

Research Article

An Outranking Approach for Gene Prioritization Using Multinetworks

 Jesús Jaime Solano Noriega¹, Juan Carlos Leyva López^{1,*}, Fiona Browne², Jun Liu³
¹Department of Economic and Management Sciences, Universidad Autónoma de Occidente, Lola Beltrán, Culiacán, México

²Artificial Intelligence Team, Datactics, Belfast, Northern Ireland, United Kingdom

³School of Computing and Mathematics, Ulster University, Jordanstown Campus, Northern Ireland, United Kingdom

ARTICLE INFO

Article History

Received 07 May 2021

Accepted 22 May 2021

Keywords

 Disease gene prioritization
 Multicriteria decision support
 Fuzzy outranking
 Multinetwork analysis
 Topological analysis
 Omic integration

ABSTRACT

High-throughput experimental techniques such as genome-wide association studies have been instrumental in the identification of disease-associated genes. These methods often produce large lists of disease candidate genes which are time-consuming and expensive to experimentally validate. Computational gene prioritization methods are required to identify relevant genes from a larger pool of candidates. Research has shown that the integration of diverse “omic” evidence can reduce the candidate-gene search space. In this paper we present a general framework that integrates “omic” data using a multinetwork approach and topological analysis to prioritize disease-candidate genes. Specifically, we propose a data integration method within a multicriteria decision analysis context using aggregation mechanisms based on decision rules identifying positive and negative criteria for judging gene-candidates ranks. The proposed *multinetwork disease gene prioritization* method is applied to the prioritization of disease genes in ovarian cancer progression. Using this approach we uncovered known ovarian cancer genes GSTA1, ERBB2, IL1A, MAGEB2, along with significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways ErbB signaling and pathways in cancer. Relatively high predictive performance (area under Receiver Operating Characteristic [ROC] curve 0.704) was observed when classifying epithelial ovarian high-grade serous carcinoma cancer early and late stage RNA-Seq expression profiles from individuals using 10-fold cross-validation.

© 2021 The Authors. Published by Atlantis Press B.V.

 This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

The identification of genes associated with disease development and progression is an important and complex challenge. Discovery of disease-associated genes has resulted in an improved understanding of gene functions, their interactions and pathways and application in medical treatments. However, diseases are rarely the result of an abnormality in a single gene. Instead they involve interactions between cellular and molecular processes [1]. Relationships between these processes can be encoded in a comprehensive network known as the interactome which is the collection of all physical interactions within a cell. These interactions include regulatory and metabolic interactions [2].

The advances in large-scale high-throughput technologies such as linkage analysis and Genomic-Wide Association Studies (GWAS) have aided in our understanding of genes and their roles in complex diseases [3]. Such approaches have been instrumental in the generation of candidate disease genes. A recent meta-analysis GWAS study on Parkinson’s disease by Nalls *et al.* [4] used a set of 7,893,274 variants across 13,708 cases and 95,282 controls and identified 6 new loci involved in Parkinson’s disease. Genome-wide linkage

analysis for congenital heart defects was performed by Flaquer *et al.* [5] using MOD-score analysis resulting in the identification of 4 novel candidate genes. A common problem with these methods is the resulting large lists of disease-candidate genes. Due to limited resources, scientists can only select a subset of genes to investigate further in the preclinical setting. To address these limitations, computational gene prioritization methods need to be developed to aid in the identification of relevant genes from a larger pool of candidates.

Networks have been for gene prioritization [6,7] whereby a set of seed genes (such as known disease genes) are used by an algorithm to rank potential candidate genes in a network. This is based on the premise that genes in close proximity to seed genes in a network are more likely to be associated with a disease [8]. A number of methods to measure proximity include PageRank, HITS [9], random walk with restart [10] and diffusion [11]. A comprehensive comparison of state-of-the-art gene prioritization methods can be found in [12]. Gene prioritization tools include ENDEAVOUR [13] and ToppGene [14] which rank candidate genes by applying algorithms to a network based on annotation similarity scores calculated from enrichment values based on a training data set of seed genes. AlignPI proposed by Wu *et al.* [15] uses a network alignment approach to predict disease genes. GeneProspector [16] prioritizes

 *Corresponding author. Email: juan.leyva@uadeo.mx

potential disease candidates using a curated literature database of genetic association studies. A review of computational tools can be found in [17].

Use of biological networks such as Protein–Protein Interaction network (PPIN) and network analysis is a commonly used approach for the uncovering of disease genes [18]. The PPIN is an important biological network widely used in areas from drug target prediction, predication of protein complexes, functions, identification of essential genes and network motif discovery [19]. Studies such as [20] have applied graph centrality measures to PPINs integrating breast cancer signatures to understand the relationship between gene expression and clinical outcome of breast cancer. Li *et al.* [21] applied centrality measures to a weighted PPIN for essential gene prediction. Taylor *et al.* [22] indicated that the alterations in the physical interaction network may be an indicator of breast cancer prognosis [23]. A study by Lui *et al.* [24] used GWAS studies, genetic variants, gene coexpression data and PPIN to identify significant interaction among Alzheimer and Parkinson Disease susceptibility genes.

Data integration is another approach which collates “omic” data from different sources such as the Gene Ontology (GO) and tissue and pathway along with PPIN data for disease gene prioritization. For example, Chen *et al.* [25] proposed a prioritization approach that integrated multiple data sources such as gene coexpression and PPIN along with a unified graphic representation of information. The study by Li *et al.* [26] constructed a weighted tissue-specific network by integrating DNA methylation and PPIN data, a pagerank algorithm was applied to rank candidate disease genes. This approach outperformed the ranking on the original PPIN. The study by Cantini *et al.* [27] applied a community detection consensus clustering algorithm to a multinetwork of layers

including mRNA cotargeting and gene coexpression to identify disease genes in cancers including lung and colorectal. Enrichment analysis confirmed candidate genes involvement in cancer related regulatory pathways. Bennett *et al.* [28] recently proposed a mixed integer nonlinear programming model, SimMod that optimizes modularity across layers of a multilayer network resulting in a single partition of composite communities. When applied to yeast networks of physical, genetic and coexpression interactions strongly enriched composite modules in terms of the GO were identified. Current multinetwork analysis methods consider the prioritized lists of disease-candidate genes obtained from different networks as a total ordered set of disease-candidate genes. In this paper we propose a novel outranking-based gene ranking method to prioritize disease-candidate genes. We present our generalized framework (Figure 1) which integrates multiple omic data sources including: (1) human PPIN, (2) RNA-Seq coexpression, (3) Transcription Factor (TF) CoTargeting and (4) tissue-specific data. We apply this framework to prioritize disease genes involved in high-grade serous carcinoma (HGSC) ovarian cancer progression. Ovarian HGSC is associated with substantial mortality and morbidity but, can be difficult to detect at its earliest stage due to its molecular complexity and clinical behavior [29]. Therefore, identification of key gene signatures at an early stage will be highly helpful. A list of significant disease-related genes derived from the integrative framework using RNA-Seq, protein interaction, tissue and TF cotargeting network data are ranked by our outranking method in decreasing order of preference. The method is based on ELECTRE III [30] to create a fuzzy outranking relation and then, it uses a multiobjective evolutionary algorithm [31] to exploit the outranking relation and to derive a prioritized list of candidate disease genes. The approach takes into consideration criteria including network centralities. The proposed framework is advantageous

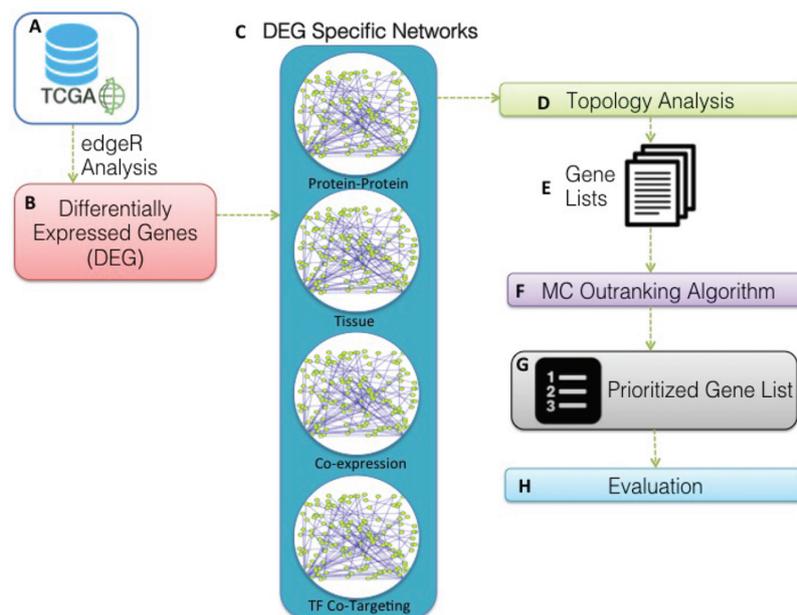


Figure 1 | Proposed framework integrating diverse networks using novel outranking approach. This framework consists of a number of steps including data selection and preprocessing (stage A and B), diverse network construction (stage C) analysis and ranking (stage D, E and F) and evaluation of prioritized lists (stage G and H).

as data from diverse “omic” networks can be combined using the proposed fuzzy outranking and multiobjective evolutionary algorithm to produce a prioritized list of disease-candidate genes. Furthermore, this generalized framework can be applied to other diseases and multinet network problem domains. The remainder of the paper is outlined as follows. In Section 2 the Methods are detailed including an overview of the framework and data sets applied. Section 3 highlights key results in terms of both classification predictive power and biological relevance. Conclusions and future work are presented in Section 4.

2. METHODS

To identify the key genes whose expression may discriminate between early and late stage ovarian HGSC (OV) samples, we developed the following methodology illustrated in Figure 1. This approach consists of a number of steps: (A) extracting RNA-Seq OV gene expression and clinical information, (B) identification of differentially expressed genes (DEG) between early stage and late stage tumor samples, (C) construction of DEG specific networks, (D) analysis of the DEG specific networks to identify stage associated hub genes and their interactors, (E) novel fuzzy outranking algorithm to combine lists of key hub genes across the networks providing a prioritized gene list, (F) evaluation of hub genes including the development of a hub gene-based classifier model to distinguish OV stages.

2.1. Datasets

Transcriptome and clinical ovarian OV data from the TCGA data portal (<https://portal.gdc.cancer.gov/>) were applied in this analysis. Batch corrected TCGA OV data was downloaded from the MD Anderson TCGA Batch Effects website (<http://bioinformatics.mdanderson.org/tcgambatch/>) using the category UNC (IlluminaHiSeqRNASeqV2) which we refer to as RNA-Seq in this paper, EB-parametric priors and the PCA assessment algorithm. The data set consists of 20,531 genes across 262 samples.

2.2. Differential Expression Analysis of Individual Genes

DEG analysis of individual genes from the RNA-Seq dataset was carried out using the edgeR Bioconductor package [32]. Construction of DEG specific networks was performed using each of the four networks described in the following section.

2.3. Networks

2.3.1. Protein–protein interaction network

High-quality human protein interactions were obtained from InWeb_IM [33]. PPIs were derived from eight sources including BioGRID [34] and IntAct [35]. This experimental interaction data came from human and model organisms and is scored against a gold standard [33]. PPI confidence scores of 0.10 and above were selected

in this paper as they are considered highly reliable [33] resulting in 625,641 pairs.

2.3.2. Coexpression network

Using the TCGA OV dataset we calculate absolute correlation coefficients between 0 and 1 using the Spearman method for gene expression levels among genes. Binary pairs were filtered to include expression values ≥ 0.5 (a similar approach used by Zhang *et al.* [36] when using coexpression networks in predicting gene functions and disease biomarkers) resulting in 145,198 pairs.

2.3.3. Tissue-specific functional interaction network

An ovarian tissue-specific functional interaction network was retrieved from the GIANT interface [37]. Greene *et al.* [37] constructed functional interaction networks of human tissues using a data-driven Bayesian approach. Integration of a large volume of experiments was performed using data from more than 14,000 publications. The tissue-specific network was filtered to only include edges with evidence supporting a tissue-specific functional interaction resulting in 64,594,988 pairs.

2.3.4. TF cotargeting network

The TF cotargeting network was obtained from the study by Cantini *et al.* [27]. A weighted network was constructed in this study using experimentally validated TF-target interactions using ChIP-seq from ENCODE [38]. The weight of the network links represent TFs shared by gene pairs. The weight of the link is the count of TFs targeting both the genes.

Construction of DEG specific networks was performed using each of the networks described. The set of DEG were mapped to the network nodes via NCBI gene IDs for each network. Each network was filtered based on these DEG and their interactors.

2.4. Topological Analysis

Studies including [39] and [40] have shown how network topological metrics can detect key biological processes. The study by Yu *et al.* [41] showed how nodes with high betweenness values control most of the information flow in a protein interaction network. Topology-based analysis of metabolic networks has been important in studying their impact on modularity [42] and function and regulation [43]. These networks are represented as an undirected graph, $G = (V, E)$, whereby V represents a set of nodes (proteins) and $E = \{(u, v) | u, v \in V\}$, the set of edges connecting the nodes. In this study Bonacich Power Centrality and PageRank topological metrics were applied to the four DEG specific networks to identify key players in biological processes using the iGraph, plyr and influence R libraries in R [44].

2.4.1. PageRank

PageRank ranks genes according to importance, that is, connection to other genes. This metric was recently applied to associate gene

connectivity with different subtypes of ovarian cancer tumors [45]. PageRank can be described as:

$$PR_{v_n} = \frac{1-d}{N} + d \sum_{v_j \in (v_n)} \frac{PR(v_j)}{L(v_j)} \quad (1)$$

where d represents the damping factor, n the index of node of interest, v_n the node, $M(v_i)$ the set of nodes linking to v_n and $L(v_j)$ the out-link counts from node v_j .

2.4.2. Bonacich power centrality

This metric is used to measure the power of the given node in the network. The Bonacich power of gene i , $c_i(\alpha, \beta)$ can be described as:

$$c_i(\alpha, \beta) = \sum_j (\alpha + \beta c_j) A_{ij} \quad (2)$$

where A_{ij} is the element of row i and column j of the adjacency matrix. Parameter α is used to normalize the measure. Parameter β sets the dependence of each nodes centrality to the adjacent nodes.

2.5. Outranking Approach for Multinetwork Disease Gene Prioritization

Topological analysis for each of the four networks described in the Networks section was performed and genes from the networks were ranked in terms of their PageRank and their Bonacich Power Centrality values from high to low. This resulted in 8 lists of ranked genes across the four networks. These lists were combined to produce a final list of ranked genes using our proposed outranking algorithm described below.

2.5.1. Preliminaries

For ease of description, some basic multicriteria elements are first defined. Let $A = \{a_1, a_2, \dots, a_{m_i}\}$ be a set of a_{m_i} genes of interest. A priority list or a ranking \succ_k is an ordering defined on $A_k \subseteq A$, $k = 1, 2, \dots, n$. Thus $a_i \succ_k a_j$ means a_i is ranked better than a_j in \succ_k . When $A_k = A$, \succ_k is said to be a *full list*. Otherwise, it is a *partial list*. If $a_i \in A_k$, r_i^k denotes the *rank* or *position* of a_i in \succ_k . We assume that the best gene is assigned the position 1 and the worst one is assigned the position $|A_k|$. Let P_A be the set of all permutations on A . A *profile* is a n -tuple of rankings or priority lists $G = (\succ_1, \succ_2, \dots, \succ_n)$. Restricting G to the rankings containing gene a_i defines G_i .

The *Multinetwork Disease Gene Prioritization* problem consists of finding a ranking function or procedure Ψ defined by:

$$\Psi : P^n \rightarrow P \quad (3)$$

$$G = (\succ_1, \succ_2, \dots, \succ_n) \rightarrow \sigma = \Psi(G) \quad (4)$$

where σ is called a *consensus ranking*.

We consider, in the following, cases in which only ordinal information is available and no other additional information is provided such as the relevance scores.

2.5.2. Specificities of the multinetwork disease gene prioritization problem

Limited Significance of the Prioritization List of Disease-Candidate Genes

The exact positions of genes in one prioritization list of disease-candidate genes have limited significance and should not be overemphasized. For instance, having three potential disease-candidate genes in the first three positions, any perturbation of these three pieces will have the same value. Indeed, in the disease gene prioritization context, the *complete order* provided by a gene prioritization method may hide *ties*. In this case, we call such rankings *pre orders*. Moreover, some genes a_i and a_j could be incomparable when there is no clear evidence in favor of some type of preference or indifference. In this case, we call such rankings *partial (pre)orders*. It is therefore important to build the consensus ranking based on *robust information*:

- Genes with near positions in \succ_k are more likely to have similar interest or relevance. Thus a slight perturbation of the initial ranking is meaningless.
- Assuming that gene a_i is better ranked than gene a_j in a ranking \succ_k , a_i is likely to be more relevant than a_j in \succ_k when the number of intermediate positions between a_i and a_j increases.

Partial Priority Lists

Disease gene prioritized lists obtained from each of the “omic” networks are often partial lists. Integration of partial lists raises four major difficulties which we state hereafter, proposing for each of them various working assumptions:

1. Partial lists can have various lengths, which can favor long lists. We thus consider the following two working hypotheses:
 - $H_{k\%}^1$: We only consider the top $k\%$ best genes from each prioritization list of disease-candidate genes.
 - H_{all}^1 : We consider all the genes from each prioritization list of disease-candidate genes.
2. Since there are different genes in the prioritization lists of disease-candidate genes, we must decide which genes should be kept in the consensus ranking. Two working hypotheses are therefore considered:
 - H_k^2 : We only consider genes which are present in at least k prioritization lists of disease-candidate genes ($k > 1$).
 - H_{all}^2 : We consider all the genes which are ranked in at least one prioritization list of disease-candidate genes.

Hereafter, we call genes which will be retained in the consensus ranking, *candidate genes*, and genes that will be excluded from the consensus ranking, *excluded genes*. We also call a candidate gene which is missing in one or more rankings, a *missing gene*.

3. Some candidate genes are missing genes in some prioritized list. Main reasons for a missing gene are that it was not ranked or it was indexed but deemed irrelevant; usually this information is not available. We consider the following two working hypotheses:

- H_{yes}^3 : Each missing gene in each \succ_k is assigned a position.
- H_{no}^3 : No assumption is made, that is each missing gene is considered neither better nor worse than any other gene.

4. When assumption H_k^2 holds, each prioritized list may contain genes which will not be considered in the consensus ranking. Regarding the positions of the candidate genes, we can consider the following working hypotheses:

- H_{init}^4 : The initial positions of candidate genes are kept in each prioritized list.
- H_{new}^4 : Candidate genes receive new positions in each prioritized list, after discarding excluded ones.

2.5.3. Outranking approach for multi-network disease gene prioritization

Positional methods like Borda count [46], linear combination methods [47], Footrule optimal aggregation [48,49] and Probabilistic methods [50], consider implicitly that the positions of the genes in the prioritized lists are *scores* giving thus a cardinal meaning to an ordinal information. This constitutes a strong assumption that is questionable, especially when the prioritized lists have different lengths. Moreover, for positional methods, assumptions H^3 and H^4 , which are often arbitrary, have a strong impact on the results. For instance, let us consider a prioritized list of 120 genes out of 250 disease-candidate genes. Whether we assign to each of the missing genes the position 3, 34, 76 or 180—corresponding to variations of H_{yes}^3 , it will give rise to very contrasted results, especially regarding the top of the consensus ranking.

Majoritarian methods like Condorcet procedure [51], Kemeny optimal aggregation, and Markov chain methods [52] do not suffer from the above-mentioned drawbacks of the positional methods since they build consensus rankings exploiting only *ordinal information* contained in the gene prioritized lists. Nevertheless, they suppose that such rankings are complete orders, ignoring that they may hide ties or incompatibilities. Therefore, majoritarian methods base consensus rankings on illusory discriminant information rather than less discriminant but more robust information. Trying to overcome the limits of current integration methods, we found that *outranking approaches*, which were initially used for multi-criteria aggregation problems [30], can also be used for the data integration purpose, where each prioritized list plays the role of a criterion. A representative method of the outranking approach is the ELECTRE III method [53,54]. The ELECTRE III method is based on a pairwise comparison of the genes, leading to fuzzy preference degrees. It is a variation of the Condorcet procedure. A comprehensive literature review on methodologies and applications on ELECTRE can be found in [55]. In this paper, we adapt the ELECTRE III method to the *Multinetwork Disease Gene Prioritization* problem.

ELECTRE III method starts by a pairwise comparison of each gene to the remaining ones with the aim of accepting, rejecting, or, more generally, assessing the credibility of the assertion gene a_i is ranked at least at the same position (rank) than gene a_j , usually called a_i outranks a_j (denoted $a_i Sa_j$) taking into account the following three aspects:

- the indifference (q) and preference (p) thresholds defined for each gene prioritized list.
- the coefficients of importance attached to each gene prioritized list.
- the possible difficulties of relative comparison of two genes when one is significantly better than the other on a subset of gene prioritized lists, but much worse on at least one gene prioritized list from a complementary subset (veto threshold (v)).

The ELECTRE III model allows, with the use of thresholds, to take into account the ill-determination, imprecision, and uncertainty that may affect positions. For instance:

- a gene a_j such that r_j^k is smaller than r_i^k but greater than $r_i^k - q$ will be considered indifferent to a_i .
- a gene a_j such that r_j^k is smaller than $r_i^k - p$ will be considered as strictly preferred to a_i .
- a gene a_j such that r_j^k is smaller than $r_i^k - q$ but greater than $r_i^k - p$, the preference will be considered as not significantly established.

where r_i^k is the position or rank of a_i in the k^{th} prioritized list, $k = 1, 2, \dots, n$.

The comparison of genes in the way that has just been described before leads to the construction of a concordance index for each ordered pair of genes (a_i, a_j) , which expresses to what extent the gene prioritized list is in harmony with the assertion a_i is ranked at least at the same position (rank) than a_j .

The three thresholds can be defined as follows [56]:

- the *indifference threshold* (q) corresponds to the largest difference of positions between two genes, compatible with an indifference situation;
- the *preference threshold* (p) corresponds to the smallest difference of positions between two genes from which the decision-maker strictly prefers the gene presenting the best position;
- the *veto threshold* (v) is the smallest difference of the positions between two genes from which the decision-maker considers that it is not possible to support the idea that the worst of the two genes under consideration on a certain gene prioritization list may be comprehensively considered as good as the better one, even if its positions on all the other gene prioritized lists are better.

The choice of thresholds intimately affects whether a particular binary relation holds. While the choice of appropriate threshold is

not easy, in disease gene prioritization decision-making situations, there are good reasons for choosing nonzero values for p and q .

Using thresholds, the ELECTRE method seeks to build an outranking relation S . a_iSa_j means that, according to the global model of DM preferences, there are good reasons to consider that “ a_i gene is ranked at least at the same position than gene a_j ”. Each ordered pair of genes (a_i, a_j) is then tested to check if assertion a_iSa_j is valid or not.

The test to accept assertion a_iSa_j is implemented using two principles:

- 1) A concordance principle, which requires that a majority of prioritized lists, after considering their relative importance, are in favor of the assertion (the majority principle), and
- 2) A nondiscordance principle, which requires that within the minority of prioritized lists, which do not support the assertion, none of them is strongly against the assertion (the respect of minorities’ principle).

The operational implementation of these two principles is now discussed. We first consider the outranking relation defined for each of the n prioritized lists, that is, a_iSa_j means that “gene a_i is ranked at least at the same position than gene a_j ” with respect to the k^{th} prioritized list, $k = 1, 2, \dots, n$.

The k^{th} prioritized list is in concordance with assertion $a_iS_ka_j$ if and only if $a_iS_ka_j$, that is, if $r_i^k - q_k \leq r_j^k$. Thus, even if r_i^k is greater than r_j^k by an amount up to q_k , it does not contravene assertion $a_iS_ka_j$ and therefore is in concordance.

The k^{th} criterion is in discordance with assertion $a_iS_ka_j$ if and only if $a_jP_ka_i$, that is, if $r_j^k \leq r_i^k - p_k$. That is, if a_j is strictly preferred to a_i for criterion k , then it is clearly not in concordance with the assertion $a_iS_ka_j$.

With these concepts, it is now possible to measure the strength of assertion $a_iS_ka_j$. The first step is to develop a measure of concordance, as contained in the concordance index $C(a_i, a_j)$, for every pair of genes $(a_i, a_j) \in A \times A$. Let w_k be the importance coefficient or weight for criterion k . We define a concordance measure as follows:

$$C(a_i, a_j) = \frac{1}{w} \sum_{k=1}^n w_k c_k(a_i, a_j), w = \sum_{k=1}^n w_k \quad (5)$$

In which

$$c_k(a_i, a_j) = \begin{cases} 1, & \text{if } r_i^k - r_j^k \leq q_k \\ 0, & \text{if } r_i^k - r_j^k > p_k \\ \frac{p_k - [r_i^k - r_j^k]}{p_k - q_k}, & \text{otherwise} \end{cases} \quad (6)$$

where $k = 1, 2, \dots, n$.

Thresholds and weights represent subjective input provided by the decision maker. Weights in ELECTRE are coefficients of importance and, as [56] notes, are like votes given to each of the gene prioritized list candidates. Criteria weights are crucial factors in any

multicriteria decision analysis (MCDA) method as they affect the final solution derived from the aggregation procedure. In this proposal, we set the same importance to each of the eight lists, however, it could be used any MCDA method to set the relative importance for each list such as the one presented in [57].

In the concordance index, we have, in a manner of speaking, a measure of the extent to which we are in harmony with the assertion that a_i is ranked at least at the same position than a_j , a_iSa_j . However, what disconfirming or disharmonious evidence do we have? In other words, is there any discordance associated with assertion $a_iS_ka_j$? To calculate discordance, a further threshold called the veto threshold is defined. The veto threshold v_k allows for the possibility of a_iSa_j to be refused totally if, for any one criterion k , $r_j^k < r_i^k - v_k$. The discordance index for each criterion k , $d_k(a_i, a_j)$ is calculated as follows:

$$d_k(a_i, a_j) = \begin{cases} 0, & \text{if } r_i^k - p_k \leq r_j^k \\ 1, & \text{if } r_i^k - v_k \geq r_j^k \\ \frac{r_j^k - [r_i^k - p_k]}{p_k - v_k}, & \text{otherwise} \end{cases} \quad (7)$$

where $k = 1, 2, \dots, n$.

For each pair of genes $(a_i, a_j) \in A \times A$, there are now a concordance and a discordance measure. The final step in the model building phase is to combine these two measures to produce a measure of the degree of outranking, that is, a credibility index $\sigma(a_i, a_j)$, ($0 \leq \sigma(a_i, a_j) \leq 1$) which assesses the strength of the assertion that a_i is ranked at least at the same position than a_j . The credibility degree for each pair $(a_i, a_j) \in A \times A$ is defined as follows:

$$\sigma(a_i, a_j) = \begin{cases} C(a_i, a_j) & , \text{if } K(a_i, a_j) = \phi \\ C(a_i, a_j) \cdot \prod_{k \in K(a_i, a_j)} \frac{1 - d_k(a_i, a_j)}{1 - C(a_i, a_j)} & , \text{if } K(a_i, a_j) \neq \phi \end{cases} \quad (8)$$

where $K(a_i, a_j)$ is the set of criteria such that $d_k(a_i, a_j) > C(a_i, a_j)$.

This formula assumes that if the strength of the concordance exceeds that of the discordance, then the concordance value should not be modified. Otherwise, we are forced to question the assertion that a_iSa_j and modify $C(a_i, a_j)$ according to the above equation. If the discordance is 1.0 for any $(a_i, a_j) \in A \times A$ and any criterion k , then we have no confidence that a_iSa_j ; therefore, $\sigma(a_i, a_j) = 0.0$. Hence, we have constructed a fuzzy outranking relation S_A^σ defined on $A \times A$; this means that we associate each ordered pair $(a_i, a_j) \in A \times A$ a real number $\sigma(a_i, a_j)$, ($0 \leq \sigma(a_i, a_j) \leq 1$) that reflects the degree of strength of the arguments favouring the crisp outranking a_iSa_j . This concludes the construction of the outranking model. The next step in the outranking approach is to exploit the model and produce a ranking of genes from the fuzzy outranking relation S_A^σ .

Outranking relations are not necessarily transitive and do not necessarily correspond to rankings since directed cycles may exist. Therefore, we need specific procedures in order to derive a consensus ranking. We propose the procedure which finds its roots in [31,58,59]. Our approach for exploitation a fuzzy outranking relation to derive a ranking is to use a multiobjective evolutionary algorithm-based heuristic method. It consists in partitioning

the set of genes into r classes. Each class contains genes with the same relevance. Then, based solely on the initially provided information, we elicit a reflexive and antisymmetric crisp outranking relation between the determined classes. After that, with the above as background, we propose a partial order of classes of genes as a recommendation for the Multinetwork Disease Gene Prioritization problem. This approach integrates partitions and relations between classes into the optimization process that the multiobjective evolutionary algorithm performs. Because of space limitations, we omitted the presentation of the ranking procedure. In order to address this gap, the reader can consult the paper [31].

3. RESULTS

In this section we present the results obtained from applying the generalized framework illustrated in Figure 1 to identify OV disease progression genes. We show how the proposed outranking approach was used to prioritize disease candidate genes obtained from multiple diverse networks. Enrichment analysis including GO, Pathway and literature analysis were then performed to demonstrate the effectiveness of the proposed approach. Furthermore, we performed ovarian cancer stage classification using the prioritized gene list obtained from our proposed integrative approach and compare with ToppNet [60].

3.1. RNA-Seq Differential Expression Analysis and Network Construction

We firstly performed differential expression gene (DEG) analysis using the OV RNA-Seq data to identify DEG genes in the data set. DEG analysis was conducted using edgeR [32]. The OV RNA-Seq samples described in the Methods section were divided into two phenotypic groups using the associated clinical data and labeled as either early stage (stage I and II) or late stage (stage III or IV). Raw counts were extracted for these samples and significantly DEG were identified using the paired data with the edgeR package in R using Fisher's exact test, False Discovery Rate (FDR) adjusted p-value < 0.01 and Fold Change > 2 . A total of 283 DEGs were identified among 20,531 genes.

These DEGs were selected for subsequent network analyses as their expression profiles between early and late state OV cancer indicated possible involvement in disease progression. Each of the four networks described in the Networks section were subsequently filtered to include only the identified DE genes a summary of the networks and their properties are described in Table 1.

3.2. Topological Analysis of DEG Networks

Topological analysis of the four different networks based on DEGs was performed. These are (i) protein–protein interaction network (N_{PPI}), (ii) transcription factor network (N_{TF}), (iii) gene coexpression network (N_{COE}) and (iv) tissue-specific functional network (N_{Tissue}). These networks are selected as they capture different interaction types between genes. Furthermore, the biological relationships modeled are not independent. For example, regulatory pathways are controlled in the cell at transcriptional and posttranscriptional (microRNA) levels, alterations at these levels involve

Table 1 | Global properties of the DE networks.

	N_{PPI}	N_{COE}	N_{TF}	N_{Tissue}
Nodes	3286	242	229	315
Edges	4860	871	307	325
Clustering coefficient	0.114	0.223	0.196	0.022
Centralization	0.107	0.162	0.621	0.346
Heterogeneity	4.890	1.279	4.149	3.985

DE, Differential Expression; NPPI, protein–protein interaction network; NCOE, gene coexpression network; NTF, transcription factor network; NTissue, tissue-specific functional network.

modification in gene expression levels [27]. All networks are based on DEGs of OV RNA-Seq data. A summary of the global topological features of the networks are presented in Table 1.

Each network was represented as a graph $G = (V, E)$ where V represents the set of nodes (genes) and E the edges (link types) where $E = \{(u, v) | u, v \in V\}$. The connectivity (degree) is a measure of the number of edges that connects a node. Genes with a high degree of connectivity within a network have large numbers of interacting partners. For a node $v \subseteq V$, the set of edges on v is denoted as N_v where $N_v = \{(u | u \subseteq V, (u, v) \subseteq E)\}$. The cardinality of N_v is the connectivity of v in G . Node betweenness is obtained through a count of the shortest paths that pass through a node. Betweenness centrality of a node n , $C_b(n)$, is calculated as $C_b(n) = \sum_{s \neq n \neq t} (\theta_{st}(n) / \theta_{st})$ where s and t are nodes in the network different to n , θ_{st} represents the number of shortest paths from s to t and $\theta_{st}(n)$ is the number of shortest paths from s to t where n occurs.

A list of nodes from all four networks were extracted based the measures PageRank and Bonacich Power Centrality. Similar to the approach applied by Taylor *et al.* [22], for each of the four networks the top 150 were selected for ranking. This resulted in eight lists containing 1071 distinct genes to be prioritized using the proposed outranking algorithm detailed in Section 2.5.

3.3. Application of Outranking Approach

Using the eight lists of ranked genes as input, our outranking algorithm was applied to prioritize disease genes across the different networks. The outcome from the applied algorithm is the list of ranked prioritized genes described in Table 2. A total of 23 prioritized genes were identified from the 1071 distinct genes across the eight lists.

3.4. Functional Enrichment Analysis

GO, pathway, and disease analysis was applied to investigate the biological implications of the integrated prioritized disease gene-candidate list. Using the GO Biological Process, KEGG and OMIN using Enrichr [61] a statistical over representation test was performed using the prioritized genes ($p < 0.05$ fisher exact test and adjusted p-value) presented in Tables 3 and 4 below. We applied this approach to measure if the prioritized genes are more enriched in GO terms or involved in pathways than what would be expected by chance.

Table 2 | The list of the 23 prioritized genes identified across the eight lists using the proposed Outranking approach.

Gene ID	Gene Symbol	Description
84107	ZIC4	Zic family member 4
22839	DLGAP4	DLG associated protein 4
1410	CRYAB	crystallin alpha B
2938	GSTA1	glutathione S-transferase alpha
12064	ERBB2	erb-b2 receptor tyrosine kinase 2
1264	CNN1	calponin 1
23704	KCNE4	potassium voltage-gated channel subfamily E
2167	FABP4	fatty acid binding protein 4
5709	PSMD3	proteasome 26S subunit, non-ATPase 3
3552	IL1A	interleukin 1 alpha
727851	RGPD8	RANBP2-like and GRIP domain containing 8
4438	MSH4	mutS homolog 4
654	BMP6	bone morphogenetic protein 6
27295	PDLIM3	PDZ and LIM domain 3
7431	VIM	vimentin
6927	HNF1A	HNF1 homeobox A
4113	MAGEB2	MAGE family member B2
6876	TAGLN	transgelin
4958	OMD	osteonmodulin
4060	LUM	lumican
11128	POLR3A	RNA polymerase III subunit A
1440	CSF3	colony stimulating factor 3
23314	SATB2	SATB homeobox 2

Table 3 | Significant KEGG pathways.

Pathway	P Value	Adjusted P-value
Epstein-Barr virus infection	0.002	0.085
Hematopoietic cell lineage	0.005	0.125
Proteoglycans in cancer	0.023	0.155
Cytokine–cytokine receptor interaction	0.037	0.155
Maturity onset diabetes of the young	0.029	0.155
RNA polymerase	0.036	0.155
MicroRNAs in cancer	0.045	0.155
Prion diseases	0.040	0.155
Bladder cancer	0.046	0.155
Graft-versus-host disease	0.046	0.155
Type I diabetes mellitus	0.048	0.155
Proteasome	0.049	0.155

KEGG, Kyoto Encyclopedia of Genes and Genomes.

Table 3 highlights significant KEGG pathways such as the Epstein-Barr virus infection. Studies including [62] and [63] provide some support that Epstein-Barr virus may play a role in the etiology of ovarian cancer. A recent study in [64] indicated that an association between the expression of Hematopoietic lineage cell-specific protein 1 and an increased risk of poor overall survival in patients with ovarian carcinoma.

From GO enrichment analysis highlighted in Table 4, it was observed that processes including regulation of epithelial cell proliferation, keratan sulfate processes and regulation of ERBB signaling pathway were enriched. ErbB2 a major target of a number of FDA-approved compounds for solid tumors [65]. GO enrichment analysis for nontop prioritized genes using our integrative approach mainly lay within GO metabolic and cellular processes. Furthermore, OMIN analysis showed the enrichment of the ranked genes in ovarian cancer.

These results highlight the potential of applying our generalized framework in integrating prioritized lists disease genes for the prediction of ovarian cancer biomarkers.

3.5. Comparison to Gene Prioritization Network Tools Using Predictive Analysis

We compare our integrative approach for prioritization to the gene prioritization tool ToppNet in the ToppGene suite [60] a state-of-the-art gene prioritization approach. ToppNet uses an extension of the PageRank and HITS algorithms, and the K-Step Markov method [14]. In order to perform this comparison we undertake predictive analysis to determine if coexpression values of prioritized genes can predict sample stage (i.e., early or late stage).

3.5.1. ToppGene training and test sets

ToppGene [60] ranks candidate genes based on their similarity to known OV genes. We constructed a training set which is a list of genes known to be involved in OV from the Cancer Genetics Web [66]. This resulted in a list of 307 known OV genes. The test sets were constructed using the same approach we applied as input to the outranking algorithm. A total of eight lists were derived from the top 150 genes from each of the four networks obtained from the application of PageRank and Bonacich Power Centrality. Using the training and test sets, gene prioritization was performed using ToppGene. For each prioritization task, the top 23 prioritized genes were selected for comparison to the proposed outranking approach.

3.5.2. Feature vector construction

Using the prioritized gene-candidate lists obtained from the novel ranking approach, ToppGene's feature vectors were constructed for the classification task of predicting OV stage (early or late). The vectors were constructed using the ranked genes along with their gene expression values from the TCGA Transcriptome and clinical ovarian serous carcinoma data set (referred to as OV). The vectors consisted of 23 features each with 260 instances containing the coexpression values of the prioritized genes across each sample in the dataset. The OV data set has a total of 262 individual samples consisting of 18 early stage and 242 late stage samples. These samples were used as labels when measuring the classification performance.

3.5.3. Predictive analysis

We used the naive Bayesian classifier in the WEKA toolbox [67] for the predictive task of sample stage classification. To evaluate the

Table 4 | Significant GO biological processes.

Term	P Value	Adjusted P-value
Keratan sulfate catabolic process	0.0001	0.0318
Glandular epithelial cell development	0.0080	0.0861
Keratan sulfate metabolic process	0.0007	0.0422
Type B pancreatic cell development	0.0080	0.0861
Regulation of aldosterone biosynthetic	0.0080	0.0861
Sulfur compound catabolic process	0.0008	0.0422
Regulation of cellular amide metabolic	0.0023	0.0678
Regulation of epithelial cell proliferation	0.0032	0.0857
Keratan sulfate biosynthetic	0.0005	0.0422
Regulation of interleukin-2 biosynthetic	0.0103	0.0861

GO, Gene Ontology.

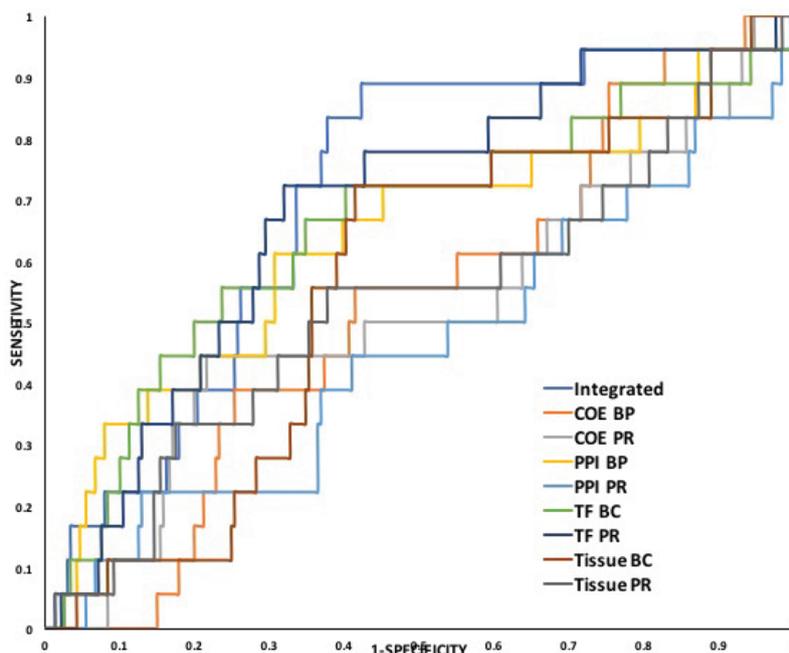


Figure 2 | Plotting of Receiver Operating Characteristic (ROC) curves comparing our integrative ranking approach with ToppGene ranking across the individual networks and topology measures to predict ovarian (OV) cancer stage.

performance, 10-fold cross-validation was carried out. To compare the performance of the prioritized gene-candidate lists on predicting sample stage, we selected the top ranked 23 genes from ToppGene prioritized using the test sets developed from the network topology analysis described above. Comparison of the outranking approach to ToppGene was performed to determine if adding additional information from “omic” networks improves classification obtained using transcription alone.

To evaluate the performance of the outranking approach we used Receiver Operating Characteristic (ROC) curves which compares our outranking approach with ToppGene detailed in Figure 2.

The outranking prioritization approach, which integrates prioritized gene lists from the four networks based on both PageRank and Bonacich Power Centrality topological measures, obtained the

highest area under ROC curve (AUC) value presented in Figure 2 achieving an AUC of 0.704. To evaluate the statistical significance of these results we performed the paired T-Test (corrected) with a significance cut-off 0.05 in Weka comparing the outranking prioritization approach to the individual network prioritization approaches in Table 4. Statistical significance was observed between the outranking approach and the individual prioritization approaches. This analysis suggests that the integration of diverse data in the prioritization of disease genes are indeed useful for sample stage prediction. Interestingly, it is observed that lists from the TF individual network achieved good predictive results using ToppGene obtaining an AUC of 0.68 when the N_{TF} to obtain a list ranked using power centrality. The N_{TF} has been constructed from weighted data obtained from ENCODE [68]. However, the N_{TF} is limited in terms of interactome coverage.

The results highlight how our integrative outranking approach has identified 23 driver genes for classification compared to the 283 DEGs that were represented as features in the RNA-Seq Baseline data set. Furthermore, we can see that the reduced gene list has an improvement in AUC compared to ToppGene, which considers the networks individually. This is important in the context for the development of prognostic assays where the measurement of 23 genes compared to 283 genes is advantageous in terms of the expense and resources required to test and validate additional genes.

3.6. Literature Analysis

A list of genes involved in ovarian cancer was obtained from literature using the Cancer Genetics Web [66], which lists diseases and gene associations from published studies. A total of 306 OV genes were identified. From the outranking prioritized list 4 genes were identified as OV literature genes including: GSTA1, ERBB2, IL1A, MAGEB2 which have been implicated in the carcinogenesis and prognosis of ovarian cancer [69–73]. To determine the statistical significance of this result, we randomly selected 9 genes from the set of genes in the OV dataset 1000 times and compared these to the list of 306 literature obtained genes. Using equation below we calculated the p value as:

$$\frac{\sum \text{outranking} \geq \text{Rand}}{1000}$$

where *outranking* represents the number of OV genes identified by the outranking prioritization process and *Rand* the randomly calculated genes sets OV genes identified. This resulted in $p = 0$ which indicates the statistical significance of the list of genes obtained using the outranking found in the 306 gene set identified in the literature search compared to the random selection of genes from the OV dataset.

3.7. Biological Overview

The outranking list of genes was analyzed in terms of biological relevance. The downstream effects of disease processes such as mutation, methylation and copy number aberration are captured in transcriptional profiling. The activity of PLAGL1 (inhibiting proliferation and tumorigenicity) is altered through loss of heterozygosity or methylation and it has been found to be down-regulated in almost two-fifths of epithelial ovarian cancer (EOC) patients [74,75]. KRAS mutations, with downstream transcriptional loss, are more associated with low-grade or mucinous EOC, rather than high-grade serous OV, which is the subject of this study. Previous research has suggested that most of the genetic alterations associated with metastatic potential are already present in early-stage EOCs, though our study suggests an accumulation of further genetic changes [76,77].

4. CONCLUSION

Recent years have witnessed the systematic investigation of cancer through the application of high-throughput experimental technologies and the development of centralized databases. These

experimental “omic” platforms target the comprehensive analysis of genes, mRNA, proteins and metabolites. These developments have been essential in (1) advancing the knowledge of biological systems; (2) understanding, predicting, diagnosing and monitoring diseases; (3) discovering biomarkers and (4) identifying drug targets. However, translation of relevant prognostic markers identified by such experiments into clinical tools for personalized patient treatment has been slow and unreliable [78].

Biological networks have proved to be effective in capturing the complexity of human disease and discerning how complexity controls disease manifestations, prognosis and therapy. In this study we have demonstrated how the integration of multiple “omic” biological networks can prioritize biologically relevant ovarian cancer progression genes. Prediction of a disease’s clinical stage is crucial for the development of appropriate therapeutics that may delay the progression of the disease. Yet the genetic basis of disease progression for many diseases are still unclear and in some cases unknown. In this study we applied our general framework that integrates “omic” data using a multinet network approach framework and ranking to the Use Case HGSC progression. HGSC is a malignant form of ovarian cancer and accounts for up to 70% of all ovarian cancer cases [79]. Patients with HGSC are often diagnosed at the late stage of the disease. This is problematic as later stages generally result in a poor prognosis [80]. Chemotherapeutic agents have improved the five-year survival rate, however, the overall mortality of ovarian cancer has not changed in over 50 years [79]. This is mainly due to the lack of success in diagnosing ovarian cancer at an early stage and effective treatments for late stage of the disease [81].

In this research we have proposed an outranking data integration method which combined information from multinet networks using MCDA with aggregation mechanisms to prioritize lists of disease-candidate genes. The outranking approach is based on ELECTRE III and uses decision rules to identify positive and negative ranking criteria along with a multiobjective evolutionary algorithm to exploit the outranking relation and derive a prioritized list. The advantage of this approach is the ability to combine information from multiple heterogeneous data sources to produce ranked lists of potential disease-candidate genes. The approach however is generic in nature and can be applied to other problem domains such as the prioritization of alternative transcripts and protein isoforms, metabolites and drugs.

Using our proposed outranking approach, known ovarian cancer genes GSTA1, ERBB2, IL1A, MAGEB2 were uncovered. KEGG pathways including ErbB signalling and pathways in cancer along with GO terms metalloendopeptidase and ErbB2 were significantly enriched. Furthermore, relatively high predictive performance (AUC: 0.704) was observed when classifying ovarian cancer early and late stage RNA-Seq expression profiles from individuals using 10-fold cross-validation.

In this work we have applied a case study to demonstrate the proof of concept where we know (some of) the underpinning biology. In future work we will apply our approach for discovery in diseases such as chemotherapy-resistant and sensitive HGSC. Furthermore, we will extend our current approach on the integration of multiple “omic” network data using multilayer network theory. We aim to model importance and reliability of layers in a biological multilayer network for predictive task disease progression and apply to other diseases such as lung carcinoma.

CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest.

AUTHORS' CONTRIBUTIONS

Jesús Jaime Solano Noriega: Conceptualization, Methodology, Validation, Investigation, Data curation, Writing original draft, Writing review and editing; Juan Carlos Leyva López: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing original draft, Writing review; Fiona Browne: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing original draft, Writing review; Jun Liu: Validation, Formal analysis, Investigation, Writing review.

Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

- [1] S.D. Ghiassian, J. Menche, A. Barabási, A disease module detection (diamond) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome, *PLOS Comput. Biol.* 11 (2015), e1004120.
- [2] J. Menche, A. Sharma, M. Kitsak, M. Vidal, J. Loscalzo, A.-L. Barabási, Uncovering disease-disease relationships through the incomplete interactome, *Science*. 347 (2015), 1257601.
- [3] D. Botstein, N. Risch, Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease, *Nat. Genet.* 33 (2003), 228–237.
- [4] M.A. Nalls, N. Pankratz, C.M. Lill, C.B. Do, D.G. Hernandez, M. Saad, *et al.*, Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease, *Nat. Genet.* 46 (2014), 989–993.
- [5] A. Flaquer, C. Baumbach, E. Piñero, F. García Algas, M.A. De La Fuente Sánchez, J. Rosell, *et al.*, Genome-wide linkage analysis of congenital heart defects using mod score analysis identifies two novel loci, *BMC Genet.* 14 (2013), 44.
- [6] X. Wu, R. Jiang, M.Q. Zhang, S. Li, Network-based global inference of human disease genes, *Mol. Syst. Biol.* 4 (2008), 189.
- [7] X. Zeng, Y. Liao, Y. Liu, Q. Zou, Prediction and validation of disease genes using hetesim scores, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 14 (2017), 687–695.
- [8] H.B. Fraser, J.B. Plotkin, Using protein complexes to predict phenotypic effects of gene mutation, *Genome Biol.* 8 (2007), R252.
- [9] J. Chen, B.J. Aronow, A.G. Jegga, Disease candidate gene identification and prioritization using protein interaction networks, *BMC Bioinform.* 10 (2009), 73.
- [10] M.Q. Xie, Y.J. Xu, Y.G. Zhang, T.H. Hwang, R. Kuang, Network-based phenome-genome association prediction by bi-random walk, *PloS One*. 10 (2015), e0125138.
- [11] J.E. Shim, S. Hwang, I. Lee, Pathway-dependent effectiveness of network algorithms for gene prioritization, *PLoS One*. 10 (2015), e0130589.
- [12] D. Guala, E.L.L. Sonnhammer, A large-scale benchmark of gene prioritization methods, *Sci. Rep.* 7 (2017), 46598.
- [13] L.-C. Tranchevent, R. Barriot, S. Yu, S. Van Vooren, P. Van Loo, B. Coessens, B. De Moor, S. Aerts, Y. Moreau, Endeavour update: a web resource for gene prioritization in multiple species, *Nucleic Acids Res.* 36 (2008), W377–W384.
- [14] J. Chen, E.E. Bardes, B.J. Aronow, A.G. Jegga, Toppgene suite for gene list enrichment analysis and candidate gene prioritization, *Nucleic Acids Res.* 37 (2009), W305–W311.
- [15] X. Wu, Q. Liu, R. Jiang, Align human interactome with phenome to identify causative genes and networks underlying disease families, *Bioinform.* 25 (2009), 98–104.
- [16] W. Yu, A. Wulf, T. Liu, M.J. Khoury, M. Gwinn, Gene prospector: an evidence gateway for evaluating potential susceptibility genes and interacting risk factors for human diseases, *BMC Bioinform.* 9 (2008), 528.
- [17] Y. Moreau, L.-C. Tranchevent, Computational tools for prioritizing candidate genes: boosting disease gene discovery, *Nat. Rev. Genet.* 13 (2012), 523–536.
- [18] E. Jadamba, S. Beom Cho, M. Shin, Netranker: a network-based gene ranking tool using protein-protein interaction and gene expression data, *BioChip J.* 9 (2015), 313–321.
- [19] W. Lan, J. Wang, M. Li, W. Peng, F. Wu, Computational approaches for prioritizing candidate disease genes based on PPI networks, *Tsinghua Sci. Technol.* 20 (2015), 500–512.
- [20] J. Wang, G. Chen, M. Li, Y. Pan, Integration of breast cancer gene signatures based on graph centrality, *BMC Syst. Biol.* 5 (2011), 1.
- [21] M. Li, J. Wang, H. Wang, Y. Pan, Essential proteins discovery from weighted protein interaction networks, in: M. Borodovsky, J.P. Gogarten, T.M. Przytycka, S. Rajasekaran (Eds.), *Bioinformatics Research and Applications*, Springer, Berlin, Heidelberg, Germany, 2010, pp. 89–100.
- [22] I.W. Taylor, R. Linding, D. Warde-Farley, Y. Liu, C. Pesquita, D. Faria, *et al.*, Dynamic modularity in protein interaction networks predicts breast cancer outcome, *Nat. Biotechnol.* 27 (2009), 199–204.
- [23] J. Wang, S. Zhang, Y. Wang, L. Chen, X.-S. Zhang, Disease-aging network reveals significant roles of aging genes in connecting genetic diseases, *PLoS Comput. Biol.* 5 (2009), e1000521.
- [24] G. Liu, X. Bao, Y. Jiang, M. Liao, Q. Jiang, R. Feng, *et al.*, Identifying the association between Alzheimer's disease and Parkinson's disease using genome-wide association studies and protein-protein interaction network, *Mol. Neurobiol.* 52 (2015), 1629–1636.
- [25] Y. Chen, W. Wang, Y. Zhou, R. Shields, S.K. Chanda, R.C. Elston, J. Li, In silico gene prioritization by integrating multiple data sources, *PloS One*. 6 (2011), e21137.
- [26] M. Li, J. Zhang, Q. Liu, J. Wang, F.-X. Wu, Prediction of disease-related genes based on weighted tissue-specific networks by using DNA methylation, *BMC Med. Genomics*. 7 (2014), S4.
- [27] L. Cantini, E. Medico, S. Fortunato, M. Caselle, Detection of gene communities in multi-networks reveals cancer drivers, *Sci. Rep.* 5 (2015), 1–10.
- [28] L. Bennett, A. Kittas, G. Muirhead, L.G. Papageorgiou, S. Tsoka, Detection of composite communities in multiplex biological networks, *Sci. Rep.* 5 (2015), 10345.
- [29] M.A. Bookman, Optimal primary therapy of ovarian cancer, *Ann. Oncol.* 27 (2016), i58–i62.

- [30] B. Roy, The outranking approach and the foundations of ELEC-TRE methods, *Theory Decis.* 31 (1991), 49–73.
- [31] J.C. Leyva Lopez, J.J. Solano Noriega, J.R. Figueira, J. Liu, D.A. Gastélum Chavira, Non-dominated sorting genetic-based algorithm for exploiting a large-sized fuzzy outranking relation, *Eur. J. Oper. Res.* 293 (2021), 615–631.
- [32] M.D. Robinson, D.J. McCarthy, G.K. Smyth, Edger: a bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics.* 26 (2010), 139–140.
- [33] T. Li, R. Wernersson, R.B. Hansen, H. Horn, J. Mercer, G. Slodkowitz, et al., A scored human protein–protein interaction network to catalyze genomic interpretation, *Nat. Methods.* 14 (2017), 61.
- [34] A. Chatr-Aryamontri, B.-J. Breitkreutz, R. Oughtred, L. Boucher, S. Heinicke, D. Chen, et al., The biogrid interaction database: 2015 update, *Nucleic Acids Res.* 43 (2015), D470–D478.
- [35] S. Kerrien, B. Aranda, L. Breuza, A. Bridge, F. Broackes-Carter, C. Chen, et al., The intact molecular interaction database in 2012, *Nucleic Acids Res.* 40 (2011), D841–D846.
- [36] J. Zhang, K. Lu, Y. Xiang, M. Islam, S. Kotian, Z. Kais, et al., Weighted frequent gene co-expression network mining to identify genes involved in genome stability, *PLoS Comput. Biol.* 8 (2012), e1002656.
- [37] C.S. Greene, A. Krishnan, A.K. Wong, E. Ricciotti, R.A. Zelaya, D.S. Himmelstein, et al., Understanding multi-cellular function and disease with human tissue-specific networks, *Nat. Genet.* 47 (2015), 569–576.
- [38] ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome, *Nature.* 489 (2012), 57–74.
- [39] J.P. González, A.J. Berenstein, A. González Pérez, A. Chernomoretz, L.I. Furlong, Uncovering disease mechanisms through network biology in the era of next generation sequencing, *Sci. Rep.* 6 (2016), 24570.
- [40] A. Sarajlić, V. Janjić, N. Stojković, D. Radak, N. Pržulj, Network topology reveals key cardiovascular disease genes, *PLoS One.* 8 (2013), e71537.
- [41] H. Yu, P.M. Kim, E. Sprecher, V. Trifonov, M. Gerstein, The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics, *PLoS Comput. Biol.* 3 (2007), e59.
- [42] K. Raosaheb Patil, J. Nielsen, Uncovering transcriptional regulation of metabolism by using metabolic network topology, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005), 2685–2689.
- [43] R. Guimera, L.A. Nunes Amaral, Functional cartography of complex metabolic networks, *Nature.* 433 (2005), 895–900.
- [44] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2017. <http://www.R-project.org/>.
- [45] C. Ma, Y. Chen, D. Wilkins, X. Chen, J. Zhang, An unsupervised learning approach to find ovarian cancer genes through integration of biological data, *BMC Genom.* 16 (2015), S3.
- [46] J.-C. de Borda, Memoire sur les elections au scrutin, *Histoire de l'Academie Royale des Sciences*, Paris, France, 1781.
- [47] C.C. Vogt, G.W. Cottrell, Fusion via a linear combination of scores, *Inf. Retr.* 1 (1999), 151–173.
- [48] J.I. Marden, Analyzing and Modeling Rank Data, *Monographs on Statistics and Applied Probability*, Chapman and Hall/CRC, 1996.
- [49] W.D. Cook, M. Kress, Ordinal ranking with intensity of preference, *Manag. Sci.* 31 (1985), 26–32.
- [50] D. Lillis, F. Toolan, R. Collier, J. Dunnion, Prob fuse: a probabilistic approach to data fusion, in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06)*, ACM, New York, NY, USA, 2006, pp. 139–146.
- [51] M. de Condorcet, *Essai sur l'Application de L'Analyse ala Probabilité des Décisions Rendues ala Pluraliste des Voix*, 1785.
- [52] C. Dwork, R. Kumar, M. Naor, D. Sivakumar, Rank aggregation methods for the web, in *Proceedings of the 10th International Conference on World Wide Web (WWW '01)*, ACM, New York, NY, USA, 2001, pp. 613–622.
- [53] B. Roy, Decision-aid and decision-making, in: C.A. Bana e Costa (Eds.), *Readings in Multiple Criteria Decision Aid SE - 2*, Springer, Berlin, Heidelberg, Germany, 1990, pp. 17–35.
- [54] B. Roy, *Multicriteria Methodology for Decision Aiding*, Kluwer Academic Publishers, The Netherlands, 1996.
- [55] K. Govindan, M.B. Jepsen, ELEC-TRE: a comprehensive literature review on methodologies and applications, *Eur. J. Oper. Res.* 250 (2016), 1–29.
- [56] J.R. Figueira, S. Greco, B. Roy, R. Slowinski, An overview of ELEC-TRE methods and their recent extensions, *J. Multià Criteria.* 20 (2013), 61–85.
- [57] B. Dutta, S.D. Dao, L. Martínez, M. Goh, An evolutionary strategic weight manipulation approach for multi-attribute decision making: TOPSIS method, *Int. J. Approx. Reason.* 129 (2021), 64–83.
- [58] J.C. Leyva López, J.J. Solano Noriega, D.A. Gastélum Chavira, M.D.S. Castaneda, A multiobjective evolutionary approach to a medium-sized multicriteria ranking problem, in: J.C. Leyva Lopez, R. Espin Andrade, R. Bello Pérez, P.A. Alvarez Carrillo (Eds.), *Studies on Knowledge Discovery, Knowledge Management, and Decision Making*, Atlantis Press, Mazatlan, Mexico, 2013, pp. 188–197.
- [59] J.C. Leyva López, J.J. Solano Noriega, J.L. Garcia Alcaraz, D.A. Gastelum Chavira, Exploitation of a medium-sized fuzzy outranking relation based on multi-objective evolutionary algorithms to derive a ranking, *Int. J. Comput. Intell. Syst.* 9 (2016), 745–764.
- [60] E. Guney, B. Oliva, Exploiting protein-protein interaction networks for genome-wide disease-gene prioritization, *PloS One.* 7 (2012), e43557.
- [61] M.V. Kuleshov, M.R. Jones, A.D. Rouillard, N.F. Fernandez, Q. Duan, Z. Wang, et al., Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, *Nucleic Acids Res.* 44 (2016), W90–W97.
- [62] A.J. Littman, M.A. Rossing, M.M. Madeleine, M.-T.C. Tang, Y. Yasui, Association between late age at infectious mononucleosis, epstein-barr virus antibodies, and ovarian cancer risk, *Scand. J. Infect. Dis.* 35 (2003), 728–735.
- [63] K. Wang, Y. Ge, C. Ni, B. Cui, J. Du, B. Zhang, et al., Epsteinbarr virus-induced up-regulation of tcab1 is involved in the dna damage response in nasopharyngeal carcinoma, *Sci. Rep.* 7 (2017), 3218.
- [64] W. Liu, H. Kajiyama, K. Shibata, Y. Koya, T. Senga, F. Kikkawa, Hematopoietic lineage cell-specific protein 1 immunoreactivity indicates an increased risk of poor overall survival in patients with ovarian carcinoma, *Oncol. Lett.* 15 (2018), 9406–9412.

- [65] M.I. Sereni, E. Baldelli, G. Gambarà, J. Deng, L. Zanotti, E. Bandiera, *et al.*, Functional characterization of epithelial ovarian cancer histotypes by drug target based protein signaling activation mapping: implications for personalized cancer therapy, *Proteomics*. 15 (2015), 365–373.
- [66] S.J. Cotterill, *Cancer Genetics*. 2021, <http://www.cancer-genetics.org/index.htm>
- [67] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The weka data mining software: an update, *ACM SIGKDD Explor. Newsl.* 11 (2009), 10–18.
- [68] K. Venkatesan, J.-F. Rual, A. Vazquez, U. Stelzl, I. Lemmens, T. Hirozane-Kishikawa, *et al.*, An empirical framework for binary interactome mapping, *Nat. Methods*. 6 (2009), 83–90.
- [69] S. Davies, A. Holmes, L. Lomo, M. Steinkamp, H. Kang, C.Y. Muller, B.S. Wilson, High incidence of *erbb3*, *erbb4* and *met* expression in ovarian cancer, *Int. J. Gynecol. Pathol.* 33 (2014), 402.
- [70] W.B. Growdon, J. Groeneweg, V. Byron, C. DiGloria, D.R. Borger, R. Tambouret, *et al.*, *Her2* over-expressing high grade endometrial cancer expresses high levels of *p95her2* variant, *Gynecol. Oncol.* 137 (2015), 160–166.
- [71] Y. Li, R. Zhang, D. Pan, B. Huang, M. Weng, X. Nie, *Kras* mutation in adenocarcinoma of the gastrointestinal type arising from a mature cystic teratoma of the ovary, *J. Ovarian Res.* 7 (2014), 1.
- [72] Z. Yang, Y. Liu, J. Liao, C. Gong, C. Sun, X. Zhou, *et al.*, Quercetin induces endoplasmic reticulum stress to enhance *cddp* cytotoxicity in ovarian cancer: involvement of *stat3* signaling, *FEBS J.* 282 (2015), 1111–1125.
- [73] T. Kamikihara, T. Arima, K. Kato, T. Matsuda, H. Kato, T. Douchi, *et al.*, Epigenetic silencing of the imprinted gene *zac* by dna methylation is an early event in the progression of human ovarian cancer, *Int. J. Cancer*. 115 (2005), 690–700.
- [74] R.C. Bast, B. Hennessy, G.B. Mills, The biology of ovarian cancer: new opportunities for translation, *Nat. Rev. Cancer*. 9 (2009), 415–428.
- [75] W. Feng, R.T. Marquez, Z. Lu, J. Liu, K.H. Lu, J.-P.J. Issa, *et al.*, Imprinted tumor suppressor genes *arhi* and *peg3* are the most frequently down-regulated in human ovarian cancers by loss of heterozygosity and promoter methylation, *Cancer*. 112 (2008), 1489–1502.
- [76] V. Shridhar, J. Lee, A. Pandita, S. Iturria, R. Avula, J. Staub, *et al.*, Genetic analysis of early-*versus* late-stage ovarian tumors, *Cancer Res.* 61 (2001), 5895–5904.
- [77] P. Cheng, C. Schmutte, K.F. Cofer, J.C. Felix, M.C. Yu, L. Dubeau, Alterations in DNA methylation are early, but not initial, events in ovarian tumorigenesis, *Br. J. Cancer*. 75 (1997), 396.
- [78] A. Dupuy, R.M. Simon, Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting, *J. Natl. Cancer Inst.* 99 (2007), 147–157.
- [79] R.J. Kurman, Origin and molecular pathogenesis of ovarian high-grade serous carcinoma, *Ann. Oncol.* 24 (2013), x16–x21.
- [80] M.E. Sherman, R.I. Drapkin, N.S. Horowitz, C.P. Crum, S. Friedman, J.S. Kwon, *et al.*, Rationale for developing a specimen bank to study the pathogenesis of high-grade serous carcinoma: A review of the evidence, *Cancer Prev. Res.* 9 (2016), 713–720.
- [81] A.M. Karst, P.M. Jones, N. Vena, A.H. Ligon, J.F. Liu, M.S. Hirsch, D. Etemadmoghadam, D.D.L. Bowtell, R. Drapkin, *Cyclin e1* deregulation occurs early in secretory cell transformation to promote formation of fallopian tube-derived high-grade serous ovarian cancers, *Cancer Res.* 74 (2014), 1141–1152.

