

Supplementary materials

Gene Set Analysis: Limitations in popular existing methods and proposed improvements

Pashupati Mishra, Petri Törönen, Yrjö Leino, Liisa Holm

Contents

1	Extreme value distributions	2
1.1	Extreme value type I distribution	2
1.2	Extreme value type II distribution	3
1.3	Extreme value type III distribution	3
1.4	General extreme value distribution	3
2	Series expansion for p-value calculation from <i>EVD</i> and <i>GEVD</i>	4
3	Schematic representation of mGSZ	5
4	Evaluation of the p-value calculation methods with p53 data	6
4.1	Asymptotic and empirical p-values for mGSZ scores	7
4.2	Asymptotic and empirical p-values for mGSA scores	8
4.3	Asymptotic and empirical p-values for mAllez scores	9
4.4	Asymptotic and empirical p-values for SUM scores	10
5	Evaluation of the p-value calculation methods with gender data	10
5.1	Asymptotic and empirical p-values for mGSZ scores	10
5.2	Asymptotic and empirical p-values for mGSA scores	11
5.3	Asymptotic and empirical p-values for mAllez scores	12
5.4	Asymptotic and empirical p-values for SUM scores	12
5.5	Asymptotic approximation of empirical null distribution with gender data .	12
6	Comparison of stability of gene set scores with gender data	14
7	Calculation of prior variance in Gene Set Z-score function	15

8	Similarity of the GSZ and GSA scoring functions	16
9	Log p-values of the top gene sets	17
9.1	Log p-values of the top gene sets in p53 dataset	17
9.2	Log p-values of the top gene sets in gender dataset	17
10	False positive signal test with p53 data	18
11	Comparison of mGSZ, mGSA and mAllez with program packages	18
12	Evaluation of mGSZ on dataset with small sample size	20
13	Top gene sets identified by standard Allez and mAllez	22
14	Top gene sets identified by standard GSA and mGSA	23
15	List of gene sets that are strongly relevant to p53 activity	23
16	List of gene sets that are strongly relevant to gender	25

1 Extreme value distributions

Extreme value distributions are limiting distributions for extreme values (maximums or minimums) of a sample of independent and identically distributed (i.i.d) random variables as the sample size increases (Kotz and Nadarajah, 2000). The extreme value distributions can be based either on the smallest extreme or the largest extreme. We used extreme value distributions to model mGSZ and mGSA scores. Both mGSZ and mGSA scores are maximum values from a group of data. For that reason, we will focus mainly on the extreme value distribution based on the largest extreme. The class of Extreme Value Distributions involves three types of extreme value distributions, type I, type II and type III.

1.1 Extreme value type I distribution

The extreme value type I distribution is also referred to as the Gumbel distribution. The probability density function of the extreme value type I distribution is,

$$f(x) = \frac{1}{\beta} e^{-\frac{(x-\mu)}{\beta}} e^{-e^{-\frac{(x-\mu)}{\beta}}} \quad (1)$$

The cumulative distribution function of the extreme value type I distribution is,

$$F(x) = e^{-e^{-(x-\mu)/\beta}} \quad (2)$$

where, μ is the location parameter and $\beta > 0$ is the scale parameter. Standard extreme value type I distribution has $\mu = 0$ and $\beta = 1$.

1.2 Extreme value type II distribution

Extreme value type II distribution is also referred as to Fréchet distribution. The probability density function of the extreme value type II distribution is,

$$f(x) = \frac{a}{\beta} \left(\frac{x - \mu}{\beta} \right)^{-1-a} e^{-\frac{x-\mu}{\beta}-a} \quad (3)$$

The cumulative distribution function of the extreme value type II distribution is,

$$F(x) = e^{-\frac{x-\mu}{\beta}-a} \quad (4)$$

where, μ is the location parameter, β is the scale parameter and a is the shape parameter.

1.3 Extreme value type III distribution

Extreme value type III distribution is also called as Weibull distribution. The probability density function of the extreme value type III distribution is,

$$f(x) = \begin{cases} \frac{a}{\beta} \left(\frac{x}{\beta} \right)^{a-1} e^{-(x/\beta)^a} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

The cumulative distribution function of the extreme value type III distribution is,

$$F(x) = \begin{cases} 1 - e^{-(x/\beta)^a} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where, $\mu > 0$ is the location parameter, $\beta > 0$ is the scale parameter and a is the shape parameter.

1.4 General extreme value distribution

General extreme value distribution (GEVD) combines the extreme value type I, II and III distributions. The probability density function of the GEVD is,

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-(1+az)^{-\frac{1}{a}}}(1+az)^{-1-1/a} & a \neq 0 \\ \frac{1}{\beta} e^{-z - e(-z)} & a = 0 \end{cases}$$

The cumulative distribution function is,

$$F(x) = \begin{cases} e^{-(1+az)^{-1/a}} & a \neq 0 \\ e^{-e(-z)} & a = 0 \end{cases}$$

where, $z = (x - \mu)/\beta$ and a, β, μ are shape, scale and location parameters respectively. The scale must be positive, the shape and location can take on any real value.

2 Series expansion for p-value calculation from *EVD* and *GEVD*

Log p-values from EVD and GEVD are calculated as,

$$F(x) = -\log(p\text{-value}_{EVD}) = -\ln(1 - e^{-e^x}) \quad (5)$$

where,

$$x = -(z - \mu)/\beta \quad (6)$$

z is the absolute mGSZ or mGSA score value for the analyzed gene set and μ and β are location and scale parameters respectively for *EVD*.

and

$$G(x) = -\log(p\text{-value}_{GEVD}) = -\ln\left(1 - e^{-(1+ax)^{-1/a}}\right), a > 0 \quad (7)$$

where,

$$x = (z - \mu)/\beta \quad (8)$$

z is the absolute mGSZ or mGSA score value for the analyzed gene set and μ , β and a are the location, scale and shape parameters respectively for *GEVD*.

We used two well known power series expansions for calculation of extremely small p-values from *EVD* and *GEVD*, namely

$$e^{-t} = 1 - t + t^2/2 - t^3/6 + t^4/24 + \dots \quad (9)$$

and

$$\ln(1 + t) = t - t^2/2 + t^3/3 - t^4/4 + \dots \quad (10)$$

Both expansions converge rapidly when $|t|$ is reasonably small.

Now, $F(x)$ and $G(x)$ are essentially of the similar form $p(t) = \ln(1 - e^{-t})$, so we shall start by finding the power series expansion for $p(t)$ near $t = 0$. Applying the first expansion above for the exponential function, the rule for the logarithm of a product $\ln(ab) = \ln(a) + \ln(b)$, and finally the second expansion with t equalling the whole power series that remains inside the logarithmic function in addition to the constant term 1, we get for our auxiliary function $p(t)$ the result

$$p(t) = \ln(t) - t/2 + t^2/24 - t^4/2880 + t^6/181440 + \dots \quad (11)$$

To get the series expansion for function $F(x)$, we replace t in $p(t)$ by e^x . We shall apply the resulting series for large *negative* values of x , which means that the series converges rapidly. Taking into account the minus sign in the definition of $F(x)$ we end up with the expansion:

$$F^s(x) = -x + e^x/2 - e^{2x}/24 + e^{4x}/2880 - e^{6x}/181440 + \dots \quad (12)$$

This result shows us that when $x \rightarrow -\infty$, the value of $F^s(x)$ approaches $-x$ monotonously from above, the difference being of order $e^x/2$.

In a completely analogous fashion it can be shown that for the function $G(x)$ the corresponding expansion takes the form

$$G^s(x) = \frac{\ln(1 + ax)}{a} + \frac{(1 + ax)^{-1/a}}{2} - \frac{(1 + ax)^{-2/a}}{24} + \frac{(1 + ax)^{-4/a}}{2880} - \frac{(1 + ax)^{-6/a}}{181440} - \dots \quad (13)$$

With $x \rightarrow \infty$, function $G^s(x)$ approaches $-(\ln(1 + ax))/a$ from below, and for large values of x , the difference is of order $((1 + ax)^{-1/a})/2$.

3 Schematic representation of mGSZ

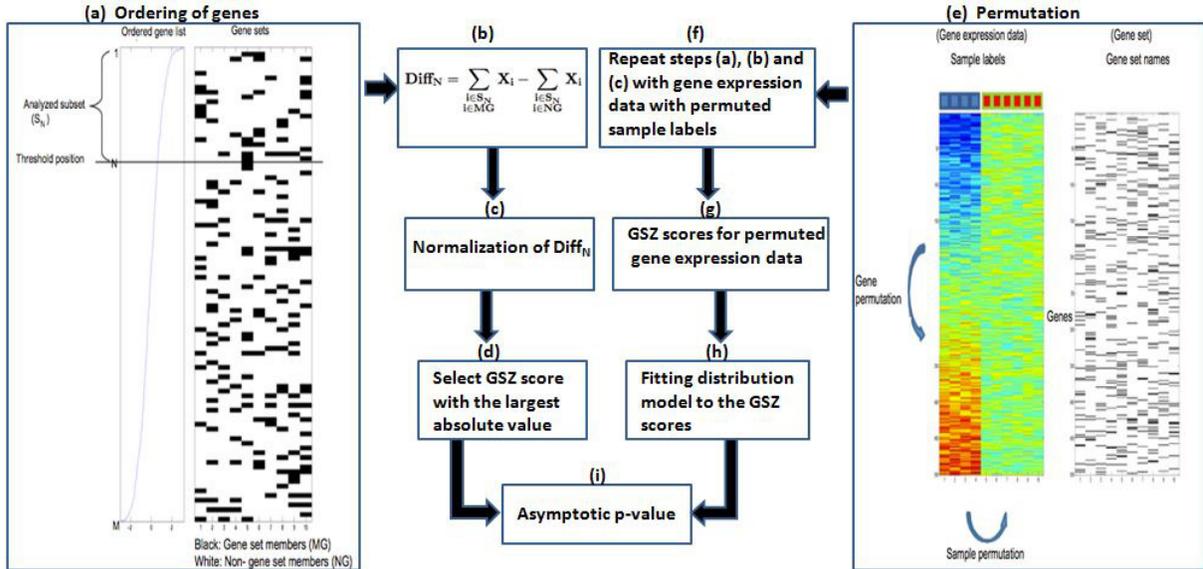


Figure 1: Schematic representation of mGSZ workflow. mGSZ takes differential gene expression test scores and list or matrix of gene sets as input. The gene list with differential gene expression test scores is (a) ordered based on differential gene expression test scores. Subsets of genes are taken at threshold positions placed in between every consecutive pair of genes, and for each subset (b) the difference between the sum of differential gene expression test scores for the member and non-member genes of the analyzed gene set is calculated, the calculated difference is (c) normalized with the estimates of mean and standard deviation, and (d) the largest absolute value is selected as the GSZ score for the analyzed gene set. The steps (a), (b), (c) and (d) is repeated with differential gene expression test scores from permuted (sample permutation) gene expression data. (h) Asymptotic distribution model is fitted to the GSZ scores from permuted gene expression data. (i) Asymptotic p-value is calculated for the analyzed gene set.

4 Evaluation of the p-value calculation methods with p53 data

In this section, we evaluate asymptotic and empirical methods of p-value estimation. Mean and standard deviation of the gene set scores were adjusted to 0 and 1 respectively for all but *GAMMA* p-value estimation. We considered empirical p-values from 100000 sample permutations as the reference of truth, against which we compared asymptotic p-values and empirical p-values from 500 sample permutations. Our analysis is based on log p-values.

4.1 Asymptotic and empirical p-values for mGSZ scores

The resolution of empirical p-values depends on the number of permutations (Figure 2). As expected, p-values obtained from Gaussian model fitted to mGSZ scores failed miserably. P-values obtained from extreme value (EVD) and general extreme value (GEVD) distributions fitted to mGSZ scores were the best estimates (Figure 2).

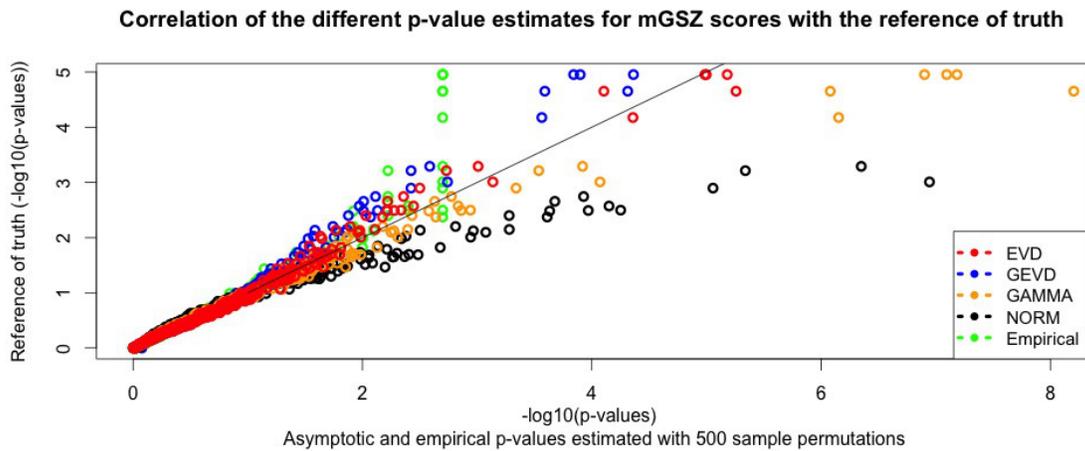


Figure 2: Correlation of empirical and asymptotic p-values (*EVD*, *GEVD*, *GAMMA* and *NORM*) (X-axis) estimated from 500 sample permutations with the reference of truth (Y-axis). The reference of truth corresponds to empirical p-values estimated from 100000 sample permutations.

4.2 Asymptotic and empirical p-values for mGSA scores

The best p-values for mGSA scores were the asymptotic p-values estimated from extreme value distribution (EVD), general extreme value distribution (GEVD) and gamma distribution (GAMMA) fitted to mGSA scores (Figure 3). Empirical p-values and asymptotic p-values obtained from Gaussian distribution fitted to mGSA scores had the worst correlation with the reference of truth (Figure 3).

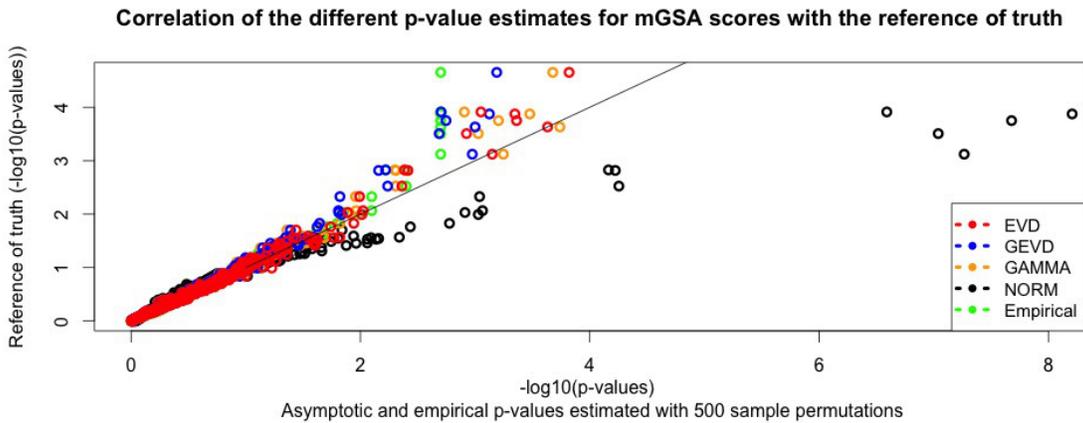


Figure 3: Correlation of empirical and asymptotic p-values (EVD, GEVD, GAMMA and NORM) (X-axis) estimated from 500 sample permutations with the reference of truth (Y-axis). The reference of truth corresponds to empirical p-values estimated from 100000 sample permutations.

4.3 Asymptotic and empirical p-values for mAllez scores

Asymptotic p-values for mAllez scores were estimated from Gaussian distribution fitted to mAllez scores. Here too, empirical p-value estimated with 500 sample permutations had the worst correlation with the reference of truth (Figure 4).

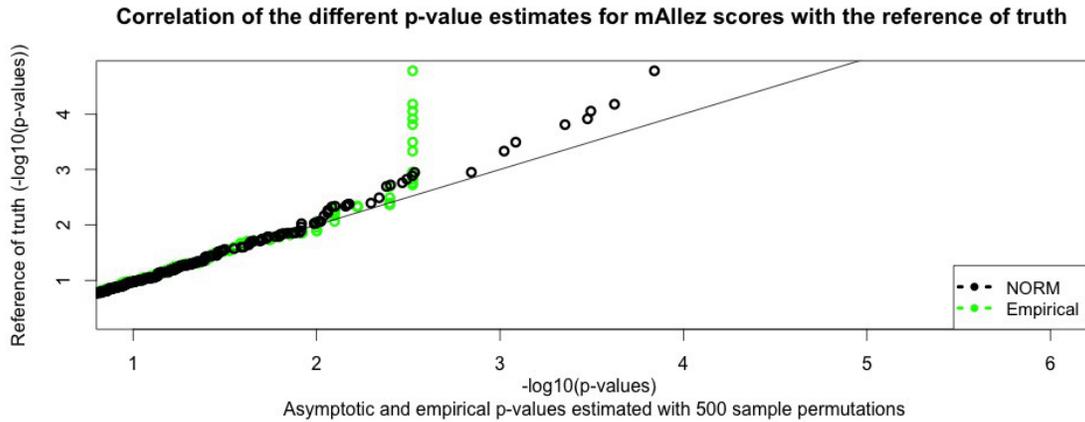


Figure 4: Correlation of empirical and asymptotic p-values (NORM) (X-axis) estimated from 500 sample permutations with the reference of truth (Y-axis). The reference of truth corresponds to empirical p-values estimated from 100000 sample permutations.

4.4 Asymptotic and empirical p-values for SUM scores

Empirical and asymptotic p-value evaluation for SUM scores showed similar results as that of mAllez (see Figure 5).

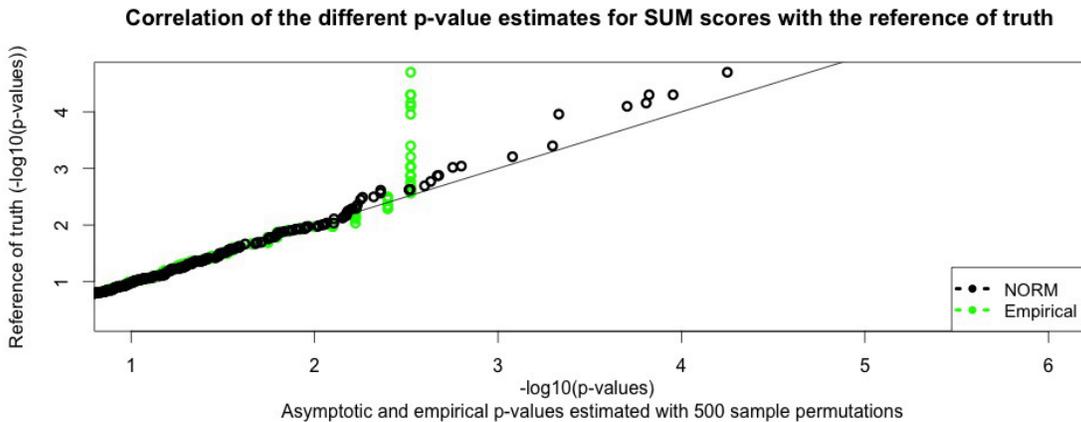


Figure 5: Correlation of empirical and asymptotic p-values (NORM) (X-axis) estimated from 500 sample permutations with the reference of truth (Y-axis). The reference of truth corresponds to empirical p-values estimated from 100000 sample permutations.

5 Evaluation of the p-value calculation methods with gender data

Null distribution of gene set scores from permuted data is mostly dependent on gene set scoring functions. In this section, we show the evaluation of asymptotic and empirical methods of p-value estimation with gender data.

5.1 Asymptotic and empirical p-values for mGSZ scores

P-values estimated with *EVD* and *GEVD* still remain the best p-values in gender data as well (Figure 6). However, in contrast to the results from p53 data, empirical p-values are equally good as asymptotic p-values calculated with *EVD* and *GEVD* (Figure 6). This is because gender data has relatively lower biological signal as compared to p53 data and thus empirical p-values reach the finest resolution already with 500 sample permutations.

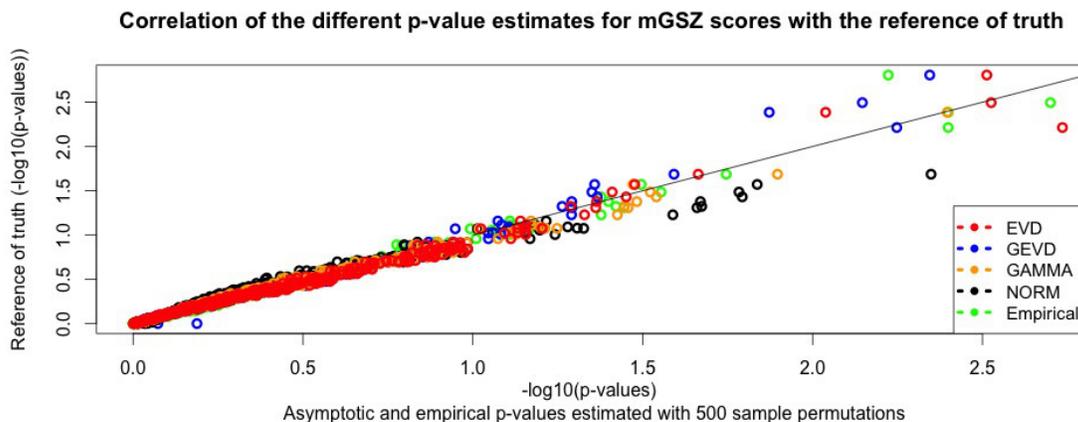


Figure 6: Correlation of empirical and asymptotic p-values (*EVD*, *GEVD*, *GAMMA* and *NORM*) (X-axis) estimated from 500 sample permutations with the reference of truth (Y-axis). The reference of truth corresponds to empirical p-values estimated from 100000 sample permutations.

5.2 Asymptotic and empirical p-values for mGSA scores

The results are similar to that of p53 data except that the empirical p-values show equally good correlation with the reference of truth for the reason described in Section 5.1 (Figure 7).

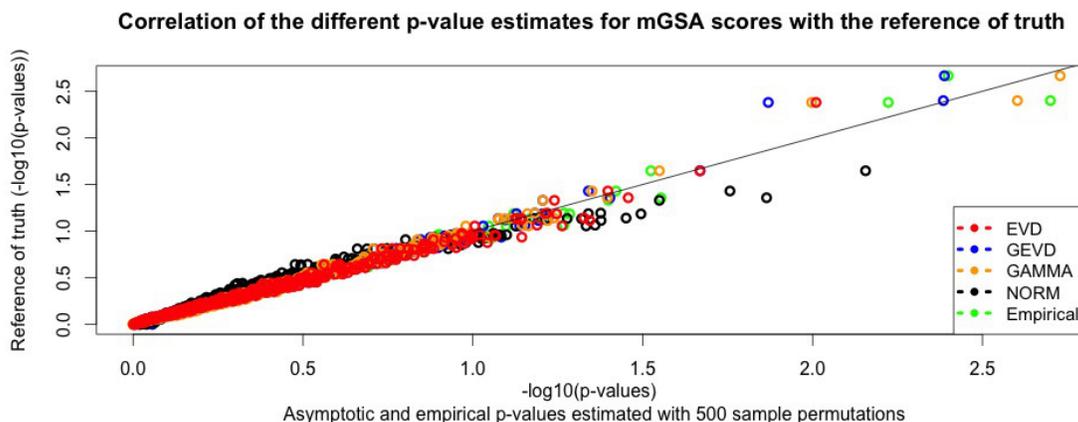


Figure 7: Correlation of empirical and asymptotic p-values (*EVD*, *GEVD*, *GAMMA* and *NORM*) (X-axis) estimated from 500 sample permutations with the reference of truth (Y-axis). The reference of truth corresponds to empirical p-values estimated from 100000 sample permutations.

5.3 Asymptotic and empirical p-values for mAllez scores

The results are similar to that of p53 data except that the empirical p-values show equally good correlation with the reference of truth for the reason described in Section 5.1 (Figure 8).

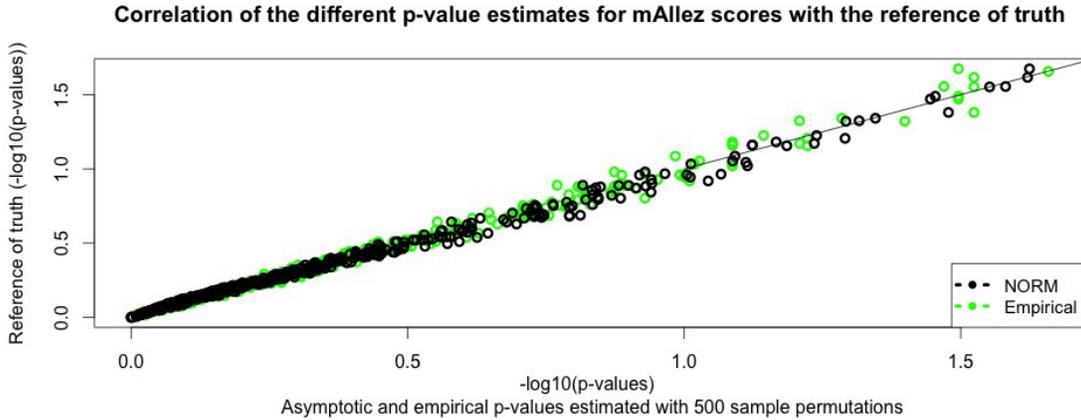


Figure 8: Correlation of empirical and asymptotic p-values (NORM) (X-axis) estimated from 500 sample permutations with the reference of truth (Y-axis). The reference of truth corresponds to empirical p-values estimated from 100000 sample permutations.

5.4 Asymptotic and empirical p-values for SUM scores

The results are similar to that of p53 data except that the empirical p-values show equally good correlation with the reference of truth for the reason described in Section 5.1 (Figure 9).

5.5 Asymptotic approximation of empirical null distribution with gender data

Table 1: Asymptotic distribution models fitted on the empirical null distribution generated by the compared methods with gender data and their results. Scores not meeting the criteria for optimal model (Section 2.5.1 from the main article) are highlight.

Methods	Models	mse	$mse(subset)$	cor	$cor(subset)$
mGSZ	EVD	0.002	0.03	0.99	0.99
	GEVD	0.002	0.03	0.99	0.99
	GAMMA	0.003	0.05	0.99	0.98
	NORM	0.03	0.73	0.96	0.99
mGSA	EVD	0.002	0.03	0.99	0.99
	GEVD	0.001	0.02	0.99	0.98
	GAMMA	0.001	0.02	0.99	0.97
	NORM	0.03	0.87	0.94	0.96
mAllez	NORM	0.0007	0.002	0.99	0.98
WRS	NORM	0.01	0.07	0.96	0.95
SS	GAMMA	0.008	0.3	0.97	0.95
SUM	NORM	0.0005	0.002	0.99	0.98
KS	–	–	–	–	–
wKS	–	–	–	–	–

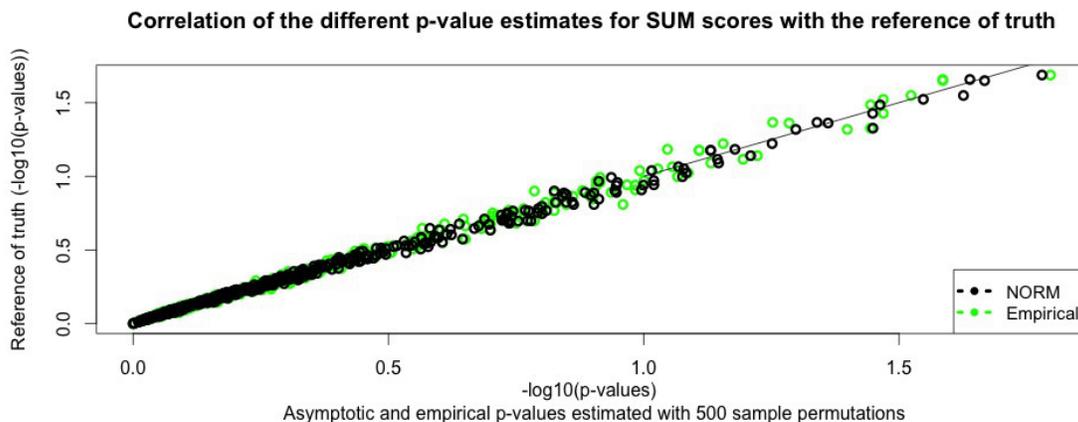


Figure 9: Correlation of empirical and asymptotic p-values (NORM) (X-axis) estimated from 500 sample permutations with the reference of truth (Y-axis). The reference of truth corresponds to empirical p-values estimated from 100000 sample permutations.

6 Comparison of stability of gene set scores with gender data

In this section, we present the comparison of stability of the gene set scores from mGSZ, KS and wKS with gender data. Similar to that of p53 data, here too the stability plot for mGSZ, based permuted data shows that middle part of the seven percentiles of the gene set score profiles of the permuted data stay quite stable across gene list positions (Figure 10a). The maximum percentile shows smaller deviation from zero than the minimum percentile (Figure 10a). This indicates that gene set scores coming from under-representation show stronger signal than those coming from over-representation. The minimum of the gene set score profile from original data comes quite early in the gene list, pointing that gene set members occur near the bottom of the ordered gene list (Figure 10a). Notice that the separation between gene set score profile from original data and permuted data is quite clear across different threshold positions (Figure 10a). The importance of our visualization is more highlighted when we analyze the behavior of wKS and KS with the same gene expression dataset and the same gene set. For wKS the gene set score profiles from permuted data show strange behavior with S shaped profile (Figure 10b) and for KS the middle area shows strong variation (Figure 10c). The separation between gene set score profiles from permuted data and original data is very weak in both wKS and KS (Figure 10c). This comparison is based on the strongest gene set reported by wKS. Note that the mGSZ software reports absolute values for gene set scores.

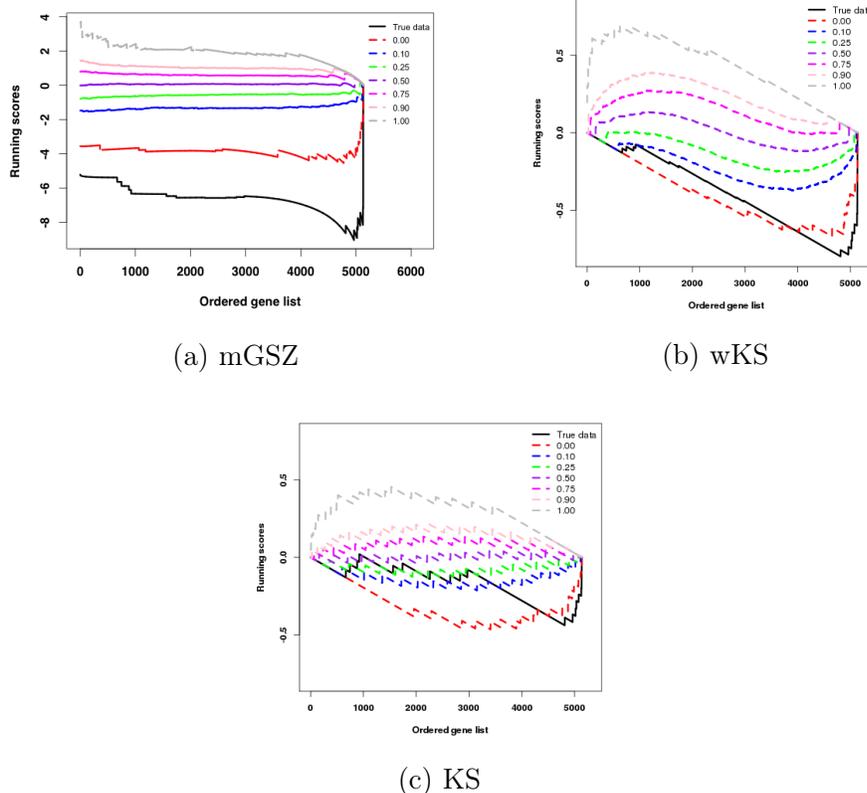


Figure 10: Visualization of scoring function profiles for mGSZ, wKS and KS for gene set "XINACT MERGED" in gender data. Figure also shows 7 percentiles (0, 10, 25, 50, 75, 90, 100) of gene set score profiles obtained from 1000 sample permutation data. Notice the clear separation of the gene set score profiles from original data with those from permuted data and stability of seven percentiles of gene set score profiles from permuted data in case of GSZ-score.

7 Calculation of prior variance in Gene Set Z-score function

For the calculation of prior variance (k) in Equation 2 of the main article, we use; 1) the median of the variance estimates obtained with the analyzed gene set across the whole gene list and multiply it with a weight, $w2$ ($0 \leq w2 \leq 1$), and 2) the median of the variance estimates obtained with the gene set of size 10 across the whole gene list and multiply it with a weight, $w1$ ($0 \leq w1 \leq 1$). Prior variance is then obtained by summing the weighted variance medians. The selection of the reference class size and the values for $w1$ and $w2$ were based on the original paper by Törönen *et al.*, 2009.

8 Similarity of the GSZ and GSA scoring functions

For the calculation of max-mean statistics in the GSA method, the differential expression test scores are summed and divided by the size of the whole gene set in both positively and negatively regulated parts of the ordered gene list. The maximum of the absolute values of the two parts is selected as the max-mean statistics (Equation 14).

$$MaxMean = \max \left(\text{abs} \left(\sum_{\substack{X_i > 0 \\ i \in GeneSet}} X_i \right), \text{abs} \left(\sum_{\substack{X_i < 0 \\ i \in GeneSet}} X_i \right) \right) / N_{GeneSet} \quad (14)$$

Where, X_i is the differential expression test score for i th gene of the analyzed gene set and $N_{GeneSet}$ is the total number of genes in the analyzed gene set.

GSZ equation uses non-member genes (genes not belonging to the analyzed gene set) in the calculation. However, the result from the list of non-member genes is essentially the result from the list of member genes subtracted from the result from the whole gene list. Thus, the $Diff_N$ estimate of GSZ (Equation 1 in the main article) is a function of sum in MaxMean score function (Equation 14). Both methods focus only on one (either positively or negatively regulated) part of the gene list. The differences are : 1) the mean and standard deviation estimates for each half in GSA are obtained via gene permutations, whereas in GSZ, they are obtained by estimates from the actual data, 2) Unlike GSA, GSZ scoring function is used to test many threshold positions. However the function can be modified by fixing the threshold position.

9 Log p-values of the top gene sets

The empirical p-values of the top 50 gene sets reported by the compared methods in p53 and gender datasets calculated from 100000 sample permutations were compared. The comparison was considered inconclusive as the ordering of the methods varied between datasets. We tested the sensitivity of the methods to false positive signals in the main article. The p-value analysis should be complemented by the sensitivity test of the methods to false positive signals. Therefore, we present the results of the p-value analysis in the supplementary.

9.1 Log p-values of the top gene sets in p53 dataset

SS and mGSZ are the best performing methods on the upper region of the ordered gene list (OGL) (Figure 11a). However, on the lower region of the list, SS is the best method followed by mGSZ, SUM and mAllez (Figure 11a).

9.2 Log p-values of the top gene sets in gender dataset

Unlike the results from p53 dataset, the best methods on the upper region of the OGL of gender dataset are mGSZ and mAllez (Figure 11b). WRS and mAllez report the best p-values on the lower region of the list (Figure 11b). Surprisingly, SS reports the weakest p-values in the gender dataset (Figure 11b).

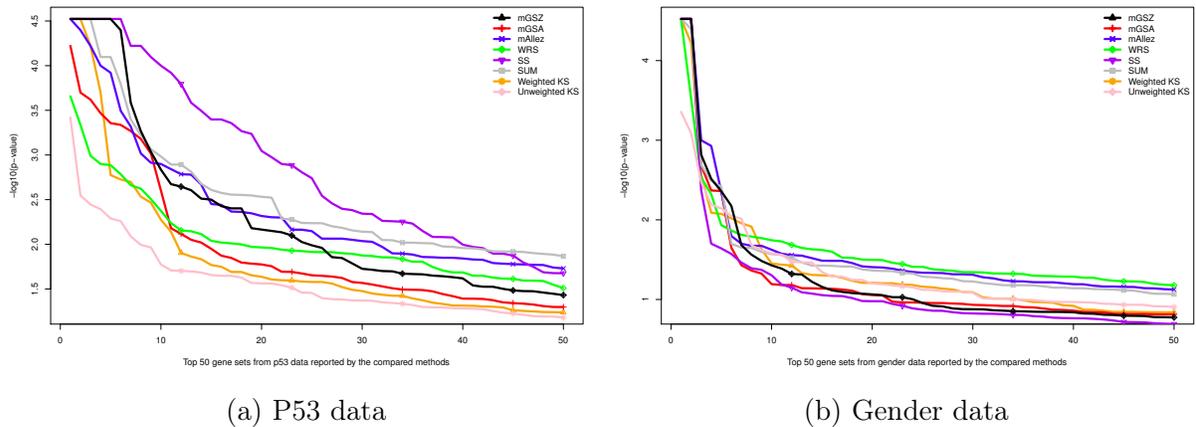


Figure 11: Empirical log p-values for the top 50 gene sets estimated by each of the compared methods in p53 and gender dataset.

10 False positive signal test with p53 data

In addition to the leukemia dataset, the compared methods were also evaluated with p53 dataset and GO gene sets for false positive signal test. Here, we selected GO gene sets instead of curated gene sets from Subramanian *et al.*, 2005 because GO gene sets have wide range of sizes. This allows to test the stability of the methods over wide range of gene sets sizes. Unlike with the leukemia dataset, even though mGSZ, GSA, weighted KS and unweighted KS are more conservative, the difference to the other methods is significantly small (Figure 12).

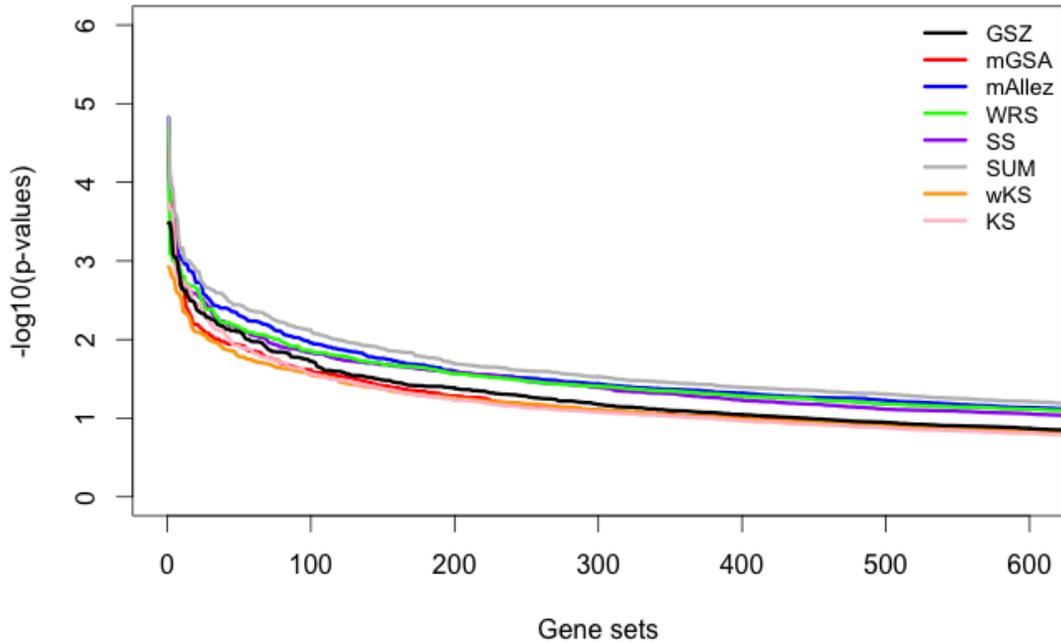


Figure 12: Empirical log p-values for GO gene sets estimated by the compared methods on randomized p53 dataset.

11 Comparison of mGSZ, mGSA and mAllez with program packages

False positive signal test shows that mGSZ and most other methods show quite similar behavior (Figure 13). While CAMERA is the most conservative, ROAST shows strong

noise signal (Figure 13). The result points out a major difference between competitive and self-contained gene set analysis methods. Self-contained gene set analysis methods calculate the gene set scores without considering the genes other than member genes. Thus, in a signal rich dataset like leukemia dataset most of the null gene sets are reported as significant by self-contained gene set analysis method like ROAST.

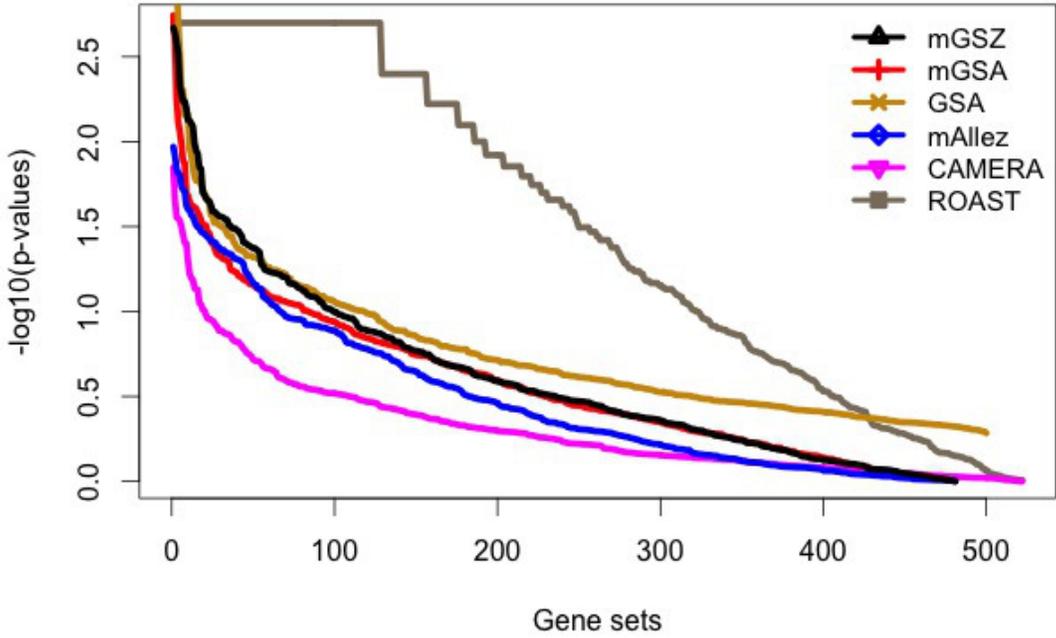


Figure 13: Comparison of mGSZ, mGSA and mAllez with the program packages with randomized gene sets.

Based on the detection of the relevant gene sets from p53 and gender data by mGSZ, mGSA, mAllez and the program packages, mGSZ is clearly the best method (Figure 14a and 14b). Moreover, mGSA and mAllez shows improved performance as compared to GSA and Allez (Figure 14a and 14b).

mGSZ reports the best p-values (based on resolution) with p53 dataset for the top 50 gene sets as compared to the other methods (Figure 15a). However, in case of gender dataset, mGSZ reports the best p-values for the upper region of the gene list and then slightly lags behind mGSA, mAllez and ROAST in the lower region of the gene list (Figure 15b).

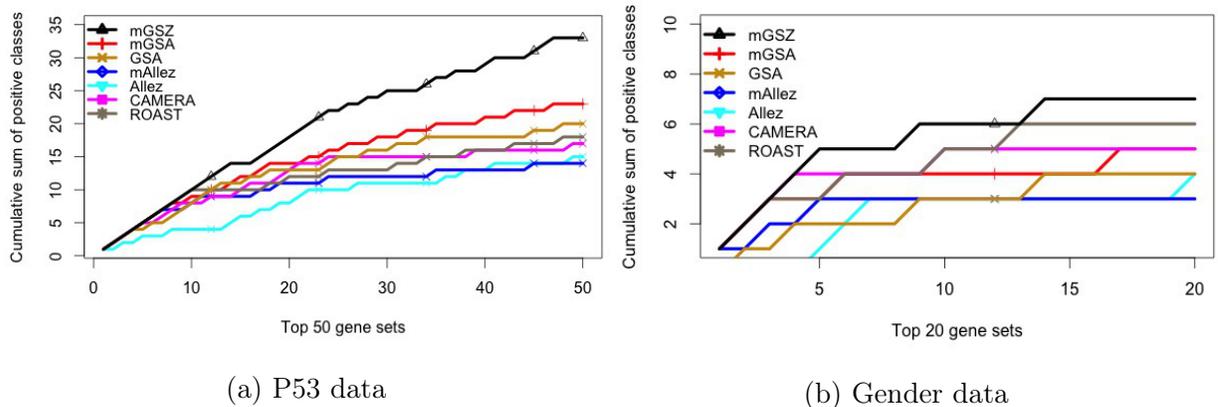


Figure 14: Relevant gene sets identified by the compared methods. Figures represent cumulative count of biologically relevant gene sets (Tables 4 and 5) over the ranked list of top 50 gene sets in case of the p53 data and top 20 gene sets in case of the gender data reported by each of the compared methods. mGSZ (black) shows the best performance.

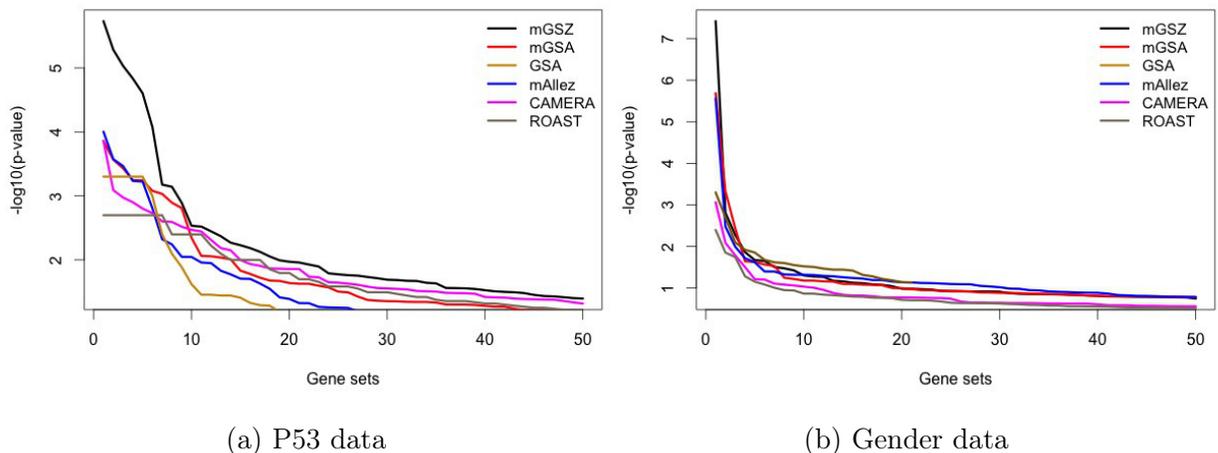
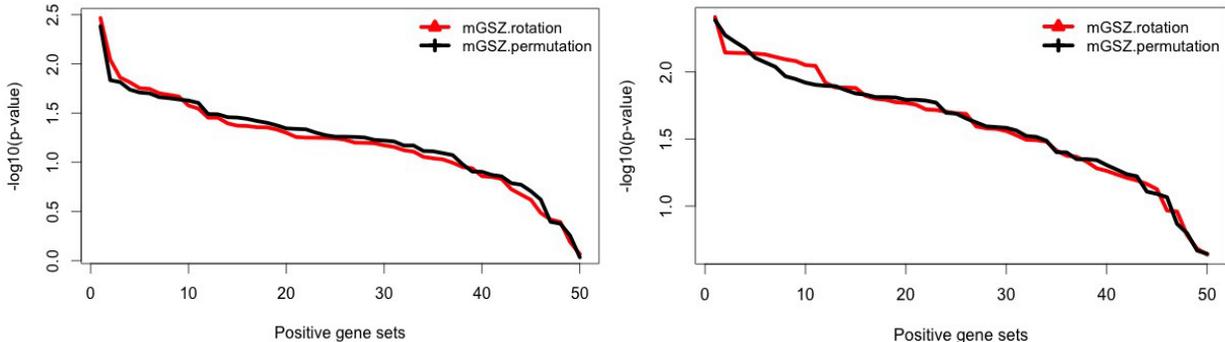


Figure 15: Log p-values for the top 50 gene sets estimated by each of the compared methods in p53 and gender datasets.

12 Evaluation of mGSZ on dataset with small sample size

In this section, we present the results obtained by implementing mGSZpermutaion and mGSZrotation on the simulated dataset described in Section 2.7 of the main article. In

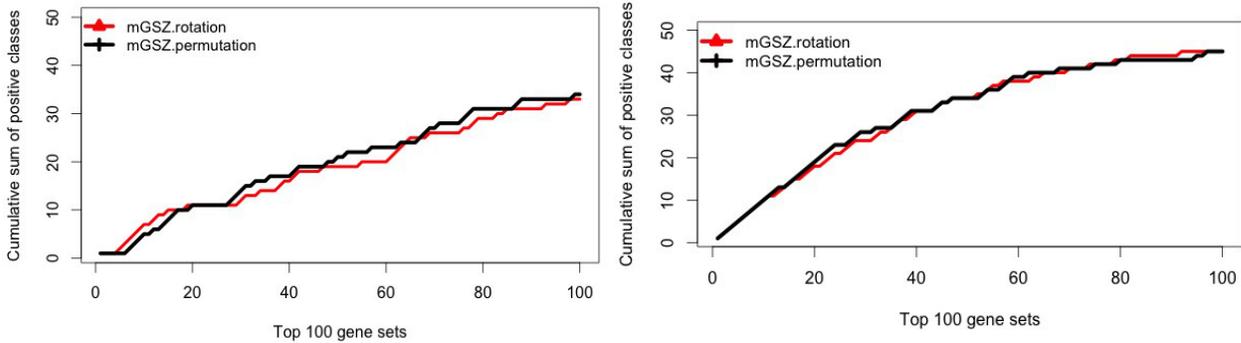
addition, mGSZpermutation and mGSZrotation were also implemented on a simulated dataset with slightly higher signal achieved by adding the constant 1.5, to see if the results vary with varying signals in datasets. The evaluation was based on comparison of log p-values and identification of the positive gene sets by the two methods. In both the comparisons no significant difference was observed (Figures 16 and 17).



(a) Simulated data with lower signal

(b) Simulated data with higher signal

Figure 16: Asymptotic log p-values for the gene sets in the simulated datasets estimated by mGSZrotation and mGSZpermutation.



(a) Simulated data with lower signal

(b) Simulated data with higher signal

Figure 17: Positive gene sets identified by mGSZrotation and mGSZpermutation. Figures represent cumulative count of positive gene sets over the ranked list of the top 50 gene sets reported by each of the compared methods.

13 Top gene sets identified by standard Allez and mAllez

Table 2: Top 10 gene sets reported by standard Allez and mAllez from p53 dataset. Standard Allez lacks sample permutation and reports z-score. Empirical and asymptotic p-values for mAllez scores were calculated from 1000 sample permutations.

Allez		mAllez		
Classes	Z-score	Classes	Empirical p-value	Asymptotic p-value
P53_UP	5.15	p53Pathway	0.001	0.00009
HTERT_UP	5.03	hsp27Pathway	0.001	0.0002
p53Pathway	4.55	p53hypoxiaPathway	0.001	0.0003
GLUT_UP	4.46	P53_UP	0.001	0.0003
rasPathway	4.19	rasPathway	0.001	0.0003
GPCRs_Class_A_Rhodopsin-like	4.01	radiation_sensitivity	0.001	0.0004
HUMAN_CD34_ENRICHED_TF_JP	3.77	SA_DAG1	0.001	0.0009
hsp27Pathway	3.72	ngfPathway	0.001	0.001
INSULIN_2F_UP	3.67	il4Pathway	0.002	0.003
LEU_DOWN	3.67	fmlpPathway	0.002	0.003

14 Top gene sets identified by standard GSA and mGSA

Table 3: Gene sets with p-value less than 0.01 reported by GSA and mGSA from p53 dataset with 1000 sample permutations. GSA calculates empirical p-value, whereas mGSA calculates asymptotic p-value.

GSA		mGSA		
Classes	Empirical p-value	Classes	Empirical p-value	Asymptotic p-value
Hsp27Pathway	0	p53Pathway	0.001	0.0001
P53hypoxiaPathway	0	hsp27Pathway	0.001	0.0003
P53Pathway	0	radiation_sensitivity	0.001	0.0003
SA_G1_AND_S_PHASES	0	P53hypoxiaPathway	0.001	0.0005
fmlppathway	0	P53_UP	0.001	0.0006
rasPathway	0	SA_G1_AND_S1_PHASES	0.001	0.001
radiation_sensitivity	0.001	fmlppathway	0.001	0.002
P53_UP	0.001	SA_DAG1	0.002	0.0007
SA_DAG1	0.002	rasPathway	0.002	0.001
BadPathway	0.003	BadPathway	0.004	0.009
ngfPathway	0.003			
MAPK_Cascade	0.006			
il4Pathway	0.009			
MAP00790_Folate_bios	0.009			

15 List of gene sets that are strongly relevant to p53 activity

Top fifty gene sets reported by each of the compared methods were pooled. Each of the pooled gene sets were then analyzed separately for their relevance to p53 activity. A total of forty gene sets highly relevant to p53 activity were selected as relevant gene sets for p53 data.

Table 4: Gene sets and their relevance to p53 gene activity

Gene sets	Relavance to P53	Gene sets	Relavance to P53
p53hypoxiaPathway	Hypoxia induces p53 mediated apoptosis ^b	Ceramide Pathway	Apoptosis ^b
p53Pathway	Pathway ^b	Ca_nf_at_signaling	Apoptosis ^b
radiation_sensitivity	Mutation in p53 affects radiation sensitivity ^b	Hivnef pathway	Apoptosis ^b
SA_PROGRAMMED_CELL_DEATH	Apoptosis ^b	ras pathway	Regulates cell growth, differentiation and death ^b
P53_UP	Apoptosis ^b	ngf pathway	Deacetylation of p53 ^b
hsp27Pathway	Modulation/regulation of p53 signalling/activity (O'Callaghan-Sunol <i>et al.</i> , 2007)	il4 pathway	p53 suppresses IL-4 ^b
atmPathway	Pathway member ^b	ST-Interleukin-4-pathway	Cytokins ^b
p53 signalling	Pathway member ^b	ccr3 pathway	Apoptosis ^b
chemical pathway	Pathway member ^b	cxcr4 pathway	Apoptosis ^b
Cr. death	Pathway member ^b	erk pathway	Apoptosis (Cagnol and Chambard, 2010)
DNA damage signalling	Pathway member ^b	MAPK cascade	Activated by p53 ^b
G1 pathway	Pathway member ^b	ck1 pathway	Phosphorylates p53 (Alsheich-Bartok <i>et al.</i> , 2008)
G2 pathway	Pathway member ^b	igf1 pathway	Regulated by p53 (Feng, 2010)
ST FAS signalling Pathway	Pathway member ^b	bcr pathway	Apoptosis (Kroesen <i>et al.</i> , 2001)
Cell cycle pathway	Pathway member ^b	MAP00790_folate_biosynthesis	Folate stress induces p53 mediated apoptosis (Hoeflerin <i>et al.</i> , 2013)
Drug resistance and metabolism	Pathway member ^b	ST_Phosphoinositide_3_kinase_pathway	p53 mediated apoptosis (Cooper and Nayak, 2012)
Breast cancer estrogen signalling	Pathway member ^b	mapk pathway	Interacts with p53 ^b
BAD Pathway	Apoptosis ^b	SA_G2_AND_M_PHASES	Cell cycle ^b
Mitochondria pathway	Apoptosis ^b	SA_G1_AND_S_PHASES	Cell cycle ^b
bcl2family and reg. network	Apoptosis ^b	SA_DAG1	Regulates cell growth and apoptosis (Wright and McMaster, 2002)

^aManually curated

^bManually curated

16 List of gene sets that are strongly relevant to gender

Top twenty gene sets reported by each of the compared methods were pooled. Each of the pooled gene sets were then analyzed separately for their relevance to gender. A total of ten gene sets highly relevant to gender were selected as relevant gene sets for gender data.

Table 5: Gene sets and their relevance to gender

Gene sets	Relavance to gender
XINACT_MERGED	Female specific ^a
TESTIS_GENES_FROM_XHX_AND_NETAFFX	Male specific ^a
GNF_FEMALE_GENES	Female specific ^a
aifPathway	Triggers cell death in males (Du <i>et al.</i> , 2004)
SIG_Regulation_of_the_actin_cytoskeleton_by_Rho_GTPases	Gender specific regulation ^a
tercPathway	Telomere length varies with gender (Barrett and Richardson, 2011)
NFKB_REDUCED	Variation with gender (Vina <i>et al.</i> , 2011)
caspase_activity	Variation with gender (Liu <i>et al.</i> , 2009)
MAP00252_Alanine_and_aspartate_metabolism	Variation with gender (Mera <i>et al.</i> , 2008)
MAP00910_Nitrogen_metabolism	Variation with gender (Cheney <i>et al.</i> , 1987)

^aManually curated

References

- Alsleich-Bartok, O., Haupt, S., Alkalay-Snir, I., Saito, S., Appella, E., and Haupt, Y. (2008). Pml enhances the regulation of p53 by ckl in response to dna damage. *Oncogene*, **27**(26), 3653–3661.
- Barrett, E. L. B. and Richardson, D. S. (2011). Sex differences in telomeres and lifespan. *Aging cell*, **10**(6), 913–21.
- Cagnol, S. and Chambard, J.-C. (2010). Erk and cell death: mechanisms of erk-induced cell death–apoptosis, autophagy and senescence. *The FEBS journal*, **277**(1), 2–21.
- Cheney, C. L., Lenssen, P., Aker, S. N., Cunningham, B. A., Gauvreau, J. M., Darbinian, J., and Barale, K. V. (1987). Sex differences in nitrogen balance following marrow grafting for leukemia. *Journal of the American College of Nutrition*, **6**(3), 223–230.
- Cooper, G. M. and Nayak, G. (2012). p53 is a major component of the transcriptional and apoptotic program regulated by pi 3-kinase/akt/gsk3 signaling. *Cell Death and Disease*, **3**(10), e400.
- Du, L., Bayir, H., Lai, Y., Zhang, X., Kochanek, P. M., Watkins, S. C., Graham, S. H., and Clark, R. S. B. (2004). Innate gender-based proclivity in response to cytotoxicity and programmed cell death pathway. *The Journal of biological chemistry*, **279**(37), 38563–38570.
- Feng, Z. (2010). p53 regulation of the igf-1/akt/mtor pathways and the endosomal compartment. *Cold Spring Harbor perspectives in biology*, **2**(2), a001057.
- Hoeflerlin, L. A., Fekry, B., Ogretmen, B., Krupenko, S. a., and Krupenko, N. I. (2013). Folate stress induces apoptosis via p53-dependent de novo ceramide synthesis and up-regulation of ceramide synthase 6. *The Journal of biological chemistry*.

- Kotz, S. and Nadarajah, S. (2000). *Extreme value distributions*, pages 1–3. Number x. Imperial College Press.
- Kroesen, B. J., Pettus, B., Luberto, C., Busman, M., Sietsma, H., Leij, L. d., and Hannun, Y. A. (2001). Induction of apoptosis through b-cell receptor cross-linking occurs via de novo generated c16-ceramide and involves mitochondria. *The Journal of biological chemistry*, **276**(17), 13606–13614.
- Liu, F., Li, Z., Li, J., Siegel, C., Yuan, R., and McCullough, L. D. (2009). Sex differences in caspase activation after stroke. *Stroke; a journal of cerebral circulation*, **40**(5), 1842–1848.
- Mera, J. R., Dickson, B., and Feldman, M. (2008). Influence of gender on the ratio of serum aspartate aminotransferase (ast) to alanine aminotransferase (alt) in patients with and without hyperbilirubinemia. *Digestive diseases and sciences*, **53**(3), 799–802.
- O’Callaghan-Sunol, C., Gabai, V. L., and Sherman, M. Y. (2007). Hsp27 modulates p53 signaling and suppresses cellular senescence. *Cancer research*, **67**(24), 11779–11788.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(43), 15545–15550.
- Törönen, P., Ojala, P. J., Marttinen, P., and Holm, L. (2009). Robust extraction of functional signals from gene set analysis using a generalized threshold free scoring function. *BMC Bioinformatics*, **10**(1), 307.
- Vina, J., Gambini, J., Lopez-Grueso, R., Abdelaziz, K. M., Jove, M., and Borras, C. (2011). Females live longer than males: role of oxidative stress. *Current pharmaceutical design*, **17**(36), 3959–65.
- Wright, M. M. and McMaster, C. R. (2002). Phospholipid synthesis, diacylglycerol compartmentation, and apoptosis. *Biological research*, **35**(2), 223–229.