1-2017

Improved Performance of Gene Set Analysis on Genome-Wide Transcriptomics Data When Using Gene Activity State Estimates

Thomas Kamp
Dordt College

Micah Adams
Dordt College

Craig Disselkoen
Dordt College

Nathan L. Tintle
Dordt College, nathan.tintle@dordt.edu

Improved Performance of Gene Set Analysis on Genome-Wide Transcriptomics Data When Using Gene Activity State Estimates

Abstract
Gene set analysis methods continue to be a popular and powerful method of evaluating genome-wide transcriptomics data. These approach require a priori grouping of genes into biologically meaningful sets, and then conducting downstream analyses at the set (instead of gene) level of analysis. Gene set analysis methods have been shown to yield more powerful statistical conclusions than single-gene analyses due to both reduced multiple testing penalties and potentially larger observed effects due to the aggregation of effects across multiple genes in the set. Traditionally, gene set analysis methods have been applied directly to normalized, log-transformed, transcriptomics data. Recently, efforts have been made to transform transcriptomics data to scales yielding more biologically interpretable results. For example, recently proposed models transform log-transformed transcriptomics data to a confidence metric (ranging between 0 and 100%) that a gene is active (roughly speaking, that the gene product is part of an active cellular mechanism). In this manuscript, we demonstrate, on both real and simulated transcriptomics data, that tests for differential expression between sets of genes using are typically more powerful when using gene activity state estimates as opposed to log-transformed gene expression data. Our analysis suggests further exploration of techniques to transform transcriptomics data to meaningful quantities for improved downstream inference.

Keywords
data sets, gene expression, transcriptomics

Disciplines
Genetics and Genomics

Comments
Presented at the 2017 Pacific Symposium on Biocomputing, held on the Big Island of Hawaii, January 3-7, 2017.

HHS Public Access

Author manuscript

Pac Symp Biocomput. Author manuscript; available in PMC 2016 December 13.

Published in final edited form as:

Pac Symp Biocomput. 2016 ; 22: 449–460.

IMPROVED PERFORMANCE OF GENE SET ANALYSIS ON GENOME-WIDE TRANSCRIPTOMICS DATA WHEN USING GENE ACTIVITY STATE ESTIMATES

Thomas Kamp,

Department of Mathematics, Statistics, and Computer Science, Dordt College Sioux Center, IA 51250, USA

Micah Adams,

Department of Mathematics, Statistics, and Computer Science, Dordt College Sioux Center, IA 51250, USA

Craig Disselkoen, and

Department of Mathematics, Statistics, and Computer Science, Dordt College Sioux Center, IA 51250, USA

Nathan Tintle

Department of Mathematics, Statistics, and Computer Science, Dordt College Sioux Center, IA 51250, USA

Thomas Kamp: Thomas.Kamp@dordt.edu; Micah Adams: Micah.Adams@dordt.edu; Craig Disselkoen: Craig.Disselkoen@dordt.edu; Nathan Tintle: Nathan.Tintle@dordt.edu

Abstract

Gene set analysis methods continue to be a popular and powerful method of evaluating genome-wide transcriptomics data. These approach require *a priori* grouping of genes into biologically meaningful sets, and then conducting downstream analyses at the set (instead of gene) level of analysis. Gene set analysis methods have been shown to yield more powerful statistical conclusions than single-gene analyses due to both reduced multiple testing penalties and potentially larger observed effects due to the aggregation of effects across multiple genes in the set. Traditionally, gene set analysis methods have been applied directly to normalized, log-transformed, transcriptomics data. Recently, efforts have been made to transform transcriptomics data to scales yielding more biologically interpretable results. For example, recently proposed models transform log-transformed transcriptomics data to a confidence metric (ranging between 0 and 100%) that a gene is active (roughly speaking, that the gene product is part of an active cellular mechanism). In this manuscript, we demonstrate, on both real and simulated transcriptomics data, that tests for differential expression between sets of genes using are typically more powerful when using gene activity state estimates as opposed to log-transformed gene expression data. Our analysis suggests further exploration of techniques to transform transcriptomics data to meaningful quantities for improved downstream inference.

1. Introduction

Gene set analysis methods are a popular approach to assessing statistical significance on *a priori*, biologically defined sets of genes, as opposed to on a gene by gene basis [1]. These approaches have now been widely applied to SNP and RNA microarrays, and, more recently, RNA and DNA sequencing. The hope and promise of these methods is a combination of both statistical and biological improvements. Statistically, by analyzing sets of genes, instead of each gene individually, multiple testing penalties can be reduced. Furthermore, by potentially aggregating multiple independent effects (in different genes in the set), the true signal may more easily rise above the 'noise' of other genes in the set. Both reduced multiple testing penalties and aggregated effects have the potential to improve the statistical power of gene set tests. Biologically, by defining gene sets using *a priori* defined sets of genes, there is the increased potential for testing specific and more complex biological hypotheses (e.g., defining a set of genes as all genes in a pathway).

Previously, we discussed application of gene set analysis methods to testing for differential levels of gene expression in a genome-wide transcriptomics setting for bacteria [2]. In particular, we evaluated the performance of novel methods of testing for differential gene expression finding that the novel methods often outperformed, other popular methods, like Fisher's Exact Test (FET) [3]. These novel methods of testing for differential gene expression between two experiments (or bacterial strains) utilize the entire vector of normalized gene expression values for all genes in the set, instead of first defining an arbitrary cutoff (as is the case in FET). By leveraging the entire vector of expression values, instead of suffering from the information loss due to defining an arbitrary cutoff, the methods are generally more powerful than FET.

While gene set analysis typically focus on analyzing 'raw' gene expression data, many current approaches to understanding genome-wide transcriptomics data attempt to further leverage the data by classifying genes into one of two states: *active* (roughly speaking, the gene product is part of an active cellular mechanism) or *inactive* (the cellular mechanism is not active) [4]–[6]. We label this classification a determination of the *gene activity state*. Recently, we published a novel approach, *MultiMM* [7], to address documented deficiencies in many of the current state of the art methods. *MultiMM* is a parametric Bayesian mixture modelling approach which addresses limitations in existing methods as demonstrated through a rigorously grounded statistical framework, better performance than existing methods on simulated and real transcriptomics data, and through improved consistency with well-accepted biological realities and fluxomics data. Full details of, and links to, software for the *MultiMM* method are available elsewhere [7]. Ultimately, the *MultiMM* method yields a confidence estimate, $a_{ij} \in [0,1]$, that gene i is active in condition j. One stated goal of the *MultiMM* method is to improve inference in downstream interpretations of gene expression data.

In this manuscript we consider the performance of a variety of gene set analysis methods on both raw gene expression data, as well as on a_{ij} values (confidence estimates that gene i, is active in experiment j) in order to determine if a_{ij} values are advantageous for use when conducting gene set analysis.

2. Methods

2.1. Methods of gene set testing

We consider three broad classes of gene set analysis methods [2], [3], [8].

First, we consider the burden test type of gene set testing method, with test statistic defined as:

$$B_m = \left| \sum_{i=1}^{k} e_{ij_1}^m - \sum_{i=1}^{k} e_{ij_2}^m \right|^{\frac{1}{m}} \quad (1)$$

Where e_{ij} is the expression value of the i^{th} gene measured in the j^{th} condition, m is a positive constant (including infinity), and k is the number of genes in the set. As is discussed elsewhere [8], the Burden (B_m) test class of methods of conducting gene set analysis assumes that the effects of the genes within the test will tend to be in the same direction. For example, all genes in the set of interest are either not changing in underlying expression values, or are increasing, but none are decreasing. In the framework of 'activity states' this means that all genes are either moving from inactive to active (across the two experiments being compared) or are in the same state in both experiments. When this assumption is not met, Burden tests tend to be low powered since effects 'cancel out.' As m increases, increasing weight is put on the most expressed genes, such that if m=∞, $\sum_{i=1}^{k} e_{ij_1}^m = \text{argmax}(e_{ij_1})$.

The Variance Components class of test methods was envisioned primarily in response to the fact that Burden tests could not appropriately handle changes in multiple directions within the same set of genes (e.g., some genes move from inactive to active and others from active to inactive when comparing two experiments) [9]. The general form of a Variance Components gene set test statistic, VC_m, is given as:

$$\text{VC}_m = \left(\sum_{i=1}^{k} |e_{ij_1} - e_{ij_2}|^m \right)^{\frac{1}{m}}$$

Similar to the behavior for Burden tests, Variance components tests put increasing weight on pairwise differences in expression values as m increases, such that when $m=\infty$, the VC statistic takes the value of the largest observed pairwise difference in expression values.

The third class of tests we considered was Fisher's Exact Test (FET). In this approach, an arbitrary cutoff, c, is first chosen, such that if $|e_{ij_1} - e_{ij_2}| > c$, then the gene is coded '1' (changing state; differentially expressed) and otherwise is coded '0' (not changing state; not differentially expressed). The proportion of genes in the set of interest which are deemed to be differentially expressed ($>c$) is compared to the proportion of genes not in the set of interest which are deemed to be differentially expressed using Fisher's Exact test, which uses a hypergeometric distribution to assess statistical significance.

2.2. Implementation of methods of gene set testing

In this manuscript we consider nine different tests, applied to both raw expression data (e_{ij}) and gene activity state estimates (a_{ij}; see next section for details). The nine tests are B_1, B_2, B_∞, VC_1, VC_2, VC_∞, $FET(1SD)$, $FET(2SD)$ *and* $FET(3SD)$. The test statistic equations for B and VC are given in the previous section, along with a description of the FET approach. For the FET approach, we use 1SD, 2SD and 3SD to denote how determine a cutoff value, c. In short, we find the average within gene SD across genes and experiments for which data is available, and then use that value (1SD), 2 times that value (2SD) or 3 times that value (3SD) to determine the cutoffs. For e_{ij} $1SD = 0.75$ and, for a_{ij}, 1SD=0.3. FET determines statistical significance using the hypergeometric distributions. All other tests are evaluated for statistical significance by comparing the observed statistic to a null distribution of 10,000 randomly generated statistics obtained by randomly choosing 10,000 sets of the same size as the gene set being evaluated and finding the fraction of randomly chosen sets with larger statistics than observed (the p-value).

2.3. Moving from raw expression values to estimates of gene activity states

The *MultiMM* algorithm takes as input a genome-wide matrix of transcriptomics data $\boldsymbol{E}$ across numerous experimental conditions, such that the entries in $\boldsymbol{E}$ are denoted e_{ij} and represent the estimated gene expression of gene i in condition j. Additionally, if available, *MultiMM* allows for *a priori* identification of sets of genes which are known to be co-regulated such that in the same experimental condition, the co-regulated genes are all active or all inactive. The *MultiMM* algorithm starts by using the Bayesian Information Criterion (*BIC*) to assess the fit of a 1-component (univariate or multivariate) Gaussian mixture distribution (gene is always active or inactive in the set of conditions represented) vs. a 2-component mixture distribution (gene is sometimes active and sometime inactive in the set of conditions represented) using the *R* package *Mclust* [10]. Following Raftery et al. [11] we require the BIC to be at least 12 points better for the 1-component model to be chosen vs. the 2-component model. Second, for all genes estimated to come from a 2-component mixture distribution, a Gaussian mixture model is fit and a Gibbs sampler is used in order to yield estimates of the means and standard deviations of the components of the mixture model, along with an estimate of the proportion of experiments for which the gene is active. In the case of co-regulated sets of genes this mixture model is multivariate, whereas for genes that are not known to be co-regulated with other genes, the mixture model is univariate. Finally, the estimated mixture distribution parameters can be used to yield a confidence estimate, $a_{ij} \in [0,1]$, that gene i is active in condition j. For genes inferred as being always active or always inactive in the dataset in step one of the algorithm, multiple imputation is used to impute a_{ij} values. Full details of, and links to, software for the *MultiMM* method are available elsewhere [7].

2.4. Simulation of gene expression data

We simulated expression data with ‘known’ gene activity states (active/inactive). The simulation of expression data was informed by the *E. coli* expression data described later. We first ran the Screening Method described above (BIC with MClust) and dropped all operons (co-regulated gene sets), including single gene operons, for which the two-

component model did not yield the highest BIC (n=697 dropped). We then randomly selected 26.3% (=697/2648) of the remaining 1951 operons to be single component in the simulated data, with each of the single component operons having an equal likelihood of being always active or always inactive.

To calculate the mixing parameter, $\boldsymbol{\pi}$, used in the simulation for the 1438 two-component operons we chose a random value for $\boldsymbol{\pi}$ between 0.2 and 0.8. Values for $\vec{\mu}_0$, $\vec{\mu}_1$, Σ_0=Σ_1 are all as estimated by the *MultiMM* method computed on the real expression data. To generate simulated expression values, ε_{ij}^{s}, we drew 907($\boldsymbol{\pi}_i$) random values from a multivariate normal distribution ($\vec{\mu}_{1i}$, Σ_{1i}) and 907(1 − $\boldsymbol{\pi}_i$) random values from a multivariate normal distribution ($\vec{\mu}_{0i}$, Σ_{0i}). Thus, we generated a 907 by 3435 matrix of ε_{ij}^{s} values. Prior analysis has shown this simulated data to have good properties and behave in reasonable ways [7].

2.5. Simulation of gene sets for analysis

We used the simulated gene expression data described above to generate random sets of genes for evaluation of different methods of gene set analysis. We selected random sets of 8, 20 or 40 genes from among genes which were not changing or changing states between the two experiments of interest. In particular, we looked at the following proportions of genes in the set which were not changing state (0, 25, 50, 75 and 100%), and either 0%, 50% or 100% of the genes in the set active in the first experiment. Thus, we explored 45 simulation settings (3 (set size) by 5 (not changing) by 3 (starting state). Of these 45 simulation settings, 9 represent settings for which we can evaluate the empirical type I rate and 36 will be used to evaluate statistical power. Each of the nine test statistics is computed for the set, and then each of the nine statistics is compared to a distribution of the same statistic across 10,000 randomly selected sets of the same size (an approach termed ‘gene sampling’ which uses a ‘competitive null hypothesis’[12]). We considered 1000 randomly selected sets at each of the 45 simulation settings. Full simulation results are available in Supplemental File #1. We also analyzed 574 *a priori* defined operon (co-regulated) sets based on operon definitions for *E. coli* as provided by Microbes Online [13]. Full results are available in Supplemental File #2. Supplemental Files are available at: http://homepages.dordt.edu/ntintle/gsa_supp.zip

2.6. Real data

We also used genome-wide gene expression data from 907 different microarray data sets collected on 4329 *Escherichia coli* genes via the M3D data repository [14]–[16] both to inform simulated data analysis and when considering the actual performance of the methods. Raw data from Affymetrix [17] CEL files were normalized using RMA [18]. Details of data processing are described elsewhere [19], [20].

2.7. Statistical analysis

Empirical power and type I error rate estimates are computed as the proportion of times that the p-value was less than the significance level for a particular test and simulation setting. We considered significance levels of 5%, 0.5% and 0.05%.

Results

Across 36 simulation settings where at least one gene in the set changed activity states, power was consistently better when using gene activity state estimates, than raw expression data (see Table 1 for overall summary). Across the 9 simulation settings where none of the genes in the set changed state (type I error setting), the Type I error rate was generally controlled for all methods (detailed results not shown). Table 1 shows that gains in power can be high across all methods, whereas when power is worse when using activity states, the reduction in power is usually quite minimal (19 to 82 average percentage point increase vs. 0.3 to 2.3 average percentage point decrease).

For each of the thirty-six simulation settings used to estimate power, the power was always highest across all 18 methods (nine different test statistics using either e_{ij} or a_{ij}) for a method using gene activity state estimates. This was true for each of the 3 different significant levels. VC_{∞} was frequently the most powerful approach (16 out of 36 times for significance level 5%; 26 out of 36 times for significance level 0.5% and 33 times for significance level 0.05%). While other B and VC methods were periodically most powerful, notably, the FET methods were never the most powerful, even when using gene activity state estimates (a_{ij}).

Figure 1 illustrates typical performance of the VC methods as the proportion of genes in the set changes, by highlighting the performance of the methods on sets of size 8. VC_{∞} is most robust to lower proportions of genes in the set changing state, while all methods perform well when the proportion of genes in the set changing state is relatively large.

Analysis of the 574 real, operon based sets of genes showed similar performance to the randomly generated gene sets, with even better performance of the activity state informed methods in many cases (detailed results not shown).

Real data example

The L-arabinose (*ara*) operon is a well-studied set of three co-located genes (*araB, araA, araD*) which encode enzymes needed for the catabolism of arabinose in *E. coli* [52]. Across the 907 experiments in our dataset, L-arabinose is present in the media in 227 cases. We randomly selected 1000 pairs of experiments where one experiment had L-arabinose present in the media and one experiment did not. We then computed different gene set analysis test statistics for the L-arabinose operon using both raw expression data and activity state estimates, as compared to 100,000 randomly selected sets of 3 genes. Table 2 illustrates that methods using activity state estimates were always more powerful than methods which were based on raw expression values.

4. Discussion

Gene set analysis remains a statistically promising and biological relevant approach to the analysis of genome-wide transcriptomics data. Here we demonstrate that, in line with previous work [2], methods which don't arbitrarily introduce a cutoff and lose information, are generally more powerful than methods that do (e.g., Fisher's exact test). We also demonstrate that using a more statistically grounded metric to quantify gene expression

(activity state estimates, a_{ij}) generally leads to more powerful tests than using raw gene expression data (e_{ij}) on simulated data, with promising results also observed on real data in well-understood biological systems.

We note that the VC_{∞} method performed particularly well, especially at low significance thresholds. This finding reflects the use of gene-sampling (a competitive null hypothesis). Briefly, when using gene sampling to assess statistical significance, test statistics generated for the gene set of interest, are compared to randomly chosen gene sets. The VC_{∞} method performs relatively better as compared to other methods as the significance level decreases because it is focused on the most extreme observed difference in activity state estimates and, thus, is more robust than other methods to small numbers of randomly selected sets of genes with extreme values of the test statistic. This performance was particularly notable in the example with the L-arabinose operon, where the VC_{∞} method using activity state estimates (a_{ij}) outperformed its performance on raw expression values (e_{ij}) by nearly 100%. While other test statistics did not show as large of a difference, in all cases the power was higher when using activity state estimates. Thus, when attempting to determine if sets of genes are differentially active in two conditions, inferring gene activity state estimates prior to applying gene set analysis methods will maximize the likelihood of identifying differential activity. In short, use of these methods will maximize our ability to identify sets of genes associated with differential activity between two conditions.

We note numerous opportunities for future work, including (1) the ability to expand these methods to incorporate information from multiple, similar experimental conditions, instead of only comparing two conditions, (2) integrating directionality and/or gene set topology, (3) potential improvements by further leveraging the statistical properties of well-calibrated a_{ij} (the posterior likelihood that gene i is active in gene j), (4) potential further improvements in power by using non-competitive null hypotheses, which may be possible through statistical quantification of the null distributions of particular methods when using well-calibrated a_{ij}'s and (5) use of this general framework to test for whether a set of genes in a single experiment shows evidence of significant 'activity' (vs. only a change in activity levels between two experiments, as we considered here).

The most notable limitation of our analysis here is the limited application to real data, though initial results are promising and performance on real (operon-based sets) was also quite encouraging. Further work is necessary to ensure transferability of these promising initial findings to additional organisms. For example, to determine if these methods will successfully distinguish sets of differentially active genes between diseased and non-diseased tissue. Furthermore, further work is necessary to explore validation in other well-understood biological systems and as compared to the results of other –omics data (e.g., genome-scale metabolic models; fluxomics, etc.).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by NSF MCB-1330734. We gratefully acknowledge the use of the Silicon Mechanics grant funded beaker computer cluster on the campus of Dordt College for computations.

References

1. de Leeuw C, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. Nat. Rev. Genet. 2016; 17:353–364. [PubMed: 27070863]
2. Tintle NL, Best AA, DeJongh M, Van Bruggen D, Heffron F, Porwollik S, Taylor RC. Gene set analyses for interpreting microarray experiments on prokaryotic organisms. BMC Bioinformatics. 2008; 9(1):469. [PubMed: 18986519]
3. Khatri P, Drăghici S. Ontological analysis of gene expression data: Current tools, limitations, and open problems. Bioinformatics. 2005; 21(18):3587–3595. [PubMed: 15994189]
4. Abel S, Bucher T, Nicollier M, Hug I, Kaever V, Abel zur Wiesch P, Jenal U. Bimodal Distribution of the Second Messenger c-di-GMP Controls Cell Fate and Asymmetry during the Caulobacter Cell Cycle. PLoS Genet. 2013; 9(9):e1003744. [PubMed: 24039597]
5. Gallo CA, Cecchini RL, Carballido JA, Micheletto S, Ponzoni I. Discretization of gene expression data revised. Brief. Bioinform. 2015 May.:1–13.
6. Ferrell JE. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. Curr. Opin. Cell Biol. 2002; 14(2):140–148. [PubMed: 11891111]
7. Disselkoen C, Greco B, Cook K, Koch K, Lerebours R, Viss C, Cape J, Held E, Ashenafi Y, Fischer K, Acosta A, Cunningham M, Best AA, DeJongh M, Tintle NL. A Bayesian framework for the classification of microbial gene activity states. Front. Microbiol. 2016; 7:1191. [PubMed: 27555837]
8. Liu K, Fast S, Zawistowski M, Tintle NL. A geometric framework for evaluating rare variant tests of association. Genet. Epidemiolgoy. 2013; 37(4):712–722.
9. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. 2011 Jul; 89(1):82–93. [PubMed: 21737059]
10. Fraley C, Raftery A, Scurcca L, Murphy TB, Fop M. mclust: Normal mixutre modelling for model-based clustering, classification and density estimation. CRAN. 2015 [Online]. Available: https://cran.r-project.org/web/packages/mclust/index.html.
11. Raftery A. Bayesian model selection in social research. Sociol. Methods. 1995; 25:111–163.
12. Goeman J, Buhlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. Bioinformatics. 2007; 23(8):980–987. [PubMed: 17303618]
13. Price MN, Huang KH, Alm EJ, Arkin AP. A novel method for accurate operon predictions in all sequenced prokaryotes. Nucleic Acids Res. 2005 Jan; 33(3):880–892. [PubMed: 15701760]
14. Many Microbes Database. [Online]. Available: http://m3d.mssm.edu
15. Faith JJ, Driscoll ME, Fusaro Va, Cosgrove EJ, Hayete B, Juhn FS, Schneider SJ, Gardner TS. Many Microbe Microarrays Database: uniformly normalized Affymetrix compendia with structured experimental metadata. Nucleic Acids Res. 2008; 36(Database issue):D866–D870. [PubMed: 17932051]
16. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol. 2007 Jan.5(1):e8. [PubMed: 17214507]
17. Affymetrix. [Online]. Available: http://www.affymetrix.com
18. Irizarry T, Bolstad Ra, Benjamin Collin, Francois Cope, Leslie Hobbs, Bridget Speed. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 2003 Feb; 31(4):15e–15e.
19. Powers S, DeJongh M, Best Aa, Tintle NL. Cautions about the reliability of pairwise gene correlations based on expression data. Front. Microbiol. 2015 Jan.6:650. no. June. [PubMed: 26167162]

20. Tintle N, Sitarik A, Boerema B, Young K, Best A, De Jongh M. Evaluating the consistency of gene sets used in the analysis of bacterial gene expression data. BMC Bioinformatics. 2012 Jan.13(1): 193. [PubMed: 22873695]

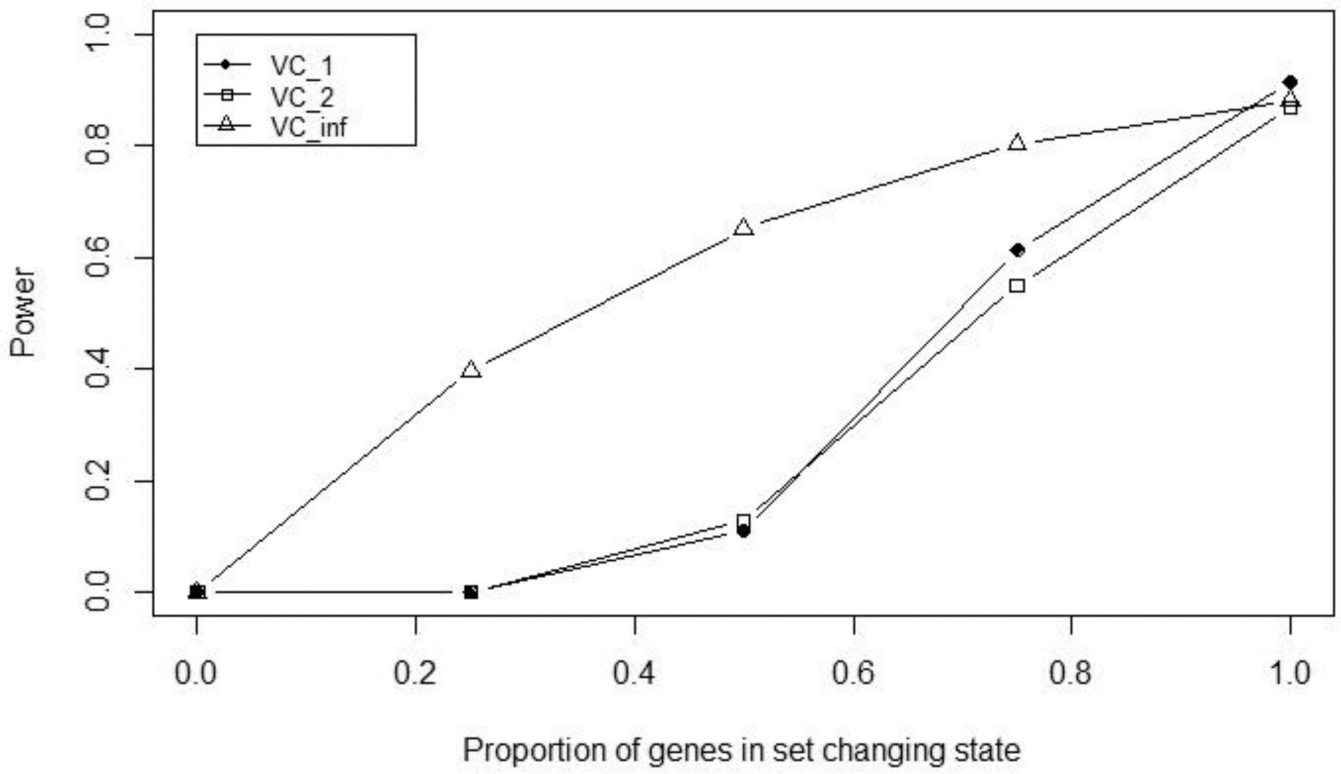

Figure 1.
Power of different VC tests as the proportion of genes in the set changing state varies

Table 1

Power improvements comparing raw expression data to gene activity state estimates using a variety of gene set analysis approaches

Gene set analysis approach		Proportion of 36 simulation settings where power is better using a_{ij}	Average (SD) power gain when power is better using a_{ij}[1]	Proportion of 36 simulation settings where power is the same using a_{ij}	Proportion of 36 simulation settings where power is worse using a_{ij}	Average (SD) power loss when power is worse using a_{ij}[2]
Fisher's exact test	Cutoff=3SD	73.1%	24.9% (21.8%)	17.6%	9.3%	0.3% (0.2%)
	Cutoff=2SD	63.9%	28.4% (20.5%)	16.7%	19.4%	0.7% (0.9%)
	Cutoff=1SD	66.7%	25.2% (21.0%	16.7%	16.7%	0.9% (0.8%)
Burden	m=1	48.1%	19.1% (16.4%)	29.6%	22.2%	1.4% (1.1%)
	m=2	46.3%	22.1% (17.7%)	25.0%	28.7%	1.2% (1.3%)
	m=∞	55.6%	39.0% (29.2%)	10.2%	34.3%	2.3% (2.9%)
Variance components	m=1	61.1%	28.9% (20.6%)	19.4%	19.4%	0.8% (0.6%)
	m=2	64.8%	40.4% (28.0%)	10.2%	25.0%	1.0% (1.1%)
	m=∞	100%	82.2% (17.4%)	0	0	-

[1] In situations when the power is better using a_{ij} vs. e_{ij}, what is the difference in power estimates between the two different methods. For example, for VC_{∞} the difference power between using a_{ij} and e_{ij} averaged 82.2% percentage points, reflecting the fact that VC_{∞} is substantially better when using a_{ij}

[2] In situations when the power is worse using a_{ij} vs. e_{ij}, what is the difference in power estimates between the two different methods. For example, for B_{∞} the difference power between using a_{ij} and e_{ij} averaged 2.3% percentage points, reflecting the fact that B_{∞} is not much worse using a_{ij} and e_{ij} in the 34.3% of cases when it is worse

Table 2

Empirical power estimates for detecting significant changes in activity for the L-arabinose operon in *E. coli* when comparing an experiment with L-arabinose present in the media vs. one without

Sig. Level	Method	B_1	B_2	B_∞	VC_1	VC_2	VC_∞
0.05%	Raw expression (e_{ij})	96.6%	98.1%	1.6%	95.7%	52.3%	1.7%
	Activity state estimates (a_{fj})	100%	100%	99.6%	100%	100%	99.6%
0.005%	Raw expression (e_{ij})	85.3%	86.1%	0%	58.0%	3.9%	0%
	Activity state estimates (a_{fj})	99.6%	99.6%	99.6%	99.6%	99.6%	99.6%