JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 28, Number 0, 2021

Pp. 1–13
DOI: 10.1089/cmb.2020.0244

Research Article

Rank-Similarity Measures for Comparing Gene Prioritizations: A Case Study in Autism

CONCETTINA GUERRA,[1] SARANG JOSHI,[1] YINQUAN LU,[1] FRANCESCO PALINI,[2] UMBERTO FERRARO PETRILLO,[2] and JAREK ROSSIGNAC[1]

ABSTRACT

We discuss the challenge of comparing three gene prioritization methods: network propagation, integer linear programming rank aggregation (RA), and statistical RA. These methods are based on different biological categories and estimate disease–gene association. Previously proposed comparison schemes are based on three measures of performance: receiver operating curve, area under the curve, and median rank ratio. Although they may capture important aspects of gene prioritization performance, they may fail to capture important differences in the rankings of individual genes. We suggest that comparison schemes could be improved by also considering recently proposed measures of similarity between gene rankings. We tested this suggestion on comparison schemes for prioritizations of genes associated with autism that were obtained using brain- and tissue-specific data. Our results show the effectiveness of our measures of similarity in clustering brain regions based on their relevance to autism.

Keywords: autism, comparative analysis, disease–gene association, gene prioritization.

1. INTRODUCTION

Gene prioritization refers to the problem of identifying genes that are implicated in a disease. Experimental approaches, such as large scale sequence experiments (Krishnan et al., 2016; Yuen et al., 2017), are the ultimate way to address this problem, but they tend to be costly and often time consuming. Complementary to experimental sequencing approaches, computational approaches exploit existing knowledge about causal genes for a specific disease to predict risk genes from a pool of test genes. Thus, computational tools help narrow down the set of prospect genes for further validation.

Most of the gene prediction methods are based on the premise that similarity in terms of biological features of the test genes to known causal genes, often called seed genes, corresponds to functional similarity and to similar roles in a disease. Existing methods use various types of biological data and annotations gathered from public websites as well as from new experiments: from expression profiles, to interaction data, to gene ontology (GO).

[1]Georgia Institute of Technology College of Computing, School of Interactive Computing, Atlanta, Georgia, USA.
[2]Dipartimento di Scienze Statistiche, Università di Roma—La Sapienza, Rome, Italy.

A recent survey reviews the goals and challenges of the computational aspects of gene prioritization (Moreau and Tranchevent, 2012). Other surveys examine a small number of the existing computational tools, discussing their availability at public web sites, ease of use, requirements in terms of input data, and their various advantages/disadvantages (Nivit et al., 2014; Zolotareva and Kleine, 2019). The variety of data sources and the lack of an established gold standard render the task of evaluating and comparing the different computational tools rather complex.

In this study, we provide a framework for the comparison of different approaches in predicting genes associated with a disease. We consider two classes of approaches widely used in biological applications and specifically in gene prioritization studies: rank aggregation (RA) and network propagation (NP).

Commonly used measures of performance of prediction methods, receiver operating curve (ROC), area under the curve (AUC), and median rank ratio (MRR; Börnigen et al., 2012), are effective in assessing the ability of a prediction tool in retrieving known causal autism genes from a pool of candidates. These measures generally provide the basis for a comparison of computational tools. However, they do not take into account the identity of genes, that is, which genes have a given rank in the outputs of two methods. To account for that in evaluating the similarity of two gene rankings, we propose to use the Jaccard index and the L1 distance, D_{rank}, which, as far as we know, have never been used in this context.

We show that, all together, the mentioned measures, AUC, MRR, Jaccard index, and L1, contribute to quantitatively analyze how the result of gene prioritization is influenced by the choice of the data categories and the choice and implementation of a computational method.

We use as a case study the challenge of discovering the genes implicated in autism. Autism spectrum disorders (ASDs) refer to a wide range of disorders, from mild to severe, mostly affecting interpersonal relations. This disorder has a significant genetic component. A large number of genes may be implicated in ASDs, but currently strong evidence exists only for ~ 100 genes, although many more may have a role (Banerjee-Basu and Packer, 2010).

We report on three sets of experiments: (1) we ran RA and NP in the same experimental setting with the same biological data categories; (2) we ran the same method, say RA, applied to different data categories; (3) we ran RA and NP on the whole brain, and on tissue-specific data.

As inputs, we use brain-specific data: gene expression profiles from *Brainspan* (Miller et al., 2014) and multiple input categories from ToppGene (Chen et al., 2009).

From these experiments, we conclude that (1) network-based propagation generally outperforms RA; (2) the impact of the different data categories is statistically significant; in fact, the similarity of the lists of prioritized genes produced by an algorithm applied to different data categories is not higher than that expected by chance; (3) all measures indicate a better performance on some tissues than on the whole brain, highlighting the importance of those tissues for ASDs; and (4) the L1 distance of prioritized lists of genes between tissues provides the basis for a valid clustering of the brain tissues based on their involvement in autism.

The article is organized as follows. In Section 2, we review the computational methods of RA and NP. In Sections 3 and 4, we present the data sets used in our experiments and the performance measures. The results on the whole brain are presented in Section 5, whereas those on tissue-specific data are presented in Section 7.

2. COMPUTATIONAL APPROACHES

The computational approaches considered here, RA and NP, are not tailored to specific data categories and do not heavily rely on big data. The software can be downloaded or easily reproduced, or the results on new candidate genes can be obtained from a public website. A description of both methods follows.

2.1. Rank aggregation

RA has been often proposed as a way to solve the gene prediction problem (Adle et al., 2009; Chen et al., 2009; Kumar and Vassilvitskii, 2010; Kolde et al., 2012; Minji et al., 2015; Li and Milenkovic, 2017). It takes in input several rankings on the same set of genes and produces a single average ranking that best summarizes them. A ranking may be viewed as a permutation on the integers $\{1, 2, \cdots, n\}$, n being the number of genes, with the integers corresponding to some lexicographical sorting of the genes. Thus this approach deals with permutations, that is, arrangements of the same set of genes, and takes advantage of the vast prior art on

distance of permutations and their ''average'' to tackle the gene prioritization problem. The fundamental questions in RA are (1) how to define the average ranking and (2) how to compute it efficiently.

There are two main families of approaches to RA: combinatorial and statistical. For our analysis, we select a representative for each family. The combinatorial method uses rankings as inputs and is framed as an optimization problem; the statistical method uses ratings, that is, rankings with scores, and the well-established Fisher inverse χ^2 method.

2.1.1. A combinatorial approach. An approach to RA is to define a suitable distance between rankings and seek the ranking that minimizes the sum of its distances to all input rankings. This distance can be chosen among the plethora of distances introduced in the literature on permutations, including the Kendall's distance, the Spearman's distance, and the Hamming distance. Here, we consider the most common one, the Kendall's distance. It counts the number of pairs of genes that appear in opposite order in two rankings.

Given a set of rankings P on n genes, the RA is solved as an optimization problem:

$$s^* = \min_{s \in S_n} \sum_{p \in P} d(s, p), \tag{1}$$

where d is the Kendall distance and S_n is the set of all $n!$ permutations of n symbols. It is known that this optimization problem is NP-complete (Dwork et al., 2001).

An exact solution to problem (1) can be obtained by integer linear programming (ILP) as proposed in Conitzer and Sandholm (2006) and Schalekampf and von Zuylen (2009). We refer to this approach as RA-ILP.

Even though much faster approaches to RA exist (Dwork et al., 2001) and provide good approximate solutions, we opted for the mentioned approach, implemented by means of the Gurobi solver (Optimization, 2020), because we wanted the exact solution for our comparison.

2.1.2. A statistical approach. Some approaches to the integration of heterogeneous information rely not only on rankings but also on a score associated with each gene in a ranking, often a p value (Chen et al., 2009; Minji et al., 2015). Integrating scored rankings can be done using a variety of statistical techniques.

The statistical approach considered in our comparative analysis is ToppGene (Chen et al., 2009), which is based on the Fisher inverse χ^2 method under the hypothesis of independent ratings. It establishes similarity between a set of test genes and the seed genes using as many as 14 categories including GO: molecular function, GO: biological process, GO: cellular component, human phenotype, mouse phenotype, pathway, PubMed, disease. For a given category, a similarity score is computed for every pair of test and seed genes along with its p value. The computed p values induce a ranking of the test genes according to a specific category, resulting in as many as 14 scored rankings, one for each category.

Then, ToppGene integrates such scored rankings using Fisher's inverse χ^2 result stating that $-2\sum_{j=1}^{m} \log p(j) \rightarrow \chi^2(2m)$, where m is the number of annotations and $\chi^2(2m)$ is the χ^2 distribution with $2m$ degrees of freedom.

2.2. Network propagation

Network-based methods have gained popularity in the past decade as a powerful tool in a variety of domains including biology, where they have been used to determine subnetworks corresponding to functional modules (Mitra et al., 2013), to find conserved subnetworks in different species (Ciriello et al., 2012), and in drug discovery (Csermely et al., 2013). Studies on network-based propagation for gene prioritization include RA et al. (2006), Vanunu et al. (2010), Lee et al. (2011), Sinan et al. (2011), Magger et al. (2012), Guala et al. (2014), Shim et al. (2015), Wong et al. (2015), and Jingchao Ni et al.(2016). A survey by Cowan et al. (2017) provides a good introduction and a description of various applications in biology. A recent survey (Guala and Sonnhammer, 2017) reviews NP for gene prioritization and uses GO terms as validation.

Molecular networks most often used in gene prioritization are protein–protein interaction networks, representing physical binding of proteins and gene coexpression networks, where edges are labeled with coexpression values of genes.

The idea underlying all these approaches is that topological proximity of proteins or genes implies a higher likelihood to be involved in the same disease. In a network-based propagation process, the information flows from a node to its neighboring nodes, and from these to their neighbors and so on, with an iterative process that stops either after a fixed number of steps or upon convergence. Among the variants of this basic strategy, random walk with restart (RWR) seems to be the popular choice (Guala and Sonnhammer, 2017). It allows at each step to restart a random walk from the initial seeds with a given probability, thus taking into account both local and global topology of the network.

We selected, as a representative of the network-based class of algorithms, the NP-RWR approach and we used the tool DADA (Sinan et al., 2011) to generate the lists of predicted genes. DADA is based on RWR and employs statistical methods to correct a common drawback of such methods, that is, that they penalize loosely connected nodes in favor of highly connected nodes, or hubs. DADA was run on coexpression networks as inputs.

3. DATA

Two sets of genes are typically considered in the computational prediction of disease genes: (1) seed genes, for which there is a strong evidence of association with autism and (2) test genes, among which to identify the ASD-risk genes. The test set may consist of the entire set of human genes, but most often is restricted to a much smaller set. For the purpose of the evaluation, the set of test genes is considered to be divided into two subsets: candidate and random genes. The candidate set consists of genes that have some ties to the specific disease but need further validation, and the random set contains genes for which no functional association exists in the brain regions.

3.1. Data sources

In our experiments, autism genes are obtained from the Simons Foundation Autism Research Initiative (SFARI) Gene database (Banerjee-Basu and Packer, 2010), Genecards (Stelzer et al., 2016), and previously published articles. SFARI assigns genes to different categories depending on the evidence that exists of their association with autism, category 1 being the one with strongest evidence. The complete list of input genes used in our experiments is given in Tables 4–7 of the Supplementary Material. It includes 10 seed nodes (Willsey et al., 2013). Most of the candidate genes are from SFARI categories 1 and 2. They are included in the sets SF-cat1 (i.e., SFARI genes falling in the phenotypic profile 1) and SF-cat2 (i.e., SFARI genes falling in the phenotypic profile 2). Genes not present in SFARI are ASD-risk genes with sequence-level mutations from experimental studies reported in Yuen et al. (2017). At the time of the publication (Yuen et al., 2017), such genes had not previously been reported in the literature. This set consisting of 18 genes is called Y-set. The set ALL includes all the 124 candidate genes belonging to SF-cat1, SF-cat2 and Y-set. A set of 137 random genes were selected from the list in Krishnan et al. (2016).

From the BrainSpan website (Miller et al., 2014), we downloaded spatiotemporal expression values of all genes originating from 16 brain regions and at 15 different developmental stages (Kang et al., 2011). Coexpression values for a pair of genes were computed as the correlation coefficient of their spatiotemporal expression profiles and were mapped into the interval [0,1] (Zhang and Horvath, 2005).

We opted for an input list of modest length because of the high computational demands of RA-ILP.

4. PERFORMANCE MEASURES

We evaluate the comparison schemes on the entire set of test genes as well as on partial sets of candidates using the following measures of performance:

- *ROC*. It is a common measure of performance, plotting the false positive rate (FPR) versus the true positive rate. The label true positive (TP) is assigned to candidate genes that are ranked above a given rank threshold and false positive (FP) to random genes above the same threshold. The AUC is a synthetic measure, one single value, of performance.
- *MRR*. It gives an indication on how highly the candidate genes are placed in the output. It is defined as the median rank of the candidate genes normalized by the length of the list of all test genes (candidate and random; Börnigen et al., 2012; Guala and Sonnhammer, 2017).

The mentioned two measures, computed separately for each prediction tool, are concerned with the ranks of test genes but do not consider the identity of the genes, in other words which genes have a specific rank. By contrast, the following are pairwise similarity/distance measures of predictive tools that look at the individual genes to establish their closeness in two prioritized list of genes. To the best of our knowledge, they have not been used so far to compare prediction tools.

- D_{rank} and ND_{rank}. They are based on the differences in ranks of all candidate genes in two complete output rankings (which include all test genes). Precisely, given two rankings p and q of the entire set of test genes, their distance D_{rank} is the L1 distance defined as the sum over all candidate genes of the absolute values of the difference in rank of the genes in p and q:

$$D_{rank} = \sum_g |rank_p(g) - rank_q(g)|,$$

where $rank_p(g)$ represents the rank of candidate gene g in p. Note that the ranks of random genes are not considered in the mentioned definition since they are not relevant for the analysis. ND_{rank} is the normalized value of D_{rank}, that is, it is equal to D_{rank} divided by the length of the list of the test genes.

We use a permutation test to establish the significance of the D_{rank} value by estimating a p value. The test is done on a large number of possible rankings of the candidate genes. It is computed as the fraction of times the distance between one ranking, say p, and q', which is obtained by randomly permuting q, is smaller than or equal to the observed distance.

- *Jaccard index.* As a measure of similarity, we compute the *Jaccard index* of the sets of TPs in the two ranked lists of genes. Given the tools A and B and the sets of TP(A) and TP(B), the Jaccard index is given by the size of the intersection of two sets divided by the size of their union:

$$Jaccard_{index} = \frac{|TP(A) \cap TP(B)|}{|TP(A) \cup TP(B)|}.$$

We determine the significance of the Jaccard index by using the bootstrap procedure in Chung et al. (2019) on binary vectors whose size is the number of candidates; a 1 is assigned to elements of the vector corresponding to TPs and 0 to FPs. The p value is computed as the number of times the Jaccard index is below the observed value.

5. RESULTS

In this section, we provide comparisons of the computational approaches already presented: RA-ILP, ToppGene (Chen et al., 2009), and NP-RWR (Jingchao Ni et al., 2016).

5.1. NP-RWR versus RA-ILP on coexpression values data

We run RA-ILP and NP-RWR utilizing the same biological data, that is, coexpression values of the seeds with the test genes. Thus a comparison of their results will assess the relative merits of the two computational approaches under similar experimental settings. Precisely, RA-ILP has in input 10 rankings of the test genes; each ranking is induced by the coexpression values of the test genes with a seed. NP-RWR has in input a complete network with nodes corresponding to all genes, that is, seed, candidate and random genes, and edges connecting all pairs of genes; the edges are labeled with the coexpression of the corresponding genes.

The ROCs and AUC values representing the global performance of the methods are shown in Figure 1. Overall, NP-RWR has a better performance than RA-ILP in terms of AUC (0.79 vs. 0.75). This difference becomes less pronounced when limiting the curves to partial data with FPR <0.2 (Fig. 2).

The values of the MRR for the four subsets of candidate genes (Table 1) show almost no difference between the methods, except for the set SF-cat2. The higher MRR value of Y-set than of SF-cat1 confirms that those 18 genes have less support when their prediction is based on prior knowledge and not on

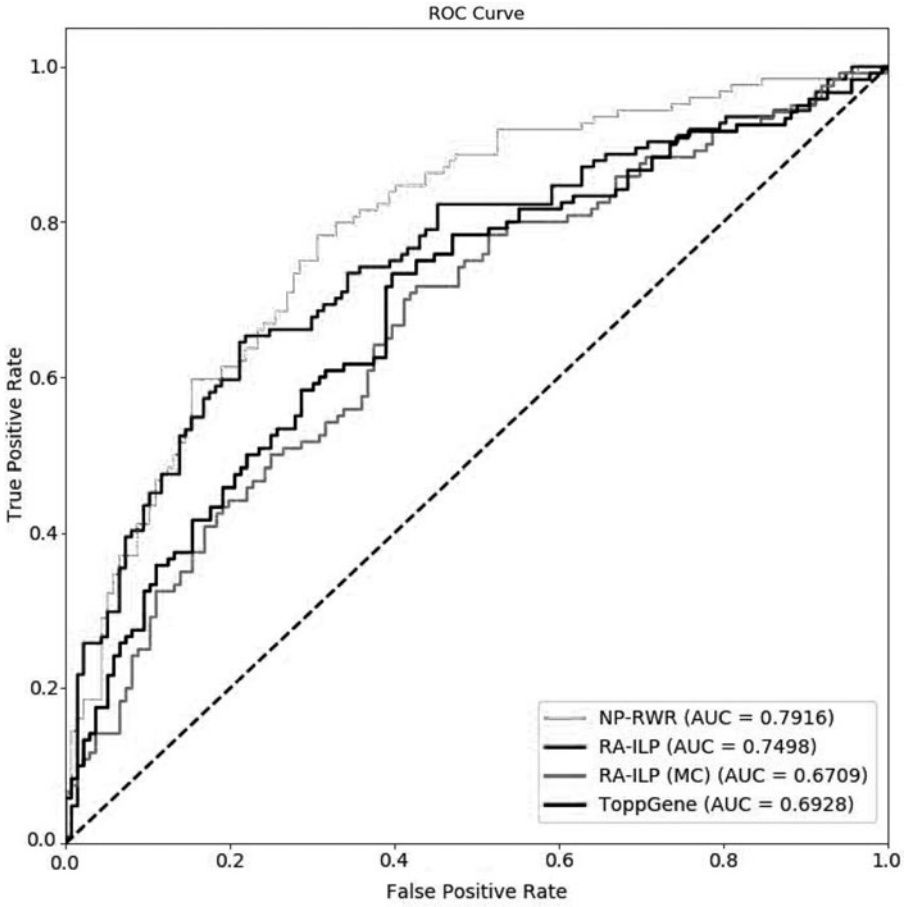


FIG. 1. The ROCs and corresponding AUC values of each of the considered approaches. AUC, area under the curve; NP-RWR, network propagation using random walk with restart; RA-ILP, rank aggregation with integer linear programming; ROCs, receiver operating curves.

experimental genomic evidence. All the mentioned measures indicate that NP-RWR outperforms RA-ILP, although the differences in the parameters are relatively small.

The Jaccard index of 0.71 ($p < 0.01$) computed for the two sets of TP reveals a somewhat different scenario. Although, as the p value indicates, this value is higher than the one that would be expected by chance, it is still true that 30% of genes reported as TP in one output are not TP in the other. Not surprisingly, the Jaccard index for the TP in the top 20% ranks is even lower (0.45).

5.2. ToppGene versus RA-ILP on multiple input categories

RA-ILP and ToppGene are both based on the aggregation of input rankings. They differ not only in their computational approach (combinatorial vs. statistical) but also in their input data types and categories. Although ToppGene takes into account scores (p values) associated with all genes, RA-ILP only considers their ranks. Furthermore, Toppgene has input rankings induced by various biological categories rather than coexpression values only. Thus a comparison of the performance of ToppGene with respect to RA-ILP is not indicative of the effectiveness of its computational process. To partially address this issue, we ran RA-ILP on the same input rankings as ToppGene. The set of candidate genes in this experiment does not include the following genes since they are missing from the ToppGene website: *FAM47A*, *MSNP1*, *UBN2*, *GNAS-AS1*, and *GRIA1*. They are not in the set SF-cat1.

According to the ROCs and AUC values (Fig. 2), ToppGene has a slightly better performance than RA-ILP (0.69 vs. 0.66). The same is true if we consider the partial AUC (0.06 vs. 0.05). The values of the MRR are somewhat mixed, but in both cases the MRR of the Y-set is about the same as that of a set of random genes of the same size.

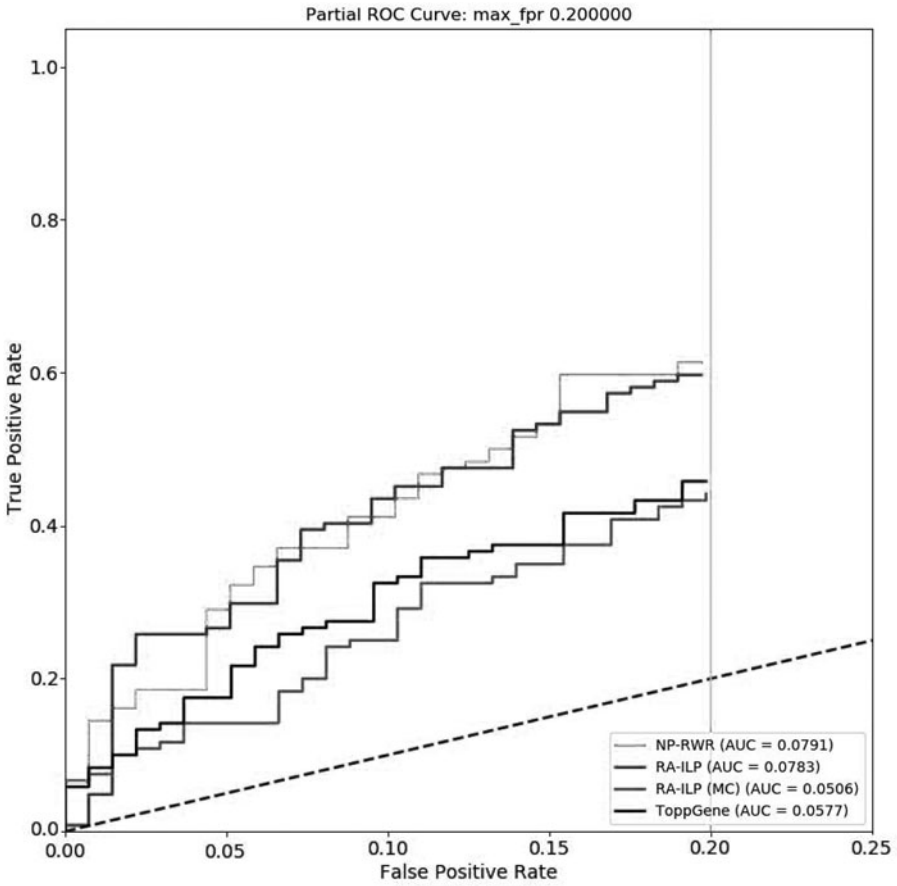


FIG. 2. The partial ROCs and corresponding AUC values of each of the considered approaches. Only values of false positive rate below or equal to 0.2 are displayed.

6. DISCUSSION

We cannot draw definitive conclusions about the various approaches due to the obvious limitation in the experiments. However, based on our results and performance measures, we can observe the following trends.

- With coexpression data as input, the NP method achieves the best performance in terms of both AUC and MRR although only marginally. For instance, the difference in MRR between NP-RWR and

Table 1. The Median Rank Ratio of the Candidate Genes of the Sets ALL, SF-cat1, SF-cat2, and Y-Set for Each of the Computational Methods and Data Type Considered

	MRR			
	Coexpression values		*Multiple categories*	
	NP-RWR	*RA-ILP*	*ToppGene**	*RA-ILP**
ALL	0.31	0.31	0.35	0.37
SF-cat1	0.20	0.20	0.25	0.40
SF-cat2	0.38	0.34	0.35	0.33
Y-set	0.26	0.27	0.55	0.52

The MRR values corresponding to the methods marked with an asterisk have a different normalization factor due to the different size of the test data. MRR, median rank ratio; NP-RWR, network propagation using random walk with restart; RA-ILP, rank aggregation with integer linear programming; SF-cat1, SFARI genes falling in the phenotypic profile 1; SF-cat2, SFARI genes falling in the phenotypic profile 2; Y-set, ASD-risk genes with sequence-level mutations from experimental studies reported in Yuen et al. (2017) and not considered in the SFARI gene database; ALL, including all the 124 candidate genes belonging to SF-cat1, SF-cat2 and Y-set.

RA-ILP is 0 when the analysis focuses on the genes with strong prior evidence of association with autism, as those in SF-cat1.

- A good practical strategy to increase the reliability of the prediction of disease genes is to combine the outputs of two or more tools and focus on the shared top-most ranked genes. This is illustrated by the fact that, when the difference in AUC and MRR values is negligible, the set of genes that are reported as TP may show remarkable differences. Based on the Jaccard index, we can conclude that the difference among the outputs is larger than the other parameters AUC and MRR seem to suggest.
- A larger difference in performance appears when a computational method is applied to different inputs, that is, coexpression values and multiple categories. This is already evident from the partial ROCs (Fig. 2). Although this is not surprising and has been observed before, the extent of it can be better appreciated by considering the following measures.

The Jaccard index of the sets of TPs of RA-ILP using coexpression values and multiple ToppGene categories is 0.4 ($p=0.89$) no larger than that expected by chance.

Although the MRR values of the set of all candidates are only slightly different, those of the sets SF-cat1 and of Y-set are remarkably different (Table 1, columns 3 and 5).

The ND_{rank} distances on all pairs of approaches along with the permutation-based p values (Table 2) confirm the strong impact of data inputs on the results. In particular, the D_{rank} distance of the ranked lists of RA-ILP on different inputs is no smaller than that expected by chance.

Furthermore, for the Y-set, all ND_{rank} distances are much larger than for the set ALL, that is, for those genes there is much less agreement among the various outputs. This suggests that with the existing biological knowledge and the selection of seed genes, the genes of the Y-set, identified by sequence experiments, likely would not be predicted.

The analysis so far has focused on the accuracy of the results. The time performance is also very important since it affects the amount of data that can be prioritized. The running times of RA are few orders of magnitude higher than those of network-based propagation. In fact, tools based on RA can only deal with partial lists of genes, whereas others are able to process the entire genome. This may be a heavy limitation since in high-throughput sequence experiments, it is useful to rely on promising genes selected from the whole genome. More practical approaches to RA exist that generate approximate solutions with a bounded error rate; however, we considered here the exact version of the optimization process for better accuracy.

In conclusion, although the standard measures ROC, AUC, and MRR may indicate minor differences in performance of the various approaches, measures that look at individual genes, such as Jaccard index and D_{rank}, or measures computed on subsets of genes (SF-cat1, Y-set), may signal major differences in the predicted genes. We argue that a comparative analysis should take that into consideration.

Table 2. The ND_{RANK} Distance of the Ranked Lists of Candidate Genes of All Pairs of Approaches

	ND_{rank} of ALL			
	Coexpression values		*Multiple categories*	
	NP-RWR	*RA-ILP*	*ToppGene*	*RA-ILP*
NP-RWR		16.1 ($p<0.01$)	37.4 ($p=0.746$)	37.6 ($p=0.902$)
RA-ILP			42 ($p=0.952$)	42 ($p=0.952$)
ToppGene				10.8 ($p<0.01$)
	ND_{rank} of Y-set			
	Coexpression values		*Multiple categories*	
	NP-RWR	*RA-ILP*	*ToppGene*	*RA-ILP*
NP-RWR		32.7 ($p<0.01$)	80 ($p=0.4$)	76 ($p=0.34$)
RA-ILP			85 ($p=0.79$)	89.1 ($p=0.88$)
ToppGene				13.7 ($p<0.01$)

p Values are in parentheses. The top table shows the ND_{rank} values over all candidates, the table at the bottom shows the ND_{rank} values over the genes of the Y-set.

7. APPROACHES USING TISSUE-SPECIFIC DATA

Much work on disease–gene associations has relied on the same data irrespective of the specific disease, cancer, or diabetes, etc., and of the fact that a given disease may manifest itself only in specific organs and specific tissues. For instance, when studying autism, expression values of the whole brain or even of the entire organism were considered even though the functions of gene products may be dependent on the specific regions of the brain where they are performed. Only recently some studies focused on tissue-specific data (Kang et al., 2011; Magger et al., 2012; Antanaviciute et al., 2015; Greene et al., 2015; Krishnan et al., 2016). As a result, most of the benchmarks on gene prioritization lack tissue-specificity analysis.

In this study, we assess the performance of prediction tools applied to tissue-specific coexpression data. Furthermore, we show that the distance D_{rank} between the predicted rankings of the candidate genes in two regions can provide an effective basis for clustering brain regions in accordance with their role in ASDs.

In our comparative analysis, we only consider NP-RWR and RA-ILP since the data categories of ToppGene are not tissue specific.

7.1. NP-RWR versus RA-ILP

Brainspan (Miller et al., 2014) provides tissue-specific temporal data from 16 brain regions for 15 time periods. In the analysis, we associated 16 profiles with each gene, one for each region, consisting of all the gene expression values in that region over the entire time span. We also conducted the same investigation by limiting the gene profiles to the time periods 3–7 corresponding to the prenatal stages. Coexpression values of gene pairs were computed as correlations of those profiles. We constructed 16 coexpression networks, one for each region, and used them as inputs to NP-RWR, thus obtaining multiple outputs. Similarly, we applied RA-ILP to each separate region.

Our results show that NP-RWR outperforms RA-ILP in terms of both AUC and MRR in the majority of tissues. Furthermore, for each region we observe a stronger agreement between the outputs of the two methods, NP-RWR and RA-ILP, measured by the Jaccard index and the D_{rank}, than when the methods are applied to the whole brain. This indicates a more reliable identification of prioritized genes based on tissue-specific data.

Details of the analysis along with ROCs, MRR, and D_{rank} values are given in Section 1 of the Supplementary Material.

The considered methods succeed in highlighting specific regions that appear relevant to ASDs. Based on the performance measures, the region S1C primary somatosensory cortex and striatum stand out as important, testifying of the role of these tissues in the disease and consistent with known results (Fuccillo, 2016; Balasco et al., 2020).

7.2. Clustering of brain tissues

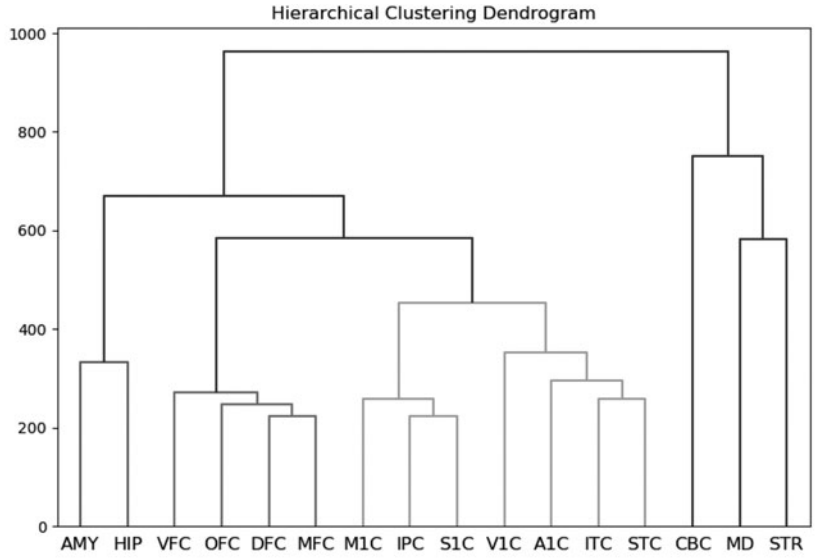
In this study, we show that the distance D_{rank} introduced in this study may serve as a basis for clustering brain tissues according to their association with known causal ASD genes. Specifically, given the prioritized lists of genes obtained by an algorithm on two tissues, the distance between two tissues is defined as the D_{rank} distance of such lists. Thus, this measure takes into account the difference among tissues in the ranks of individual candidate genes.

We used the prioritized lists of genes obtained as outputs of NP-RWR and hierarchical clustering to generate the clusters reported in Fig. 3. As it can be observed, the clusters accurately separate the regions of the prefrontal cortex from those of the primary motor-somatosensory cortex.

A comparison with other studies based on brain- and tissue-specific data shows the effectiveness of this clustering approach.

Based on Brainspan data (Willsey et al., 2013), clusters of brain regions with high transcriptional similarity during the fetal developmental stage were identified. They were obtained by calculating pairwise Pearson correlation coefficients between each of 9 seed genes and 16,947 genes from the exon array data set. The main difference with our results is in the primary motor-somatosensory cortex that in Willsey et al. (2013) is split across two clusters: {*M1C*, *S1C*, *VFC*, *MFC*, *DFC*, *OFC*} (prefrontal and primary motor-somatosensory cortex) and {*V1C*, *ITC*, *IPC*, *A1C*, *STC*}.

	Cluster 1
A1C	Primary auditory cortex
IPC	Posterior inferior parietal cortex
ITC	Inferior temporal cortex
M1C	Primary motor cortex
S1C	Primary somatosensory cortex
STC	Superior temporal cortex
V1C	Primary visual cortex
	Cluster 2
AMY	Amygdala
HIP	Hippocampus
	Cluster 3
OFC	Orbital prefrontal cortex
DFC	Dorsal prefrontal cortex
VFC	Ventral prefrontal cortex
MFC	Medial prefrontal cortex



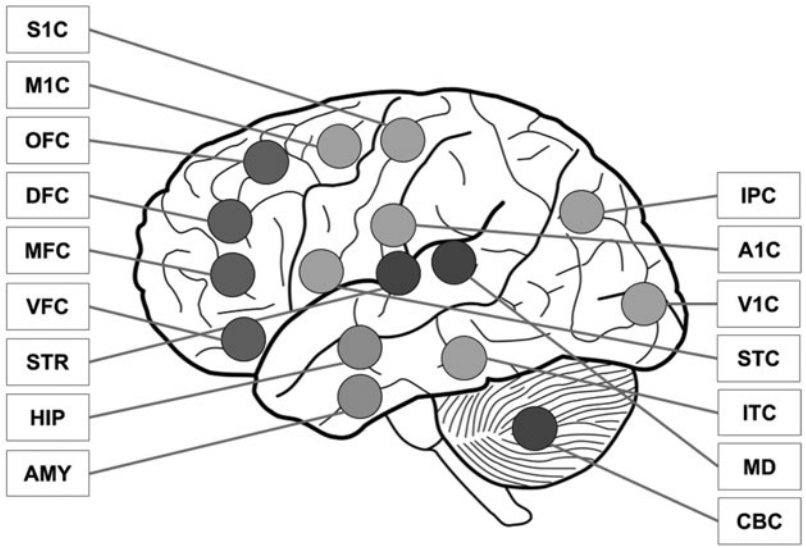


FIG. 3. Hierarchical Clustering of Brain Regions Based on the Distance D_{rank} of Their Prioritized Lists of Causal Genes from Network Propagation Using Random Walk with Restart.

The D_{rank} distance is able to produce clusters that are as accurate as previous methods by only considering a small set of test genes rather than the entire genome.

As a last note, the clusters generated starting from the outputs of RA-ILP did not match previously identified clusters and were difficult to interpret (not reported here).

8. CONCLUSIONS

We conducted a systematic comparison of two main computational approaches to gene prioritization. We evaluated the approaches on a small set of test genes including high-confidence ASD genes to validate their association with autism. The comparison was done separately on data of the whole brain and on 16 distinct brain regions. For a more thorough comparison, in addition to standard measures of performance (i.e., ROC, AUC, and MRR), we proposed the use of the Jaccard index and of the D_{rank} distance. Such distances take into consideration the ranks that each individual gene has in the prioritized lists by different methods and different input data types. Moreover, they allow to quantitatively evaluate the impact of the data categories on a given computational approach.

Overall, based on all the mentioned measures, the results of our experiments clearly show that the network-based propagation approach outperforms the RA approach. We observed that, although parameters such as AUC and MRR show minor differences in performance between the two methods on the entire set of test genes, they signal a remarkably different behavior when focusing on small subsets of test genes such as SF-cat1, SF-cat2, and Y-set.

On tissue-specific data, all performance parameters suggested a more important role of some regions in ASDs. Furthermore, we found that the distance D_{rank} was effective in identifying meaningful clusters of brain regions.

ACKNOWLEDGMENT

All authors thank the Department of Statistical Sciences of University of Rome—La Sapienza for computing time on the TeraStat cluster.

AUTHOR DISCLOSURE STATEMENT

The authors declare they have no conflicting financial interests.

FUNDING INFORMATION

Guerra was partially supported by US–Israel Binational Grant (BSF) n. 2018141. U.F.P. was partially supported by GNCS Project 2019 Innovative methods for the solution of medical and biological big data. U.F.P. and F.P. are partially supported by Università di Roma—La Sapienza Research Project 2018 Analisi, sviluppo e sperimentazione di algoritmi praticamente efficienti.

SUPPLEMENTARY MATERIAL

Supplementary Material

REFERENCES

Adle, R., Kolde, R., Kull, M., et al. 2009. Mining for coexpression across hundreds of datasets using novel rank aggregation and visualization methods. *Genome Biol*. 10, R139.

Antanaviciute, A., Daly, C., Crinnion, L., et al. 2015. Genetier: Prioritization of candidate disease genes using tissue-specific gene expression profiles. *Bioinformatics*. 31, 2728–2735.

Balasco, L., Provenzano, G., and Bozzi, Y. 2020. Sensory abnormalities in autism spectrum disorders: A focus on the tactile domain, from genetic mouse models to the clinic. *Front. Psychiatry*. 43, 1016.

Banerjee-Basu, S., and Packer, A. 2010. SFARI Gene: An evolving database for the autism research community. *Dis. Model Mech.* 3, 133–135.

Börnigen, D., Tranchevent, L., and Bonachela-Capdevila, F. 2012. An unbiased evaluation of gene prioritization tools. *Bioinformatics*. 28, 3081–3088.

Chen, J., Bardes, E.E., Aronow, B.J., et al. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* 45, W305–W311.

Chung, N., Miasojedow, B., Startek, M., et al. 2019. Jaccard/tanimoto similarity test and estimation methods for biological presence-absence data. *BMC Bioinformatics*. 20, 644.

Ciriello, G., Mina, M., Guzzi, P.H., et al. 2012. Alignnemo: A local network alignment method to integrate homology and topology. *PLoS One*. 7, e38107.

Conitzer, V., and Sandholm, T. 2006. Computing the optimal strategy to commit to. Proceedings of the 7th ACM Conference on Electronic Commerce, Ann Arbor, MI, USA, pp. 82–90.

Cowen, L., Ideker, T., Raphael, B., et al. 2017. Network propagation: A universal amplifier of genetic associations. *Nat. Rev. Genet. Rev*. 18, 551–562.

Csermely, P., Korcsmáros, T., Kiss, H.J., et al. 2013. Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. *Pharmaco. Ther.* 138, 333–408.

Dwork, C., Kumar, R., Naor, M., et al. 2001. Rank aggregation methods for the web. Proceedings of the 10th International Conference on World Wide Web, Hong Kong, pp. 613–622.

Fuccillo, M.V. 2016. Striatal circuits as a common node for autism pathophysiology. *Front. Neurosci.* 10, 27.

Greene, C., Krishnan, A., Wong, A., et al. 2015. Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* 47, 569–576.

Guala, D., Sjölund, L., and Sonnhammer, E. 2014. Maxlink: Network-based prioritization of genes tightly linked to a disease seed set. *Bioinformatics*. 30, 2689–2690.

Guala, D., and Sonnhammer, E. 2017. A large-scale benchmark of gene prioritization methods. *Sci. Rep*. 7, 1–10.

Jingchao Ni, J., Koyuturk, M., Tong, H., et al. 2016. Disease gene prioritization by integrating tissue-specific molecular networks using a robust multi-network model. *BMC Bioinformatics*. 17, 453.

Kang, H.J., Kawasawa, Y.I., Cheng, F., et al. 2011. Spatio-temporal transcriptome of the human brain. *Nature*. 478, 483–489.

Kolde, R., Laur, S., Adler, P., et al. 2012. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 28, 573–580.

Krishnan, A., Zhang, R., Yao, V., et al. 2016. Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat. Neurosci.* 19, 1454–1462.

Kumar, R., and Vassilvitskii, S. 2010. Generalized distances between rankings. Proceedings of the 2010 International World Wide Web Conference, Raleigh, NC, USA, pp. 571–580.

Lee, I., Blom, U., Wang, P., et al. 2011. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.* 21, 1109–1121.

Li, P., and Milenkovic, O. 2017. Multiclass minmax rank aggregation. 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, pp. 3000–3004.

Magger, O., Waldman, Y., Ruppin, E., et al. 2012. Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. *PLoS Comput. Biol.* 8, e1002690.

Miller, J.A., Ding, S.L., Sunkin, S.M., et al. 2014. Transcriptional landscape of the prenatal human brain. *Nature*. 508, 199–206.

Minji, K., Farnoud, F., and Milenkovic, O. 2015. Hydra: Gene prioritization via hybrid distance-score rank aggregation. *Bioinformatics*. 31, 1034–1043.

Mitra, K., Carvunis, A., Ramesh, S.K., et al. 2013. Integrative approaches for finding modular structure in biological networks. *Nat. Rev. Genet*. 14, 719–732.

Moreau, Y., and Tranchevent, L. 2012. Computational tools for prioritizing candidate genes: Boosting disease gene discovery. *Nat. Rev. Genet.* 13, 523–526.

Nivit, G., Singh, S., and Aseri, T.C. 2014. Computational disease gene prioritization: An appraisal. *J. Comput. Biol.* 21, 456–465.

Optimization. Inc., ''gurobi optimizer reference manual,'' 2020. Available at: https://www.gurobi.com/documentation/9.0/refman/index.html. Last viewed on May 25, 2019.

Ra, G., Liu, J., Feng, L., et al. 2006. Analysis of protein sequence and interaction data for candidate disease gene prediction. *Nucleic Acids Res.* 34, 313–323.

Schalekampf, F., and von Zuylen, A. 2009. Rank aggregation: Together we're strong. Proceedings of the 11th SIAM Workshop on Algorithm Engineering and Experiments (ALENEX), New York, NY, USA, pp. 38–51.

Shim, J., Hwang, S., and Lee, I. 2015. Pathway-dependent effectiveness of network algorithms for gene prioritization. *PLoS One*. 10, e0130589.

Sinan, E., Bebek, G., Ewing, R.M., et al. 2011. DADA: Degree-aware algorithms for network-based disease gene prioritization. *BioData Mining*. 4, 19.

Stelzer, G., Rosen, N., Plaschkes, I., et al. 2016. The GeneCards suite: From gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics*. 54, 1–30.

Vanunu, O., Magger, O., Ruppin, E., et al. 2010. Associating genes and protein complexes with disease via network propagation. *PLoS Comput. Biol.* 6, e1000641.

Willsey, A.J., Sanders, S.J., Li, M., et al. 2013. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*. 155, 997–1007.

Wong, A., Krishnan, A., Yao, V., et al. 2015. Imp 2.0: A multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks. *Nucleic Acids Res.* 43, W128–W133.

Yuen, C., Merico, R., Bookman, D., et al. 2017. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* 20, 602–611.

Zhang, B., and Horvath, S. 2005. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Molec. Biol.* 4, 17.

Zolotareva, O., and Kleine, M. 2019. A survey of gene prioritization tools for mendelian and complex human diseases. *J. Integr. Bioinform*. 16, 20180069.

Address correspondence to:
Dr. Concettina Guerra
Georgia Institute of Technology College of Computing
School of Interactive Computing
85 5th Street NW
Atlanta, GA 30308
USA

E-mail: guerra@cc.gatech.edu