Cancer Res. *69*, 9183–9191.

65. Ruest, L.B., Hammer, R.E., Yanagisawa, M., and Clouthier, D.E. (2003). Dlx5/6-enhancer directed expression of Cre recombinase in the pharyngeal arches and brain. Genesis *37*, 188–194.

66. Su, P., Yuan, Y., Huang, Y., Wang, W., and Zhang, Z. (2013). Anorectal malformation associated with a mutation in the P63 gene in a family with split hand-foot malformation. Int. J. Colorectal Dis. *28*, 1621–1627.

67. Rakowiecki, S., and Epstein, D.J. (2013). Divergent roles for Wnt/β-catenin signaling in epithelial maintenance and breakdown during semicircular canal formation. Development *140*, 1730–1739.

68. Zhao, M., Sun, J., and Zhao, Z. (2013). TSGene: a web resource for tumor suppressor genes. Nucleic Acids Res. *41*, D970–D976.

69. Sharan, R., Ulitsky, I., and Shamir, R. (2007). Network-based prediction of protein function. Mol. Syst. Biol. *3*, 88.

70. Ng, K.L., Ciou, J.S., and Huang, C.H. (2010). Prediction of protein functions based on function-function correlation relations. Comput. Biol. Med. *40*, 300–305.

71. Chen, L., Yang, J., Xing, Z., Yuan, F., Shu, Y., Zhang, Y., Kong, X., Huang, T., Li, H., and Cai, Y.D. (2017). An integrated method for the identification of novel genes related to oral cancer. PLoS ONE *12*, e0175185.

72. Li, J., Chen, L., Wang, S., Zhang, Y., Kong, X., Huang, T., and Cai, Y.D. (2018). A computational method using the random walk with restart algorithm for identifying novel epigenetic factors. Mol. Genet. Genomics *293*, 293–301.

73. Zhang, J., Yang, J., Huang, T., Shu, Y., and Chen, L. (2016). Identification of novel proliferative diabetic retinopathy related genes on protein-protein interaction network. Neurocomputing *217*, 63–72.

74. Chen, L., Xing, Z., Huang, T., Shu, Y., Huang, G., and Li, H.-P. (2016). Application of the Shortest Path Algorithm for the Discovery of Breast Cancer-Related Genes. Curr. Bioinform. *11*, 51–58.

75. Shen, S., Gui, T., and Ma, C. (2017). Identification of molecular biomarkers for pancreatic cancer with mRMR shortest path method. Oncotarget *8*, 41432–41439.

76. Chen, L., Zhang, Y.H., Huang, T., and Cai, Y.D. (2016). Identifying novel protein phenotype annotations by hybridizing protein-protein interactions and protein sequence similarities. Mol. Genet. Genomics *291*, 913–934.

77. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

78. Leiserson, M.D., Vandin, F., Wu, H.T., Dobson, J.R., Eldridge, J.V., Thomas, J.L., Papoutsaki, A., Kim, Y., Niu, B., McLellan, M., et al. (2015). Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nat. Genet. *47*, 106–114.

79. Köhler, S., Bauer, S., Horn, D., and Robinson, P.N. (2008). Walking the interactome for prioritization of candidate disease genes. Am. J. Hum. Genet. *82*, 949–958.

80. Lu, S., Yan, Y., Li, Z., Chen, L., Yang, J., Zhang, Y., Wang, S., and Liu, L. (2017). Determination of genes related to uveitis by utilization of the random walk with restart algorithm on a protein–protein interaction network. Int. J. Mol. Sci. *18*, 1045.

81. Chen, L., Liu, T., and Zhao, X. (2018). Inferring anatomical therapeutic chemical (ATC) class of drugs using shortest path and random walk with restart algorithms. Biochim. Biophys. Acta *1864*, 2228–2240.

82. Li, Y., and Patra, J.C. (2010). Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network. Bioinformatics *26*, 1219–1224.

83. Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J.M., and Pascual-Montano, A. (2007). GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome Biol. *8*, R3.

84. Chen, L., Zhang, Y.-H., Lu, G., Huang, T., and Cai, Y.-D. (2017). Analysis of cancer-related lncRNAs using gene ontology and KEGG pathways. Artif. Intell. Med. *76*, 27–36.

85. Huang, T., Chen, L., Cai, Y.D., and Chou, K.C. (2011). Classification and analysis of regulatory pathways using graph property, biochemical and physicochemical property, and functional property. PLoS ONE *6*, e25297.

86. Chen, L., Zhang, Y.H., Zheng, M., Huang, T., and Cai, Y.D. (2016). Identification of compound-protein interactions through the analysis of gene ontology, KEGG enrichment for proteins and molecular fragments of compounds. Mol. Genet. Genomics *291*, 2065–2079.