Reviews and Comments on Paper 71

Submission details
Add comment
Add new review
Request review
Revise review 2
Edit note

Paper information

Paper: Gang Fang, Michael Steinbach, Chad Myers and Vipin Kumar. Integration of Differential Gene-combination Search and Gene Set Enrichment Analysis: A General Approach
Conflict of Interest: Chad Myers
Current decision: **REJECT** (reject)

Summary of received reviews and comments

Reviews superseded by other reviews are shown in the grey color in the table. All times are GMT.

	date	PC member	subreviewer	score	confidence
Review 1	Feb 19	Jason Ernst		**-2**	3
Review 2	Feb 21	Josh Stuart		**-1**	3
Review 3	Feb 21	Haiyan Huang	Ci-Ren Jiang	**-1**	3
Comment 1	Feb 23	Eric Xing			
Comment 2	Feb 23	Haiyan Huang			
Comment 3	Feb 23	Josh Stuart			
Comment 4	Feb 24	Jason Ernst			

Review 1

PC member: Jason Ernst
Overall rating: **-2** (reject)
Confidence: **3** (high)

In Fang et al the authors propose and demonstrate a framework for extensions to the Gene Set Enrichment Analysis procedure from (Subrahmanian et al, 2005) that in addition to scoring genes based on univariate measures can also score genes based on functions defined from combination of genes. The authors argue that while it is difficult to be statistically confident of the importance of any specific gene combination, in aggregate statistics derived from gene combinations can be useful to identify relevant pre-defined gene sets to some condition. The authors have cast their framework as being very general, and while it could potentially be useful, for the instantiation of it for which they report results the benefits of it over

simpler alternatives previously applied to the same data was not demonstrated (see below).

Specific Comments:
1. Terry Speed and colleagues in
Irizarry RA, Wang C, Zhou Y, Speed TP.
Gene set enrichment analysis made simple. Stat Methods Med Res. 2009;18(6):565-7.

argued that the GSEA procedure is less statistically powerful and more complex than the standard z- and chi-square tests. Fang et al uses the original GSEA procedure as the baseline for comparison and proposes extensions making the procedure even more complex. Irizarry et al. in their analysis used some of the same datasets as Fang et al used. In comparing Fig 4 of Fang et al with Table of 1 of Irizarry, et al. for categories that both papers report FDR, it seems in many cases the FDR reported by Irizarry et al are much more significant. For instance on the Boston dataset in table 1 of Irizarry, et al. they report an FDR <0.001 for all 8 categories reported in the original GSEA procedure experiment, while Fang et al only reports one of these categories to have an FDR <0.01. The Fang et al claim of improved consistency across the three lung cancer data sets is also previously reported by Irizarry et al using their simpler approach.

2. In Figure 4 the authors should consider including the FDR for each test and category they are considering. As the authors do not report the FDR for categories that were more significant in the baseline GSEA procedure it is difficult to determine the extent to which the additional test the authors are proposing are
also producing worse FDRs for some categories. Also for the categories which were significant in the new proposed tests, but not in the original GSEA procedure, it would be useful to know the FDR in the original GSEA procedure to get a sense if
these additional categories discovered were close to the cutoff previously.

3. Related to the second point, the authors at the end of Section 4.4.1 do not advocate simply using the test statistic for which they developed to integrate all the various test statistics. Instead they suggest using multiple variants based on different combination of test statistics collectively in a rather ill-defined way. The authors were vague on which combinations they were advocating
one should use as for their four tests there were 15 possible combinations, eight combinations are shown in Figure 4, and in section 4.4.4. on multiply hypothesis testing they only consider four

Review:

of them. The authors also do not formalize
how they are adjusting the FDR when considering these multiple-test collectively. While they suggest a few of their most significant categories would withstand a multiple-test correction, the corrected FDRs for these are not reported nor do they systematically evaluate how many categories as a whole would remain significant after a correction. The authors should perhaps consider making this meta-level collective test step a formal part of their method, and then focus directly comparing the final corrected FDRs of this meta-level collective step with the other tests they are comparing with.

4. How does the computational runtime of the authors procedure compare to the original GSEA procedure? This was not discussed, but is potentially
a practical issue since gene enrichments are often run interactively and considering higher order combinations could lead to orders of magnitude increases in running time.

5. The procedure the authors describe uses a set of permutations to convert the various raw test scores for genes to p-value. The procedure for the gene score to raw test score seem to be reusing the permutations which are already used by the
main outer GSEA procedure. This re-use of permutations seems to make the outer permutations of GSEA no-longer independent and thus a gene which gets a significant p-value in the raw score to p-value procedure in one permutation would be less likely to get a high score for other outer-GSEA permutations. An alternative approach would be to have a new set of permutations for converting raw scores to p-values for each outer GSEA permutation. Perhaps this alternative would be computationally prohibitive with only minor differences, but this is an issue that is perhaps worthy of comment and/or further investigation.

6. The data sets for which the author evaluate have a relatively large number samples as compared to typical microarray experiments. As the number of samples decreases there can be many chance correlations between pairs of genes potentially making the statistics based on combinations of genes less useful. It would perhaps be worth assessing the improvement of using the combinatorial measures
over just univariate measures as a function of the number of samples in a dataset.

Minor
7. It was not clear what exactly was meant by and/or the reason for including on page 2 the phrase "partitioned into k equal parts" and

page 3 by "divided into k equal parts". If the authors are suggesting dividing all raw scores by k, it wasn't clear the point since k seemed to be a constant in the context, otherwise it seems the phrases were extraneous and the sentences they are in would have the same meaning without them.

8. In defining R_a and R_B of M_3 it was not clear how co-expressed was defined

9. In Fig 4. instead of reporting an FDR of 0 report "<X" where X is the minimum significance level

10. Fig 4. legend "non-trivially decreased FDR" - perhaps simply state what non-trivially is

11. Page 9 "are mostly insignificant" - this should be made more precise

12. Typo in calling everything table 4.3

13. Typo in equation 2 in section 2: corrB should be corr_B

PC only:

Time: Feb 19, 01:31

Review 2

PC member: Josh Stuart

Overall rating: **-1** (weak reject)

Confidence: **3** (high)

The authors propose a method to produce different differential expression scores for genes that accounts for gene combinations. The idea is that gene combinations may exhibit a pattern of expression that is more discriminative between two cellular states than any single gene alone. If such combinations exist, it could increase the power of tools like Gene Set Enrichment Analysis (GSEA) to detect informative biological patterns.

Major Comments

The main methodology here is an attempt to coerce correlation measures on gene sets into a framework that can be utilized by downstream GSEA analysis. As such, the method should be viewed as a GSEA preprocessing method which lowers the generalizability and impact of the proposed method. By definition, it is an incremental

Review:

step forward since methods for combinatorial differential expression scoring already exist. Thus, the main contribution has to then be how such scores derived from combinations are summarized to a single gene level so that a method such as GSEA can be applied. However, the methodology is completely trivial: a max is used as the aggregation method. Very little justification for this choice is provided. In fact the authors readily admit their aggregation method is simple: "Since the focus of this paper is in the overall integrative framework, we use max for simplicity." However, since the framework is also straightforward and provides a preprocessing step to GSEA, one must conclude that the methodological contributions of the work are minor at best.

The interpretation of the biological results is also very anecdotal, amounting to comparing lists of pathways found to be significantly associated with differentially expressed genes. No gold standard is available here presumably so the arguments that one method is better than the other are very subjective. Perhaps a simulation could help in this case. The overall conclusions of the work are vague.

Minor Comments

Fig 2. The authors should explain parts A and B. While its clear that they are two different gene pairs, the authors should reiterate the genes and discuss (briefly!) why each example is shown. Is (b) an example of one where even the combination is not discriminative while (a) is?

Page 3. Para 3. Last sentence. "almost-zero modification" is vague. Just give a clue about what kind of code modification is necessary to run the software.

Page 3. Eqn 3. The definition of R_A(alpha) is vague. The authors say it is the fraction of gene pairs in alpha that are coexpressed. How is this fraction computed? Is there a correlation cutoff above which is considered correlated? If so, this cutoff should be shown as an explicit parameter to this function or at least described in the methods.

Page 3: The B in corrB for Formula (2) should be a subscript.

Page 6 mentions Table 4.3. I only see Table 1 and Table 2 on page 4.

Fig 4. Presenting these results graphically would have been very helpful.

What is the running time of this pipeline? It seems pretty computationally intensive. What size combinations (k) can it handle?

PC only:

Time: Feb 21, 04:54

Review 3

PC member: Haiyan Huang

Reviewer: Ci-Ren Jiang

Overall rating: **-1** (weak reject)

Confidence: **3** (high)

Review:

The authors proposed a general approach integrating DGCS measures and GSEA to search for combinations of genes highly differentiating when individual genes are not. The challenges of using gene combination techniques with the GSEA approach were discussed. Four datasets available on the GSEA website were used to illustrate the proposed method.

However, I do not find the propose method very interesting as only heuristic procedures were introduced. Also there lacks detailed description and discussion on the used DGCS measures. For example, how to measure the co-expression of 3 genes in computing M3? This would be critical for the effectiveness of M3. Also it is not very clear why M3 is a generalization of M2 since M2 is the measure of correlation difference while M3 is the measure of fraction difference (in M3, the level of co-expression was not considered). The authors should clarify this. Basically, the impact or significance of the work does not seem strong enough to be published in ISMB.

This manuscript is also very hard to read: 1) the English needs to be smoothed out. For instance, some of the sentences read awkward. I listed two sentences below (there should be more): "A gene-combination-to-gene score summarization procedure..."; "a procedure that reduces the number of scores to be handled by GSEA to the number of genes by summarizing the scores of the gene combinations involving a particular gene in a single score ..." 2) Some figures are not very easy to be understood. For example, I only understood Figure 2 after reading page 3 (after knowing the data structure). Also there is no Table 4.3. Should it be Table 2?

Minor comments:
1. The intuition of procedure A is not given. The aggregation

functions in procedure A needs to be discussed. Is the choice sensitive to datasets?
2. The format of citation is not very consistent. The paper by Subramanian et. al sometimes is cited with authors' names.
3. Typo in the third sentences in the introduction section. (two such as's)

PC only:
Time: Feb 21, 07:47

Comment 1

By: Eric Xing
Comment: There seems to be a consensus here. Are we fine with a rejection?
Time: Feb 23, 22:08

Comment 2

By: Haiyan Huang
Comment: Yes, I am fine with the rejection.
Time: Feb 23, 22:13

Comment 3

By: Josh Stuart
Comment: Yes, rejection is fine with me too.
Time: Feb 23, 22:34

Comment 4

By: Jason Ernst
Comment: Yes, I am fine with rejection as well.
Time: Feb 24, 00:48

Add comment

Please type your comments in the text area below. You comments will only be visible to PC members having access to this paper. It will not be sent to the authors.

Add Comment