

RESEARCH ARTICLE

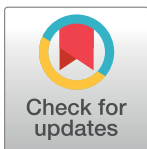
An integrated method for the identification of novel genes related to oral cancer

Lei Chen^{1,2}, Jing Yang¹, Zhihao Xing³, Fei Yuan⁴, Yang Shu³, YunHua Zhang⁵, XiangYin Kong³, Tao Huang^{3*}, HaiPeng Li^{6*}, Yu-Dong Cai^{1*}

1 School of Life Sciences, Shanghai University, Shanghai, People's Republic of China, **2** College of Information Engineering, Shanghai Maritime University, Shanghai, People's Republic of China, **3** Institute of Health Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, People's Republic of China, **4** Department of Science & Technology, Binzhou Medical University Hospital, Binzhou, Shandong, People's Republic of China, **5** School of Resources and Environment, Anhui Agricultural University, Hefei, Anhui, People's Republic of China, **6** CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, People's Republic of China

☯ These authors contributed equally to this work.

* tohuangtao@126.com (TH); lihaipeng@picb.ac.cn (HPL); cai_yud@126.com (YDC)



OPEN ACCESS

Citation: Chen L, Yang J, Xing Z, Yuan F, Shu Y, Zhang Y, et al. (2017) An integrated method for the identification of novel genes related to oral cancer. PLoS ONE 12(4): e0175185. <https://doi.org/10.1371/journal.pone.0175185>

Editor: Quan Zou, Tianjin University, CHINA

Received: January 5, 2017

Accepted: March 21, 2017

Published: April 6, 2017

Copyright: © 2017 Chen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by the National Natural Science Foundation of China (31371335), Natural Science Foundation of Shanghai (17ZR1412500), Shanghai Sailing Program and The Youth Innovation Promotion Association of Chinese Academy of Sciences (CAS) (2016245).

Competing interests: The authors have declared that no competing interests exist.

Abstract

Cancer is a significant public health problem worldwide. Complete identification of genes related to one type of cancer facilitates earlier diagnosis and effective treatments. In this study, two widely used algorithms, the random walk with restart algorithm and the shortest path algorithm, were adopted to construct two parameterized computational methods, namely, an RWR-based method and an SP-based method; based on these methods, an integrated method was constructed for identifying novel disease genes. To validate the utility of the integrated method, data for oral cancer were used, on which the RWR-based and SP-based methods were trained, thereby building two optimal methods. The integrated method combining these optimal methods was further adopted to identify the novel genes of oral cancer. As a result, 85 novel genes were inferred, among which eleven genes (*e.g.*, MYD88, FGFR2, NF- κ BIA) were identified by both the RWR-based and SP-based methods, 70 genes (*e.g.*, BMP4, IFNG, KITLG) were discovered only by the RWR-based method and four genes (L1R1, MCM6, NOG and CXCR3) were predicted only by the SP-based method. Extensive analyses indicate that several novel genes have strong associations with cancers, indicating the effectiveness of the integrated method for identifying disease genes.

Introduction

Cancer is a significant public health problem worldwide. Oral cancer (OC) is a subgroup of head and neck cancer; it develops on the lips, tongue, salivary glands, gingiva, oropharynx, and on buccal surfaces [1]. Oral squamous cell carcinoma (OSCC) accounts for more than 90% of all OC [2]. OC is estimated by the WHO (World Health Organization) to be the eleventh most common cancer in the world, accounting for 300,000 new cases and 145,000 deaths in 2012 [3, 4]. The incidence of OC exhibits significant local variation and continues to be high in

India, East Asia, Eastern Europe, and parts of South America [5]. Tobacco and alcohol are the most important risk factors for OC [6]. Poor nutrition, genetic factors and viral infection may be potential risk factors for OC [7, 8].

Carcinogenesis is a multi-step process, and a variety of alterations accumulate, driving and gradually increasing tumorigenesis [9]. During the past decade, a large variety of genomic variations has been implicated in OC. EGFR (epidermal growth factor receptor) amplification and over-expression were found in a large proportion of oral tumors [10–12]. ErbB2 amplification and overexpression appear to occur frequently in OC specimens, and high levels of ErbB2 may be related to a worse prognosis of patients [12]. A strong correlation has been detected between c-erbB-2 overexpression and overall survival of patients with oral squamous cell carcinoma [13, 14]. The cyclin family plays a critical role in cell cycle progression. Aberrant up-expression of cyclin D accounts for 36–66% of OC. Amplification and an SNP of cyclin D may be associated with a worse prognosis and susceptibility to OC, respectively [15, 16]. Several lines of evidence suggest that the p53 tumor suppression network is altered in OC [17, 18]. In addition, many other target genes have been reported, such as ras, VEGF (vascular endothelial growth factor), and the MMP (matrix metalloproteinases) family [19–21]. To the best of our knowledge, the mechanism that underlies OC is still unclear. A search for new genes related to OC may facilitate earlier diagnosis and effective treatment.

In general, many standard methods have been used for detection of virulence genes. In the hospital and laboratory, cancer samples are sliced into pathological sections and stained to determine the disease pathology type. Quantitative reverse transcription-PCR is typically used to detect the mRNA level of genes in cancer. IHC (immunohistochemical) and western blotting measure the expression of related proteins in tissues and cells, respectively. However, it is difficult to analyze genes synthetically and comprehensively. In addition, some large-scale experiments have been exploited to screen virulence genes, such as microarray, GWAS (genome-wide association study) and NGS (next-generation sequencing). The most common techniques are time-consuming and costly, and thus new methods must be explored to identify tumor genes. In recent years, with the development of computer techniques [22–28], some of them can be applied to tackle this problem. Up to now, several computational methods have been proposed to identify disease genes. Many of them are based on guilt-by-association [29], i.e., the assumption that genes are similar to their neighbors in a gene network. Thus, the neighbors of the disease genes are more likely to be disease genes. However, these types of methods are local methods that use only part of the network. Thus, they do not always yield good performance. Many other methods employ the Random Walk with Restart (RWR) algorithm to identify disease genes [30–32]. This algorithm simulates a walker starting from a seed node or a set of seed nodes that represent disease genes and random walking on the network. The probability of a node being a disease gene is updated until the probabilities of all nodes become stable. The genes corresponding to nodes with high probabilities are selected as novel candidate disease genes. Recently, computational methods have adopted the shortest path (SP) algorithm to address the problem [33–39]. This algorithm assumes that genes lying in the shortest paths connecting any two disease genes may also be disease genes. Clearly, methods based on RWR or SP algorithms take full advantage of the network compared with those based on guilt-by-association. Thus, they can yield clues for the discovery of novel disease genes.

In this study, we used the RWR and SP algorithms to construct two novel computational methods, the RWR-based method and the SP-based method, respectively, for the identification of novel disease genes. Additionally, an integrated method was constructed by combining these two methods. To indicate the effectiveness of the integrated method, disease genes of OC were employed. It has been reported in some studies that using only the RWR algorithm and SP algorithm consistently produces several false discoveries [33, 40], which may be caused by

the structure of the network or lack of consideration of the essential properties of genes. Thus, further rules, produced by the permutation test, associations between candidates and the validation of genes according to their properties, were integrated into the RWR-based and SP-based methods. These methods were executed using the corresponding large network that was built using protein-protein interaction (PPI) information and that was trained on validated OC genes to determine the optimal parameters. The obtained optimal methods were used to infer novel genes related to OC, and the integrated method combined the predicted genes using these two optimal methods to yield 85 novel genes. Among them, eleven genes were obtained by both methods, 70 genes were obtained only by the RWR-based method and four genes were obtained only by the SP-based method. According to the analyses, several genes show stimulative or suppressive effects on cancers by experiments or have a certain relationship with cancers reported in published papers, indicating the utility of the integrated method. It is also clear that the integrated method can provide more comprehensive analysis of various diseases because it is capable of producing more possible disease genes than the RWR-based method or the SP-based method.

Materials and methods

Genes related to oral cancer

The OC-related genes were collected from the following three sources: (1) 44 genes were retrieved from UniProt (<http://www.uniprot.org/>, accessed in July, 2015) [41] after ‘human oral cancer reviewed’ was input as a keyword; (2) seven genes were chosen from the catalogue of oral cancer from the TSGene (<https://bioinfo.uth.edu/TSGene/>, accessed in July, 2015) database; and (3) 156 genes were retrieved from the NCI (National Cancer Institute, <https://gforge.nci.nih.gov>, accessed in July, 2015) database using ‘Homo sapiens’ as a keyword. After combining the OC-related genes mentioned above, 202 OC-related genes were finally obtained; these genes are provided in [S1 Table](#). Because our methods are based on the network constructed from the PPI information retrieved from Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) [42], these 202 OC-related genes were mapped into their Ensembl IDs, and those not occurring in the network were discarded. One hundred eighty-four Ensembl IDs of OC-related genes were accessed and composed the set S_{oc} .

Protein-protein interactions

PPIs play important roles in several intracellular and intercellular biochemical processes. The known PPI information has been widely used to investigate several protein-related problems, such as protein function prediction [43, 44] and disease gene identification [33, 34, 36, 39]. Based on the results reported in these previous studies, it can be concluded that two proteins that can interact with each other always have a functional relationship. In this study, we attempted to discover novel candidate OC-related genes using validated OC-related genes. Thus, we can utilize PPIs to search for proteins that have a functional relationship with proteins encoded by validated OC-related genes, thereby mining novel OC-related genes.

The PPI information was retrieved from STRING (<http://string-db.org/>, version 9.1) [42], a well-known online public database that contains known and predicted protein interactions. The interactions reported in STRING are derived from the following four types of sources: (I) Genomic Context; (II) High-throughput Experiments; (III) (Conserved) Coexpression; (IV) Previous Knowledge, which include direct (physical) and indirect (functional) associations between proteins and thus offer more chances to mine hidden protein information. From the database, a file called ‘[protein.links.v9.1.txt.gz](#)’ was retrieved, which contained 2,425,314 human PPIs. Each PPI was represented by two Ensembl IDs and a score ranging between 150

and 999, which indicates the strength of the interaction; i.e., proteins in interactions with higher scores are regarded as more likely to interact with each other. Let us denote the score of a PPI between proteins p_1 and p_2 by $S(p_1, p_2)$.

RWR-based method

RWR is a ranking algorithm [30]. In a network, this algorithm simulates a walker starting from a seed node or a set of seed nodes and random walking on the network. In each step, the probability of the walker walking to each node is updated, and it stabilizes after several steps. Nodes in the network are ranked by the final probabilities assigned to them. Based on the validated disease genes (used as seed nodes), the RWR algorithm can be used to discover novel disease genes by investigating novel genes with high ranks. This algorithm has therefore been used to mine novel disease genes in recent years [30–32]. However, the RWR algorithm consistently provides false discoveries [40], supporting the need for additional screening rules. In this study, we used the RWR algorithm as the basic algorithm and added some rules to build the present RWR-based learning method for the identification of novel OC-related genes. Furthermore, to obtain a better RWR-based method, some parameters were employed that are determined by training the method.

Network construction for RWR algorithm. An accurate network is important for the identification of novel OC-related genes. Here, we adopted the PPI information mentioned in the Section “Protein-protein interactions” to build the network. The constructed network defined the 20,770 proteins occurring in the 2,425,314 human PPIs as nodes, and two nodes were adjacent if and only if the corresponding proteins could interact with each other. It can be observed that each edge represented a PPI. To employ the interaction score in the network, each edge was assigned a weight that was defined as the score of the corresponding PPI. For convenience, let us denote the constructed network by N_{RWR} .

Searching for new candidate genes using the RWR algorithm. As mentioned in the Section “Genes related to oral cancer”, a set S_{oc} , consisting of 184 Ensembl IDs of OC-related genes was used in this study. The RWR algorithm simulated a random walker starting from S_{oc} . Before executing the RWR algorithm on N_{RWR} , each node in N_{RWR} was assigned a probability, with $1/184$ set for the Ensembl IDs in S_{oc} and 0 set for the remaining nodes. The initial probability of each node constituted the probability vector P_0 . The RWR algorithm updated the probability vector in each step. Let P_t denote the probability vector after performing the t -th step, which can be updated by

$$P_{t+1} = (1 - r)A^T P_t + rP_0 \quad (1)$$

where r is set to 0.8 and A is the column-wise normalized adjacency matrix of N_{RWR} . The update procedure was repeated until the change between P_t and P_{t+1} , measured by L_1 norm, was less than $1e-6$. Ensembl IDs with high probabilities were considered to be encoded by OC-related genes. Because we did not know which probability was a suitable threshold for selecting candidate genes, we established a parameter, p_{RWR}^p , for the threshold, which can be determined by training the RWR-based method.

SP-based method

The SP algorithm is a classic graph algorithm. In recent years, some investigators have applied the SP algorithm for the identification of disease genes [33–39]. The new candidate genes were extracted from the shortest paths connecting any two validated disease genes. Here, we built an SP-based machine learning method to identify novel OC-related genes. Like the RWR-

based method, further screening rules were established to discard non-essential candidate genes and select important ones.

Network construction for SP algorithm. Similar to N_{RWR} for the RWR-based method, we also constructed the network N_{SP} . As mentioned in the Section “Network construction for RWR-based method”, each edge in N_{RWR} represented a PPI. In addition, each PPI had an interaction score as described in the Section “Protein-protein interactions”. This score was used to define the weight of the edge. The range of the interaction score was between 150 and 999, and an edge with a low weight indicated strong correlations between its endpoints in the SP-based model. Thus, for an edge e with endpoints n_1 and n_2 , weight was defined in the following manner:

$$w(e) = 1000 - S(p_1, p_2) \tag{2}$$

where p_1 and p_2 were two corresponding proteins of nodes n_1 and n_2 . The constructed network N_{SP} consisted of 20,770 nodes and 2,425,314 edges.

Searching for new candidate genes using the SP algorithm. The SP algorithm, Dijkstra’s algorithm [45], was applied to the network N_{SP} to search for all of the shortest paths connecting any two Ensembl IDs in S_{oc} . From the obtained shortest paths, we extracted the inner nodes that did not represent validated OC-related genes. Genes corresponding to the extracted nodes were considered to be related to OC and denoted candidate genes. In addition, each candidate gene was assigned a measurement, called betweenness, which was defined as the number of shortest paths within which it was contained.

Screening rules

The RWR and SP algorithms can produce a number of candidate genes for OC when the probability threshold is provided for the RWR algorithm. However, several false discoveries are inevitable, as mentioned in previous studies [33, 40]. To screen out these false discoveries, a series of screening rules were added to this section and integrated into the RWR-based and SP-based methods. Some parameters were employed to build the rules. Their optimal values in RWR-based and SP-based methods are determined by training these two methods, respectively.

The probability that a candidate gene produced by the RWR algorithm was influenced by the structure of the network N_{RWR} and the betweenness for a candidate gene produced by the SP algorithm was also clearly influenced by the structure of the N_{SP} network. Some candidate genes receiving high probabilities (betweenness) were not specific to OC. Thus, a permutation test was designed that first randomly constructed 1,000 sets of Ensembl IDs such that each set had the same size of S_{oc} , denoted by $S_1, S_2, \dots, S_{1000}$. For each set, the RWR algorithm and SP algorithm were executed on N_{RWR} and N_{SP} , respectively, by setting the Ensembl IDs in the set as the input, yielding a probability for each candidate gene produced by the RWR algorithm and a betweenness for each candidate gene produced by the SP algorithm. Finally, for each candidate gene, there was one probability (betweenness) for S_{oc} and 1,000 probabilities (betweenness) for $S_1, S_2, \dots, S_{1000}$. If the candidate gene was specific to OC, its probability (betweenness) for S_{oc} should clearly be larger than most probabilities (betweenness) for $S_1, S_2, \dots, S_{1000}$. Thus, we calculated the p-value for each candidate gene g , which was defined by

$$p - \text{value}(g) = \frac{\sum_{i=1}^{1000} \delta_i}{1000} \tag{3}$$

where $\delta_i = 1$ if the probability (betweenness) for S_i was larger than that of S_{oc} ; $\delta_i = 0$ otherwise.

Because 0.05 was always used as an important cut-off for the significance level of the test, it was set to be the threshold of the p-value; *i.e.*, candidate genes with p-values less than 0.05 were selected.

The network-based method is useful for complicated problems. However, this type of method seldom considers the essential properties of nodes in the network, leading to several false discoveries. Rules utilizing the essential properties of candidate and validated genes are needed to exclude false discoveries and select important ones.

As mentioned in the Section “Protein-protein interactions”, two proteins that can interact with each other always have a functional relationship. Furthermore, considering the interaction scores, proteins in an interaction with a high score are more likely to have a strong functional relationship than those in an interaction with a low score. Thus, the interaction score can be used to measure the associations between candidate genes and OC-related genes. For each candidate gene g , we calculated the maximum interaction score (MIS), which was defined by

$$MIS(g) = \max\{S(g, g') : g' \in S_{oc}\} \quad (4)$$

Clearly, a candidate gene with a high MIS is more likely to be a novel OC-related gene and thus should be selected. However, the threshold of MIS was not easy to determine. Thus, it was set as a parameter, p_{MIS} . In the RWR-based method, the optimal value of p_{MIS} is determined by training this method. In addition, the optimal value of p_{MIS} in the SP-based method is also accessed by training the SP-based method.

The MIS measures the associations between candidate genes and OC-related genes. The following measurement evaluates the associations between them in another way. It is known that OC-related genes must be highly related to some gene ontology (GO) terms or biological pathways. Thus, we can use the annotated GO term and KEGG pathway information for candidate genes and OC-related genes to evaluate their correlations. To achieve this goal, each candidate gene or OC-related gene was encoded by its GO enrichment scores and KEGG enrichment scores. For a candidate gene g and an OC-related gene g' , their associations can be measured by

$$M(g, g') = \frac{ES(g) \cdot ES(g')}{\|ES(g)\| \cdot \|ES(g')\|} \quad (5)$$

where $ES(g)$ ($ES(g')$, respectively) is a vector consisting of the GO enrichment scores and KEGG enrichment scores of g (g' , respectively). A high value for Eq 5 indicated a strong association. Similar to the definition of MIS , we calculated the maximum function score (MFS) of each candidate gene g by

$$MFS(g) = \max\{M(g, g') : g' \in S_{oc}\} \quad (6)$$

Similarly, a candidate gene with a high MFS might be a novel OC-related gene with a high probability and should be selected. Additionally, it was difficult to determine the threshold of MFS . The parameter p_{MFS} was also established for this threshold. Similar to parameter p_{MIS} mentioned above, the optimal value of p_{MFS} in the RWR-based and SP-based methods is determined by training these two methods, respectively.

Integrated method

The RWR-based and SP-based methods, together with the screen rules mentioned in the Section “Screening rules”, were constructed. The pseudo-codes of these two methods are listed in Tables 1 and 2, respectively. The integrated method encompassed these two methods by combining their results.

Table 1. The pseudo-code of the RWR-based method.

RWR-based method
Input: An OC-related gene set, S_{oc} ; a network, N_{RWR}
Output: A number of putative OC-related genes
1. Execute the RWR algorithm on N_{RWR} using S_{oc} as the input, producing a probability for each gene in N_{RWR} ; select candidate genes with a probability higher than p_{RWR}^o ;
2. Execute the permutation test, producing the p-value for each gene; select candidate genes with a p-value less than 0.05;
3. For each candidate gene, calculate its <i>MIS</i> and select candidate genes with an <i>MIS</i> no less than p_{MIS} ;
4. For each candidate gene, calculate its <i>MFS</i> and select candidate genes with an <i>MFS</i> larger than p_{MFS} ;
5. Output the remaining candidate genes as the putative OC-related genes.

<https://doi.org/10.1371/journal.pone.0175185.t001>

Evaluation methods

The RWR-based and SP-based methods were applied to identify novel OC-related genes. However, there are some parameters in these two methods that must be determined before they are used to identify novel OC-related genes. Thus, these two methods were trained on the validated OC-related gene set S_{oc} , through which the optimal parameters can be determined. We used the jackknife test [46, 47], which is one of the classic cross-validation methods [48, 49], to evaluate the performance of these two methods, *i.e.*, each OC-related gene in S_{oc} was singled out sequentially, and the remaining genes in S_{oc} were used to generate predictions under various combinations of parameters. When training the methods, we supposed that all genes in the network other than the validated OC-related genes were negative; *i.e.*, they were not OC-related genes. The performance of the method can be measured according to the following two features: (1) whether the selected OC-related gene can be recovered by executing the method on the remaining OC-related genes; (2) the predicted genes other than the selected OC-related gene should be as low as possible. Thus, we considered the following two measurements: precision and recall, which are always used to evaluate the performance of the methods on a binary classification problem in the fields of [pattern recognition](#) and information retrieval. Recall is defined as the proportion of retrieved OC-related genes among all OC-related genes, and precision represents the proportion of retrieved OC-related genes among all retrieved genes. Furthermore, another measurement, the F1-measure, is often used to evaluate overall performance, and it can be calculated by

$$F1 - measure = \frac{2 \cdot recall \cdot precision}{recall + precision} \quad (7)$$

Table 2. The pseudo-code of the SP-based method.

SP-based method
Input: An OC-related gene set, S_{oc} ; a network, N_{SP}
Output: A number of putative OC-related genes
1. Execute the SP algorithm on N_{SP} using S_{oc} as the input, extracting candidate genes lying on the obtained shortest paths;
2. Execute the permutation test, producing the p-value for each candidate gene; select candidate genes with p-values less than 0.05;
3. For each candidate gene, calculate its <i>MIS</i> and select candidate genes with an <i>MIS</i> no less than p_{MIS} ;
4. For each candidate gene, calculate its <i>MFS</i> and select candidate genes with an <i>MFS</i> larger than p_{MFS} ;
5. Output the remaining candidate genes as the putative OC-related genes.

<https://doi.org/10.1371/journal.pone.0175185.t002>

Eq 7 shows that recall and precision have the same role. However, in this study, we considered recall to be more important than precision because the method with low recall and high precision could not reliably produce the predicted results. Thus, we simply revised Eq 7 and defined a new measurement, namely, F1-measure-R, which can be computed by

$$F1 - measure - R = recall \cdot F1 - measure = \frac{2 \cdot recall^p \cdot precision}{recall + precision} \quad (8)$$

Because the retrieved genes differ when the selected OC-related gene was not the same, the recall, precision and F1-measure-R must be considered each time for the predicted results. Thus, under a combination of parameters, the RWR-based and SP-based methods can produce a series of recall, precision and F1-measure-R values. We calculated the average values to indicate the performance of the method under this combination of parameters. For convenience, the recall, precision and F1-measure-R presented in the rest of this paper represent the average values.

Results

Optimized parameters for the RWR-based and SP-based methods

As mentioned in the Section “RWR-based method”, “SP-based method” and “Screening rule”, some parameters should be optimized for the RWR-based method and the SP-base method. To extract an optimal combination of parameters for each method, the RWR-based method and the SP-base method were trained on S_{oc} , and their performance was evaluated by the Jack-knife test.

For the RWR-based method, three parameters, the threshold of probability p_{RWR}^p , the threshold of MIS p_{MIS} , and the threshold of MFS p_{MFS} , should be optimized. For p_{RWR}^p , we tried various values ranging from 1E-05 to 1E-04; for p_{MIS} , we tried three values, 400, 700, 900, which are reported in STRING for thresholds of medium confidence, high confidence and highest confidence, respectively; for p_{MFS} , we tried various values ranging from 0 to 0.9. The measurements mentioned in the Section “Evaluation methods” for the RWR-based method with different combinations of parameters are listed in S2 Table. For ease of observation, the values of F1-measure-R obtained using the RWR-based method with different parameters are illustrated in Fig 1; we can see the same level of performance when the parameters p_{RWR}^p and p_{MFS} are equivalent. The maximum F1-measure-R was 1.677E-03 when the parameters were set to be $p_{RWR}^p = 0.00006$, $p_{MFS} = 0.8$ and $p_{MIS} = 400$ or 700. Because this combination of parameters yielded the best performance, they were used to build the optimal RWR-based method, which is adopted to identify novel OC-related genes.

The SP-based method was also trained to extract the optimal parameters for p_{MIS} and p_{MFS} . We tested three values for p_{MIS} , as mentioned in the above paragraph, and various values ranging from 0 to 0.9 for p_{MFS} . For the results obtained using the SP-based method with different combinations of parameters, the measurements mentioned in the Section “Evaluation methods” were counted and are provided in S3 Table. Additionally, three curves are plotted in Fig 2 to show the values of F1-measure-R obtained using the SP-based method with different values of p_{MFS} and a fixed value of p_{MIS} . When the values of p_{MFS} were small, the values of F1-measure-R were proportional to the value of p_{MIS} . However, the values of F1-measure-R were almost the same when p_{MFS} were large. The maximum F1-measure-R was 2.693E-04 when the parameters were set to be $p_{MFS} = 0.8$ and $p_{MIS} = 400$ or 700 or 900. Similarly, we used these values to build the optimal SP-based method for the identification of novel OC-related genes.

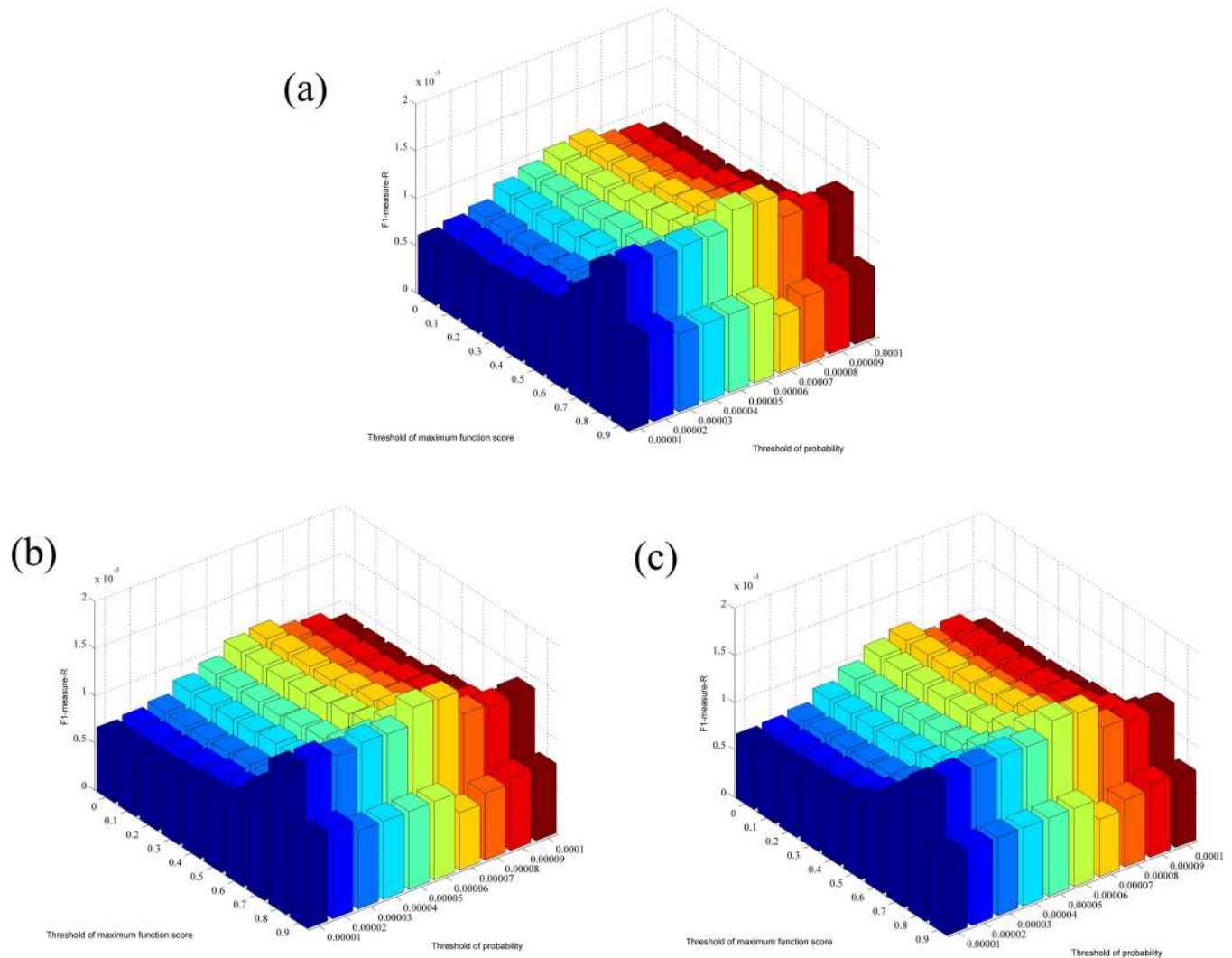


Fig 1. The performance of the RWR-based method under different combinations of parameters. (a) The performance of the RWR-based method setting $p_{MIS} = 400$. (b) The performance of the RWR-based method setting $p_{MIS} = 700$. (c) The performance of the RWR-based method setting $p_{MIS} = 900$.

<https://doi.org/10.1371/journal.pone.0175185.g001>

Inferred results of the RWR-based method

As mentioned in the Section “Optimized parameters for the RWR-based and SP-based methods”, the optimal RWR-based method was built using $p_{RWR}^p = 0.00006$, $p_{MFS} = 0.8$ and $p_{MIS} = 400$ or 700 . This method was further used to identify novel OC-related genes based on all validated OC-related genes mentioned in the Section “Genes related to oral cancer”. Because the optimal RWR-based method contains two options for the parameter p_{MIS} , this method was executed twice. First, p_{MIS} was set to 400, while in the second iteration, it was set to 700. The genes identified using this method with a different parameter, p_{MIS} , are listed in [S4 Table](#). A careful review of the results showed that the identified genes produced using these two optimal RWR-based methods were equivalent. They all identified 81 novel genes. Because all of these genes were viewed as high probability by the RWR method and had strong associations with validated OC-related genes, we believe that they are highly related to OC, and they are termed putative OC-related genes.

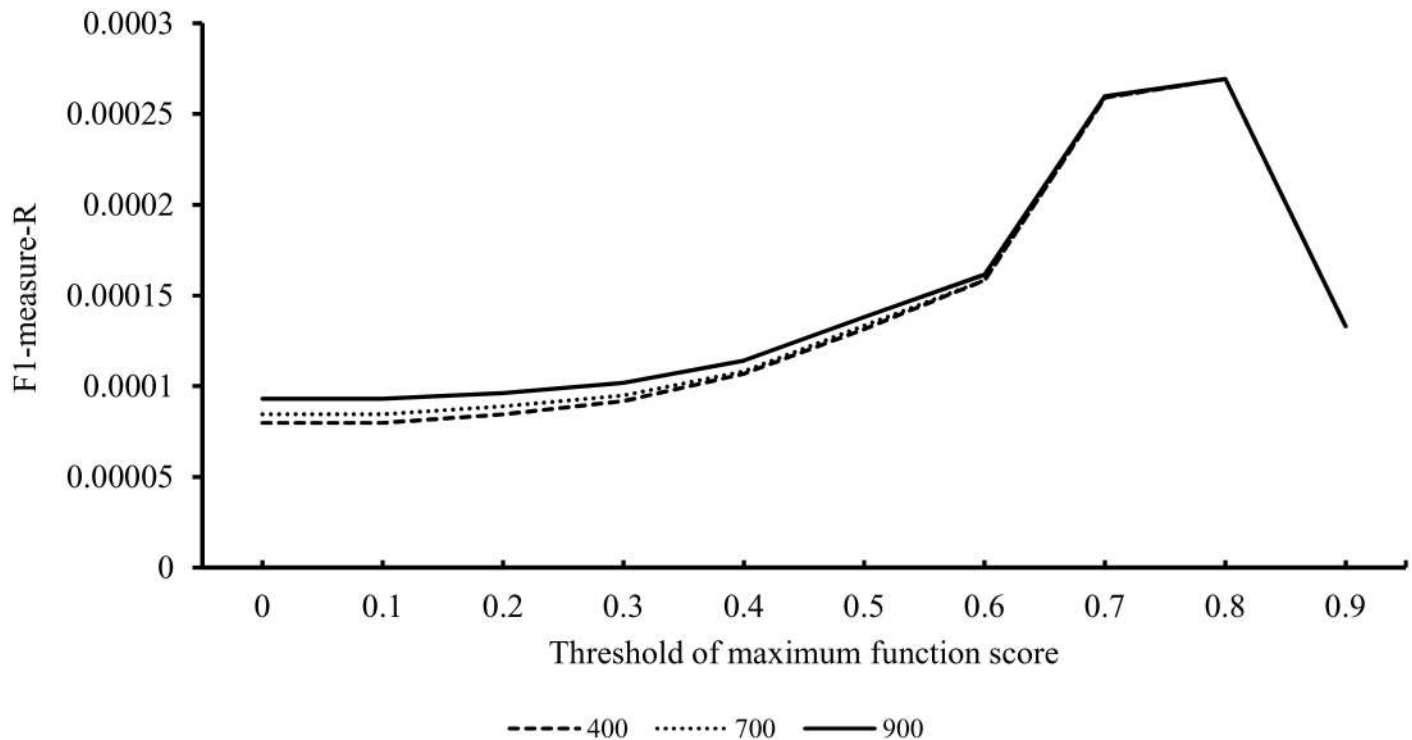


Fig 2. The performance of the SP-based method under different combinations of parameters. There are three lines in this figure, which represent the performance of the SP-based method with different thresholds of maximum interaction score. In detail, the full line represents the performance of the SP-based method with the threshold of maximum interaction score 900, the dot line represents the performance of the SP-based method with the threshold of maximum interaction score 700, the dash line represents the performance of the SP-based method with the threshold of maximum interaction score 400.

<https://doi.org/10.1371/journal.pone.0175185.g002>

Inferred results of the SP-based method

The optimal SP-based method was built in the Section “Optimized parameters for the RWR-based and SP-based methods”, in which p_{MFS} was set to 0.8 and p_{MIS} was set to 400, 700 or 900. This method was also adopted to identify novel OC-related genes. By setting p_{MIS} to 400, 700 or 900, we can build three optimal SP-based methods. They were executed sequentially to identify novel OC-related genes. Similarly, they all yielded fifteen identical genes, which are listed in Table 3. It is believed that these genes are closely related to OC, and they are termed putative OC-related genes.

Inferred results of the integrated method

As mentioned in the Sections “Inferred results of the RWR-based method” and “Inferred results of the SP-based method”, the RWR-based method yielded 81 putative genes, and the SP-based method yielded fifteen putative genes. The union of these two putative gene sets provided the results of the integrated method, in which 85 putative genes were obtained. Their distribution is illustrated in Fig 3; we can see that eleven putative genes were identified by both the RWR-based and the SP-based methods. Among the remaining putative genes, 70 were identified using the RWR-based method, and four putative genes were identified using the SP-based method. Because the principles of the RWR-based and SP-based methods are very different, the identified novel OC-related genes were not the same. By considering the identified genes produced by either of them, we can obtain more putative genes and have an opportunity

Table 3. Genes identified by the optimal SP-based method.

Ensembl ID	Gene symbol	Betweenness	P-value	MIS	MFS	Function
ENSP00000354394	STAT1 ^b	1443	<0.001	999	0.852	functions as a key factor in cell viability in response to different cell stimuli and pathogens [125]
ENSP00000263341	IL1B ^b	543	<0.001	994	0.873	a member of the interleukin 1 cytokine family
ENSP00000379625	MYD88 ^a	528	0.006	999	0.880	an essential signal transducer in the IL1 and Toll-like receptor signaling pathways [126, 127]
ENSP00000233946	IL1R1 ^b	528	0.001	920	0.843	interleukin 1 receptor type 1 [128]
ENSP00000216797	NFKBIA ^b	201	0.018	999	0.825	a member of the NF-kappa-B inhibitor family, which is involved in inflammatory responses [61]
ENSP00000222382	CYP3A43 ^b	183	<0.001	958	0.988	a member of the cytochrome P450 superfamily of enzymes [129]
ENSP00000328181	NOG ^b	183	0.005	999	0.862	binds and inactivates members of the TGF-beta superfamily signaling proteins [130]
ENSP00000410294	FGFR2 ^a	183	0.01	999	0.846	a tyrosine protein kinase that functions as a receptor for fibroblast growth factors and plays key roles in cell proliferation, differentiation, migration and apoptosis [131]
ENSP00000362795	CXCR3 ^a	179	0.021	999	0.808	a G protein-coupled receptor with selectivity for chemokines [132, 133]
ENSP00000260356	THBS1 ^a	10	0.033	984	0.807	an adhesive glycoprotein that mediates cell-cell and cell-matrix interactions [134, 135]
ENSP00000264156	MCM6 ^b	8	0.014	999	0.822	be involved in the formation of replication forks [136]
ENSP00000301141	CYP2A6 ^b	3	0.016	950	0.948	a member of the cytochrome P450 superfamily of enzymes [129]
ENSP00000331736	SELE ^b	1	0.006	978	0.852	responsible for the accumulation of blood leukocytes at sites of inflammation [67]
ENSP00000168712	FGF4 ^b	1	0.016	999	0.847	fibroblast growth factor 4 which are involved in various biological processes such as cell growth and morphogenesis
ENSP00000286758	CXCL13 ^b	1	0.005	992	0.838	C-X-C motif chemokine ligand 13

^a: Genes that have shown stimulative or suppressive effects on cancer as validated by experiments.

^b: Genes that have been reported to have a certain relationship with cancer but that have not been validated by experiments.

<https://doi.org/10.1371/journal.pone.0175185.t003>

to extensively study disease genes in OC. To indicate the obtained 85 putative genes are highly related to OC, we extracted a sub-network containing these putative genes and OC-related genes from N_{RWR} and N_{SP} as shown in Fig 4.

Discussion

Using the RWR-based method, 81 genes were obtained. Using the SP-based method, fifteen genes were accessed. All of these genes were deemed to be significantly associated with OC. Furthermore, eleven genes were identified by both methods, which may be more important than the others. In this section, some important putative genes (listed in Tables 4–6) were extensively analyzed to confirm their associations with OC.

Putative genes identified using both RWR-based and SP-based methods

The following eleven proteins were identified using both RWR-based and SP-based methods: CYP3A43, FGF4, NFKBIA, THBS1, IL1B, CXCL13, CYP2A6, SELE, STAT1, MYD88 and FGFR2, as listed in Table 4.

MYD88 (myeloid differentiation primary response 88) encodes a cytosolic adapter protein that functions as a key signal transducer in the interleukin-1 and Toll-like receptor (TLRs) signaling pathways. TLRs and their ligands play a crucial role in inflammation and host defense

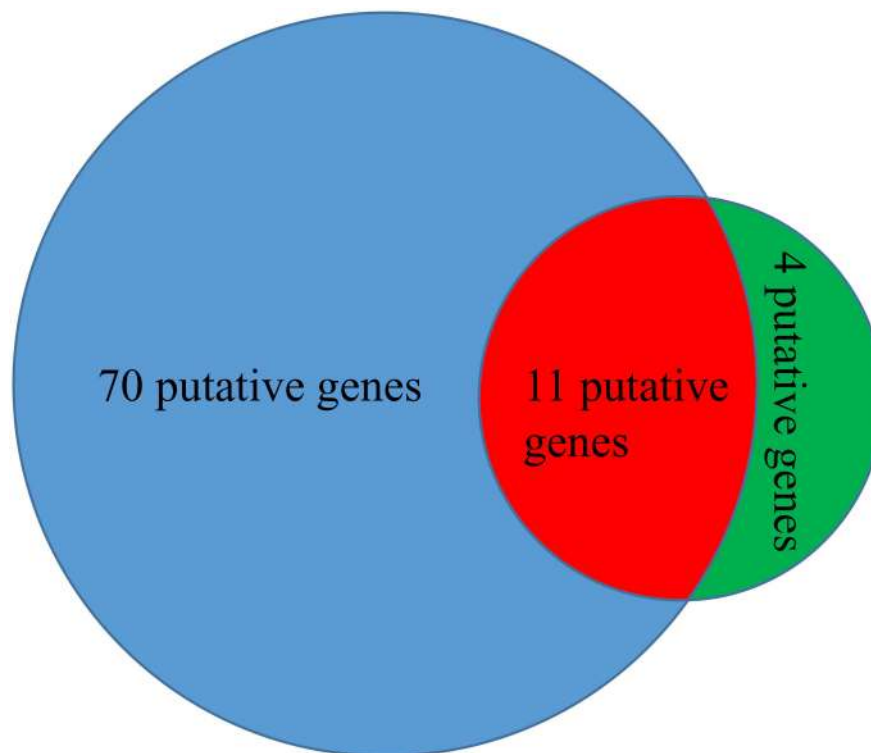


Fig 3. The distribution of the 85 putative OC-related genes obtained in this study. The blue part represents the set consisting of 70 putative genes obtained using only the RWR-based method. The red part represents the set consisting of 11 putative genes obtained using both the RWR-based and SP-based methods. The green part represents the set consisting of 4 putative genes obtained using only the SP-based method.

<https://doi.org/10.1371/journal.pone.0175185.g003>

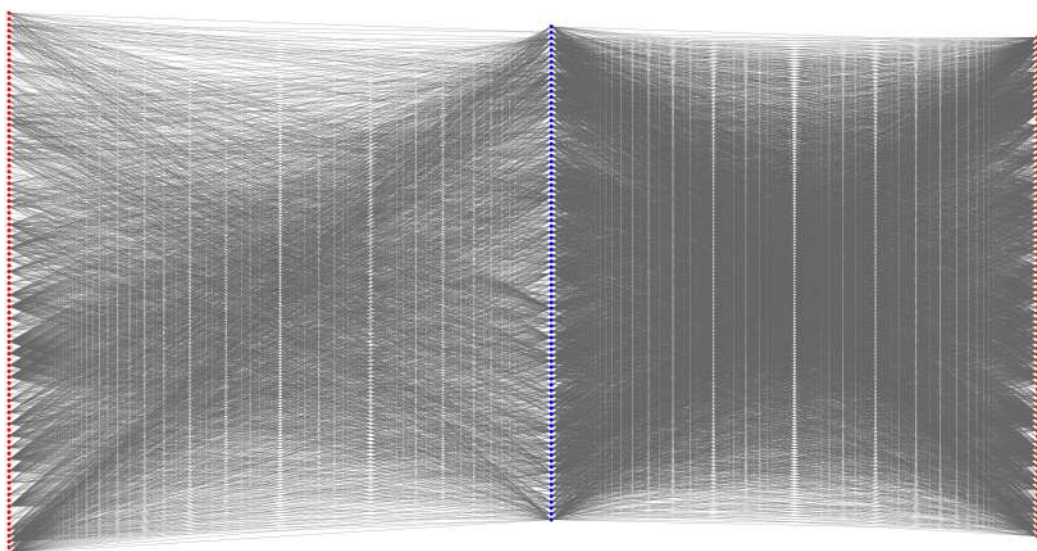


Fig 4. The sub-network containing the putative genes and OC-related genes that was extracted from the network for RWR-based and SP-based methods. The blue nodes represent putative genes and red nodes represent OC-related genes.

<https://doi.org/10.1371/journal.pone.0175185.g004>

Table 4. Eleven putative genes identified using both RWR-based and SP-based methods.

Ensembl ID	Gene symbol	RWR-based method				SP-based method				Function
		Probability	P-value	MIS	MFS	Betweenness	P-value	MIS	MFS	
ENSP00000379625	MYD88 ^a	6.67E-05	0.032	999	0.880	528	0.006	999	0.880	an essential signal transducer in the IL1 and Toll-like receptor signaling pathways [126, 127]
ENSP00000410294	FGFR2 ^a	9.04E-05	0.021	999	0.846	183	0.01	999	0.846	a tyrosine protein kinase that functions as a receptor for fibroblast growth factors and plays key roles in cell proliferation, differentiation, migration and apoptosis [131]
ENSP00000216797	NFKBIA ^b	7.56E-05	0.026	999	0.825	201	0.018	999	0.825	a member of the NF-kappa-B inhibitor family, which is involved in inflammatory responses [61]
ENSP00000331736	SELE ^b	9.73E-05	<0.001	978	0.852	1	0.006	978	0.852	responsible for the accumulation of blood leukocytes at sites of inflammation [67]
ENSP00000260356	THBS1 ^a	6.94E-05	<0.001	984	0.807	10	0.033	984	0.807	an adhesive glycoprotein that mediates cell-cell and cell-matrix interactions [134, 135]
ENSP00000354394	STAT1 ^b	4.36E-04	<0.001	999	0.852	1443	<0.001	999	0.852	functions as a key factor in cell viability in response to different cell stimuli and pathogens [125]
ENSP00000301141	CYP2A6 ^b	8.85E-05	<0.001	950	0.948	3	0.016	950	0.948	a member of the cytochrome P450 superfamily of enzymes [129]
ENSP00000222382	CYP3A43 ^b	3.35E-04	<0.001	958	0.988	183	<0.001	958	0.988	a member of the cytochrome P450 superfamily of enzymes [129]
ENSP00000286758	CXCL13 ^b	6.63E-05	0.006	992	0.837	1	0.005	992	0.838	C-X-C motif chemokine ligand 13
ENSP00000168712	FGF4 ^b	8.21E-05	0.001	999	0.847	1	0.016	999	0.847	fibroblast growth factor 4 which are involved in various biological processes such as cell growth and morphogenesis
ENSP00000263341	IL1B ^b	1.92E-04	<0.001	994	0.873	543	<0.001	994	0.873	a member of the interleukin 1 cytokine family

^a: Genes that have shown stimulative or suppressive effects on cancer as validated by experiments.

^b: Genes that have been reported to have a certain relationship with cancer but that have not been validated by experiments.

<https://doi.org/10.1371/journal.pone.0175185.t004>

[50]. In a broad variety of tumor tissues and cell lines, aberrant expression of TLRs plays a role in tumor immune escape or resistance to apoptosis [51, 52]. The expression of TLR4 and MyD88 is also aberrant and affects the down-stream signaling pathway [53]. In this study, MyD88 was predicted using both RWR-based and SP-based methods, and thus it may be a potential target for OC.

FGFR2 (fibroblast growth factor receptor 2) belongs to the FGFR tyrosine kinase family, which is one of the most frequently altered kinase families in some types of cancer [54]. Point mutations have been observed in 12% of endometrial carcinomas [55]. These studies suggest that FGFR2 may act together with the regulation of the PI3K/AKT/mTOR pathway to drive endometrial cancer growth in a subset of patients [56, 57]. FGFR2 mutations also occur in oral squamous cell carcinoma, and a patient with OSCC was found to respond to pazopanib, a multiple tyrosine kinase inhibitor [58]. These findings support the potential diagnostic and therapeutic value of FGFR2 as an effective strategy for the treatment of patients with OC with defined molecular characteristics. In addition, our results showed that FGF4 had a significant probability in both methods. Some large-scale experiments have suggested that aberrant

Table 5. Important genes among the seventy putative genes identified using the RWR-based method.

Ensembl ID	Gene symbol	Probability	P-value	MIS	MFS	Function
ENSP00000245451	BMP4 ^a	9.57E-05	0.026	981	0.905	bind TGF-beta receptor leading to recruitment and activation of transcription factor [137]
ENSP00000225831	CCL2 ^b	1.21E-04	0.012	984	0.869	C-C motif chemokine ligand 2
ENSP00000351671	CCL20 ^b	6.69E-05	0.003	965	0.804	C-C motif chemokine ligand 20
ENSP00000293272	CCL5 ^a	7.70E-05	0.002	994	0.891	C-C motif chemokine ligand 5
ENSP00000292303	CCR5 ^a	1.01E-04	0.003	996	0.839	C-C motif chemokine receptor 5
ENSP00000246657	CCR7 ^a	9.82E-05	<0.001	998	0.823	C-C motif chemokine receptor 7
ENSP00000229135	IFNG ^a	1.70E-04	0.03	994	0.839	binds to the interferon gamma receptor to response to infection [138]
ENSP00000228280	KITLG ^a	1.05E-04	0.026	948	0.816	the ligand of the tyrosine-kinase receptor
ENSP00000162749	TNFRSF1A ^b	9.69E-05	0.013	999	0.825	a member of the TNF receptor superfamily which plays a role in various biological processes
ENSP00000289153	PIK3CB ^a	1.23E-04	<0.001	997	0.926	an isoform of the catalytic subunit of PI3K
ENSP00000366563	PIK3CD ^b	1.15E-04	<0.001	997	0.919	PI3Ks phosphorylate inositol lipids and it is involved in the immune response [139]
ENSP00000352121	PIK3CG ^a	1.21E-04	<0.001	996	0.921	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma
ENSP00000324648	CYP2B6 ^b	9.73E-05	<0.001	962	0.943	a member of the cytochrome P450 superfamily
ENSP00000360372	CYP2C19 ^b	9.32E-05	<0.001	962	0.966	cytochrome P450 family 2 subfamily C member 19
ENSP00000360247	CYP2J2 ^a	7.48E-05	0.004	912	0.969	cytochrome P450 family 2 subfamily J member 2
ENSP00000337915	CYP3A4 ^b	3.32E-04	<0.001	963	0.941	cytochrome P450 family 3 subfamily A member 4
ENSP00000360968	CYP4X1 ^b	7.43E-05	0.018	939	0.959	cytochrome P450 family 4 subfamily X member 1
ENSP00000304283	RAC3 ^a	1.08E-04	0.006	990	0.981	a GTPase regulates cell growth, cytoskeletal reorganization, and the activation of kinases [140–142]

^a: Genes that have shown stimulative or suppressive effects on cancer as validated by experiments.

^b: Genes that have been reported to have a certain relationship with cancer but that have not been validated by experiments.

<https://doi.org/10.1371/journal.pone.0175185.t005>

amplification of FGF4 occurs in several types of cancer, including lymph node metastasis and urinary bladder cancer [59, 60].

NF-κBIA (nuclear factor-kappa B inhibitor alpha) inhibits NF-κB, which is involved in inflammatory responses and is a hallmark linking inflammation to tumor development and progression [61, 62]. The polymorphic variations in NFKBIA were associated with the risk of various cancers, including gastric cancer, prostate cancer and melanoma [63–65]. NFKBIA polymorphisms have significant associations with OSCC [66]. In our study, NFKBIA had a close relationship with OC using both methods.

SELE (selectin E), which was found in cytokine-stimulated endothelial cells, functions in the accumulation of blood leukocytes at sites of inflammation [67]. The level of SELE-

Table 6. Four putative genes identified using the SP-based method.

Ensembl ID	Gene symbol	Betweenness	P-value	MIS	MFS	Function
ENSP00000264156	MCM6 ^b	8	0.014	999	0.822	be involved in the formation of replication forks [136]
ENSP00000328181	NOG ^b	183	0.005	999	0.862	binds and inactivates members of the TGF-beta superfamily signaling proteins [130]
ENSP00000362795	CXCR3 ^a	179	0.021	999	0.808	a G protein-coupled receptor with selectivity for chemokines [132, 133]
ENSP00000233946	IL1R1 ^b	528	0.001	920	0.843	interleukin 1 receptor type 1 [128]

^a: Genes that have shown stimulative or suppressive effects on cancer as validated by experiments.

^b: Genes that have been reported to have a certain relationship with cancer but that have not been validated by experiments.

<https://doi.org/10.1371/journal.pone.0175185.t006>

mediated adhesion of colon cancer and head and neck squamous cell cancer cells to the endothelium has been implied to be involved in metastasis [68, 69]. Our analysis revealed that SELE might be a putative marker for tumorigenesis in OC. A study in Taiwan supports the idea that SELE-related inflammation plays a crucial role in the pathogenesis process of OSCC [70]. Future investigations of the function of SELE may clarify the mechanism of tumorigenesis and metastasis.

THBS1 (thrombospondin 1) is a subunit of a disulfide-linked homotrimeric protein, which mediates cell-cell and cell-matrix interactions [71]. The significant relationship between THBS1 and OC was detected using both RWR-based and SP-based methods. Several studies have indicated that THBS1 has crucial functions in oral tumorigenesis [72, 73]. THBS1 might be a potential diagnostic and therapeutic target for OC.

STAT1 (signal transducer and activator of transcription 1), which belongs to the STAT protein family, mediates the expression of a variety of genes and cell viability in response to stimuli and pathogens [74]. Aberrant activation of STAT1 has frequently been found in various cancers, such as head and neck cancer [75]. A similar STAT1 activation status was detected in patients with OSCC [76]. In this study, STAT1 displayed a significant association with OC using both the RWR-based and SP-based methods, and this molecular marker could help in guiding diagnostic and therapeutic decisions in patients with OC.

Among these eleven putative genes, including genes such as MYD88, FGFR2 and THBS1, stimulative or suppressive effects on cancers have been shown by the experiments. We speculate that these three genes have certain functions in OC. In addition, it has been reported that other genes, including NFKBIA, SELE, STAT1, IL1B, CYP2A6, CYP3A43, FGF4 and CXCL13 [77–80], have a certain relationship with various forms cancer. However, the mechanism has not been thoroughly studied. In our analysis, they have a significant relationship with OC, and the mechanism should be explored.

Putative genes identified only by the RWR-based method

Seventy proteins were identified only by the RWR-based method. The important genes are BMP4, CCL2, CCL20, CCL5, CCR7, IFNG, KITLG, TNFRSF1A, PIK3CB, PIK3CD, PIK3CG, CYP2B6, CYP2C19, CYP2J2, CYP3A4, CYP4X1 and RAC3; these genes are listed in [Table 5](#). Some of them participate in tumorigenesis by regulating cell growth or apoptosis, such as BMP4 and RAC3. Some genes are important factors in the immune system of cancer patients, such as CCL2, CCL20, CCL5, CCR7, IFNG and P450 family. Detailed function analyses of candidate genes are shown in [S1 File](#).

We identified several novel putative genes using only the RWR-based method, which have strong associations with the tumorigenesis of OC. The functions of some genes have been explored in experiments, such as BMP4, IFNG, KITLG, CCL5, CCR5, CCR7, CYP2J2, PIK3CB, PIK3CG and RAC3; others revealed mutations or aberrant expression in cancers, but the mechanism is not clear. Future research is required to replicate and validate the effects of these genes.

Putative genes identified only by the SP-based method

Four proteins were predicted to be closely associated with OC using the SP-based method but not the RWR-based method. These genes were IL1R1, MCM6, NOG and CXCR3; they are listed in [Table 6](#). It has been reported that CXCR3 can promote metastasis in some types of cancer [81–86]. Other genes showed aberrant expression or mutations in cancers, which merits attention. Detailed function analyses of candidate genes are shown in [S1 File](#).

Based on the analyses of the novel genes obtained by the integrated method, we found that several genes have shown stimulative or suppressive effects on cancers by experiments or aberrant expression or mutations in cancers reported in published papers, implying the effectiveness of the integrated method. In addition, the integrated method can provide more potential disease genes for investigating OC because it combines the results yielded by the RWR-based and SP-based methods; i.e., simultaneous usage of the RWR-based and SP-based method can help us mine more information about disease genes in oral cancer. We believe that it is helpful to investigate various diseases using both the RWR-based and the SP-based methods.

GO-term and pathway function enrichment analysis of putative genes using DAVID

By the integrated method, 85 putative genes were obtained. To uncover the biological meaning behind these genes, the functional annotation tool, the Database for Annotation, Visualization and Integrated Discovery (DAVID) [87], was adopted to analyze them. The obtained results are provided in [S5 Table](#).

For the GO enrichment results yielded by DAVID (see [S5 Table](#)), we can see that seven-two biological process (BP) GO terms, fifteen molecular function (MF) GO terms and nine cellular component (CC) GO terms are with statistically significance ($FDR < 0.05$). The biological process mostly focused on cell proliferation, apoptosis, transcription and signal transduction. Malignant proliferation and anti-apoptosis are the major characteristics of many cancers. In this study, it has been found many proliferation related genes such as EGF, IGF1 and CDKN1B. EGF (epidermal growth factor) and its receptor EGFR affect cellular processes like proliferation, motility and adhesion. EGF has a higher expression and is correlated with the progression of cancer such as breast cancer [88]. The circulating IGF1 (insulin-like growth factor-1) level is associated with the risk to develop breast cancer [89, 90]. CDKN1B functions as a key cell cycle gatekeeper to prevent or slow down cell division [91, 92]. Aberrant downregulating CDKN1B may promote the proliferation of multiple myeloma cells [93]. Un-controlled expanded number of cancers is determined not only by cell proliferation but also by the cell evading apoptosis. In our study, many genes are enriched in apoptosis process such as TP53, CASP3 and *cmyc*. The mutations of TP53 are the most common alteration in cancer and are related to cell apoptosis, cell cycle and malignancy [94–97]. TP53 could arrest cell cycle and induced cell apoptosis in response to DNA damage [98]. It has been reported that over-expression of *c-Myc* drives the level of BAX and other apoptosis-related genes and is involved in cell apoptosis [99–103]. CASPs is a kind of cysteine-dependent aspartate-specific proteases, and is associated with the initiation and execution of apoptosis. As an activate effector, CASP3 receives the apoptotic signals to perform the cell death process [104]. In our result, a series of genes were enriched in the GO term related apoptosis, which proved process of apoptosis is a key process in oral cancer. In addition, several other biological process GO terms were shown such as transcription and signal transduction. In proliferation, apoptosis and other processes, functional genes were selectively transcribed and expressed. For example, *c-Myc* is a transcription factor which could mediate a series of down-stream genes express to drive cellular proliferation and apoptosis [105]. It can be seen from [S5 Table](#) that putative genes were enriched in several CC GO terms and MF GO terms such as growth factor activity (MF), cytokine activity (MF), enzyme or protein binding (MF), extracellular region (CC), extracellular space (CC) and plasma membrane (CC). These results suggest that the functional activity and localization of the protein are directly or indirectly related to oral cancer.

The DAVID also produced the KEGG pathway enrichment on putative genes (see [S5 Table](#)), several pathways were highlighted such as TNF signaling pathway, PI3K-Akt pathway,

Pap1 pathway, MAPK pathway, TLR pathway, NF- κ B pathway, JaK-STAT, Ras pathway and cytokine-cytokine receptor interaction. TNF pathway could be activated by various signals to affect the immunity, cell growth, apoptosis and other biological behaviors of tumor cells [106–110]. Protein-protein interaction and pathways is a cross-talk network. TNF can also activate the NF- κ B pathway. The NF- κ B pathway is activated in various cancers [111, 112]. Activation of NF- κ B can be inhibited by the blockade of PI3K/Akt and ERK pathway to suppress tumor metastasis [113]. It was reported that PI3K/AKT-NF- κ B is an axis which promotes bone metastasis in prostate cancer [114]. The Jak/STAT pathway is critical in normal tissues and tumors and the Jak kinase family includes JAK1, JAK2, JAK3 and TYK2 [115]. Jak mediated STAT phosphorylation leads to their nuclear translocation. STAT molecules bind specific promoter of genes and result in the transcription in nucleus, which regulate the cell proliferation, differentiation and apoptosis [115, 116]. It was reported that over-activation of the JAK/STAT pathway is related to subsets of patients with certain solid tumors and chronic myeloid leukemia [117, 118]. MAPK pathway is induced by activation of TLR (toll like receptors) and NOD (nucleotide-binding oligomerization domain receptors), which initiates of inflammation and are involved in cancer proliferation and control [119]. Ras pathway is a highly conserved pathway in cell function, including cell proliferation, differentiation and signaling transduction. This pathway is commonly deregulated in cancers, making the components in the pathway as targets for therapeutic interventions [120]. In these pathways, several putative genes were enriched, which indicates these pathway and gene need more attention in tumorigenesis of oral cancer.

Conclusions

In this study, we investigated genes related to oral cancer. Two popular algorithms, the random walk with a restart algorithm and the shortest path algorithm, which are often used to identify novel disease genes, were integrated with some further rules to build two computational methods. An integrated method was further built by combining these two methods. To access an optimal prediction method for the identification of genes related to oral cancer using the integrated method, these two methods were trained on validated genes. The optimal prediction method was further adopted to identify novel genes related to oral cancer. The results indicated the following facts: (1) The integrated method is effective for identification of disease genes of oral cancer; (2) Candidate genes produced by the integrated method provide an opportunity to achieve a more extensive investigation of oral cancer. We hope that the integrated method can be useful for identifying novel disease genes. In view of the utility of the RWR-based and SP-based methods for identification of disease genes, it is hopeful that these methods can be applied to investigate other problems, such as DNA-binding protein prediction [121], protein fold recognition [122, 123], detection of tubule boundary [124], *etc.*

Supporting information

S1 File. Detailed analysis of the putative genes obtained by RWR-based method or SP-based method.

(DOCX)

S1 Table. 202 genes related to oral cancer and their sources.

(DOCX)

S2 Table. The performance of the RWR-based method with different combinations of parameters.

(DOCX)

S3 Table. The performance of the SP-based method with different combinations of parameters.

(DOCX)

S4 Table. Genes identified by the optimal RWR-based method.

(DOCX)

S5 Table. The enrichment analysis of putative genes yielded by DAVID.

(XLSX)

Acknowledgments

This study was supported by the National Natural Science Foundation of China (31371335), Natural Science Foundation of Shanghai (17ZR1412500), Shanghai Sailing Program and The Youth Innovation Promotion Association of Chinese Academy of Sciences (CAS) (2016245).

Author Contributions

Conceptualization: TH HL YDC.

Data curation: ZX FY YS.

Formal analysis: LC JY YZ XK.

Funding acquisition: TH YDC.

Methodology: LC YDC.

Supervision: LC YDC.

Validation: JY ZX FY TH.

Writing – original draft: LC JY.

Writing – review & editing: TH YDC.

References

1. Tsantoulis PK, Kastrinakis NG, Tourvas AD, Laskaris G, Gorgoulis VG. Advances in the biology of oral cancer. *Oral Oncol.* 2007; 43(6):523–34. <https://doi.org/10.1016/j.oraloncology.2006.11.010> PMID: [17258495](https://pubmed.ncbi.nlm.nih.gov/17258495/)
2. Shridhar K, Walia GK, Aggarwal A, Gulati S, Geetha AV, Prabhakaran D, et al. DNA methylation markers for oral pre-cancer progression: A critical review. *Oral Oncol.* 2016; 53:1–9. <https://doi.org/10.1016/j.oraloncology.2015.11.012> PMID: [26690652](https://pubmed.ncbi.nlm.nih.gov/26690652/)
3. Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *International journal of cancer Journal international du cancer.* 2013; 132(5):1133–45. <https://doi.org/10.1002/ijc.27711> PMID: [22752881](https://pubmed.ncbi.nlm.nih.gov/22752881/)
4. Noda M. International Agency for Research on Cancer (<http://www.iarc.fr>). *Japanese journal of clinical oncology.* 1999; 29(11):592. PMID: [10678568](https://pubmed.ncbi.nlm.nih.gov/10678568/)
5. Cancer incidence in five continents. Volume IX. IARC Sci Publ. 2008;(160):1–837. PMID: [19388204](https://pubmed.ncbi.nlm.nih.gov/19388204/)
6. Warnakulasuriya S, Sutherland G, Scully C. Tobacco, oral cancer, and treatment of dependence. *Oral Oncol.* 2005; 41(3):244–60. <https://doi.org/10.1016/j.oraloncology.2004.08.010> PMID: [15743687](https://pubmed.ncbi.nlm.nih.gov/15743687/)
7. Lucenteforte E, Garavello W, Bosetti C, La Vecchia C. Dietary factors and oral and pharyngeal cancer risk. *Oral Oncol.* 2009; 45(6):461–7. <https://doi.org/10.1016/j.oraloncology.2008.09.002> PMID: [18990606](https://pubmed.ncbi.nlm.nih.gov/18990606/)
8. Pavia M, Pileggi C, Nobile CG, Angelillo IF. Association between fruit and vegetable consumption and oral cancer: a meta-analysis of observational studies. *The American journal of clinical nutrition.* 2006; 83(5):1126–34. PMID: [16685056](https://pubmed.ncbi.nlm.nih.gov/16685056/)

9. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990; 61(5):759–67. PMID: [2188735](#)
10. Nagatsuka H, Ishiwari Y, Tsujigiwa H, Nakano K, Nagai N. Quantitation of epidermal growth factor receptor gene amplification by competitive polymerase chain reaction in pre-malignant and malignant oral epithelial lesions. *Oral Oncol*. 2001; 37(7):599–604. PMID: [11564582](#)
11. Ishitoya J, Toriyama M, Oguchi N, Kitamura K, Ohshima M, Asano K, et al. Gene amplification and overexpression of EGF receptor in squamous cell carcinomas of the head and neck. *British journal of cancer*. 1989; 59(4):559–62. PMID: [2713242](#)
12. Shiraki M, Odajima T, Ikeda T, Sasaki A, Satoh M, Yamaguchi A, et al. Combined expression of p53, cyclin D1 and epidermal growth factor receptor improves estimation of prognosis in curatively resected oral cancer. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc.* 2005; 18(11):1482–9.
13. Xia W, Lau YK, Zhang HZ, Liu AR, Li L, Kiyokawa N, et al. Strong correlation between c-erbB-2 overexpression and overall survival of patients with oral squamous cell carcinoma. *Clin Cancer Res*. 1997; 3(1):3–9. PMID: [9815530](#)
14. Werkmeister R, Brandt B, Joos U. Clinical relevance of erbB-1 and -2 oncogenes in oral carcinomas. *Oral Oncol*. 2000; 36(1):100–5. PMID: [10889928](#)
15. Miyamoto R, Uzawa N, Nagaoka S, Hirata Y, Amagasa T. Prognostic significance of cyclin D1 amplification and overexpression in oral squamous cell carcinomas. *Oral Oncol*. 2003; 39(6):610–8. PMID: [12798405](#)
16. Koontongkaew S, Chareonkitkajorn L, Chanvitan A, Leelakriangsak M, Amornphimoltham P. Alterations of p53, pRb, cyclin D(1) and cdk4 in human oral and pharyngeal squamous cell carcinomas. *Oral Oncol*. 2000; 36(4):334–9. PMID: [10899671](#)
17. Huang MF, Chang YC, Liao PS, Huang TH, Tsay CH, Chou MY. Loss of heterozygosity of p53 gene of oral cancer detected by exfoliative cytology. *Oral Oncol*. 1999; 35(3):296–301. PMID: [10621851](#)
18. Largey JS, Meltzer SJ, Yin J, Norris K, Sauk JJ, Archibald DW. Loss of heterozygosity of p53 in oral cancers demonstrated by the polymerase chain reaction. *Cancer*. 1993; 71(6):1933–7. PMID: [8443743](#)
19. Sakata K. Alterations of tumor suppressor genes and the H-ras oncogene in oral squamous cell carcinoma. *Journal of oral pathology & medicine: official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology*. 1996; 25(6):302–7.
20. Shang ZJ, Li ZB, Li JR. VEGF is up-regulated by hypoxic stimulation and related to tumour angiogenesis and severity of disease in oral squamous cell carcinoma: in vitro and in vivo studies. *Int J Oral Maxillofac Surg*. 2006; 35(6):533–8. <https://doi.org/10.1016/j.ijom.2005.09.006> PMID: [16388929](#)
21. Patel BP, Shah PM, Rawal UM, Desai AA, Shah SV, Rawal RM, et al. Activation of MMP-2 and MMP-9 in patients with oral squamous cell carcinoma. *J Surg Oncol*. 2005; 90(2):81–8. <https://doi.org/10.1002/jso.20240> PMID: [15844188](#)
22. Zeng X, Liao Y, Liu Y, Zou Q. Prediction and validation of disease genes using HeteSim Scores. *IEEE/ACM Trans Comput Biol Bioinform*. 2016.
23. Chen X, Perez-Jimenez MJ, Valencia-Cabrera L, Wang BZ, Zeng XX. Computing with viruses. *Theoretical Computer Science*. 2016; 623:146–59.
24. Zou Q, Wan S, Ju Y, Tang J, Zeng X. Pretata: predicting TATA binding proteins with novel features and dimensionality reduction strategy. *BMC Syst Biol*. 2016; 10(Suppl 4):114. <https://doi.org/10.1186/s12918-016-0353-5> PMID: [28155714](#)
25. Gu B, Sheng VS, Tay KY, Romano W, Li S. Incremental Support Vector Learning for Ordinal Regression. *Ieee T Neur Net Lear*. 2015; 26(7):1403–16.
26. Gu B, Sheng VS, Li S. Bi-parameter space partition for cost-sensitive SVM. *Proceedings of the 24th International Conference on Artificial Intelligence; Buenos Aires, Argentina*. 2832741: AAAI Press; 2015. p. 3532–9.
27. Tang W, Liao Z, Zou Q. Which statistical significance test best detects oncomiRNAs in cancer tissues? An exploratory analysis. *Oncotarget*. 2016; 7(51):85613–23. <https://doi.org/10.18632/oncotarget.12828> PMID: [27784000](#)
28. Zhang JR, Ju Y, Lu HJ, Xuan P, Zou Q. Accurate Identification of Cancerlectins through Hybrid Machine Learning Technology. *Int J Genomics*. 2016.
29. Oliver S. Guilt-by-association goes global. *Nature*. 2000; 403(6770):601–3. <https://doi.org/10.1038/35001165> PMID: [10688178](#)
30. Kohler S, Bauer S, Horn D, Robinson PN. Walking the interactome for prioritization of candidate disease genes. *The American Journal of Human Genetics*. 2008; 82(4):949–58.

31. Navlakha S, Kingsford C. The power of protein interaction networks for associating genes with diseases. *Bioinformatics*. 2010; 26(8):1057–63. <https://doi.org/10.1093/bioinformatics/btq076> PMID: [20185403](https://pubmed.ncbi.nlm.nih.gov/20185403/)
32. Care MA, Bradford JR, Needham CJ, Bulpitt AJ, Westhead DR. Combining the interactome and deleterious SNP predictions to improve disease gene identification. *Hum Mutat*. 2009; 30(3):485–92. <https://doi.org/10.1002/humu.20917> PMID: [19156842](https://pubmed.ncbi.nlm.nih.gov/19156842/)
33. Chen L, Xing Z, Huang T, Shu Y, Huang G, Li H-P. Application of the shortest path algorithm for the discovery of breast cancer related genes. *Current Bioinformatics*. 2016; 11(1):51–8.
34. Gao Y-F, Shu Y, Yang L, He Y-C, Li L-P, Huang G, et al. A Graphic Method for Identification of Novel Glioma Related Genes. *BioMed Research International*. 2014; 2014:891945. <https://doi.org/10.1155/2014/891945> PMID: [25050377](https://pubmed.ncbi.nlm.nih.gov/25050377/)
35. Zhang J, Yang J, Huang T, Shu Y, Chen L. Identification of novel proliferative diabetic retinopathy related genes on protein-protein interaction network. *Neurocomputing*. 2016; 217:63–72.
36. Chen L, Wang B, Wang S, Yang J, Hu J, Xie Z, et al. OPMS: A computational method integrating protein interaction and sequence information for the identification of novel putative oncogenes. *Protein Pept Lett*. 2016; 23(12):1081–94. PMID: [27774893](https://pubmed.ncbi.nlm.nih.gov/27774893/)
37. Chen L, Huang T, Zhang YH, Jiang Y, Zheng M, Cai YD. Identification of novel candidate drivers connecting different dysfunctional levels for lung adenocarcinoma using protein-protein interactions and a shortest path approach. *Sci Rep*. 2016; 6:29849. Epub 2016/07/15. <https://doi.org/10.1038/srep29849> PMID: [27412431](https://pubmed.ncbi.nlm.nih.gov/27412431/)
38. Chen L, Yang J, Huang T, Kong X, Lu L, Cai Y-D. Mining for novel tumor suppressor genes using a shortest path approach. *Journal of Biomolecular Structure and Dynamics*. 2016; 34(3):664–75. <https://doi.org/10.1080/07391102.2015.1042915> PMID: [26209080](https://pubmed.ncbi.nlm.nih.gov/26209080/)
39. Cai Y-D, Zhang Q, Zhang Y-H, Chen L, Huang T. Identification of genes associated with breast cancer metastasis to bone on a protein-protein interaction network with a shortest path algorithm. *J Proteome Res*. 2017; 16(2):1027–38. <https://doi.org/10.1021/acs.jproteome.6b00950> PMID: [28076954](https://pubmed.ncbi.nlm.nih.gov/28076954/)
40. Chen L, Zhang YH, Huang T, Cai YD. Identifying novel protein phenotype annotations by hybridizing protein-protein interactions and protein sequence similarities. *Molecular Genetics and Genomics*. 2016; 291(2):913–34. <https://doi.org/10.1007/s00438-015-1157-9> PMID: [26728152](https://pubmed.ncbi.nlm.nih.gov/26728152/)
41. The UniProt Consortium. Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res*. 2013; 41(Database issue):D43–7. Epub 2012/11/20. <https://doi.org/10.1093/nar/gks1068> PMID: [23161681](https://pubmed.ncbi.nlm.nih.gov/23161681/)
42. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research*. 2013; 41(D1):D808–D15.
43. Ng KL, Ciou JS, Huang CH. Prediction of protein functions based on function-function correlation relations. *Comput Biol Med*. 2010; 40(3):300–5. <https://doi.org/10.1016/j.compbiomed.2010.01.001> PMID: [20089249](https://pubmed.ncbi.nlm.nih.gov/20089249/)
44. Hu LL, Huang T, Shi X, Lu WC, Cai YD, Chou KC. Predicting functions of proteins in mouse based on weighted protein-protein interaction network and protein hybrid properties. *PLoS ONE*. 2011; 6(1): e14556. <https://doi.org/10.1371/journal.pone.0014556> PMID: [21283518](https://pubmed.ncbi.nlm.nih.gov/21283518/)
45. Gormen TH, Leiserson CE, Rivest RL, Stein C, editors. *Introduction to algorithms*: MIT press Cambridge, MA; 1990.
46. Chen L, Zeng W-M, Cai Y-D, Feng K-Y, Chou K-C. Predicting Anatomical Therapeutic Chemical (ATC) Classification of Drugs by Integrating Chemical-Chemical Interactions and Similarities. *PLoS ONE*. 2012; 7(4):e35254. <https://doi.org/10.1371/journal.pone.0035254> PMID: [22514724](https://pubmed.ncbi.nlm.nih.gov/22514724/)
47. Wei L, Xing P, Shi G, Ji ZL, Zou Q. Fast prediction of protein methylation sites using a sequence-based feature selection technique. *IEEE/ACM Trans Comput Biol Bioinform*. 2017.
48. Wei L, Xing P, Tang J, Zou Q. PhosPred-RF: a novel sequence-based predictor for phosphorylation sites using sequential information only. *IEEE Trans Nanobioscience*. 2017.
49. Kohavi R, editor *A study of cross-validation and bootstrap for accuracy estimation and model selection*. International joint Conference on artificial intelligence; 1995: Lawrence Erlbaum Associates Ltd.
50. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. *Cell*. 2006; 124(4):783–801. <https://doi.org/10.1016/j.cell.2006.02.015> PMID: [16497588](https://pubmed.ncbi.nlm.nih.gov/16497588/)
51. Szajnik M, Szczepanski MJ, Czystowska M, Elishaev E, Mandapathil M, Nowak-Markwitz E, et al. TLR4 signaling induced by lipopolysaccharide or paclitaxel regulates tumor survival and chemoresistance in ovarian cancer. *Oncogene*. 2009; 28(49):4353–63. <https://doi.org/10.1038/nc.2009.289> PMID: [19826413](https://pubmed.ncbi.nlm.nih.gov/19826413/)

52. Szczepanski MJ, Czystowska M, Szajnik M, Harasymczuk M, Boyiadzis M, Kruk-Zagajewska A, et al. Triggering of Toll-like receptor 4 expressed on human head and neck squamous cell carcinoma promotes tumor development and protects the tumor from immune attack. *Cancer Res.* 2009; 69(7):3105–13. <https://doi.org/10.1158/0008-5472.CAN-08-3838> PMID: 19318560
53. Tucci M, Stucci S, Passarelli A, Giudice G, Dammacco F, Silvestris F. The immune escape in melanoma: role of the impaired dendritic cell function. Expert review of clinical immunology. 2014; 10(10):1395–404. <https://doi.org/10.1586/1744666X.2014.955851> PMID: 25225774
54. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 2012; 489(7417):519–25. <https://doi.org/10.1038/nature11404> PMID: 22960745
55. Dutt A, Salvesen HB, Chen TH, Ramos AH, Onofrio RC, Hatton C, et al. Drug-sensitive FGFR2 mutations in endometrial carcinoma. *Proc Natl Acad Sci U S A.* 2008; 105(25):8713–7. <https://doi.org/10.1073/pnas.0803379105> PMID: 18552176
56. Byron SA, Gartside MG, Wellens CL, Mallon MA, Keenan JB, Powell MA, et al. Inhibition of activated fibroblast growth factor receptor 2 in endometrial cancer cells induces cell death despite PTEN abrogation. *Cancer Res.* 2008; 68(17):6902–7. <https://doi.org/10.1158/0008-5472.CAN-08-0770> PMID: 18757403
57. Gozgit JM, Squillace RM, Wongchenko MJ, Miller D, Wardwell S, Mohemmad Q, et al. Combined targeting of FGFR2 and mTOR by ponatinib and ridaforolimus results in synergistic antitumor activity in FGFR2 mutant endometrial cancer models. *Cancer Chemother Pharmacol.* 2013; 71(5):1315–23. <https://doi.org/10.1007/s00280-013-2131-z> PMID: 23468082
58. Liao RG, Jung J, Tchaicha J, Wilkerson MD, Sivachenko A, Beauchamp EM, et al. Inhibitor-sensitive FGFR2 and FGFR3 mutations in lung squamous cell carcinoma. *Cancer Res.* 2013; 73(16):5195–205. <https://doi.org/10.1158/0008-5472.CAN-12-3950> PMID: 23786770
59. Zaharieva BM, Simon R, Diener PA, Ackermann D, Maurer R, Alund G, et al. High-throughput tissue microarray analysis of 11q13 gene amplification (CCND1, FGF3, FGF4, EMS1) in urinary bladder cancer. *The Journal of pathology.* 2003; 201(4):603–8. <https://doi.org/10.1002/path.1481> PMID: 14648664
60. Sugahara K, Michikawa Y, Ishikawa K, Shoji Y, Iwakawa M, Shibahara T, et al. Combination effects of distinct cores in 11q13 amplification region on cervical lymph node metastasis of oral squamous cell carcinoma. *International journal of oncology.* 2011; 39(4):761–9. <https://doi.org/10.3892/ijo.2011.1094> PMID: 21701773
61. Karin M, Greten FR. NF-kappaB: linking inflammation and immunity to cancer development and progression. *Nat Rev Immunol.* 2005; 5(10):749–59. Epub 2005/09/22. <https://doi.org/10.1038/nri1703> PMID: 16175180
62. Pikarsky E, Porat RM, Stein I, Abramovitch R, Amit S, Kasem S, et al. NF-kappaB functions as a tumour promoter in inflammation-associated cancer. *Nature.* 2004; 431(7007):461–6. <https://doi.org/10.1038/nature02924> PMID: 15329734
63. Bu H, Rosdahl I, Sun XF, Zhang H. Importance of polymorphisms in NF-kappaB1 and NF-kappaB1alpha genes for melanoma risk, clinicopathological features and tumor progression in Swedish melanoma patients. *J Cancer Res Clin Oncol.* 2007; 133(11):859–66. <https://doi.org/10.1007/s00432-007-0228-7> PMID: 17492467
64. Zhang P, Wei Q, Li X, Wang K, Zeng H, Bu H, et al. A functional insertion/deletion polymorphism in the promoter region of the NFKB1 gene increases susceptibility for prostate cancer. *Cancer Genet Cytogenet.* 2009; 191(2):73–7. <https://doi.org/10.1016/j.cancergencyto.2009.01.017> PMID: 19446741
65. Lo SS, Chen JH, Wu CW, Lui WY. Functional polymorphism of NFKB1 promoter may correlate to the susceptibility of gastric cancer in aged patients. *Surgery.* 2009; 145(3):280–5. <https://doi.org/10.1016/j.surg.2008.11.001> PMID: 19231580
66. Lin CW, Hsieh YS, Hsin CH, Su CW, Lin CH, Wei LH, et al. Effects of NFKB1 and NFKBIA gene polymorphisms on susceptibility to environmental factors and the clinicopathologic development of oral cancer. *PLoS One.* 2012; 7(4):e35078. <https://doi.org/10.1371/journal.pone.0035078> PMID: 22509384
67. Bevilacqua MP, Nelson RM. Selectins. *J Clin Invest.* 1993; 91(2):379–87. <https://doi.org/10.1172/JCI116210> PMID: 7679406
68. Lauri D, Needham L, Martin-Padura I, Dejana E. Tumor cell adhesion to endothelial cells: endothelial leukocyte adhesion molecule-1 as an inducible adhesive receptor specific for colon carcinoma cells. *Journal of the National Cancer Institute.* 1991; 83(18):1321–4. PMID: 1715924
69. Wenzel CT, Scher RL, Richtsmeier WJ. Adhesion of head and neck squamous cell carcinoma to endothelial cells. The missing links. *Archives of otolaryngology—head & neck surgery.* 1995; 121(11):1279–86.

70. Chang PY, Kuo YB, Wu TL, Liao CT, Sun YC, Yen TC, et al. Association and prognostic value of serum inflammation markers in patients with leukoplakia and oral cavity cancer. *Clinical chemistry and laboratory medicine: CCLM / FESCC*. 2013; 51(6):1291–300.
71. Lahav J. The functions of thrombospondin and its involvement in physiology and pathophysiology. *Biochim Biophys Acta*. 1993; 1182(1):1–14. PMID: [8347680](https://pubmed.ncbi.nlm.nih.gov/8347680/)
72. Yao L, Zhao YL, Itoh S, Wada S, Yue L, Furuta I. Thrombospondin-1 expression in oral squamous cell carcinomas: correlations with tumor vascularity, clinicopathological features and survival. *Oral Oncol*. 2000; 36(6):539–44. PMID: [11036248](https://pubmed.ncbi.nlm.nih.gov/11036248/)
73. Pal SK, Nguyen CT, Morita KI, Miki Y, Kayamori K, Yamaguchi A, et al. THBS1 is induced by TGFB1 in the cancer stroma and promotes invasion of oral squamous cell carcinoma. *Journal of oral pathology & medicine: official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology*. 2016.
74. Durbin JE, Hackenmiller R, Simon MC, Levy DE. Targeted disruption of the mouse Stat1 gene results in compromised innate immunity to viral disease. *Cell*. 1996; 84(3):443–50. PMID: [8608598](https://pubmed.ncbi.nlm.nih.gov/8608598/)
75. Xi S, Dyer KF, Kimak M, Zhang Q, Gooding WE, Chaillet JR, et al. Decreased STAT1 expression by promoter methylation in squamous cell carcinogenesis. *Journal of the National Cancer Institute*. 2006; 98(3):181–9. <https://doi.org/10.1093/jnci/djj020> PMID: [16449678](https://pubmed.ncbi.nlm.nih.gov/16449678/)
76. Laimer K, Spizzo G, Obrist P, Gastl G, Brunhuber T, Schafer G, et al. STAT1 activation in squamous cell cancer of the oral cavity: a potential predictive marker of response to adjuvant chemotherapy. *Cancer*. 2007; 110(2):326–33. <https://doi.org/10.1002/cncr.22813> PMID: [17559122](https://pubmed.ncbi.nlm.nih.gov/17559122/)
77. Haroun F, Al-Shaar L, Habib RH, El-Saghir N, Tfayli A, Bazarbachi A, et al. Effects of CYP2B6 genetic polymorphisms in patients receiving cyclophosphamide combination chemotherapy for breast cancer. *Cancer Chemother Pharmacol*. 2015; 75(1):207–14. <https://doi.org/10.1007/s00280-014-2632-4> PMID: [25428516](https://pubmed.ncbi.nlm.nih.gov/25428516/)
78. Li Y, Zhao W, Zhao Z, Wu J, Chen L, Ma Y, et al. IL1B gene polymorphisms, age and the risk of non-small cell lung cancer in a Chinese population. *Lung Cancer*. 2015; 89(3):232–7. <https://doi.org/10.1016/j.lungcan.2015.06.009> PMID: [26141218](https://pubmed.ncbi.nlm.nih.gov/26141218/)
79. Zeigler-Johnson C, Friebe T, Walker AH, Wang Y, Spangler E, Panossian S, et al. CYP3A4, CYP3A5, and CYP3A43 genotypes and haplotypes in the etiology and severity of prostate cancer. *Cancer Res*. 2004; 64(22):8461–7. <https://doi.org/10.1158/0008-5472.CAN-04-1651> PMID: [15548719](https://pubmed.ncbi.nlm.nih.gov/15548719/)
80. Chen L, Huang Z, Yao G, Lyu X, Li J, Hu X, et al. The expression of CXCL13 and its relation to unfavorable clinical characteristics in young breast cancer. *J Transl Med*. 2015; 13:168. <https://doi.org/10.1186/s12967-015-0521-1> PMID: [25990390](https://pubmed.ncbi.nlm.nih.gov/25990390/)
81. Walser TC, Ma X, Kundu N, Dorsey R, Goloubeva O, Fulton AM. Immune-mediated modulation of breast cancer growth and metastasis by the chemokine Mig (CXCL9) in a murine model. *Journal of immunotherapy*. 2007; 30(5):490–8. <https://doi.org/10.1097/CJ1.0b013e318031b551> PMID: [17589289](https://pubmed.ncbi.nlm.nih.gov/17589289/)
82. Walser TC, Rifat S, Ma X, Kundu N, Ward C, Goloubeva O, et al. Antagonism of CXCR3 inhibits lung metastasis in a murine model of metastatic breast cancer. *Cancer Res*. 2006; 66(15):7701–7. <https://doi.org/10.1158/0008-5472.CAN-06-0709> PMID: [16885372](https://pubmed.ncbi.nlm.nih.gov/16885372/)
83. Kawada K, Hosogi H, Sonoshita M, Sakashita H, Manabe T, Shimahara Y, et al. Chemokine receptor CXCR3 promotes colon cancer metastasis to lymph nodes. *Oncogene*. 2007; 26(32):4679–88. <https://doi.org/10.1038/sj.onc.1210267> PMID: [17297455](https://pubmed.ncbi.nlm.nih.gov/17297455/)
84. Ma X, Norsworthy K, Kundu N, Rodgers WH, Gimotty PA, Goloubeva O, et al. CXCR3 expression is associated with poor survival in breast cancer and promotes metastasis in a murine model. *Mol Cancer Ther*. 2009; 8(3):490–8. <https://doi.org/10.1158/1535-7163.MCT-08-0485> PMID: [19276169](https://pubmed.ncbi.nlm.nih.gov/19276169/)
85. Cambien B, Karimjee BF, Richard-Fiardo P, Bziouech H, Barthel R, Millet MA, et al. Organ-specific inhibition of metastatic colon carcinoma by CXCR3 antagonism. *British journal of cancer*. 2009; 100(11):1755–64. <https://doi.org/10.1038/sj.bjc.6605078> PMID: [19436305](https://pubmed.ncbi.nlm.nih.gov/19436305/)
86. Pradelli E, Karimjee-Soilihi B, Michiels JF, Ricci JE, Millet MA, Vandenbos F, et al. Antagonism of chemokine receptor CXCR3 inhibits osteosarcoma metastasis to lungs. *International journal of cancer Journal international du cancer*. 2009; 125(11):2586–94. <https://doi.org/10.1002/ijc.24665> PMID: [19544560](https://pubmed.ncbi.nlm.nih.gov/19544560/)
87. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4(1):44–57. Epub 2009/01/10. <https://doi.org/10.1038/nprot.2008.211> PMID: [19131956](https://pubmed.ncbi.nlm.nih.gov/19131956/)
88. Kim BK, Lee JW, Park PJ, Shin YS, Lee WY, Lee KA, et al. The multiplex bead array approach to identifying serum biomarkers associated with breast cancer. *Breast cancer research: BCR*. 2009; 11(2):R22. <https://doi.org/10.1186/bcr2247> PMID: [19400944](https://pubmed.ncbi.nlm.nih.gov/19400944/)

89. Key TJ, Appleby PN, Reeves GK, Roddam AW, Breast TEH. Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. *Lancet Oncology*. 2010; 11(6):530–42. [https://doi.org/10.1016/S1470-2045\(10\)70095-4](https://doi.org/10.1016/S1470-2045(10)70095-4) PMID: [20472501](#)
90. Karamouzis MV, Papavassiliou AG. Targeting insulin-like growth factor in breast cancer therapeutics. *Crit Rev Oncol Hematol*. 2012; 84(1):8–17. <https://doi.org/10.1016/j.critrevonc.2012.02.010> PMID: [22424863](#)
91. Polyak K, Lee MH, Erdjumentbromage H, Koff A, Roberts JM, Tempst P, et al. Cloning of P27(Kip1), a Cyclin-Dependent Kinase Inhibitor and a Potential Mediator of Extracellular Antimitogenic Signals. *Cell*. 1994; 78(1):59–66. PMID: [8033212](#)
92. Chang BL, Zheng SQL, Isaacs SD, Wiley KE, Turner A, Li G, et al. A polymorphism in the CDKN1B gene is associated with increased risk of hereditary prostate cancer. *Cancer Research*. 2004; 64(6):1997–9. PMID: [15026335](#)
93. Lang T, Nie YL. MiR-148a participates in the growth of RPMI8226 multiple myeloma cells by regulating CDKN1B. *Biomedicine & Pharmacotherapy*. 2016; 84:1967–71.
94. Nanda MS, Shah ZA, Sameer AS, Syeed N, Murtaza I, Siddiqi MA, et al. TP53-Molecular Soldier's Mutations in Bladder Cancer in the Kashmiri Population. *Asian Pacific Journal of Cancer Prevention*. 2011; 12(1):67–72. PMID: [21517233](#)
95. Cheng LA, Zhang SB, MacLennan GT, Williamson SR, Lopez-Beltran A, Montironi R. Bladder cancer: translating molecular genetic insights into clinical practice. *Human Pathology*. 2011; 42(4):455–81. <https://doi.org/10.1016/j.humpath.2010.07.007> PMID: [21106220](#)
96. Levine AJ, Oren M. The first 30 years of p53: growing ever more complex. *Nat Rev Cancer*. 2009; 9(10):749–58. <https://doi.org/10.1038/nrc2723> PMID: [19776744](#)
97. Nishiyama H, Watanabe J, Ogawa O. p53 and chemosensitivity in bladder cancer. *Int J Clin Oncol*. 2008; 13(4):282–6. <https://doi.org/10.1007/s10147-008-0815-x> PMID: [18704627](#)
98. Wang SJ, Gu W. To be, or not to be: functional dilemma of p53 metabolic regulation. *Curr Opin Oncol*. 2014; 26(1):78–85. <https://doi.org/10.1097/CCO.000000000000024> PMID: [24240177](#)
99. Evan GI, Wyllie AH, Gilbert CS, Littlewood TD, Land H, Brooks M, et al. Induction of Apoptosis in Fibroblasts by C-Myc Protein. *Cell*. 1992; 69(1):19–28. PMID: [1555236](#)
100. Bissonnette RP, Echeverri F, Mahboubi A, Green DR. Apoptotic cell death induced by c-myc is inhibited by bcl-2. *Nature*. 1992; 359(6395):552–4. <https://doi.org/10.1038/359552a0> PMID: [1406975](#)
101. Egle A, Harris AW, Bouillet P, Cory S. Bim is a suppressor of Myc-induced mouse B cell leukemia. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(16):6164–9. <https://doi.org/10.1073/pnas.0401471101> PMID: [15079075](#)
102. Murphy DJ, Junttila MR, Pouyet L, Karnezis A, Shchors K, Bui DA, et al. Distinct thresholds govern Myc's biological output in vivo. *Cancer Cell*. 2008; 14(6):447–57. <https://doi.org/10.1016/j.ccr.2008.10.018> PMID: [19061836](#)
103. Mitchell KO, Ricci MS, Miyashita T, Dicker DT, Jin ZY, Reed JC, et al. Bax is a transcriptional target and mediator of c-Myc-induced apoptosis. *Cancer Research*. 2000; 60(22):6318–25. PMID: [11103792](#)
104. Nicholson DW, Thornberry NA. Caspases: killer proteases. *Trends Biochem Sci*. 1997; 22(8):299–306. PMID: [9270303](#)
105. Dang CV, O'Donnell KA, Zeller KI, Nguyen T, Osthus RC, Li F. The c-Myc target gene network. *Seminars in cancer biology*. 2006; 16(4):253–64. <https://doi.org/10.1016/j.semcancer.2006.07.014> PMID: [16904903](#)
106. Zhao X, Mohaupt M, Jiang J, Liu S, Li B, Qin Z. Tumor necrosis factor receptor 2-mediated tumor suppression is nitric oxide dependent and involves angiostasis. *Cancer Res*. 2007; 67(9):4443–50. <https://doi.org/10.1158/0008-5472.CAN-07-0185> PMID: [17483359](#)
107. Qin Z, Kruger-Krasagakes S, Kunzendorf U, Hock H, Diamantstein T, Blankenstein T. Expression of tumor necrosis factor by different tumor cell lines results either in tumor suppression or augmented metastasis. *J Exp Med*. 1993; 178(1):355–60. PMID: [8315391](#)
108. Blankenstein T, Qin ZH, Uberla K, Muller W, Rosen H, Volk HD, et al. Tumor Suppression after Tumor-Cell Targeted Tumor-Necrosis-Factor-Alpha Gene-Transfer. *Journal of Experimental Medicine*. 1991; 173(5):1047–52. PMID: [2022919](#)
109. Oliff A, Defeo-Jones D, Boyer M, Martinez D, Kiefer D, Vuocolo G, et al. Tumors secreting human TNF/cachectin induce cachexia in mice. *Cell*. 1987; 50(4):555–63. PMID: [3607879](#)
110. Kulbe H, Thompson R, Wilson JL, Robinson S, Hagemann T, Fatah R, et al. The inflammatory cytokine tumor necrosis factor-alpha generates an autocrine tumor-promoting network in epithelial ovarian

- cancer cells. *Cancer Res.* 2007; 67(2):585–92. <https://doi.org/10.1158/0008-5472.CAN-06-2941> PMID: [17234767](https://pubmed.ncbi.nlm.nih.gov/17234767/)
111. Ben-Neriah Y, Karin M. Inflammation meets cancer, with NF-kappaB as the matchmaker. *Nat Immunol.* 2011; 12(8):715–23. <https://doi.org/10.1038/ni.2060> PMID: [21772280](https://pubmed.ncbi.nlm.nih.gov/21772280/)
 112. DiDonato JA, Mercurio F, Karin M. NF-kappa B and the link between inflammation and cancer. *Immunol Rev.* 2012; 246:379–400. <https://doi.org/10.1111/j.1600-065X.2012.01099.x> PMID: [22435567](https://pubmed.ncbi.nlm.nih.gov/22435567/)
 113. Kang MH, Oh SC, Lee HJ, Kang HN, Kim JL, Kim JS, et al. Metastatic function of BMP-2 in gastric cancer cells: the role of PI3K/AKT, MAPK, the NF-kappaB pathway, and MMP-9 expression. *Exp Cell Res.* 2011; 317(12):1746–62. <https://doi.org/10.1016/j.yexcr.2011.04.006> PMID: [21570392](https://pubmed.ncbi.nlm.nih.gov/21570392/)
 114. Graham TR, Odero-Marah VA, Chung LW, Agrawal KC, Davis R, Abdel-Mageed AB. PI3K/Akt-dependent transcriptional regulation and activation of BMP-2-Smad signaling by NF-kappaB in metastatic prostate cancer cells. *The Prostate.* 2009; 69(2):168–80. <https://doi.org/10.1002/pros.20870> PMID: [18942118](https://pubmed.ncbi.nlm.nih.gov/18942118/)
 115. Levine RL, Pardanani A, Tefferi A, Gilliland DG. Role of JAK2 in the pathogenesis and therapy of myeloproliferative disorders. *Nat Rev Cancer.* 2007; 7(9):673–83. <https://doi.org/10.1038/nrc2210> PMID: [17721432](https://pubmed.ncbi.nlm.nih.gov/17721432/)
 116. Silva M, Richard C, Benito A, Sanz C, Olalla I, Fernandez-Luna JL. Expression of Bcl-x in erythroid precursors from patients with polycythemia vera. *N Engl J Med.* 1998; 338(9):564–71. <https://doi.org/10.1056/NEJM199802263380902> PMID: [9475763](https://pubmed.ncbi.nlm.nih.gov/9475763/)
 117. Sansone P, Bromberg J. Targeting the Interleukin-6/Jak/Stat Pathway in Human Malignancies. *Journal of Clinical Oncology.* 2012; 30(9):1005–14. <https://doi.org/10.1200/JCO.2010.31.8907> PMID: [22355058](https://pubmed.ncbi.nlm.nih.gov/22355058/)
 118. Traer E, MacKenzie R, Snead J, Agarwal A, Eiring AM, O'Hare T, et al. Blockade of JAK2-mediated extrinsic survival signals restores sensitivity of CML cells to ABL inhibitors. *Leukemia.* 2012; 26(5):1140–3. <https://doi.org/10.1038/leu.2011.325> PMID: [22094585](https://pubmed.ncbi.nlm.nih.gov/22094585/)
 119. Groeger S, Jarzina F, Domann E, Meyle J. Porphyromonas gingivalis activates NFkappaB and MAPK pathways in human oral epithelial cells. *BMC Immunol.* 2017; 18(1):1. <https://doi.org/10.1186/s12865-016-0185-5> PMID: [28056810](https://pubmed.ncbi.nlm.nih.gov/28056810/)
 120. McCubrey JA, Steelman LS, Chappell WH, Abrams SL, Franklin RA, Montalto G, et al. Ras/Raf/MEK/ERK and PI3K/PTEN/Akt/mTOR Cascade Inhibitors: How Mutations Can Result in Therapy Resistance and How to Overcome Resistance. *Oncotarget.* 2012; 3(10):1068–111. <https://doi.org/10.18632/oncotarget.659> PMID: [23085539](https://pubmed.ncbi.nlm.nih.gov/23085539/)
 121. Wei L, Tang J, Zou Q. Local-DPP: An improved DNA-binding protein prediction method by exploring local evolutionary information. *Inform Sciences.* 2017; 384:135–44.
 122. Wei L, Zou Q. Recent Progress in Machine Learning-Based Methods for Protein Fold Recognition. *International journal of molecular sciences.* 2016; 17(12).
 123. Wei L, Liao M, Gao X, Zou Q. Enhanced Protein Fold Prediction Method Through a Novel Feature Extraction Technique. *IEEE Trans Nanobioscience.* 2015; 14(6):649–59. <https://doi.org/10.1109/TNB.2015.2450233> PMID: [26335556](https://pubmed.ncbi.nlm.nih.gov/26335556/)
 124. Su R, Zhang C, Pham TD, Davey R, Bischof L, Vallotton P, et al. Detection of tubule boundaries based on circular shortest path and polar-transformation of arbitrary shapes. *Journal of microscopy.* 2016; 264(2):127–42. <https://doi.org/10.1111/jmi.12421> PMID: [27172164](https://pubmed.ncbi.nlm.nih.gov/27172164/)
 125. Bromberg J. Stat proteins and oncogenesis. *J Clin Invest.* 2002; 109(9):1139–42. <https://doi.org/10.1172/JCI15617> PMID: [11994401](https://pubmed.ncbi.nlm.nih.gov/11994401/)
 126. Wu YW, Tang W, Zuo JP. Toll-like receptors: potential targets for lupus treatment. *Acta pharmacologica Sinica.* 2015; 36(12):1395–407. <https://doi.org/10.1038/aps.2015.91> PMID: [26592511](https://pubmed.ncbi.nlm.nih.gov/26592511/)
 127. Xiang W, Chao ZY, Feng DY. Role of Toll-like receptor/MYD88 signaling in neurodegenerative diseases. *Reviews in the neurosciences.* 2015; 26(4):407–14. <https://doi.org/10.1515/revneuro-2014-0067> PMID: [25870959](https://pubmed.ncbi.nlm.nih.gov/25870959/)
 128. Muller M, Herrath J, Malmstrom V. IL-1R1 is expressed on both Helios(+) and Helios(-)FoxP3(+)CD4(+) T cells in the rheumatic joint. *Clinical and Experimental Immunology.* 2015; 182(1):90–100. <https://doi.org/10.1111/cei.12668> PMID: [26076982](https://pubmed.ncbi.nlm.nih.gov/26076982/)
 129. Danielson PB. The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Curr Drug Metab.* 2002; 3(6):561–97. PMID: [12369887](https://pubmed.ncbi.nlm.nih.gov/12369887/)
 130. Haudenschild DR, Palmer SM, Moseley TA, You Z, Reddi AH. Bone morphogenetic protein (BMP)-6 signaling and BMP antagonist noggin in prostate cancer. *Cancer Res.* 2004; 64(22):8276–84. <https://doi.org/10.1158/0008-5472.CAN-04-2251> PMID: [15548695](https://pubmed.ncbi.nlm.nih.gov/15548695/)

131. Zhang X, Ibrahimi OA, Olsen SK, Umemori H, Mohammadi M, Ornitz DM. Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family. *J Biol Chem.* 2006; 281(23):15694–700. <https://doi.org/10.1074/jbc.M601252200> PMID: 16597617
132. Zhou H, Wu J, Wang T, Zhang X, Liu D. CXCL10/CXCR3 axis promotes the invasion of gastric cancer via PI3K/AKT pathway-dependent MMPs production. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie.* 2016; 82:479–88.
133. Pranzatelli MR, Tate ED, McGee NR, Travelstead AL, Verhulst SJ, Ransohoff RM. Expression of CXCR3 and its ligands CXCL9, -10 and -11 in paediatric opsoclonus-myoclonus syndrome. *Clin Exp Immunol.* 2013; 172(3):427–36. <https://doi.org/10.1111/cei.12065> PMID: 23600831
134. Baenziger NL, Brodie GN, Majerus PW. Isolation and properties of a thrombin-sensitive protein of human platelets. *J Biol Chem.* 1972; 247(9):2723–31. PMID: 4260214
135. Bornstein P. Diversity of function is inherent in matricellular proteins: an appraisal of thrombospondin 1. *The Journal of cell biology.* 1995; 130(3):503–6. PMID: 7542656
136. Labib K, Tercero JA, Diffley JF. Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science.* 2000; 288(5471):1643–7. PMID: 10834843
137. Elcin AE, Parmaksiz M, Dogan A, Seker S, Durkut S, Dalva K, et al. Differential gene expression profiling of human adipose stem cells differentiating into smooth muscle-like cells by TGFbeta1/BMP4. *Exp Cell Res.* 2017.
138. Vallinoto AC, Graca ES, Araujo MS, Azevedo VN, Cayres-Vallinoto I, Machado LF, et al. IFNG +874T/A polymorphism and cytokine plasma levels are associated with susceptibility to Mycobacterium tuberculosis infection and clinical manifestation of tuberculosis. *Human immunology.* 2010; 71(7):692–6. <https://doi.org/10.1016/j.humimm.2010.03.008> PMID: 20353805
139. Elgizouli M, Lowe DM, Speckmann C, Schubert D, Hulsdunker J, Eskandarian Z, et al. Activating PI3Kdelta mutations in a cohort of 669 patients with primary immunodeficiency. *Clin Exp Immunol.* 2016; 183(2):221–9. <https://doi.org/10.1111/cei.12706> PMID: 26437962
140. Wang G, Wang H, Zhang C, Liu T, Li Q, Lin X, et al. Rac3 regulates cell proliferation through cell cycle pathway and predicts prognosis in lung adenocarcinoma. *Tumour Biol.* 2016; 37(9):12597–607. <https://doi.org/10.1007/s13277-016-5126-7> PMID: 27402308
141. Haataja L, Kaartinen V, Groffen J, Heisterkamp N. The small GTPase Rac3 interacts with the integrin-binding protein CIB and promotes integrin alpha(IIb)beta(3)-mediated adhesion and spreading. *J Biol Chem.* 2002; 277(10):8321–8. <https://doi.org/10.1074/jbc.M105363200> PMID: 11756406
142. Mira JP, Benard V, Groffen J, Sanders LC, Knaus UG. Endogenous, hyperactive Rac3 controls proliferation of breast cancer cells by a p21-activated kinase-dependent pathway. *Proc Natl Acad Sci U S A.* 2000; 97(1):185–9. PMID: 10618392