A Network-Centric Framework for the Evaluation of Mutual Exclusivity Tests on Cancer Drivers

Rafsan Ahmed[1], Cesim Erten[2], Aissa Houdjedj[2], Hilal Kazan[2]* and Cansu Yalcin[2]

[1]Electrical and Computer Engineering Graduate Program, Antalya Bilim University, Antalya, Turkey, [2]Department of Computer Engineering, Antalya Bilim University, Antalya, Turkey

One of the key concepts employed in cancer driver gene identification is that of mutual exclusivity (ME); a driver mutation is less likely to occur in case of an earlier mutation that has common functionality in the same molecular pathway. Several ME tests have been proposed recently, however the current protocols to evaluate ME tests have two main limitations. Firstly the evaluations are mostly with respect to simulated data and secondly the evaluation metrics lack a network-centric view. The latter is especially crucial as the notion of common functionality can be achieved through searching for interaction patterns in relevant networks. We propose a network-centric framework to evaluate the pairwise significances found by statistical ME tests. It has three main components. The first component consists of metrics employed in the network-centric ME evaluations. Such metrics are designed so that network knowledge and the reference set of known cancer genes are incorporated in ME evaluations under a careful definition of proper control groups. The other two components are designed as further mechanisms to avoid confounders inherent in ME detection on top of the network-centric view. To this end, our second objective is to dissect the side effects caused by mutation load artifacts where mutations driving tumor subtypes with low mutation load might be incorrectly diagnosed as mutually exclusive. Finally, as part of the third main component, the confounding issue stemming from the use of nonspecific interaction networks generated as combinations of interactions from different tissues is resolved through the creation and use of tissue-specific networks in the proposed framework. The data, the source code and useful scripts are available at: https://github.com/abu-compbio/NetCentric.

Keywords: mutual exclusivity, network-centric mutual exclusivity evaluation, cancer drivers, cancer genomics, tumor mutation load

1 INTRODUCTION

Cancer is a disease caused mostly due to a gradual accumulation of somatic alterations that give rise to pathway dysregulation through alterations in copy number, DNA methylation, gene expression, and molecular function. An important challenge in cancer genomics is to distinguish driver mutations from passenger mutations. The former are those determined to be causal for cancer progression, whereas the latter are characterized as those not leading to any selective advantage. Several computational methods have been proposed for the identification of cancer driver genes or

driver modules of genes by integrating mutations data with various other types of genetic data; see Dimitrakopoulos and Beerenwinkel. (2017), Zhang and Zhang. (2018), Bailey et al., 2018, Tokheim et al., 2016 for recent comprehensive evaluations and surveys on the topic.

A phenomenon observed frequently in the data pertaining to the alterations that the tumors acquire is mutual exclusivity (ME); a driver mutation is less likely to occur in case of an earlier mutation that has common functionality in the same molecular pathway (Thomas et al., 2007; Yeang et al., 2008; Leiserson et al., 2016; van de Haar et al., 2019). Therefore several driver gene or module identification approaches employ ME detection as part of their problem definitions and optimization goals (Babur et al., 2015; Ciriello et al., 2012; Leiserson et al., 2013; Kim et al., 2015; Ahmed et al., 2019; Baali et al., 2020). Such a central role in driver gene and module identification has led to the design of many different approaches for defining and computing mutual exclusivity. Some of these approaches are based on combinatorial definitions of mutual exclusivity (Vandin et al., 2012; Leiserson et al., 2013; Sarto Basso et al., 2019; Ahmed et al., 2019; Song et al., 2020; Baali et al., 2020). In most cases the combinatorial definitions are incorporated and tested within a driver gene or module identification framework, rather than as stand-alone ME tests. On the other hand, the vast majority of the ME detection approaches are based on statistical tests (Ciriello et al., 2012; Szczurek and Beerenwinkel, 2014; Leiserson et al., 2015; Constantinescu et al., 2015; Hua et al., 2016; Canisius et al., 2016; Leiserson et al., 2016; Kim et al., 2017; Liu et al., 2020; Zhang et al., 2020) and in most cases for such approaches the specific goal is to provide ME significance results. Therefore the focus of the proposed framework is the evaluation of the latter set of approaches consisting of the statistical ME tests.

Among such approaches, MEMo builds a graph based on gene similarities and extracts cliques from this graph. To determine whether each clique has significant mutual exclusivity, it then proposes a null model generated by randomly permuting the set of genomic events, while preserving the overall distribution of observed alterations across both genes and samples, and introduces a Markov Chain Monte Carlo (MCMC) permutation strategy based on random network generation models (Ciriello et al., 2012). Szczurek and Beerenwinkel. (2014) propose a probabilistic, generative model of mutual exclusivity, explicitly taking coverage, impurity, and error rates into account. Based on such a model, they provide a statistical test of mutual exclusivity by comparing its likelihood to the null model that assumes independent gene alterations. Mutex defines the alteration of two genes to be mutually exclusive if their overlap in samples is significantly less than expected by chance, where the statistical significance of the overlaps are calculated using a hypergeometric test with the assumption of a uniform alteration frequency among samples (Constantinescu et al., 2015). This may not always be the case as in many data sources there are hyper-mutated samples. The problem is resolved partially by simply excluding such samples from the analysis. CoMEt (Leiserson et al., 2015) on the other hand provides an exact statistical test for mutual exclusivity conditional on the observed frequency of each alteration with the goal of introducing less bias towards high frequency alterations. Based on this it provides

a tail enumeration procedure to compute the exact test, as well as a binomial approximation. DISCOVER provides a statistical independence test that makes no assumption of identical gene alteration probabilities across tumors (Canisius et al., 2016). The alteration probabilities are estimated by solving a constrained optimization problem guaranteeing the probabilities are consistent with both the observed number of alterations per gene and the observed number of alterations per tumor. The tumor-specific gene alteration probabilities are then used to compute the probability of concurrent alterations which in turn are used to decide whether the number of tumors altered in both genes deviates from the expectation through an analytical test based on the Poisson-binomial distribution. WeXT provides a weighted exact test that conditions simultaneously on the number of samples with a mutation and the per-event, per-sample mutation probabilities (Leiserson et al., 2016). A recursive formulation to compute p-values for this weighted test exactly and a saddle-point approximation of the test are proposed. WeSMe provides a permutation-based test and an approximation of significance through a weighted sampling technique that enables further improvements in running time spent for sampling and a way to obtain a better precision without increasing the computational time significantly (Kim et al., 2017). Mina et al. propose the SELECT method which uses a weighted version of mutual information to identify significant mutual exclusivity or co-occurrence patterns where significance is estimated by comparing against patterns observed in random permutations of the data (Mina et al., 2017). Two recently suggested ME tests are FSME (Zhang et al., 2020) and MEScan (Liu et al., 2020). The former proposes a seed-and-extend strategy to alleviate the computational cost of a permutation-based test. The seed pairs are constructed by a combinatorial formulation incorporating both ME and the coverage of the pair. The seeds are then grown with new genes by employing an independence test. MEScan provides a test statistic that incorporates a patient and gene-specific background mutation rate in the calculation to adjust for the background noise, and that includes a gene-specific weight to down-weigh genes with high mutation rates. Such a statistic is then employed in an MCMC algorithm followed by a false discovery rate control.

We propose a network-centric framework to evaluate the pairwise significances found by statistical ME tests. It is important to make a distinction between the network-centric view of the current study and that of the previous studies employing both network data and the concept of ME (Ciriello et al., 2012; Leiserson et al., 2013; Kim et al., 2015; Ahmed et al., 2019; Baali et al., 2020). The latter are network-centric in the sense that the proposed ME tests are applied on interacting pairs or subnetworks as part of a more general goal of identifying cancer driver genes/modules. Thus due to the nature of the set objectives their evaluations focus on the success of output genes/modules matching reference cancer-related drivers/pathways. The proposed study takes on an approach in the opposite direction; we assume the interaction network and the reference cancer-related drivers to be inputs to our framework which evaluates the success of various ME tests. The focus of the proposed framework is on pairwise significances since one of the major application areas where ME tests are commonly

employed is knowledge-based cancer driver identification where pairwise ME significances are of major essence. In terms of the general objectives our work is most similar to that of Deng et al., 2017, where a framework for performance comparisons of statistical ME detection approaches is proposed and executed on six such tests. An important distinction is that the performance analysis of Deng et al. is based on experiments with simulated data and the framework does not suggest any mechanism to avoid confounders inherent in ME detection. One such confounder is due to the alterations specific to cancer subtypes (Deng et al., 2017; van de Haar et al., 2019). Alterations in different subtypes may be incorrectly diagnosed with ME, although the alterations are not due to any natural root causes of ME such as redundant functionality. Inspired by the observation that mutual exclusivity is enriched among physically interacting pairs of genes (Dao et al., 2017), our network-centric view aims to recognize such false positives by constructing reference sets based on known drivers gathered from neighborhoods of interaction networks. Furthermore, inspired by the mutation load confounding concept of van de Haar et al., 2019, we extend our network-centric framework to dissect side effects caused by mutation load artifacts; mutations that drive tumor subtypes with low mutation load might be incorrectly diagnosed as mutually exclusive. A possible drawback of the proposed network-centric evaluation framework would be due to the use of nonspecific interaction networks that are generated as combinations of interactions from different tissues and are thus suboptimal in resolving confounding issues of mutual exclusivity. In order to detect whether there exists such discrepancies or to limit their effect if they do, we therefore refine the network-centric approach by designing further tests on tissue-specific networks (TSN) we construct based on gene co-expression.

2 METHODS

The overall network-centric ME evaluations framework has three main components. The first one consists of definitions of the metrics employed in the network-centric ME evaluations. Such metrics are designed so that network knowledge and the reference set of known cancer genes are incorporated in ME evaluations under a careful definition of proper control groups. The second component detects whether the use of the interactome information provides similar advantages in ME corrections of pairwise mutual exclusivity findings as the subtype-stratification idea suggested by van de Haar et al., 2019. Finally, the third component extends our framework to incorporate tissue-specific networks with the aim of reducing the possible side effects of using nonspecific interaction networks.

2.1 Metrics for the Network-Centric Mutual Exclusivity Evaluations

Assuming that cancer driver genes in the same pathway are more likely to show mutually exclusive mutation profiles, we utilize the interactome to devise a strategy for evaluating the ME methods and the effects of the interactome information on quantifying ME. Let $\mathcal{G}, \mathcal{C}, \mathcal{T}, \mathcal{S}, p_t, c$ denote respectively the input Protein-Protein Interaction (PPI) network, the employed cohort, the statistical ME test undergoing the network-centric ME evaluations, the golden standard reference gene set of known cancer drivers, the p-value threshold for significance, and the type of the control group to be employed. Let $N_{\mathcal{S}}(g_i)$ denote the set of genes from $\mathcal{S}$ that are in the neighborhood of the node corresponding to gene g_i in the PPI network $\mathcal{G}$. For a gene $g_i \in \mathcal{S}$, corresponding to each neighbor $g_j \in N_{\mathcal{S}}(g_i)$, we randomly select a gene g_r from a control group $\mathcal{X}_c(g_i)$, and compute TP^{cur}, FP^{cur} based on the $-$ log-transformed p-values $p_{i,j}$ and $p_{i,r}$ as computed by the ME test $\mathcal{T}$. Here $p_{i,j}$ denotes the significance of the mutual exclusivity of the pair g_i, g_j for $g_i \in \mathcal{S}$ and $g_j \in N_{\mathcal{S}}(g_i)$, and $p_{i,r}$ denotes the significance of the mutual exclusivity of the pair g_i, g_r for a random gene g_r from the control group. Based on the premise that cancer driver genes interacting in the PPI network are likely to exhibit ME, a pair g_i, g_j belongs to the set of True Positives if $p_{i,j}$ is significant and a pair g_i, g_r belongs to the set of False Positives if $p_{i,r}$ is significant.

To obtain robust results, the selection of the random genes from the control group is repeated robustness_iterations number of times, which is set to 100 in all the evaluations, except for those testing the robustness of the framework with respect to various parameter settings. Finally the medians of these 100 instances are summed over all genes $g_i \in \mathcal{S}$ to provide the necessary statistics TP, FP. Thus precision, sensitivity, and the F1 scores are computed based on these statistics. Precision is calculated as $|TP|/(|TP| + |FP|)$. Sensitivity is calculated with the formula $|TP|/|P|$ where P corresponds to condition positives which are defined as the gene pairs $g_i, g_j \in \mathcal{S}$ where g_i, g_j interact in $\mathcal{G}$.

We note that limiting our focus solely on these conventionally formed TP, FP classes may be misleading as each one considers the significance of $p_{i,j}$ and $p_{i,r}$ individually. A more detailed inspection with a simultaneous consideration of their values could prove more insightful in certain cases since they both involve a common gene g_i. Towards this aim we introduce the *strict* versions of these conventional classes. More specifically TP_{strict} consists of g_i, g_j pairs where $p_{i,j}$ is significant not only with respect to the given threshold but also as compared to the p-value of the control pair g_i, g_r. Similarly FP_{strict} consists of the control pairs g_i, g_r, where $p_{i,r}$ is more significant than both the threshold value and $p_{i,j}$. Based on these strict classes we can compute three metrics: precision$_{strict}$, sensitivity$_{strict}$, and F1$_{strict}$. Precision$_{strict}$ is defined as $|TP_{strict}|/(|TP_{strict}| + |FP_{strict}|)$ and sensitivity$_{strict}$ is defined as $|TP_{strict}|/|P|$. Such a consideration is especially convenient in reducing any potential bias inherent in genes like TP53 which have large mutation frequencies almost exclusively in tumors with small numbers of mutations; both $p_{i,j}$ and $p_{i,r}$ are likely to be significant in such a scenario giving rise to vagueness in the conventional F1 score. A comparison of F1$_{strict}$ values based on the two statistics simultaneous by their nature, TP_{strict} and FP_{strict} provides a more rigorous evaluation in such cases.

For the network-centric ME evaluations we employ two different definitions for the control groups. For the first one, the control group $\mathcal{X}_1(g_i)$ consists of genes in $\mathcal{S}$ that do not interact with g_i in the PPI network. For the second one, $\mathcal{X}_2(g_i)$

consists of neighbors of g_i in the PPI network that are not in $\mathcal{S}$. In the latter case only the genes $g_i \in \mathcal{S}$ for which the number of neighbors not in $\mathcal{S}$ is larger than or equal to the number of neighbors in $\mathcal{S}$ are taken into account.

2.2 Network-Centric Mutual Exclusivity Corrections in Relation to Mutation Load Association

Some statistical mutual exclusivity tests are based on the assumption that gene's alterations across tumors are identically distributed. Among the approaches considered in this study Fisher's Exact Test and MEGSA belong to this category. However, it has been observed that the number of alterations per tumor can vary quite considerably, even in tumors of the same type; colorectal tumors with microsatellite stability have a median of 66 non-synonymous mutations, but colorectal tumors with microsatellite instability have a median of 777 mutations (Vogelstein et al., 2013; Leiserson et al., 2016). It has been shown that under such settings the mutual exclusivity tests relying on identical alteration probabilities across tumors may lead to reduced sensitivity for mutual exclusivity analysis (Canisius et al., 2016). The effects of varying alteration probabilities on pairwise mutual exclusivity calculations have been formalized within the context of the so-called mutation load confounding (MLC) in a recent study by van de Haar et al., 2019. MLC is a correlation between the number of statistically significant mutual exclusivity findings and the mutation load association (MLA) of a gene. MLA of a gene is calculated by running a logistic regression where a gene's binary mutation status indicating whether the gene is mutated or not in a tumor is used as the only feature to predict the mutation load of that tumor. Mutation load is defined as the number of genes that are mutated in a tumor. Once the coefficient of the feature is obtained by fitting the logistic regression model, it is standardized by dividing by the standard error to make it comparable across the genes. This standardized coefficient value is defined as the MLA value. Note that negative MLA values correspond to higher mutation frequencies in tumors with low mutation loads, whereas positive values correspond to higher mutation frequencies in tumors with high mutation loads. Strong negative correlations between the MLA of a gene and the number of statistically significant pairwise mutual exclusivities have been observed, implicating the finding that the more negative a gene's MLA, the higher the number of other genes that show mutual exclusivity with that particular gene (van de Haar et al., 2019). However, such a negative correlation does not always imply true ME since a gene that exclusively shows large mutation frequency in tumors with low mutation loads, naturally has a better chance of forming mutually exclusive pairs with other genes. Thus extra sources of information are necessary to filter out the pairs with true ME relations among a set of statistically significant pairwise mutual exclusivities postulated by some exclusivity test. van de Haar et al., 2019 make use of the subtype information for such a purpose and show that MLC can be reduced by correcting via tumor subtype stratification. Such a correction greatly reduces the number of gene pairs reported to show mutual exclusivity, especially for pairs that include genes with low MLA. A major drawback is the absence of subtype information for many tumors. As part of our network-centric ME framework, we suggest that such a correction can be efficiently done with the interaction network data, rather than or better yet on top of the subtype information. For this purpose we calculate the correlation between the number of statistically significant pairwise ME findings and the MLA for two settings; one where pairwise mutual exclusivities are sought between a gene in $\mathcal{S}$ and all other genes in $\mathcal{S}$, and the other where a gene in $\mathcal{S}$ is checked against only its PPI neighbors that are in $\mathcal{S}$. The computations of the two settings are repeated with the subtype-stratified data as well, to see the added value of the network-centric ME corrections on top of the subtype-based corrections on statistically significant pairwise MEs.

2.3 Network-Centric Mutual Exclusivity Evaluations in Relation to Tissue-Specific Networks

Rather than using a common nonspecific network for all the cancer types, in this component of our evaluation framework we employ TSN based on the tissue in which the tumor develops. To construct the TSN for a particular tissue, we start with the original PPI network and remove the edges between the pairs of genes that are not co-expressed in the corresponding tissue. For this purpose, we download RNA-seq datasets from GTEX portal (GTEXConsortium, 2020). See **Supplementary Table S49** for the total number of available samples for each tissue. To determine the co-expressed genes, we follow the procedure described in Luck et al., 2020. For each pair of genes that have an edge in the original PPI network, we identify the number of samples where both genes have Transcripts Per Kilobase Million (TPM) values ≥ 1. We then divide this number with the total number of samples where either gene has a TPM value ≥ 1. The resulting value is called the co-expression ratio. Gene pairs interacting in the original network are included in the TSN_{cor} if the co-expression ratio is $\geq cor$, for a given threshold cor.

In addition to applying the network-centric metrics introduced in **Section 3.1** on the constructed TSNs, we also propose a more detailed evaluation in terms of ROC analysis based on tissue-specificity. For this purpose, we define the gene pairs with co-expression ratio value of 1 as tissue-specific gene pairs. Similarly, the gene pairs with co-expression ratio values ≤ 0.5 are called non-tissue-specific gene pairs. To test whether a specific ME test identifies stronger mutual exclusivities for the tissue-specific gene pairs in $\mathcal{S}$, we rank the gene pairs in $\mathcal{S}$ in increasing order of p-values. To construct the control group, we rank the same number of random samples of gene pairs not in $\mathcal{S}$ with respect to the p-values making sure that the sizes of the positive (or negative) sets of gene pairs not in $\mathcal{S}$ are exactly the same as those that are found for the gene pairs in $\mathcal{S}$. For both gene pairs in $\mathcal{S}$ and gene pairs not in $\mathcal{S}$, the set of positives consists of the tissue-specific gene pairs, whereas non-tissue-specific gene pairs are labelled as negatives. We then compute the True Positive Rate (TPR) and the False Positive Rate (FPR) for each case. Note that for robustness considerations the control group

Method	Precision	Sensitivity	F1 Score	Precision$_{strict}$	Sensitivity$_{strict}$	F1 Score$_{strict}$
DISCOVER	0.661	0.220	0.331	0.708	0.183	0.291
DISCOVER Strat	0.727	0.041	0.078	0.727	0.041	0.078
Fisher's Exact Test	0.500	0.031	0.058	0.500	0.031	0.058
MEGSA	0.611	0.056	0.103	0.588	0.051	0.094
MEMO	0.658	0.329	0.439	0.647	0.237	0.347
WExT	0.676	0.403	0.505	0.725	0.329	0.453

computations are repeated 100 times and the median TPR and FPR values are reported.

3 RESULTS

3.1 Input Data and Parameter Settings

The somatic mutation data from TCGA was preprocessed and provided by van de Haar et al., 2019. The 8 different cancer types and their corresponding tumor samples within the dataset is as follows: BLCA (411), BRCA (1026), COADREAD (498), LUAD (568), LUSC (485), SKCM (468), STAD (438) and UCEC (531). The preprocessing step involves the removal of all mutations with "variant_classification" of "Silent," "3'UTR," "Intron," "5'UTR," "RNA," "3'Flank" and "5'Flank" from the TCGA data. The input data is then further filtered by mutation frequency threshold, t, to include genes with $> t$ mutations across the cohort. More specifically, with $t = 20$ we include the genes that are mutated in more than 20 samples within the cancer type under study. Regarding subtypes, we download subtype information for BRCA from the cBioPortal (Cerami et al., 2012; Gao et al., 2013) and the CMS stratification for COADREAD from (Guinney et al., 2015). We use the COSMIC Cancer Gene Census database to compile the set of known cancer genes (Sondka et al., 2018).

For the results presented in the main document we employ the IntAct PPI network as it is a comprehensive and well-characterized database (Orchard et al., 2014). As a preprocessing step, we remove duplicate edges and edges below the confidence threshold of 0.35 from the network. The final network contains 15,079 nodes and 103,520 edges. For the gene expression data employed in the construction of TSNs, we download RNA-Seq data from the Genotype-Tissue Expression (GTEx) portal (GTEXConsortium, 2020) (05-06-2017).

For the comparative evaluations of our network-centric framework described in the previous section, we choose six popular statistical mutual exclusivity methods: DISCOVER (Canisius et al., 2016), DISCOVER Strat (Canisius et al., 2016; van de Haar et al., 2019), Fisher's Exact Test, WeXT (Leiserson et al., 2016), MEMo (Ciriello et al., 2012) and MEGSA (Hua et al., 2016). Among these, MEMo and MEGSA are originally designed to output p-values for a set of genes with size > 2. For MEMo, we re-implement the first part of the algorithm where pairwise ME p-values are estimated. We use $Q = 100$ and $N = 10,000$ as suggested by the original paper (Ciriello et al., 2012). For MEGSA, pairwise ME p-values are calculated by applying chi-square cumulative probability less than or equal to the value of the

log likelihood calculated by the funestimate function. With regards to the parameter settings of our proposed framework, we employ the values of 5 and 20 for t.

3.2 Mutual Exclusivity Evaluations Based on Defined Metrics

Table 1 and **2** show the results of evaluating the 6 ME detection methods on COADREAD data where $t = 20$ and we use the data from 498 patients for which subtype information is available. We use $\mathcal{X}_1$ and $\mathcal{X}_2$ as the control group in **Table 1** and **2**, respectively. We first discuss the results of $\mathcal{X}_1$. We observe that DISCOVER Strat gives the highest precision and precision$_{strict}$ values. The ranking of the other methods from best to worst in terms of precision or precision$_{strict}$ is as follows: WeXT, DISCOVER, MEMo, MEGSA and Fisher's Exact Test. A comparison of the precision and precision$_{strict}$ values distinguishes two groups of ME methods; for DISCOVER, DISCOVER Strat, Fisher's Exact Test, and WexT the precision$_{strict}$ values are greater than or equal to the precision values, whereas the exact opposite is observed for MEGSA and MEMo. This suggests that the performance of the methods in the latter group gets worse when random control gene pair is considered simultaneously in the precision calculation, that is precision$_{strict}$. Compared to the precision, we observe much larger differences among the sensitivity or the sensitivity$_{strict}$ values output by the employed methods. We can group the methods into two where the first group contains WeXT, MEMo and DISCOVER, and the second group contains the rest of the methods. The first group of methods give much larger sensitivity or sensitivity$_{strict}$ values than the second. For instance, the sensitivity value obtained with WeXT is an order of magnitude larger than that of Fisher's Exact Test. This also shows that the second group of methods are more conservative than the first group of methods. WeXT is the least conservative approach based on its high sensitivity value. Even though WexT predicts many significant p-values, it still has a competitive precision$_{strict}$ value which is slightly lower than the maximum observed value (0.725 vs 0.727). Accordingly, WeXT obtains the best F1 score and F1$_{strict}$ score which is followed by MEMo and DISCOVER. The remaining three methods give much smaller F1 scores and they rank as follows from highest to lowest: MEGSA, DISCOVER Strat and Fisher's Exact Test. Comparing the conventional F1 score with the F1$_{strict}$ score of each ME method, the largest difference is observed for MEMo indicating that the consideration of the random pair as a control affects its performance dramatically. Another interesting observation is

TABLE 2 | Results of network-centric ME evaluation framework with control group $\mathcal{X}_2$ COADREAD t20 (498 samples, 107 CGC-CGC pairs).

Method	Precision	Sensitivity	F1 Score	Precision$_{strict}$	Sensitivity$_{strict}$	F1 Score$_{strict}$
DISCOVER	0.537	0.276	0.365	0.579	0.210	0.308
DISCOVER Strat	0.455	0.048	0.086	0.400	0.038	0.069
Fisher's Exact Test	0.444	0.038	0.069	0.375	0.028	0.052
MEGSA	0.571	0.075	0.133	0.538	0.066	0.118
MEMO	0.566	0.388	0.460	0.495	0.215	0.300
WExT	0.575	0.438	0.497	0.596	0.295	0.395

TABLE 3 | Results of network-centric ME evaluation framework with control group $\mathcal{X}_1$ COADREAD t5 (498 samples, 1748 CGC-CGC pairs).

Method	Precision	Sensitivity	F1 Score	Precision$_{strict}$	Sensitivity$_{strict}$	F1 Score$_{strict}$
DISCOVER	0.647	0.052	0.096	0.658	0.046	0.086
DISCOVER Strat	0.618	0.012	0.024	0.618	0.012	0.024
Fisher's Exact Test	0.583	0.008	0.016	0.565	0.007	0.014
WExT	0.645	0.121	0.203	0.668	0.102	0.177

TABLE 4 | Results of network-centric ME evaluation framework with control group $\mathcal{X}_2$ COADREAD t5 (498 samples, 1625 CGC-CGC pairs).
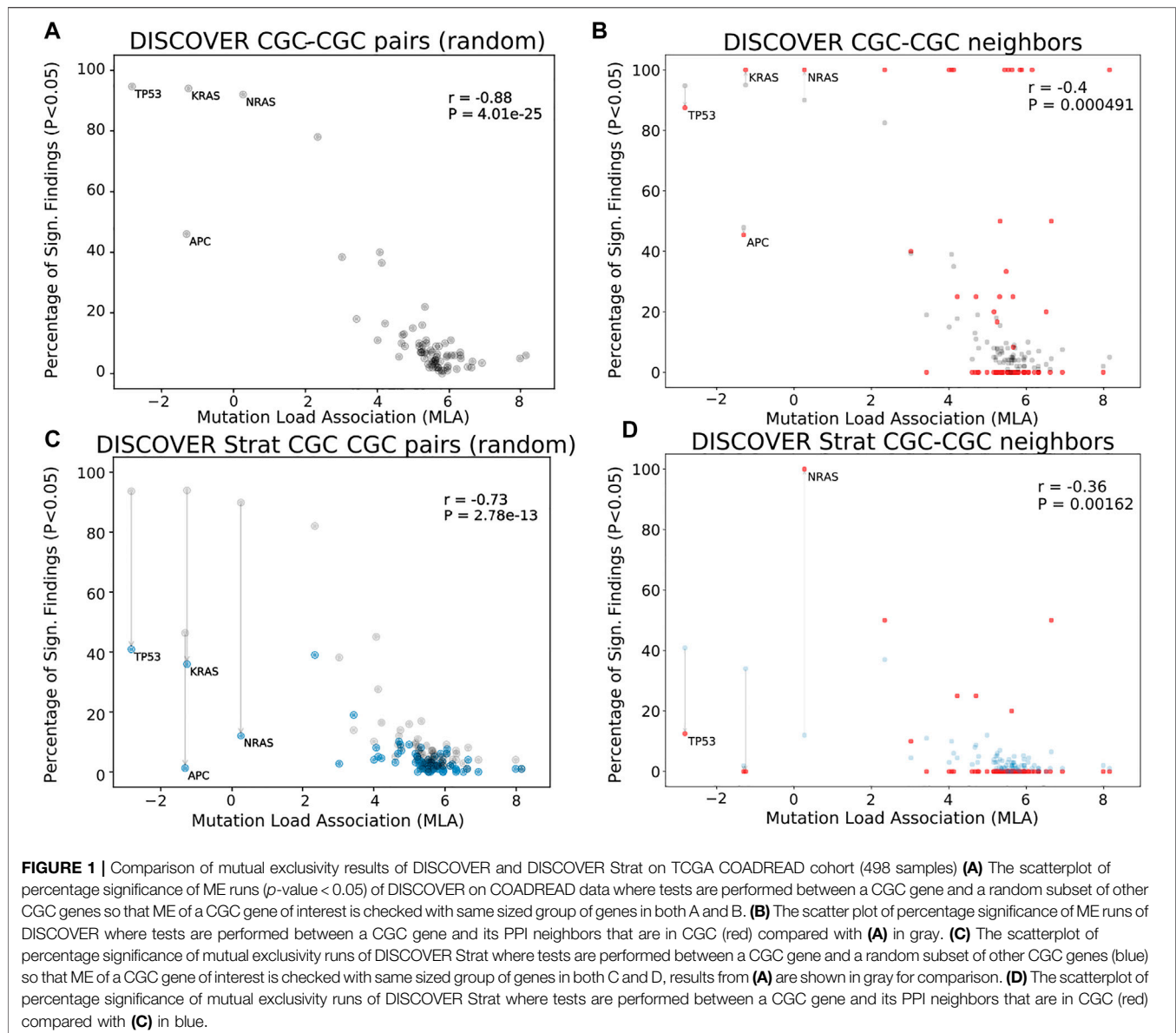
Method	Precision	Sensitivity	F1 Score	Precision$_{strict}$	Sensitivity$_{strict}$	F1 Score$_{strict}$
DISCOVER	0.721	0.052	0.097	0.746	0.048	0.090
DISCOVER Strat	0.641	0.013	0.025	0.641	0.013	0.025
Fisher's Exact Test	0.619	0.008	0.016	0.619	0.008	0.016
WExT	0.670	0.118	0.200	0.712	0.103	0.180

the lower performance of DISCOVER Strat compared to DISCOVER which suggests that the use of subtype information is not useful for COADREAD. **Table 2** shows the results where $\mathcal{X}_2$ is used as the control group. Since $\mathcal{X}_2(g_i)$ is defined as the non-CGC neighbors of g_i in the PPI network, we can only consider the CGC genes that have more non-CGC neighbors than CGC neighbors. As such, the number of pairs included in this analysis is much smaller than that of **Table 1** (107 vs 196). The ranking of the methods in **Table 2** with respect to F1 score and sensitivity remain the same as **Table 1**. However, there are differences in the ranking with respect to other metrics. For instance, WeXT ranks best in terms of precision whereas the best ranking method in **Table 1**, DISCOVER Strat, ranks the fifth. Compared to **Table 1**, the precision values of all the methods are smaller in **Table 2**. We see the opposite trend for sensitivity values. These changes are in parallel with the increase in percent significant p-values output by the methods. For instance, the percentage of significant p-values output by DISCOVER is 12% in **Table 1** and 18% in **Table 2**. We also observe differences between the conventional and the strict versions of the employed metrics. WeXT and DISCOVER have increased precision$_{strict}$ values compared to precision whereas we observe the opposite trend for the rest of the methods. Additionally, the ranking of the methods with respect to F1 score and F1$_{strict}$ score is different. Namely, MEMo's ranking decreases from second highest to third highest when we switch from F1 score to F1$_{strict}$ score. Accordingly, DISCOVER's ranking improves from third highest to second highest based on F1 score. This increases

the confidence of DISCOVER results as F1$_{strict}$ requires a stricter definition of true and false positives. **Supplementary Table S1** shows the results with $\mathcal{X}_1$ control group and $t = 20$ filtering for the other cancer types. A detailed discussion of these results are available in the **Supplementary Material**.

Table 3 and **4** show the COADREAD results of $t = 5$ setting with $c = X_1$ and $c = X_2$, respectively. Using a lower value for t increases the number of gene pairs tested in our analysis. When we compare these results with the results we obtained when $t = 20$, we observe few differences. Though the number of tested gene pairs is larger, the percentage of significant p-values obtained by the methods decreases. For instance, the percentage of significant p-values output by WeXT for COADREAD data decreases from 42 to 14% when t is changed from 20 to 5. This is likely related to the larger inclusion of low mutation frequency genes when $t = 5$. An interesting observation for $t = 5$ results is the decrease in DISCOVER Strat's performance. For COADREAD, DISCOVER Strat's precision and precision$_{strict}$ value is the highest for $t = 20$ when $\mathcal{X}_1$ is used as the control group. However, when $t = 5$, we observe that it ranks after WeXT and DISCOVER in terms of precision/precision$_{strict}$ value. Similarly, for BRCA dataset, DISCOVER Strat ranks after WeXT for both control groups $\mathcal{X}_1$ and $\mathcal{X}_2$ (**Supplementary Table S25B**, **Supplementary Table S37B**).

Lastly, we investigate the robustness of our results with respect to robustness_iterations value, the p-value significance threshold value, the reference gene set and the employed PPI network. The results together with a discussion of these results are available in

FIGURE 1 | Comparison of mutual exclusivity results of DISCOVER and DISCOVER Strat on TCGA COADREAD cohort (498 samples) **(A)** The scatterplot of percentage significance of ME runs (p-value < 0.05) of DISCOVER on COADREAD data where tests are performed between a CGC gene and a random subset of other CGC genes so that ME of a CGC gene of interest is checked with same sized group of genes in both A and B. **(B)** The scatter plot of percentage significance of ME runs of DISCOVER where tests are performed between a CGC gene and its PPI neighbors that are in CGC (red) compared with **(A)** in gray. **(C)** The scatterplot of percentage significance of mutual exclusivity runs of DISCOVER Strat where tests are performed between a CGC gene and a random subset of other CGC genes (blue) so that ME of a CGC gene of interest is checked with same sized group of genes in both C and D, results from **(A)** are shown in gray for comparison. **(D)** The scatterplot of percentage significance of mutual exclusivity runs of DISCOVER Strat where tests are performed between a CGC gene and its PPI neighbors that are in CGC (red) compared with **(C)** in blue.
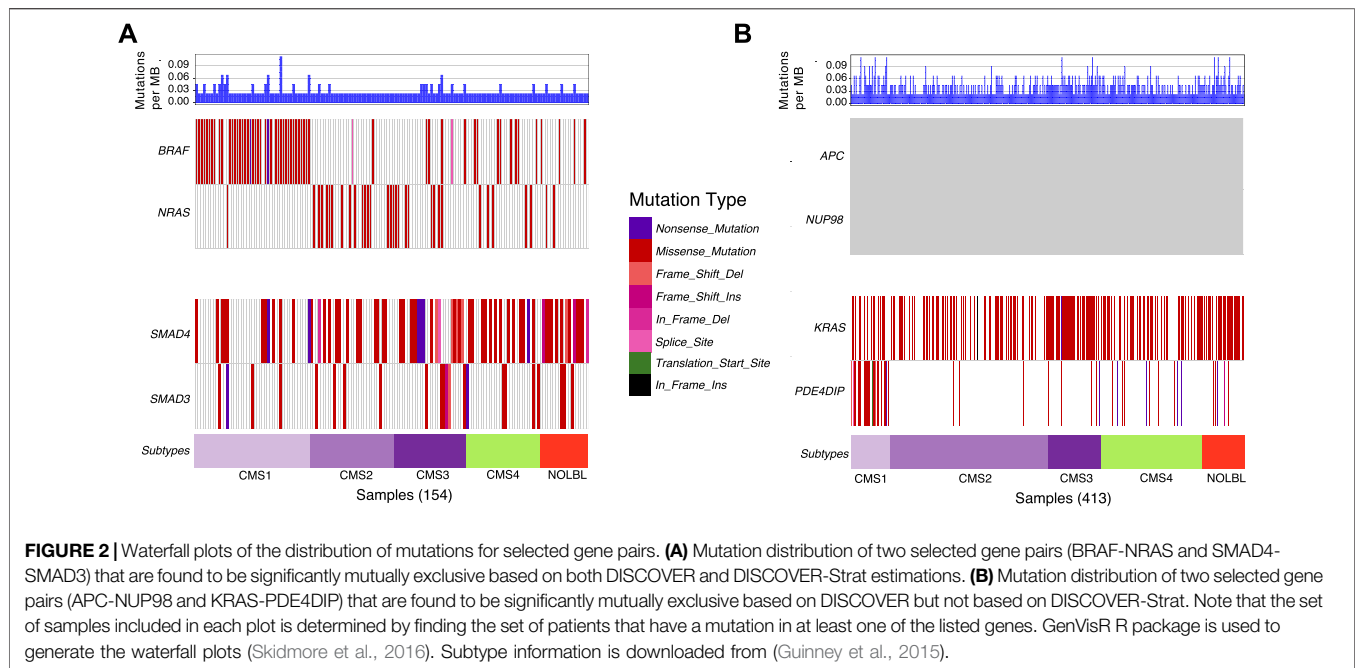
Supplementary Table S2-S47. As a summary, our conclusions remain the same in these different settings and the largest differences are observed when the employed PPI network is changed.

3.3 Mutual Exclusivity Evaluations Based on Corrections via Mutation Load Association

Having compared the ME tests with respect to our novel network-centric evaluation framework, we now assess whether including network knowledge reduces the mutation load confounding (MLC) problem introduced by van de Haar et al., 2019. van de Haar et al. identified a strong negative correlation between the MLAs of genes and their percent significant findings in mutual exclusivity tests. In van de Haar et al., 2019, these statistics are computed for a set of 341 genes from an established cancer gene

panel (Cheng et al., 2015) where, for each gene, mutual exclusivity tests are performed with all the other genes in the panel. Here, we first perform a similar analysis where we use the COSMIC CGC database (Forbes et al., 2017) to define the reference cancer gene set as it is more comprehensive and up to date.

Figures 1A shows the MLA of the reference cancer genes vs the percent significant findings in mutual exclusivity tests performed with DISCOVER for the TCGA COADREAD cohort (498 tumors). We observe a strong negative correlation between MLA values and percent significant findings in mutual exclusivity tests (Pearson correlation -0.88, p-value $4.0e - 25$) similar to van de Haar et al., 2019. In Figures 1B, we take into account the PPI information to calculate percent significant findings. Namely, for each CGC gene, we perform mutual exclusivity tests only with its PPI neighbors that are also in CGC. Note that CGC genes which do not have any CGC

FIGURE 2 | Waterfall plots of the distribution of mutations for selected gene pairs. **(A)** Mutation distribution of two selected gene pairs (BRAF-NRAS and SMAD4-SMAD3) that are found to be significantly mutually exclusive based on both DISCOVER and DISCOVER-Strat estimations. **(B)** Mutation distribution of two selected gene pairs (APC-NUP98 and KRAS-PDE4DIP) that are found to be significantly mutually exclusive based on DISCOVER but not based on DISCOVER-Strat. Note that the set of samples included in each plot is determined by finding the set of patients that have a mutation in at least one of the listed genes. GenVisR R package is used to generate the waterfall plots (Skidmore et al., 2016). Subtype information is downloaded from (Guinney et al., 2015).

neighbors are excluded from this analysis. To make a fair comparison between **Figures 1A,B**, only the CGC genes that have CGC neighbors are shown in **Figures 1A**. We also ensure that the mutual exclusivity of a gene of interest is checked with same sized group of genes in both **Figures 1A,B**. To achieve this in **Figures 1A**, for each gene, we compute mutual exclusivity with a random subsample of the CGC reference set, the same size as the set of CGC neighbors of that gene. We repeat this random sampling 100 times and plot the mean percent significant findings value. For reference, **Supplementary Figure S3A, S3D** contains versions of **Figures 1A,C**, where all CGC genes (i.e., with and without CGC neighbors) are plotted and mutual exclusivities are checked between all CGC pairs, as it was done in van de Haar et al., 2019.

In **Figures 1B**, we observe a reduced correlation when network information is included (Pearson correlation -0.4, *p*-value $4.91e − 4$). We also run DISCOVER Strat where stratification is based on CMS subtypes (Guinney et al., 2015). We plot these results in **Figures 1C** where we again ensure comparability with **Figures 1D** where both subtype and network information are considered. Comparing **Figures 1A** and **Figures 1C**, we verify the findings of van de Haar et al., although with less significance in correlation difference (Pearson correlation $−0.73$, *p*-value $2.8e − 13$). It should be noted that the subtype stratification inherently causes an overall decrease in percent significant findings, not specific to genes with low MLA. On the contrary the idea of ME corrections through network incorporation, materialized in the comparison of **Figures 1A** and **Figures 1B**, inherently leads to an increase in percent significant findings. Most of the decreases occur in genes with small number of CGC neighbors. When we compare **Figures 1D** to **Figures 1B**, the decrease in correlation from $−0.4$ to $−0.36$ indicates that including subtype information

is still useful when used on top of network-based corrections we propose.

Next, we utilize waterfall plots to compare the outputs of DISCOVER and DISCOVER-Strat to assess how MLA and subtype information can affect mutual exclusivity findings. **Figures 2A** shows two selected gene pairs that display significant mutual exclusivity based on both DISCOVER and DISCOVER-Strat estimations on TCGA COADREAD dataset. The mutual exclusivity between BRAF and NRAS, two members of the MAPK pathway, is well-known and has been detected in multiple cancer types including melanoma, myeloma and colorectal cancer (Samowitz et al., 2006; Roth et al., 2010; Popovici et al., 2012) BRAF is frequently mutated in patients from CMS1 subtype whereas NRAS shows almost no mutation across these patients. However, since BRAF and NRAS mutations are mutually exclusive across not only CMS1 subtype but also across the other subtypes, DISCOVER-Strat identifies this pair as significantly mutually exclusive. Similarly, SMAD3 and SMAD4 are two members of the TGF-β pathway and the mutual exclusivity between these two transcription factors is previously reported in colorectal cancer (Fleming et al., 2013). Mutations on SMAD3 and SMAD4 are distributed almost uniformly across the subtypes. As such, the mutual exclusivity between the mutations of these two genes is still significant when subtype information is incorporated. **Figures 2B** similarly shows two selected gene pairs that display significant mutual exclusivity based on DISCOVER but not based on DISCOVER Strat. For the first pair, we observe that NUP98 is mutated almost exclusively in patients from the CMS1 subtype which shows hypermutation due to microsatellite instability. On the other hand, there is a depletion of APC mutations among the patients from the CMS1 subtype which results in a low MLA value. As such,
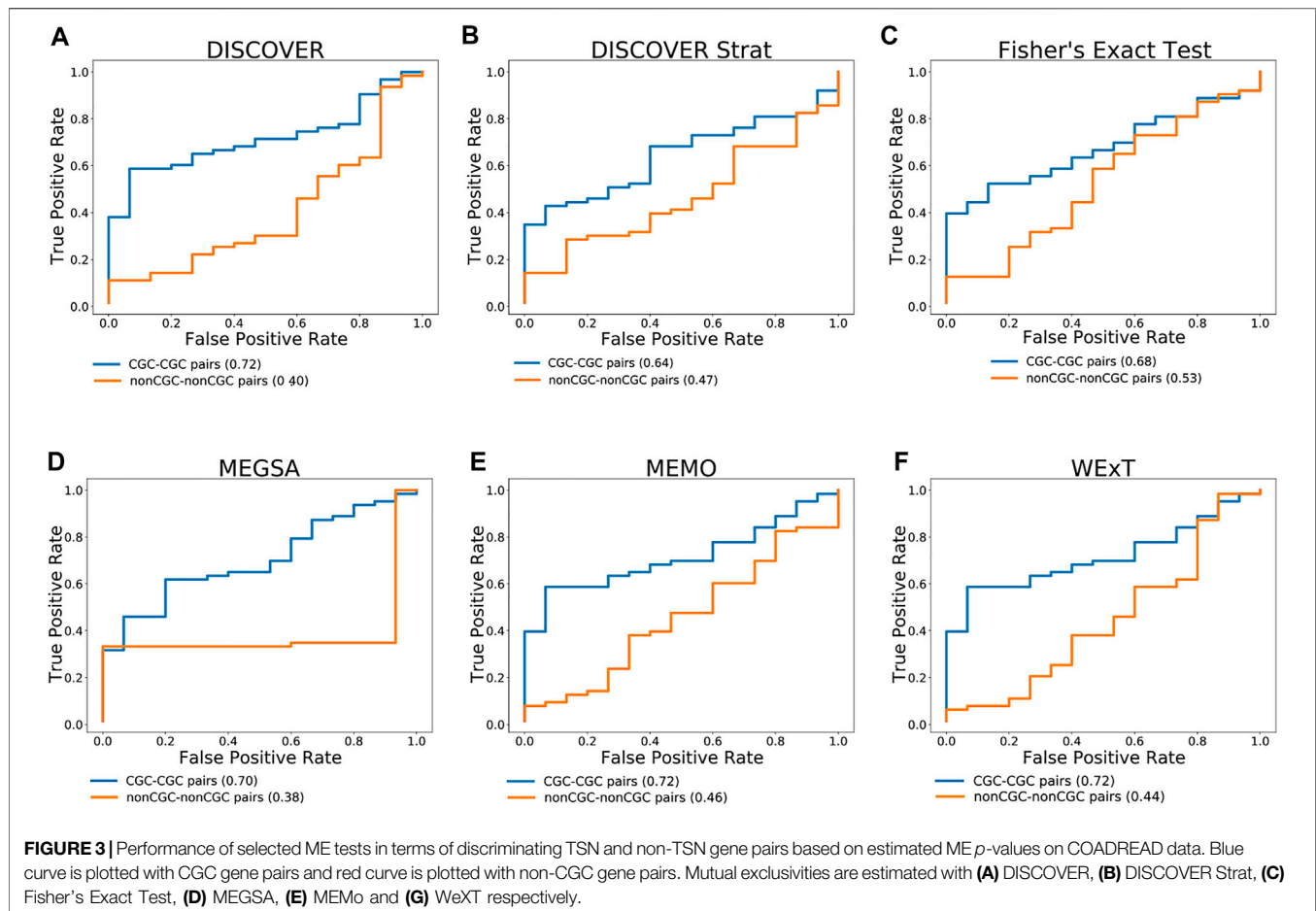
DISCOVER Strat fails to detect a significant ME between these two genes since it explores ME within each subtype separately. A similar observation can also be made for the KRAS-PDE4DIP pair where the former has a low MLA and the latter has a high MLA.

Supplementary Figure S2 compares the MLA of the reference cancer genes with the percent significant findings in mutual exclusivity tests for BRCA. Similar to the results that we obtain for COADREAD data, including network information reduces the correlation between MLA and ME detection rate (**Supplementary Figures S2B vs S2C**). The magnitude of reduction is even more significant than what we observe for COADREAD data (Pearson correlation −0.93 vs −0.27). Interestingly, including subtype information results in a very slight decrease in correlation coefficient (−0.93 to −0.91) (**Supplementary Figures S2B vs S2E**) as opposed to what we observe for COADREAD. We observe that including subtype information on top of network information results in no decrease in correlation (**Supplementary Figures S2C vs S2F**). This difference in the effect of including subtype information for BRCA and COADREAD datasets could be related to the average tumor mutation load of subtypes. BRCA subtypes have comparable average TML values (Her2: 146, LumA:65, LumB: 71, Normal: 55) whereas the CMS1 subtype in COADREAD has a dramatically larger average TML value compared to the other subtypes of COADREAD (CMS1: 1387, CMS2:93, CMS3: 272, CMS4: 212) We repeat the same analysis with the other ME detection methods as well as for other cancer types when t is set to 20 (**Supplementary Figures S1-S8**). We observe that the percent significant finding values can vary remarkably across the tumor types. Compared to other cancer types, we observe smaller percent significant findings for LUSC (**Supplementary Figures S5A,S5D, S5G**). Similarly, very few pairs have percent significance value ≥ 20 when we consider network information in LUSC (**Supplementary Figures S5C,S5F,S5I**). On the contrary, we observe many pairs with large percent significant values for CGC-CGC neighbors in UCEC data. This is particularly true for DISCOVER and WeXT results (**Supplementary Figures S8C–S8L**).

When we consider the correlation between MLA and percent significant values, we observe that adding network information decreases the correlation coefficient values for all cancer types and for all ME detection methods except for Fisher's Exact Test. Fisher's Exact Test results show an increased correlation with the addition of network information for LUSC and SKCM (**Supplementary Figures S5-S6** D vs F). Also, the correlation coefficient can not be computed for LUAD and STAD since Fisher's Exact Test gives a value of 0 for the percent significant findings of all considered genes (**Supplementary Figures S4D-S7F**). Another interesting observation is the variance in magnitude of decrease in correlation values across different tumor types. In particular, we observe a smaller decrease in correlation values for LUAD compared to other cancer types. The analogous results are also available for $t = 5$ setting (**Supplementary Figures S9-S16**). For all the cancer types, the correlation between MLA values and percent significant findings decreases and becomes non-significant for most cases.

We should also note that the majority of CGC genes have only one neighbor within the data setting of the cancer type under consideration. This leads to percentage significant findings of either 0 or 1 in many cases simply because these are the only possible values; for COADREAD see **Figures 1B** and **Figures 1D** where 41 out of 74 genes under study have only one CGC neighbor in the COADREAD data settings. To avoid any such possible biases, we repeat the same evaluations after filtering out those CGC genes with only one neighbor. The evaluations still provide significant decreases in correlation coefficient values analogous to the decreases observed in **Figures 1B** as compared to **Figures 1A** and **Figures 1D** as compared to **Figures 1C**. For detailed results, see **Supplementary Figures S17-S24** for $t = 20$ and **Supplementary Figures S25-S32** for $t = 5$.

Individual genes of interest are those that have increased percent significant findings when network neigborhood information is incorporated while at the same have significant number of CGC neighbors. More specifically, for the former constraint, we identify the CGC genes with at least 0.1 increase in percentage of significant findings value of WeXT, DISCOVER and MEMo when the network information is included as opposed to the scenario when it is not (e.g., for COADREAD, **Figures 1A** vs **Figures 1B**). We choose these 3 ME methods since they are top performers based on the defined metrics in **Section 3.1**. For STAD, SKCM and UCEC, since MEMo results are unavailable, we only consider WeXT and DISCOVER results. For the second constraint, we include the CGC genes with at least 3 CGC neighbors. For COADREAD, this selection procedure results in four genes: EP300, CREBBP, NCOA2 and NCOR2. Among these, EP300 is a well-known tumor suppressor in epithelial cancer types including COADREAD (Gayther et al., 2000). For BRCA, the only identified gene is PIK3R1. PIK3R1 is found to be significantly mutually exclusive with PIK3CA and SPEN based on both WeXT, DISCOVER and MEMo results. PIK3R1 and PIK3CA are members of the PI3K pathway and their mutual exclusivity has been previously established in the literature (Chen et al., 2018). For LUAD, PTPRB is the only identified gene and is found to be mutually exclusive with EGFR, a well-known oncogene in non-small cell lung cancer (Bethune et al., 2010). The set of identified genes for STAD are NCOA2, NCOR2 and CREBBP; all of which are found to be mutually exclusive with TP53. For SKCM, we identify ERBB4, RAC1, EP300 and ITK. ERBB4 is a well-known oncogene in skin cancer and found to be mutually exclusive with ERBB2 (Prickett et al., 2009; Nielsen et al., 2014). ERBB2 and ERBB4 indeed belong to the same family (i.e. ErbB family of receptor tyrosine kinases) and form a heterodimer receptor for Heparin-binding EGF-like growth factor (HB-EGF) (Iwamoto et al., 2017). RAC1 mutation P29S is an established driver in melanoma (Jiang et al., 2018). RAC1 is found to be mutually exclusive with MYH9, a tumor suppressor in melanoma (Singh et al., 2020). Lastly, ITK has been shown to be an oncogene in melanoma (Carson et al., 2015). For UCEC, we identify 33 genes in total. Among these, KIT and PTEN have established roles in UCEC cancer development (Chang et al., 2015; Wang et al., 2020). Moreover, PTEN is found to be strongly mutually exclusive with SPOP, whose mutations are also associated with endometrial cancer (Clark and Burleson,

FIGURE 3 | Performance of selected ME tests in terms of discriminating TSN and non-TSN gene pairs based on estimated ME p-values on COADREAD data. Blue curve is plotted with CGC gene pairs and red curve is plotted with non-CGC gene pairs. Mutual exclusivities are estimated with **(A)** DISCOVER, **(B)** DISCOVER Strat, **(C)** Fisher's Exact Test, **(D)** MEGSA, **(E)** MEMo and **(G)** WeXT respectively.

2020). Lastly, for BLCA and LUSC, no gene satisfies the abovementioned criteria. Overall these results suggest that the CGC genes that show increased ME with network incorporation as well as their mutually exclusive partner genes often have established roles in the development of the particular cancer type.

3.4 Mutual Exclusivity Evaluations Based on Corrections via Tissue-Specific Networks

We first provide our ME evaluations with respect to the metrics defined in **Section 3.1** by replacing the non-specific networks with TSNs. We provide two types of comparisons; one where we compare $TSN_{0.5}$ with the original non-tissue specific Intact network and one where results of $TSN_{0.5}$ are compared against TSN_0. We do the latter to avoid artifacts that may be introduced due to the fact that some genes in the original Intact network might be simply missing from even TSN_0 since they may be nonexistent in the GTEX database. For the BLCA dataset, comparing the F1 scores of the ME methods under TSN_0 and $TSN_{0.5}$ settings, we observe that the scores of all methods are higher for the latter network. The largest percent increase of 10% is observed for WeXT when the control group is $\mathcal{X}_1$. Similarly, the largest percent increase of 12% is observed for MEMo when the control group is $\mathcal{X}_2$. On the other hand, when we compare the

scores of TSN_0 against the original network, the differences are negligible. The next largest difference between the F1 scores obtained under $TSN_{0.5}$ as compared to TSN_0 is observed in STAD where we see a 7% increase in DISCOVER's score for $\mathcal{X}_1$, and a 10% increase in WeXT's score for $\mathcal{X}_2$. For the rest of the cancer types under study, for LUSC and UCEC we observe slight increase in performances of all the ME methods comparing the metrics under $TSN_{0.5}$ against TSN_0. For COADREAD, BRCA and SKCM we observe both increases and decreases in performances but the differences are almost negligible; see **Supplementary Tables S50-S81** for detailed results.

Figure 3 compares the ROC curves of CGC gene pairs and non-CGC gene pairs for COADREAD data where mutual exclusivities are estimated with DISCOVER, DISCOVER Strat, Fisher's Exact Test, MEGSA, MEMo and WeXT with t = 20. We observe that all the ME methods estimate stronger mutual exclusivities for tissue-specific CGC gene pairs compared to non-tissue-specific CGC gene pairs since AUROCs are greater than 0.5. Additionally, we observe much smaller AUROCs for the control group where we repeat the same analysis with non-CGC gene pairs. Analogous results are available for the other cancer types where both the positive and negative set contains at least 10 number of pairs when t is set to 20. (**Supplementary Figures S33–S35**). We observe a similar result for SKCM where CGC
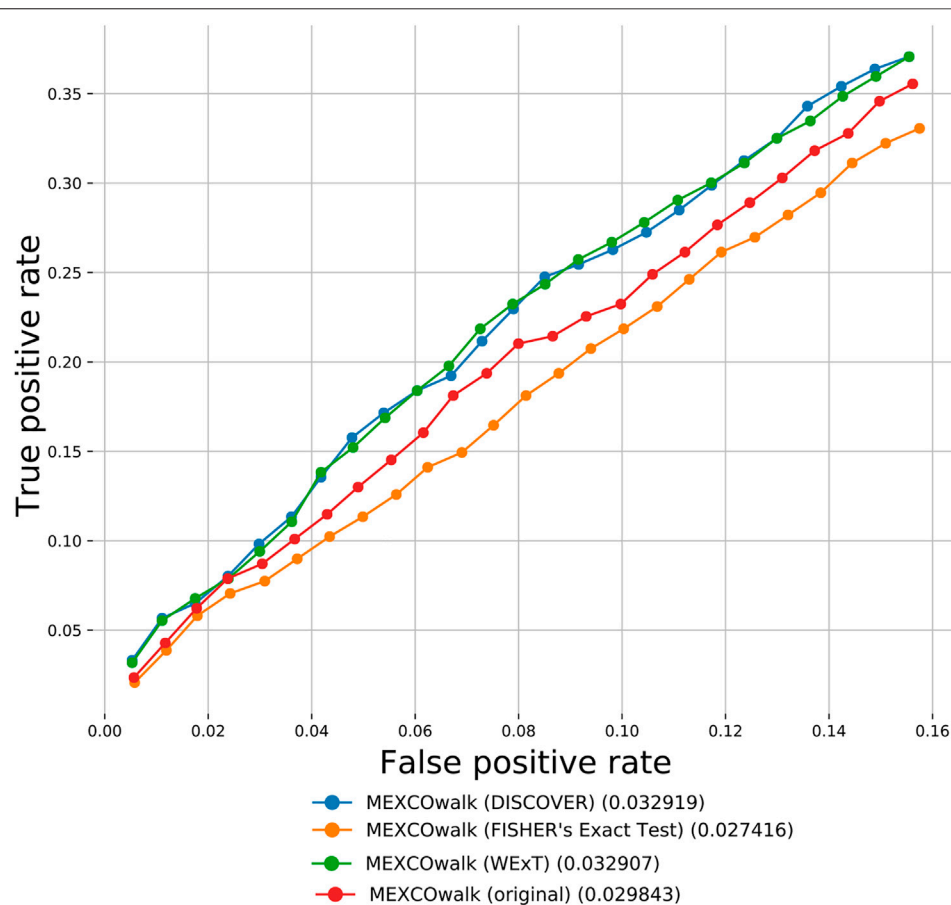
FIGURE 4 | The number of recovered CGC genes for the original MEXCOwalk as well as for its modified versions where mutual exclusivity values are estimated with DISCOVER, Fisher's Exact Test and WeXT. COADREAD dataset is used with $t = 5$ setting. The numbers in parentheses indicate the area under the ROC curve for the corresponding curve.

pairs result in larger AUROCs compared to non-CGC pairs for all ME methods (**Supplementary Figure S34**). We observe a steep increase in the ROC curves plotted for MEGSA results. This is due to the utilized likelihood ratio test that results in a p-value of 0.5 when the likelihood values are equal to each other. For UCEC, we see a significant difference between the ROC curves of CGC-pairs vs non-CGC pairs for Fisher's Exact Test and MEGSA; whereas the corresponding difference is negligible for DISCOVER and WeXT.

4 CASE STUDY

Apart from the defined network-centric ME evaluation framework, we discuss a case study where we assess whether mutual exclusivities estimated by the considered ME methods improve the performance of driver identification methods that utilize mutual exclusivity information. To this end, we compare the original version of MEXCOwalk with its alternatives where mutual exclusivity estimates are provided by the employed ME methods. Assuming that g_i and g_j genes are mutated in patient sets S_i and S_j, respectively; MEXCOWalk simply computes the mutual exclusivity between

these two genes with the following formula: $|S_i \cup S_j|/(|S_i| + |S_j|)$. MEXCOwalk uses the estimated mutual exclusivity values as part of edge weights. As such, to utilize the p-values output by ME detection methods in MEXCOwalk, we first compute $-\log(p\text{-value})$ and then convert the resulting values between 0 and 1. To this end, we replace all $-\log(p\text{-value})$'s larger than 10 with 1. We then find the maximum $-\log(p\text{-value})$ less than 10 and divide all other $-\log(p\text{-value})$'s with this value. The reason why we set a threshold for finding the maximum is the large differences across the smallest p-values output by different ME methods. For instance, WeXT outputs a very large range of p-values and if we use the smallest p-value to scale, all other $-\log(p\text{-value})$s will be converted to values that are very close to 0. In the original MEXCOWalk study, a threshold of 0.7 is applied to ME values such that all values ≤0.7 are clamped to 0. This conversion is equivalent to removing those edges from the network since the edge weights include a multiplicative term for ME values. We find that the removal of these edges correspond to a 0.035 percent reduction in graph density. For the current analysis, we determine the threshold value for each ME detection method to achieve the same percent density reduction in the graph. **Figure 4** shows the number of recovered CGC genes for fixed output gene sizes from 100 to

2,500 as a ROC curve for original MEXCOwalk as well as for versions of MEXCOwalk where mutual exclusivity values are estimated with DISCOVER, Fisher's Exact Test and WeXT, respectively. We observe that MEXCOwalk with WeXT's ME values results in the best AUROC value for COADREAD. **Supplementary Figures S36** shows the analogous results for the other cancer types. For, LUSC, STAD and UCEC, MEXCOwalk with DISCOVER gives the best AUROC whereas for BLCA, LUAD and SKCM MEXCOwalk with Fisher's Exact Test performs the best. An important observation is the worse performance of MEXCOwalk with Fisher's Exact Test compared to the original MEXCOwalk for COADREAD, STAD and UCEC. As such, using Fisher's Exact Test in place of MEXCOwalk's original ME values does have the potential to decrease the performance whereas for the other ME methods we do not observe such a risk. Note that for these analysis we employ $t = 5$ since $t = 20$ filtering does not provide enough number of genes to be evaluated.

5 DISCUSSION

It is important to investigate whether the employment of an interaction network within our ME evaluation framework causes any ascertainment bias in the findings and to elaborate on how any such potential bias is mediated within the framework. It is established that known cancer genes have larger number of interactions compared to other genes in the network (Hou and Ma, 2014a). This implies a potential bias that needs to be resolved in cancer driver gene identification methods employing interaction network data. Such a bias is less of a problem for the current study, since our aim is not to identify novel cancer driver genes but to utilize the interaction network and known cancer genes to form a ground truth of mutually exclusive interactions for evaluating existing ME methods. On the contrary, the fact that most known cancer genes have well-characterized interactions in the network provides a benefit for our work as it supports the confidence of our true positive examples. Additionally, our framework makes use of not only genes from the reference set $\mathcal{S}$ but also genes not in $\mathcal{S}$ to create random controls. Nevertheless, the fact that some known cancer genes have significantly larger number of interactions compared to other known cancer genes could lead to a bias. For instance, for our analysis of the COADREAD data ($t = 20, \mathcal{S} = CGC$), there are 74 CGC genes among which five CGC genes have more than ten CGC neighbors whereas 41 have exactly one CGC neighbor. This could lead to a bias as CGC genes with large number of CGC neighbors contribute to the aggregate statistics and metrics much more than those CGC genes with small number of CGC neighbors. To mediate this bias, our framework includes additional results where all the statistics and the traditional measures such as the F1 score are calculated in a degree-normalized way for each gene and the gene-level results are then aggregated by taking an average across the genes. These results are available in the Supplementary Document; **Supplementary Tables S12, S24, S36, S48**. To summarize, the degree-normalized results are in agreement with those of the previous settings in almost all the cases in terms of ranking based on F1 score.

Another important point worth emphasizing is that apart from the aggregate statistics provided in the previous sections as part of the metrics for the network-centric ME evaluations, our proposed framework also provides analogous statistics at the gene-level as well. Such statistics may in fact be of more interest to cancer biologists than the aggregate statistics in certain cases. Several interesting observations can be made through an inspection of these gene-level evaluations, especially for the settings where the conventionally defined F1 score fails in quantifying ME. Genes with low MLA comprise an example setting, where TP53 is a leading member. Consider the case of TP53 in COADREAD evaluations for instance. With respect to the degree-normalized setting, the values of precision, sensitivity, $precision_{strict}$ and $sensitivity_{strict}$ for WeXT are respectively 0.5, 1, 0.25, 0.25 which gives rise to an F1 score of 0.66 and $F1_{strict}$ score of 0.25. On the other hand, MEMo provides the same precision, sensitivity and F1 scores as WeXT whereas its $precision_{strict}$, $sensitivity_{strict}$ and $F1_{strict}$ scores are all 0. To summarize, although the inspection of the F1 scores does not provide a distinction between the two results, an inspection of the $F1_{strict}$ scores establishes that MEMo is worse than WeXT in this setting. We note that the advantages of inspections based on the strict definitions of the metrics rather than the conventional ones are also apparent in the aggregate analysis as well. In addition to the COADREAD evaluations shown in **Table 2**, BRCA also contains an example instance where the conventional and the strict versions of the metrics provide different conclusions; see **Supplementary Table S7B**. In terms of the F1 scores, DISCOVER Strat ranks fourth, whereas comparing $F1_{strict}$ scores it ranks the second. Also, overall we observe that MEMo's performance gets severely affected when the strict versions of the metrics are employed.

Next, our robustness analysis results reveal some suggestions for potential users of our framework. We recommend using a p-value threshold smaller than 0.1 but larger than 0.05 as lower threshold values are too stringent and lead to too few predicted positives. Regarding robustness_iterations, we tested values both smaller than and higher than the default value of 100 for COADREAD evaluations: 5, 50, 100, 300 and 500. We repeated each experiment 20 times and calculated the standard deviation of the obtained set of F1 and $F1_{strict}$ scores. For the majority of the cases, we observe a large decrease in the standard deviation values when robustness_iterations is increased from 5 to 50. (**Supplementary Table S82**). This analysis suggests that the robustness_iterations should be set to a at least 50. Lastly, we observe that different PPI networks can lead to large differences in both the $F1/F1_{strict}$ scores and the ranking of the methods. As such, exploring different PPI sources would be beneficial.

To assess whether our findings extend to other datasets other than TCGA, we repeat our evaluations on somatic mutation data of 402 colon cancer patients within the Pancancer Analysis of Whole Genomes (PCAWG) study (Campbell, 2020). **Supplementary Tables S83-S86** shows the ME evaluations with respect to the metrics defined in section 4.2. We observe an overall decrease in $F1_{strict}$ scores of the methods. Compared to analogous results in TCGA data, WexT still performs the best in terms of $F1_{strict}$ score whereas the second best performing method

is changed from MEMo to DISCOVER. The changes with respect to varying the p-value threshold, robustness_iterations value, input PPI, reference cancer gene set are consistent with the changes that we previously observe for the TCGA COADREAD dataset. When we switch from IntAct to its TSN version, we observe that all the ME methods estimate stronger mutual exclusivities for the tissue-specific CGC gene pairs compared to the non-tissue-specific CGC gene pairs as evident from AUROC values greater than 0.5; see **Supplementary Figures S37, S38** shows the results of MLA where we observe slightly smaller correlation values for DISCOVER (−0.84 vs −0.88) as compared to the results obtained from TCGA COADREAD dataset. We observe findings similar to those obtained from the TCGA COADREAD data in that the correlation values drop when the network information is incorporated. To summarize, our conclusions remain the same when we repeat our analyses on an entirely different cohort from the PCAWG study.

The majority of the somatic mutations observed in cancer genomics are passenger mutations. In the evaluations provided in the Results section we employ a simple filtering strategy where we remove silent mutations and mutations on non-coding regions of the genes. Additionally, we also assess the effects of employing a more elaborate mutation filtering procedure. To this end, we download the predictions of the Muiños et al. study on COREAD type (Muinos et al., 2021). This includes the classification of all possible mutations on 12 genes as driver or passenger mutations. Accordingly, we filter out the proposed passenger mutations from our mutation data and repeat all of our relevant analyses. We observe that the ranking of the methods according to the metrics proposed in section 4.2 remain the same where WExT, MEMO, and DISCOVER Strat show reduced $F1_{strict}$ scores, and DISCOVER and Fisher's Exact Test show higher $F1_{strict}$ scores (**Supplementary Tables S87-S90**). The TSN results and the MLA analysis results are also similar to our original results (**Supplementary Figures S39-S40**). Muiños et al. provides classifications of mutations on a subset of genes which have training data larger than a certain size. If such classifications become available for a larger set of genes in the future we can provide a better assessment regarding the filtering procedures employing these classifications.

Mutated genes in cancer prevalently exhibit a long tail phenomenon where few genes are mutated in many patients and large number of genes are mutated in few patients. To check whether assessing the mutual exclusivity of gene pairs with very different mutation frequencies bias the evaluations of the compared ME methods, we repeat our analyses after filtering out the genes with mutation frequencies < 5% and > 30%. The results after this filtering step are available in **Supplementary Tables S91-S93**. We observe that the ranking of the methods remain the same where we see a significant increase in Precision/$Precision_{strict}$ values and a slight decrease in Sensitivity/$Sensitivity_{strict}$ values. When we look at the $Precision_{strict}$ values in more detail, we observe that the FP_{strict} values drop dramatically when we apply the filtering. This suggests that the control gene pairs that include genes with very low or very high mutation frequencies can have more significant p-values as

compared to the p-values obtained for the corresponding CGC-CGC pair.

We also evaluate a more general ME detection method SELECT, which investigates both types of relationships among pairs, co-occurence and ME simultaneously. SELECT outputs ME associated scores to only a subset of the input gene pairs. Thus one strategy for comparing the results of SELECT against other methods is to focus only on such subsets. The relevant results where we use this strategy are available in **Supplementary Tables S94-S97**. We report evaluations on two subsets of TCGA COADREAD dataset: 1) the set of CGC-CGC pairs where SELECT results are available, 2) the set of CGC-CGC pairs where SELECT's version which uses subtype information (i.e., $SELECT_{subtype}$) are available. For the former, we observe that SELECT and $SELECT_{subtype}$ rank the fourth after WExT, MEMO, and DISCOVER. For the latter evaluation, $SELECT_{subtype}$ performs better than SELECT although both of them still rank the fourth among the other ME methods. Another strategy to fix this problem is to assign the worst ASC score to such pairs without specific ASC scores in the ME direction. We employ this approach as well and observe that it gives no significant difference in the comparisons.

Lastly, it is important to mention certain limitations of the proposed framework. Our framework is based on the presupposition that ME is likely to occur between interacting known cancer genes. Although rare, there may exist two different types of exceptions to this assumption; ME can be observed between non-interacting known drivers and the relationship between an interacting pair of known drivers can be that of co-occurrence rather than that of ME. These constitute respectively the false negative and the false positive events in our framework. An example instance of the former is the mutually exclusive mutations of APC and RNF43 observed in colorectal cancer [Mina et al., 2017] and example instance of the latter is the co-occurrence of CCNE1 and TP53 alterations [Zhang et al., 2014]. Both of these patterns are currently ignored by our framework and incorporation of mechanisms to dissect each such pattern to increase the performance of true ME detection is an important future step. Another limitation of the current framework is that it requires the availability of whole-genome or whole-exome sequencing data.

6 CONCLUSION

We propose a network-centric framework to evaluate pairwise mutual exclusivity findings reported by different ME algorithms. The first component of our framework consists of useful definitions of statistics employed in the network-centric ME evaluations. We observe that for the majority of the cancer types under study WeXT outperforms the other methods in terms of F1 score measured with respect to appropriately defined control groups. In half of the cancer types DISCOVER and in the other half MEMo perform as the second best methods. When comparing different cancer types we observe that BRCA and COADREAD are among the top two types leading to maximum F1 scores with at least one of the ME methods

providing a score greater than 0.5. We note that DISCOVER Strat is only applicable in two cancer types among a total of eight since these are the only cancer types with well-defined subtypes. Furthermore, among these two cancer types, DISCOVER Strat outperforms original DISCOVER algorithm in BRCA, whereas it is the second worst method after Fisher's Exact Test in COADREAD. This is noteworthy since van de Haar et al. propose subtype stratification as employed by DISCOVER Strat as a way to emphasize true mutual exclusivity by reducing mutation load confounding (van de Haar et al., 2019). We also observe that Fisher's exact test and MEGSA are more conservative compared to DISCOVER and WeXT, where from the latter group, WeXT outputs notably larger number of significant p-values. The second component of our framework evaluates ME tests by comparing two types of measures obtained with and without network information. First measure is with respect to the percent significant findings of mutually exclusive gene pairs, whereas the second is based on MLC values. In most of the cancer types and for most of the genes we observe an increase with respect to the former whereas a decrease with respect to the latter measure. Finally, we repeat the same analysis by considering TSNs in the network-centric framework. Considerable improvements achieved due to the use of TSNs as opposed tissue nonspecific interaction network are only observed for BLCA and STAD datasets. A more detailed analysis in terms of comparing ROCs of CGC gene pairs and non-CGC gene pairs on cancer types with considerable number of tissue-specific gene pairs indicate the advantages of employing tissue specificity in detecting mutual exclusivity in COADREAD, SKCM, and UCEC. Finally we extend out network-centric evaluation framework to assess whether including network knowledge reduces the mutation load confounding problem.

As noted earlier the proposed framework is intended for the network-centric evaluations of mutual exclusivities of pairs of genes rather than groups of genes. Such a choice stems form the fact that the mutual exclusivities are commonly made use of in driver gene/module identification algorithms which mostly employ pairwise mutual exclusivities. Furthermore the extensive evaluation settings proposed, the number of ME methods under study and their own computational requirements, and the potentially exponential computational complexity inherent in handling groups of genes limits the scope of the current study to evaluations of pairwise ME scorings. Nonetheless most statistical ME methods are capable of providing ME results for groups of genes as well. Regarding the ME tests considered in this study, the main ME test provided by DISCOVER is based on a pairwise test definition but it also extends the definition for possible use in quantifying the ME of a group of genes, although the experiments involving the latter are based only on simulation data. The remaining tests MEGSA, MEMo, and WeXT are all ME tests specifically designed for

groups of genes. An important direction for future work is to design a suitable extension of the proposed network-centric framework to evaluate the results of ME tests on groups of genes. Design choices relevant for such an extension would involve an appropriate and computationally efficient definition of the reference groups of genes analogous to a pair of interacting genes from the set $\mathcal{S}$ in the current setting and the definitions of control groups analogous to $\mathcal{X}_1$ and $\mathcal{X}_2$. Another future direction is to apply our network-centric framework on heterogeneous biological networks incorporating biological pathway information with PPI network data. Such incorporations have been successfully applied in other bioinformatics domains such as cancer driver identification (Hou and Ma, 2014b; Dinstag and Shamir, 2020).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.746495/full#supplementary-material

REFERENCES

Ahmed, R., Baali, I., Erten, C., Hoxha, E., and Kazan, H. (2019). MEXCOwalk: Mutual Exclusion and Coverage Based Random Walk to Identify Cancer

Modules. *Bioinformatics* 36 (3), 872–879. ISSN 1367-4803. doi:10.1093/bioinformatics/btz655

Baali, I., Erten, C., and Kazan, H. (2020). Driveways: A Method for Identifying Possibly Overlapping Driver Pathways in Cancer. *Sci. Rep.* 10. doi:10.1101/2020.04.01.015388

Babur, Ö., Gönen, M., Aksoy, B. A., Schultz, N., Ciriello, G., Sander, C., et al.B; ü; lent Arman Aksoy (2015). Systematic Identification of Cancer Driving Signaling Pathways Based on Mutual Exclusivity of Genomic Alterations. *Genome Biol.* 16 (1), 45. doi:10.1186/s13059-015-0612-6

Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* 173, 371–e18. e18. doi:10.1016/j.cell.2018.02.060

Bethune, G., Bethune, D., Ridgway, N., and Xu, Z. (2010). Epidermal Growth Factor Receptor (Egfr) in Lung Cancer: an Overview and Update. *J. Thorac. Dis.* 2, 48–51.

Campbell, G. (2020). Pan-cancer Analysis of Whole Genomes. *Nature* 578, 82–93. doi:10.1038/s41586-020-1969-6

Canisius, S., Martens, J. W. M., Wessels, L. F. A., and Martens, M. (2016). A Novel independence Test for Somatic Alterations in Cancer Shows that Biology Drives Mutual Exclusivity but Chance Explains Most Co-occurrence. *Genome Biol.* 17 (261), 1–17. ISSN 1474-760X. doi:10.1186/s13059-016-1114-x

Carson, C. C., Moschos, S. J., Edmiston, S. N., Darr, D. B., Nikolaishvili-Feinberg, N., Groben, P. A., et al. (2015). Il2 Inducible T-Cell Kinase, a Novel Therapeutic Target in Melanoma. *Clin. Cancer Res.* 21 (9), 2167–2176. doi:10.1158/1078-0432.CCR-14-1826

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* 2 (5), 401–404. ISSN 2159-8274. doi:10.1158/2159-8290.cd-12-0095

Chang, S.-W., Chao, W.-R., Ruan, A., Wang, P.-H., Lin, J.-C., and Han, C.-P. (2015). A Promising Hypothesis of C-KIT Methylation/Expression Paradox in C-KIT (+) Squamous Cell Carcinoma of Uterine Cervix ----- CTCF Transcriptional Repressor Regulates C-KIT Proto-Oncogene Expression. *Diagn. Pathol.* 10 (1). doi:10.1186/s13000-015-0438-2

Chen, L., Yang, L., Yao, L., Kuang, X.-Y., Zuo, W.-J., Li, S., et al. (2018). Characterization of Pik3ca and Pik3r1 Somatic Mutations in Chinese Breast Cancer Patients. *Nat. Commun.* 9 (1). doi:10.1038/s41467-018-03867-9

Cheng, D. T., Mitchell, T. N., Zehir, A., Shah, R. H., Benayed, R., Syed, A., et al. (2015). Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *J. Mol. Diagn.* 17 (3), 251–264. doi:10.1016/j.jmoldx.2014.12.006

Ciriello, G., Cerami, E., Sander, C., and Schultz, N. (2012). Mutual Exclusivity Analysis Identifies Oncogenic Network Modules. *Genome Res.* 22 (2), 398–406. doi:10.1101/gr.125567.111

Clark, A., and Burleson, M. (2020). Spop and Cancer: a Systematic Review. *Am. J. Cancer Res.* 10 (3), 704–726.

Constantinescu, S., Szczurek, E., Mohammadi, P., Rahnenführer, J., and Beerenwinkel, N. (2015). TiMEx: a Waiting Time Model for Mutually Exclusive Cancer Alterations. *Bioinformatics* 32 (7), 968–975. doi:10.1093/bioinformatics/btv400

Dao, P., Kim, Y.-A., Wojtowicz, D., Madan, S., Sharan, R., and Przytycka, T. M. (2017). BeWith: A Between-Within Method to Discover Relationships between Cancer Modules via Integrated Analysis of Mutual Exclusivity, Co-occurrence and Functional Interactions. *Plos Comput. Biol.* 13, e1005695. doi:10.1371/journal.pcbi.1005695

Deng, Y., Luo, S., Deng, C., Luo, T., Yin, W., Zhang, H., et al. (2017). Identifying Mutual Exclusivity across Cancer Genomes: Computational Approaches to Discover Genetic Interaction and Reveal Tumor Vulnerability. *Brief in Bionform* 20, 254–266. doi:10.1093/bib/bbx109

Dimitrakopoulos, C. M., and Beerenwinkel, N. (2017). Computational Approaches for the Identification of Cancer Genes and Pathways. *Wires Syst. Biol. Med.* 9 (1), e1364, 2017 . ISSN 1939-5094. doi:10.1002/wsbm.1364

Dinstag, G., and Shamir, R. (2020). PRODIGY: Personalized Prioritization of Driver Genes. *Bioinformatics* 36 (6), 18311367–18394803. doi:10.1093/bioinformatics/btz815

Fleming, N. I., Jorissen, R. N., Mouradov, D., Christie, M., Sakthianandeswaren, A., Palmieri, M., et al. (2013). SMAD2, SMAD3 and SMAD4 Mutations in Colorectal Cancer. *Cancer Res.* 73 (2), 7251538–7357445. ISSN 0008-5472. doi:10.1158/0008-5472.CAN-12-2706

Forbes, S. A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., et al. (2017). Cosmic: Somatic Cancer Genetics at High-Resolution. *Nucleic Acids Res.* 45, D777–D783. doi:10.1093/nar/gkw1121

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the Cbioportal. *Sci. Signal.* 6 (269), pl1, 2013 . ISSN 1945-0877. doi:10.1126/scisignal.2004088

Gayther, S. A., Batley, S. J., Linger, L., Bannister, A., Thorpe, K., Chin, S.-F., et al. (2000). Mutations Truncating the Ep300 Acetylase in Human Cancers. *Nat. Genet.* 24, 300–303. doi:10.1038/73536

GTEX Consortium (2020). The Gtex Consortium Atlas of Genetic Regulatory Effects across Human Tissues. *Science* 369 (6509), 1318–1330. ISSN 0036-8075. doi:10.1126/science.aaz1776

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., et al. (2015). The Consensus Molecular Subtypes of Colorectal Cancer. *Nat. Med.* 21, 1350–1356. doi:10.1038/nm.3967

Hou, J. P., and Ma, J. (2014a). Dawnrank: Discovering Personalized Driver Genes in Cancer. *Genome Med.* 6 (56), 16. doi:10.1186/s13073-014-0056-8

Hou, J. P., and Ma, J. (2014b). DawnRank: Discovering Personalized Driver Genes in Cancer. *Genome Med.* 6 (7), 56, 2014b . ISSN 1756-994X. doi:10.1186/s13073-014-0056-8

Hua, X., Hyland, P. L., Huang, J., Song, L., Zhu, B., Caporaso, N. E., et al. (2016). Megsa: A Powerful and Flexible Framework for Analyzing Mutual Exclusivity of Tumor Mutations. *Am. J. Hum. Genet.* 98, 442–455. doi:10.1016/j.ajhg.2015.12.021

Iwamoto, R., Mine, N., Mizushima, H., and Mekada, E. (2017). Erbb1 and Erbb4 Generate Opposing Signals Regulating Mesenchymal Cell Proliferation during Valvulogenesis. *Development* 144 (8), 2–e1. doi:10.1242/dev.152710

Jiang, Z. B., Ma, B. Q., Liu, S. G., Li, J., Yang, G. M., Hou, Y. B., et al. (2018). miR-365 Regulates Liver Cancer Stem Cells via RAC1 Pathway. *Mol. Carcinogenesis* 58 (1), 55–65. doi:10.1002/mc.22906

Kim, Y.-A., Cho, D.-Y., Dao, P., and Przytycka, T. M. (2015). MEMCover: Integrated Analysis of Mutual Exclusivity and Functional Network Reveals Dysregulated Pathways across Multiple Cancer Types. *Bioinformatics* 31 (12), i284–i292. doi:10.1093/bioinformatics/btv247

Kim, Y. A., Madan, S., and Przytycka, T. M. (2017). WeSME: Uncovering Mutual Exclusivity of Cancer Drivers and beyond. *Bioinformatics* 33 (6), 814–821. ISSN 1367-4803. doi:10.1093/bioinformatics/btw242

Leiserson, M. D. M., Blokh, D., Sharan, R., and Raphael, B. J. (2013). Simultaneous Identification of Multiple Driver Pathways in Cancer. *Plos Comput. Biol.* 9 (5), e1003054. doi:10.1371/journal.pcbi.1003054

Leiserson, M. D. M., Reyna, M. A., and Raphael, B. J. (2016). A Weighted Exact Test for Mutually Exclusive Mutations in Cancer. *Bioinformatics* 32, i736–i745. doi:10.1093/bioinformatics/btw462

Leiserson, M. D., Wu, H.-T., Vandin, F., and Raphael, B. J. (2015). Comet: A Statistical Approach to Identify Combinations of Mutually Exclusive Alterations in Cancer. *Genome Biol.* 16, 160, 2015 , ISSN 1474-7596. doi:10.1186/s13059-015-0700-7

Liu, S., Liu, J., Xie, Y., Zhai, T., Hinderer, E. W., Stromberg, A. J., et al. (2020). MEScan: a Powerful Statistical Framework for Genome-Scale Mutual Exclusivity Analysis of Cancer Mutations. *Bioinformatics* 11. doi:10.1093/bioinformatics/btaa957

Luck, K., Kim, D. K., Lambourne, L., Spirohn, K., Begg, B. E., Bian, W., et al. (2020). A Reference Map of the Human Binary Protein Interactome. *Nature* 580, 402–408. doi:10.1038/s41586-020-2188-x

Mina, M., Raynaud, F., Tavernari, D., Battistello, E., Sungalee, S., Saghafinia, S., et al. (2017). Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies. *Cancer Cell* 32 (2), 155–168. e6ISSN 1535-6108. doi:10.1016/j.ccell.2017.06.010

Muiños, F., Martínez-Jiménez, F., Pich, O., Gonzalez-Perez, A., and Lopez-Bigas, N. (2021). In Silico saturation Mutagenesis of Cancer Genes. *Nature* 596, 428–432. doi:10.1038/s41586-021-03771-1

Nielsen, T. O., Poulsen, S. S., Journe, F., Ghanem, G., and Sorensen, B. S. (2014). Her4 and its Cytoplasmic Isoforms Are Associated with Progression-free Survival of Malignant Melanoma. *Melanoma Res.* 24 (1), 88–91. doi:10.1097/cmr.0000000000000040

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., et al. (2014). The MIntAct Project-IntAct as a Common Curation Platform for 11 Molecular Interaction Databases. *Nucl. Acids Res.* 42 (D), D358–D363. ISSN 0305-1048. doi:10.1093/nar/gkt1115

Popovici, V., Budinska, E., Tejpar, S., Weinrich, S., Estrella, H., Hodgson, G., et al. (2012). Identification of a Poor-Prognosis BRAF-mutant-like Population of

Patients with colon Cancer. *Jco* 30 (12), 1288–1295. ISSN 1527-7755. doi:10.1200/JCO.2011.39.5814

Prickett, T. D., NeenaAgrawal, S., Agrawal, N. S., Wei, X., Yates, K. E., Lin, J. C., et al. (2009). Analysis of the Tyrosine Kinome in Melanoma Reveals Recurrent Mutations in Erbb4. *Nat. Genet.* 41 (10), 1127–1132. doi:10.1038/ng.438

Roth, A. D., Tejpar, S., Delorenzi, M., Yan, P., Fiocca, R., Klingbiel, D., et al. (2010). Prognostic Role of KRAS and BRAF in Stage II and III Resected Colon Cancer: Results of the Translational Study on the PETACC-3, EORTC 40993, SAKK 60-00 Trial. *Jco* 28 (3), 466–474. ISSN 1527-7755. doi:10.1200/JCO.2009.23.3452

Samowitz, W. S., Albertsen, H., Sweeney, C., Herrick, J., Caan, B. J., Anderson, K. E., et al. (2006). Association of smoking, CpG island methylator phenotype, and v600e BRAF mutations in colon cancer. *JNCI J. Natl. Cancer Inst.* 98 (23), 1731–1738. doi:10.1093/jnci/djj468

Sarto Basso, R., Hochbaum, D. S., and Vandin, F. (2019). Efficient Algorithms to Discover Alterations with Complementary Functional Association in Cancer. *PLOS Comput. Biol.* 15, e1006802. doi:10.1371/journal.pcbi.1006802

Singh, S. K., Sinha, S., Padhan, J., Jangde, N., Ray, R., and Rai, V. (2020). Myh9 Suppresses Melanoma Tumorigenesis, Metastasis and Regulates Tumor Microenvironment. *Med. Oncol.* 37 (10), 2020. doi:10.1007/s12032-020-01413-6

Skidmore, Z. L., Wagner, A. H., Lesurf, R., Campbell, K. M., Kunisaki, J., Griffith, O. L., et al. (2016). GenVisR: Genomic Visualizations in R. *Bioinformatics* 32 (19), 3012–3014. doi:10.1093/bioinformatics/btw325

Sondka, Z., Bamford, S., Cole, C. G., Ward, S. A., Dunham, I., and Forbes, S. A. (2018). The Cosmic Cancer Gene Census: Describing Genetic Dysfunction across All Human Cancers. *Nat. Rev. Cancer* 18 (11), 696–705. doi:10.1038/s41568-018-0060-1

Song, J., Peng, W., and Wang, F. (2020). An Entropy-Based Method for Identifying Mutual Exclusive Driver Genes in Cancer. *Ieee/acm Trans. Comput. Biol. Bioinf.* 17 (3), 758–768. doi:10.1109/tcbb.2019.2897931

Szczurek, E., and Beerenwinkel, N. (2014). Modeling Mutual Exclusivity of Cancer Mutations. *Plos Comput. Biol.* 10 (3), e1003503–12. doi:10.1371/journal.pcbi.1003503

Thomas, R. K., Baker, A. C., Debiasi, R. M., Winckler, W., Laframboise, T., Lin, W. M., et al. (2007). High-throughput Oncogene Mutation Profiling in Human Cancer. *Nat. Genet.* 39, 347–351. doi:10.1038/ng1975

Tokheim, C. J., Papadopoulos, N., Kinzler, K. W., Vogelstein, B., and Karchin, R. (2016). Evaluating the Evaluation of Cancer Driver Genes. *Proc. Natl. Acad. Sci. USA* 113, 14330–14335. doi:10.1073/pnas.1616440113

van de Haar, J., Canisius, S., Yu, M. K., Voest, E. E., Wessels, L. F. A., and Ideker, T. (2019). Identifying Epistasis in Cancer Genomes: A Delicate Affair. *Cell* 177, 1375–1383. doi:10.1016/j.cell.2019.05.005

Vandin, F., Upfal, E., and Raphael, B. J. (2012). De Novo discovery of Mutated Driver Pathways in Cancer. *Genome Res.* 22 (2), 375–385. doi:10.1101/gr.120477.111

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., and Kinzler, K. W. (2013). Cancer Genome Landscapes. *Science* 339 (6127), 1546–1558. doi:10.1126/science.1235122

Wang, T., Ruan, S., Zhao, X., Shi, X., Teng, H., Zhong, J., et al. (2020). Oncovar: an Integrated Database and Analysis Platform for Oncogenic Driver Variants in Cancers. *NAR* 49 (D1), D1289–D1301. doi:10.1093/nar/gkaa1033

Yeang, C.-H., McCormick, F., and Levine, A. (2008). Combinatorial Patterns of Somatic Gene Mutations in Cancer. *FASEB j.* 22, 2605–2622. doi:10.1096/fj.08-108985

Zhang, J., Wu, L.-Y., Zhang, X.-S., and Zhang, S. (2014). Discovery of Co-occurring Driver Pathways in Cancer. *BMC bioinformatics* 15, 271. doi:10.1186/1471-2105-15-271

Zhang, J., and Zhang, S. (2018). The Discovery of Mutated Driver Pathways in Cancer: Models and Algorithms. *Ieee/acm Trans. Comput. Biol. Bioinf.* 15 (3), 988–998. doi:10.1109/TCBB.2016.264096310.1109/tcbb.2016.2640963

Zhang, Z., Yang, Y., Zhou, Y., Fang, H., Yuan, M., Sasser, K., et al. (2020). A Forward Selection Algorithm to Identify Mutually Exclusive Alterations in Cancer Studies. *J. Hum. Genet.* 66, 509–518. doi:10.1038/s10038-020-00870-1