The GSReg package allows to analyze pathways based on the variability of the expression of sets of genes that are targets of those pathways. Basing this set statistic on variability enables inference of dysregulated pathways in diseases, including notably cancers. The first set statistic for gene variability was in the work of Eddy and his colleagues (see [1]) which used a ranked based methodology called DIRAC. DIRAC calculates a measure of variability of the ordering of the expression of genes in a pathway for specific phenotype. The basic idea behind DIRAC is to generate a template for the pair-wise comparisons of gene expressions.
of a pathway within a phenotype. DIRAC calculates a measure of the variability of the ordering within the phenotype, i.e. the expected distance of a sample from the phenotype and the template of the phenotype. In mathematical terms, if we denote denote two i.i.d. samples from the same phenotypes by \( X \) and \( X' \) and \( D \) Kendall-\( \tau \)-distance on the specific pathways, then the EVA statistic is \( E(D(X,X')) \). It identifies significantly dysregulated pathways by estimating p-values from a permutation test. Eddy et al. found that more pathological phenotypes usually have more pathways with higher variability compared to less pathological phenotypes.

However, the permutation test in DIRAC is computationally intensive and reaching low p-values may be impractical since they require a huge number of permutations. Low p-values are required for multiple hypothesis correction. A similar measure of variability of the orderings of gene sets was proposed in [2]. This method approximates the p-value theoretically, without a permutation test. This method is based on Kendall-\( \tau \) distance [3] and the theory of U-Statistics, thus we call this method Gene Set Expression Variation Analysis (or in short EVA). Specifically, Kendall-\( \tau \) distance between two expression profiles counts the number of disagreeing pairwise comparisons between two profiles. The EVA measures the variability of the gene expression of pathway genes from a phenotype by calculating the expectation of Kendall-\( \tau \) distance between two random samples from the phenotype. EVA then identifies if the variability is significantly different across two phenotypes. To approximate this p-value, EVA applies a U-Statistic Theory approach.

The GReg package contains two following utilities:

1. Identifying the dysregulated pathways with DIRAC measure of variability. The significance is calculated using permutation test. This is the first time that DIRAC analysis has been implemented in \( R \). It also is more adaptable to new datasets than the original Matlab code in [1].

2. Identifying the dysregulated pathways with EVA measure of variability. The significance is approximated through applying U-statistics theory. This is very time efficient and consistent with both DIRAC and applying permutation test on EVA.

2 Input Data

2.1 Data structure

In short, the GSReg package requires the following data in the following format:

1. Gene Expression Data
(a) The expression be in the form of a matrix where rows represent genes (or probes) and columns represent samples.
(b) The expression matrix cannot have NAs.
(c) The expression matrix rows must have names of genes or the probes.

2. Pathways
   (a) The list of pathways must contain character vectors. Only the elements of the vectors which appear in rownames of the expression matrix are considered for analysis.
   (b) The list of the pathways must have names for each vectors.

3. Phenotypes
   (a) A factor with binary levels.

We used the data provided in the GSBenchMark package to reproduce the results in Eddy et al. [1]. The GSBenchMark contains data for the pathways as well as the gene expression and phenotype data from twelve studies. We load the information about the pathways from GSBenchMark:

```r
> library(GSBenchMark)
> data(diracpathways)
> class(diracpathways)
[1] "list"
> names(diracpathways)[1:5]
[1] "DEATHPATHWAY" "TCAOPTOSISPATHWAY" "CCR3PATHWAY" 
[4] "NEUTROPHILPATHWAY" "ALTERNATIVEPATHWAY"
> class(diracpathways[[1]])
[1] "character"
```

As mentioned, GSReg package requires the information of the pathways to be as a list of character vectors. Also, GSReg requires the pathways to have names. The variable diracpathways contains gene pathways. It is a list. Each element represents a pathway with its name. Each elements contains a list of characters which represent the genes in the pathway. e.g. diracpathways[["DEATHPATHWAY"]].

Now, we load the datasets’ names:
> data(GSBenchMarkDatasets)
> print(GSBenchMark.Dataset.names)

[1] "leukemia_GSEA"    "marfan_GDS2960"    "melanoma_GDS2735"    
[4] "parkinsons_GDS2519" "prostate_GDS2545_m_nf" "prostate_GDS2545_m_p"    
[7] "prostate_GDS2545_p_nf" "sarcoma_data"    "squamous_GDS2520"    
[10] "breast_GDS807"    "bipolar_GDS2190"

The remaining examples in this vignette rely on one of the datasets, i.e. “squamous GDS2520.” Similar analyses may be reproduced for other datasets by selecting a different element of “GS-BenchMark.Dataset.names.”

> DataSetStudy = GSBenchMark.Dataset.names[[9]]
> print(DataSetStudy)

[1] "squamous_GDS2520"

> data(list=DataSetStudy)

The data consists of two variables: exprsdata and phenotypes. exprsdata consists of a gene expression matrix where the rows and columns represent genes and the samples respectively. GSReg requires the rownames of gene expression variable represent the gene names, i.e. they are represented in the pathway information variable.

The GSReg does not allow any missing data. To comply with the requirements we remove genes with NAs. The user may use any imputation to resolve this issue:

> if(sum(apply(is.nan(exprsdata),1,sum)>0))
   exprsdata = exprsdata[-which(apply(is.nan(exprsdata),1,sum)>0),];

One can extract the gene names by:

> genenames = rownames(exprsdata);
> genenames[1:10]

[1] "MAPK3"    "TIE1"    "CYP2C19"    "CXCR5"    "CXCR5"    "DUSP1"    "MMP10"    "DDR1"    
[9] "EIF2AK2"    "HINT1"

3 Analysis of the pathways

Here, we demonstrate how to use the GSReg package to compute DIRAC and EVA statistics.
3.1 DIRAC Analysis

First, we load the library:

> library(GSReg)

The package also implements the alternative EVA statistic in the function `GSReg.GeneSets.DIRAC`. This function receives gene expression as `geneexpres`, the pathway information as `pathways` and phenotypes of samples as a factor with two levels and length equal to column number of `geneexpres`. `DIRAC` uses a permutation test for p-value calculation; so, `GSReg.GeneSets.DIRAC` receives the number of permutations through (Nperm) with default value equal to 1000.

> Nperm = 10
> system.time({DIRACAn = GSReg.GeneSets.DIRAC(exprsdata, diracpathways, phenotypes, Nperm=Nperm)})

user  system elapsed
2.592 0.000 2.594

Here is the histogram of the DIRAC p-values:

> hist(DIRACAn$pvalues, xlab="pvalue", main="Hist of pvalues applying DIRAC Analysis.")
3.2 EVA

The package also implements the alternative EVA statistic in the function `GSReg.GeneSets.EVA`. The function requires the similar inputs as `GSReg.GeneSets.DIRAC` (i.e. `geneexpres`, `pathways`, `phenotypes`) except it does not need `Nperm` since the p-value is not calculated through permutation test but through the mentioned U-statistic theory approach.

> #Calculating the variance for the pathways
> #Calculate how much it takes to calculate the statistics and their p-value for all pathways
The output consists of a list. Each element of the list corresponds to a pathway. The element itself is a list. $E_1$ and $E_2$ are two fields which contain the measure of variability for phenotype levels(phenotypes)[1] and levels(phenotypes)[2] respectively. Other list elements are $pvalue$ and $zscore$ which are calculated through the theory of U-statistics and indicate the statistical significance of the difference between $E_1$ and $E_2$.

### 3.3 Comparison of DIRAC and EVA

We ran the following code to compare statistics from DIRAC and from EVA.

```r
> Nperm = 10;
> VarAnPerm = vector(mode="list",length=Nperm)
> for( i in seq_len(Nperm)) {
```
VarAnPerm[[i]] = GSReg.GeneSets.EVA(geneexpres=exprsdata, pathways=diracpathways,
    phenotypes=sample(phenotypes))

> pvaluesperm = vector(mode="numeric",length=length(VarAnPerm[[1]]))
> for( i in seq_along(VarAnPerm[[1]]))
  {
    z = sapply(VarAnPerm,function(x) x[[i]]$E1 - x[[i]]$E2)
    pvaluesperm[i] = mean(abs(VarAnKendallV[[i]]$E1-VarAnKendallV[[i]]$E2)<abs(z))
  }
> zscore = sapply(VarAnKendallV,function(x) x$zscore);
> pvalustat = sapply(VarAnKendallV,function(x) x$pvalue);

The figure represents that the theoretical p-value and p-value calculated from permutation test in EVA are very similar and we can use the theoretical p-value as a surrogate for p-value. Here is the histogram.

> hist(x=pvalustat,breaks=20,main="P-value Hist of U-Stat",xlim=c(0,1))
Figure 1: Comparing p-value from permutation test and U-statistic theory with only 10 permutations.
Figure 2 shows the result of comparing p-value EVA computing from 1000 permutation test and approximation using U-statistics theory (offline generated).

To compare with the p-value of the DIRAC analysis, we show the p-values of DIRAC versus U-Statistic methodology:

```r
> plot(x=DIRACAn$pvalues,y=pvalustat,xlab ="DIRAC",
      ylab="EVA",main=sprintf("P-value Comparison corr=\%2.2g",cor(x=DIRACAn$pvalues,y=pvalustat)))
> lmfit = lm(pvalustat~DIRACAn$pvalues-1)
> abline(lmfit)
> cor.test(x=DIRACAn$pvalues,y=pvalustat)
```
Figure 2: Theoretical p-value versus empirical p-value using 1000 permutations.
Pearson's product-moment correlation

data: DIRAC$n$ pvalues and pvalustat
t = 22.726, df = 238, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.7827537 0.8635317
sample estimates:
cor
0.8273743

Also, the correlation of the p-values of DIRAC and U-Statistics is very high:
If we use 1000 permutations instead of 10 permutations, we can see that the correlation is higher (0.88) as seen in Figure (3). The dysregulated pathways identified by DIRAC are the following pathways:

- "DEATHPATHWAY"
- "NEUTROPHILPATHWAY"
- "PGC1APATHWAY"
- "SKP2E2FPATHWAY"
- "KERATINOCYTEPATHWAY"
- "TGFBPATHWAY"
- "PROTEASOMEPATHWAY"
- "MAPKPATHWAY"
- "PDGFPATHWAY"
- "BIOPEPTIDESPATHWAY"
- "PYK2PATHWAY"
- "MYOSINPATHWAY"
- "IL7PATHWAY"
- "FMLPPATHWAY"
- "CD40PATHWAY"
- "CDC25PATHWAY"
- "RARRXRPATHWAY"
- "SKP2E2FPATHWAY"
- "KERATINOCYTEPATHWAY"
- "CHEMICALPATHWAY"
- "TGFBPATHWAY"
- "PROTEASOMEPATHWAY"
- "MAPKPATHWAY"
- "PDGFPATHWAY"
- "BIOPEPTIDESPATHWAY"
- "PYK2PATHWAY"
- "MYOSINPATHWAY"
- "IL7PATHWAY"
- "FMLPPATHWAY"
- "CD40PATHWAY"
- "CDC25PATHWAY"
- "RARRXRPATHWAY"
- "SKP2E2FPATHWAY"
- "KERATINOCYTEPATHWAY"
- "CHEMICALPATHWAY"
- "TGFBPATHWAY"
- "PROTEASOMEPATHWAY"
- "MAPKPATHWAY"
- "PDGFPATHWAY"
- "BIOPEPTIDESPATHWAY"
- "PYK2PATHWAY"
- "MYOSINPATHWAY"
- "IL7PATHWAY"
- "FMLPPATHWAY"
- "CD40PATHWAY"
- "CDC25PATHWAY"
- "RARRXRPATHWAY"
- "SKP2E2FPATHWAY"
- "KERATINOCYTEPATHWAY"
- "CHEMICALPATHWAY"
- "TGFBPATHWAY"
- "PROTEASOMEPATHWAY"
- "MAPKPATHWAY"
- "PDGFPATHWAY"
- "BIOPEPTIDESPATHWAY"
- "PYK2PATHWAY"
- "MYOSINPATHWAY"
- "IL7PATHWAY"
- "FMLPPATHWAY"
- "CD40PATHWAY"
- "CDC25PATHWAY"
- "RARRXRPATHWAY"
- "SKP2E2FPATHWAY"
- "KERATINOCYTEPATHWAY"
- "CHEMICALPATHWAY"
- "TGFBPATHWAY"
- "PROTEASOMEPATHWAY"
- "MAPKPATHWAY"
- "PDGFPATHWAY"
- "BIOPEPTIDESPATHWAY"
- "PYK2PATHWAY"
- "MYOSINPATHWAY"
- "IL7PATHWAY"
- "FMLPPATHWAY"
- "CD40PATHWAY"
- "CDC25PATHWAY"
- "RARRXRPATHWAY"
- "SKP2E2FPATHWAY"
- "KERATINOCYTEPATHWAY"
- "CHEMICALPATHWAY"
- "TGFBPATHWAY"
- "PROTEASOMEPATHWAY"
- "MAPKPATHWAY"
- "PDGFPATHWAY"
- "BIOPEPTIDESPATHWAY"
- "PYK2PATHWAY"
- "MYOSINPATHWAY"
- "IL7PATHWAY"
- "FMLPPATHWAY"
- "CD40PATHWAY"
- "CDC25PATHWAY"
- "RARRXRPATHWAY"
- "SKP2E2FPATHWAY"
- "KERATINOCYTEPATHWAY"
- "CHEMICALPATHWAY"
- "TGFBPATHWAY"
- "PROTEASOMEPATHWAY"
- "MAPKPATHWAY"
- "PDGFPATHWAY"
- "BIOPEPTIDESPATHWAY"
- "PYK2PATHWAY"
- "MYOSINPATHWAY"
- "IL7PATHWAY"
- "FMLPPATHWAY"
- "CD40PATHWAY"
- "CDC25PATHWAY"
- "RARRXRPATHWAY"
- "SKP2E2FPATHWAY"
- "KERATINOCYTEPATHWAY"
- "CHEMICALPATHWAY"
- "TGFBPATHWAY"
- "PROTEASOMEPATHWAY"
- "MAPKPATHWAY"
- "PDGFPATHWAY"
- "BIOPEPTIDESPATHWAY"
- "PYK2PATHWAY"
- "MYOSINPATHWAY"
- "IL7PATHWAY"
- "FMLPPATHWAY"
- "CD40PATHWAY"
- "CDC25PATHWAY"
- "RARRXRPATHWAY"
- "SKP2E2FPATHWAY"
- "KERATINOCYTEPATHWAY"
- "CHEMICALPATHWAY"
- "TGFBPATHWAY"
- "PROTEASOMEPATHWAY"
Figure 3: Comparing p-values EVA versus DIRAC. The correlation is 0.88.
> significantPathwaysDIRAC = names(DIRACAn$mu1)[which(DIRACAn$pvalues<0.05)];
> mu1 = DIRACAn$mu1[significantPathwaysDIRAC];
> mu2 = DIRACAn$mu2[significantPathwaysDIRAC];
> #The dysregulated pathways
> names(mu1)

[1] "DEATHPATHWAY"  "NEUTROPHILPATHWAY"  "PGC1APATHWAY"
[4] "RARRXRPATHWAY"  "SKP2E2FPATHWAY"  "KERATINOCYTEPATHWAY"
[7] "CHEMICALPATHWAY"  "TGFPATHWAY"  "PROTEASESOMEPATHWAY"
[10] "MAPKPATHWAY"  "PDGFPATHWAY"  "BIOPEPTIDESPATHWAY"
[13] "SPPAPATHWAY"  "PYK2PATHWAY"  "MYOSINPATHWAY"
[16] "BETAOXIDATIONPATHWAY"  "IL7PATHWAY"  "FMLPPATHWAY"
[19] "VITCBPATHWAY"  "CD4OPATHWAY"  "CDC25PATHWAY"
[22] "MTORPATHWAY"  "RNAPATHWAY"  "FBW7PATHWAY"
[25] "LYMPHOCTYPATHWAY"  "LAIRPATHWAY"  "HIVNEFPATHWAY"
[28] "ALKPATHWAY"  "P35ALZHEIMERSPATHWAY"  "MSPPATHWAY"
[31] "GSK3PATHWAY"  "RELAPATHWAY"  "METPATHWAY"
[34] "TNFR2PATHWAY"  "AT1RPATHWAY"  "FREEPATHWAY"
[37] "ARAPPATHWAY"  "MRPPATHWAY"  "P53HYPOXIAPATHWAY"
[40] "IL18PATHWAY"  "STRESSPATHWAY"  "MEF2DPATHWAY"
[43] "STAT3PATHWAY"  "HSP2PATHWAY"  "EPONFKBPATHWAY"
[46] "NKCELLSPATHWAY"  "MONOCYTEPATHWAY"  "CARM_ERPATHWAY"

> plot(x=mu1,y=mu2,
>     xlim=c(0,max(mu1,mu2)),ylim=c(0,max(mu1,mu2)),xlab="normal",ylab="disease",
>     main="(a) DIRAC significantly dysregulated pathways")
> lines(x=c(0,max(mu1,mu2)),y=c(0,max(mu1,mu2)))
Now, if we do the analysis using EVA, we have:

```r
> significantPathwaysGSV = names(which(pvalustat<0.05));
```

```
[1] "DEATHPATHWAY" "TCAPOPTOSISPATHWAY" "NEUTROPHILPATHWAY"
[4] "PGC1APATHWAY"  "TERCPATHWAY"  "RARRXRPATHWAY"
[7] "SKP2E2FPATHWAY" "KERATINOCYTEPATHWAY" "CHEMICALPATHWAY"
[10] "METHIONINEPATHWAY" "TGFPATHWAY"  "PS1PATHWAY"
[13] "PROTEASOMEPATHWAY" "CDK5PATHWAY"  "MAPKPATHWAY"
[16] "NTHIPATHWAY"   "PDGFPATHWAY"   "BIOPEPTIDESPATHWAY"
[19] "SPPAPATHWAY"   "PYK2PATHWAY"   "CDC42RACPATHWAY"
```
```
> eta1 = sapply(VarAnKendallV,function(x) x$E1)[significantPathwaysGSV];

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEATHPATHWAY</td>
<td>0.09441352</td>
</tr>
<tr>
<td>TCAPOPTOSISPATHWAY</td>
<td>0.08600289</td>
</tr>
<tr>
<td>NEUTROPHILPATHWAY</td>
<td>0.3559678</td>
</tr>
<tr>
<td>PGC1APATHWAY</td>
<td>0.04477053</td>
</tr>
<tr>
<td>TERCPATHWAY</td>
<td>0.07316017</td>
</tr>
<tr>
<td>RARXRPATHWAY</td>
<td>0.06914038</td>
</tr>
<tr>
<td>SKP2E2FAPATHWAY</td>
<td>0.03367003</td>
</tr>
<tr>
<td>KERATINOCYTEPATHWAY</td>
<td>0.07404055</td>
</tr>
<tr>
<td>CHEMICALPATHWAY</td>
<td>0.08521303</td>
</tr>
<tr>
<td>METHIONINEPATHWAY</td>
<td>0.09913420</td>
</tr>
<tr>
<td>TGFBPATHWAY</td>
<td>0.05028305</td>
</tr>
<tr>
<td>PS1PATHWAY</td>
<td>0.08080808</td>
</tr>
<tr>
<td>PROTEASOMEPATHWAY</td>
<td>0.09617180</td>
</tr>
<tr>
<td>CDK5PATHWAY</td>
<td>0.07975863</td>
</tr>
<tr>
<td>MAPKPATHWAY</td>
<td>0.08504987</td>
</tr>
<tr>
<td>NTHIPATHWAY</td>
<td>0.08091115</td>
</tr>
<tr>
<td>PDGFPATHWAY</td>
<td>0.08698709</td>
</tr>
<tr>
<td>BIOPEPTIDESPATWAY</td>
<td>0.09678077</td>
</tr>
<tr>
<td>SPPAPATHWAY</td>
<td>0.09690598</td>
</tr>
<tr>
<td>PYK2PATHWAY</td>
<td>0.04711514</td>
</tr>
<tr>
<td>CDC42RACPATHWAY</td>
<td>0.08948195</td>
</tr>
<tr>
<td>MYOSINPATHWAY</td>
<td>0.09005280</td>
</tr>
<tr>
<td>BETAOXIDATIONPATHWAY</td>
<td>0.02683983</td>
</tr>
<tr>
<td>IL7PATHWAY</td>
<td>0.10609668</td>
</tr>
<tr>
<td>FMLPATHWAY</td>
<td>0.05736961</td>
</tr>
<tr>
<td>FASPATHWAY</td>
<td>0.08676830</td>
</tr>
<tr>
<td>VITCBPATHWAY</td>
<td>0.07888408</td>
</tr>
<tr>
<td>CD40PATHWAY</td>
<td>0.05294705</td>
</tr>
<tr>
<td>IGF1PATHWAY</td>
<td>0.09115972</td>
</tr>
<tr>
<td>CDC25PATHWAY</td>
<td>0.06265031</td>
</tr>
<tr>
<td>MTORPATHWAY</td>
<td>0.04188827</td>
</tr>
<tr>
<td>RNAPATHWAY</td>
<td>0.03009689</td>
</tr>
<tr>
<td>FBW7PATHWAY</td>
<td>0.03174603</td>
</tr>
<tr>
<td>LYMHC10PATHWAY</td>
<td>0.22390872</td>
</tr>
<tr>
<td>LAMPATHWAY</td>
<td>0.15222872</td>
</tr>
<tr>
<td>P53PATHWAY</td>
<td>0.0920600</td>
</tr>
<tr>
<td>ALKPATHWAY</td>
<td>0.09130361</td>
</tr>
<tr>
<td>PEP1PATHWAY</td>
<td>0.05194805</td>
</tr>
<tr>
<td>MSPPATHWAY</td>
<td>0.12150072</td>
</tr>
<tr>
<td>EDG1PATHWAY</td>
<td>0.08801738</td>
</tr>
<tr>
<td>GS3PATHWAY</td>
<td>0.11946928</td>
</tr>
<tr>
<td>RELAPATHWAY</td>
<td>0.06302309</td>
</tr>
<tr>
<td>METPATHWAY</td>
<td>0.10561315</td>
</tr>
<tr>
<td>TNFR2PATHWAY</td>
<td>0.05051666</td>
</tr>
<tr>
<td>AT1RPATHWAY</td>
<td>0.05943314</td>
</tr>
<tr>
<td>ATRBRCAPATHWAY</td>
<td>0.07166907</td>
</tr>
<tr>
<td>GLYCOLYSISPATHWAY</td>
<td>0.02308802</td>
</tr>
<tr>
<td>TIDPATHWAY</td>
<td>0.07331013</td>
</tr>
<tr>
<td>EPopathway</td>
<td>0.07319696</td>
</tr>
<tr>
<td>WNTPATHWAY</td>
<td>0.07166907</td>
</tr>
<tr>
<td>ARAPPATHWAY</td>
<td>0.0.05988456</td>
</tr>
<tr>
<td>MRPPATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>P53PATHWAY</td>
<td>0.07708666</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.09375387</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.13015873</td>
</tr>
<tr>
<td>STRESSPATHWAY</td>
<td>0.05505364</td>
</tr>
<tr>
<td>P53HYPOX1A1PATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>EPopathway</td>
<td>0.07708666</td>
</tr>
<tr>
<td>WNTPATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>ARAPPATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>MRPPATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>P53PATHWAY</td>
<td>0.07708666</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.09375387</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.13015873</td>
</tr>
<tr>
<td>STRESSPATHWAY</td>
<td>0.05505364</td>
</tr>
<tr>
<td>P53HYPOX1A1PATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>EPopathway</td>
<td>0.07708666</td>
</tr>
<tr>
<td>WNTPATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>ARAPPATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>MRPPATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>P53PATHWAY</td>
<td>0.07708666</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.09375387</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.13015873</td>
</tr>
<tr>
<td>STRESSPATHWAY</td>
<td>0.05505364</td>
</tr>
<tr>
<td>P53HYPOX1A1PATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>EPopathway</td>
<td>0.07708666</td>
</tr>
<tr>
<td>WNTPATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>ARAPPATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>MRPPATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>P53PATHWAY</td>
<td>0.07708666</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.09375387</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.13015873</td>
</tr>
<tr>
<td>STRESSPATHWAY</td>
<td>0.05505364</td>
</tr>
<tr>
<td>P53HYPOX1A1PATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>EPopathway</td>
<td>0.07708666</td>
</tr>
<tr>
<td>WNTPATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>ARAPPATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>MRPPATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>P53PATHWAY</td>
<td>0.07708666</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.09375387</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.13015873</td>
</tr>
<tr>
<td>STRESSPATHWAY</td>
<td>0.05505364</td>
</tr>
<tr>
<td>P53HYPOX1A1PATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>EPopathway</td>
<td>0.07708666</td>
</tr>
<tr>
<td>WNTPATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>ARAPPATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>MRPPATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>P53PATHWAY</td>
<td>0.07708666</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.09375387</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.13015873</td>
</tr>
<tr>
<td>STRESSPATHWAY</td>
<td>0.05505364</td>
</tr>
<tr>
<td>P53HYPOX1A1PATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>EPopathway</td>
<td>0.07708666</td>
</tr>
<tr>
<td>WNTPATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>ARAPPATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>MRPPATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>P53PATHWAY</td>
<td>0.07708666</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.09375387</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.13015873</td>
</tr>
<tr>
<td>STRESSPATHWAY</td>
<td>0.05505364</td>
</tr>
<tr>
<td>P53HYPOX1A1PATHWAY</td>
<td>0.10526414</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
<td>0.05988456</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
<td>0.04877345</td>
</tr>
<tr>
<td>EPopathway</td>
<td>0.07708666</td>
</tr>
</tbody>
</table>
```

> eta2 = sapply(VarAnKendallV,function(x) x$E2)[significantPathwaysGSV];

> eta1 = sapply(VarAnKendallV,function(x) x$E1)[significantPathwaysGSV];

> eta2 = sapply(VarAnKendallV,function(x) x$E2)[significantPathwaysGSV];
## The dysregulated pathways

```
> names(eta1)
[1] "DEATHPATHWAY"  "TCAPOPTOSISPATHWAY"  "NEUTROPHILPATHWAY"  "PGC1APATHWAY"
[2] "TERCPATHWAY"  "RARRXRPATHWAY"  "SKP2E2FPATHWAY"  "KERATINOCTYPEPATHWAY"
[3] "CHEMICALPATHWAY"  "METHIONINEPATHWAY"  "TGFBPATHWAY"  "PS1PATHWAY"
[4] "PROTEASOMEPATHWAY"  "CDK5PATHWAY"  "MAPKPATHWAY"  "NTHIPATHWAY"
[5] "PDGFPATHWAY"  "BIOPEPTIDESPATHWAY"  "SPPAPATHWAY"  "PYK2PATHWAY"
[6] "CDC42RACPATHWAY"  "MYOSINPATHWAY"  "BETAOXIDATIONPATHWAY"  "IL7PATHWAY"
[7] "FMLPPATHWAY"  "FASPATHWAY"  "VITCBPATHWAY"  "CD40PATHWAY"
[8] "IGF1PATHWAY"  "CDC25PATHWAY"  "MTORPATHWAY"  "RNAPATHWAY"
[9] "ALKPATHWAY"  "PEPIPATHWAY"  "MSPPATHWAY"  "EDG1PATHWAY"
[10] "GSK3PATHWAY"  "RELAPATHWAY"  "METPATHWAY"  "TNFR2PATHWAY"
[11] "ATIRPATHWAY"  "ATRBCRAPATHWAY"  "GLYCOLYSISISPATHWAY"  "TIDPATHWAY"
[12] "EP0PATHWAY"  "WNTPATHWAY"  "ARAPPATHWAY"  "MRPPATHWAY"
[13] "P53HYPOX1APATHWAY"  "PITX2PATHWAY"  "IL18PATHWAY"  "STRESSPATHWAY"
[14] "MEF2DPATHWAY"  "MITOCHONDRIAPATHWAY"  "STAT3PATHWAY"  "EPONFKBPATHWAY"
[15] "NKCELLSPATHWAY"  "MONOCYTEPATHWAY"  "CARM_ERPATHWAY"  "HCMVPATHWAY"
```

### References

1. "DEATHPATHWAY"  "TCAPOPTOSISPATHWAY"  "NEUTROPHILPATHWAY"
2. "TERCPATHWAY"  "RARRXRPATHWAY"  "SKP2E2FPATHWAY"  "KERATINOCTYPEPATHWAY"
3. "CHEMICALPATHWAY"  "METHIONINEPATHWAY"  "TGFBPATHWAY"  "PS1PATHWAY"
4. "PROTEASOMEPATHWAY"  "CDK5PATHWAY"  "MAPKPATHWAY"  "NTHIPATHWAY"
5. "PDGFPATHWAY"  "BIOPEPTIDESPATHWAY"  "SPPAPATHWAY"  "PYK2PATHWAY"
6. "CDC42RACPATHWAY"  "MYOSINPATHWAY"  "BETAOXIDATIONPATHWAY"  "IL7PATHWAY"
7. "FMLPPATHWAY"  "FASPATHWAY"  "VITCBPATHWAY"  "CD40PATHWAY"
8. "IGF1PATHWAY"  "CDC25PATHWAY"  "MTORPATHWAY"  "RNAPATHWAY"
9. "ALKPATHWAY"  "PEPIPATHWAY"  "MSPPATHWAY"  "EDG1PATHWAY"
10. "GSK3PATHWAY"  "RELAPATHWAY"  "METPATHWAY"  "TNFR2PATHWAY"
11. "ATIRPATHWAY"  "ATRBCRAPATHWAY"  "GLYCOLYSISISPATHWAY"  "TIDPATHWAY"
12. "EP0PATHWAY"  "WNTPATHWAY"  "ARAPPATHWAY"  "MRPPATHWAY"
13. "P53HYPOX1APATHWAY"  "PITX2PATHWAY"  "IL18PATHWAY"  "STRESSPATHWAY"
14. "MEF2DPATHWAY"  "MITOCHONDRIAPATHWAY"  "STAT3PATHWAY"  "EPONFKBPATHWAY"
15. "NKCELLSPATHWAY"  "MONOCYTEPATHWAY"  "CARM_ERPATHWAY"  "HCMVPATHWAY"

---

**Note:** The dysregulated pathways are represented by their names and weights. The weights are not shown in the table but are important for further analysis. The list includes various pathways such as death, apoptosis, neutrophil, and others. The dysregulation of these pathways can be critical in understanding the underlying mechanisms of various biological processes or diseases. Further research and analysis are required to explore the implications of these dysregulated pathways in specific contexts.
<table>
<thead>
<tr>
<th>Pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPOPATHWAY</td>
</tr>
<tr>
<td>WNTPATHWAY</td>
</tr>
<tr>
<td>ARAPPATHWAY</td>
</tr>
<tr>
<td>MRPPATHWAY</td>
</tr>
<tr>
<td>P53HYPOXIAPATHWAY</td>
</tr>
<tr>
<td>PITX2PATHWAY</td>
</tr>
<tr>
<td>IL18PATHWAY</td>
</tr>
<tr>
<td>STRESSPATHWAY</td>
</tr>
<tr>
<td>MEF2DPATHWAY</td>
</tr>
<tr>
<td>MITOCHONDRIAPATHWAY</td>
</tr>
<tr>
<td>STAT3PATHWAY</td>
</tr>
<tr>
<td>EPONFKBPATHWAY</td>
</tr>
<tr>
<td>NKCELLSPATHWAY</td>
</tr>
<tr>
<td>MONOCYTEPATHWAY</td>
</tr>
<tr>
<td>CARM_ERPATHWAY</td>
</tr>
<tr>
<td>HCMVPATHWAY</td>
</tr>
</tbody>
</table>

```R
> plot(x=eta1,y=eta2,xlim=c(0,max(eta1,eta2)),ylim=c(0,max(eta1,eta2)),xlab="normal",ylab="disease",main="(b) EVA: Dysregulated pathways")

NULL

> lines(x=c(0,max(eta1,eta2)),y=c(0,max(eta1,eta2)))

NULL
```
Although there is discrepancy in identified dysregulated pathways (p-value < 0.05), the general trend found in [1] holds still true. The trend is that usually the dysregulated pathways have higher variability measure in more dangerous phenotypes. The figures reveal that both DIRAC and EVA have this property. DIRAC found 48 dysregulated pathways and EVA discovered 64 pathways, 45 pathways showed up in both analysis, and 67 pathways were discovered totally.

> print(significantPathwaysGSV)
> print(significantPathwaysDIRAC)

[1] "DEATHPATHWAY" "TCAPOPTOSISPATHWAY" "NEUTROPHILPATHWAY"
[4] "PGC1APATHWAY" "TERCPATHWAY" "RARRXRPATHWAY"
[7] "SKP2E2FPATHWAY" "KERATINOCYTEPATHWAY" "CHEMICALPATHWAY"
[10] "METHIONINEPATHWAY" "TGFBPATHWAY" "PS1PATHWAY"
[13] "PROTEASOMEPATHWAY" "CDK5PATHWAY" "MAPKPATHWAY"
[16] "NTHIPATHWAY" "PDGFPATHWAY" "BIOPEPTIDESPATHWAY"
[19] "SPAPPATHWAY" "PYK2PATHWAY" "CDC42RACPATHWAY"
[22] "MYOSINPATHWAY" "BETAOXIDATIONPATHWAY" "IL7PATHWAY"
[25] "FMLPPATHWAY" "FASPATHWAY" "VITCBPATHWAY"
[28] "CD40PATHWAY" "IGF1PATHWAY" "CDC25PATHWAY"
[31] "MOTORPATHWAY" "RNASPATHWAY" "FBW7PATHWAY"
[34] "LYMPHO CYTEPATHWAY" "LAIRPATHWAY" "HIVNEFPATHWAY"
[37] "ALKPATHWAY" "PEPIPATHWAY" "MSPPATHWAY"
[40] "EDG1PATHWAY" "GSK3PATHWAY" "RELAPATHWAY"
[43] "METPATHWAY" "TNFR2PATHWAY" "AT1RPATHWAY"
[46] "ATRBRCAPATHWAY" "GLYCOLYSIS PATHWAY" "TIDPATHWAY"
[49] "EPOPATHWAY" "WNTPATHWAY" "ARAPPATHWAY"
[52] "MRPPATHWAY" "P53HYPOXIAPATHWAY" "PITX2PATHWAY"
[55] "IL18PATHWAY" "STRESSPATHWAY" "MEF2DPATHWAY"
[58] "MITOCHONDRIAPATHWAY" "STAT3PATHWAY" "EPONFKBPATHWAY"
[61] "NKCELLSPATHWAY" "MONOCYTEPATHWAY" "CARM_ERPATHWAY"
[64] "HCMVPATHWAY"
4 System Information

Session information:

> toLatex(sessionInfo())

- R version 3.3.1 (2016-06-21), x86_64-pc-linux-gnu
- Locale: LC_CTYPE=en_US.UTF-8, LC_NUMERIC=C, LC_TIME=en_US.UTF-8, LC_COLLATE=C, LC_MONETARY=en_US.UTF-8, LC_MESSAGES=en_US.UTF-8, LC_PAPER=en_US.UTF-8, LC_NAME=C, LC_ADDRESS=C, LC_TELEPHONE=C, LC_MEASUREMENT=en_US.UTF-8, LC_IDENTIFICATION=C
- Base packages: base, datasets, grDevices, graphics, methods, stats, utils
- Other packages: GSBenchMark 0.107.0, GSReg 1.8.0
- Loaded via a namespace (and not attached): tools 3.3.1

5 Literature Cited

References

