Welcome to the TAPES wiki!

Have a look at the Quick start section.

You will find here a detailed explanation of each function TAPES has to offer.

- sort
- annotate
- db
- decompose
- analyse

You can also check this example of TAPES workflow.
Please check the warning section for a few caveats. ### Download

Either use the download button from the github website in a browser or use the command:

```
git clone https://github.com/a-xavier/tapes
```

Resolve dependencies

run `pip3 install -r requirements.txt --user` on Linux/macOS

run `pip3 install -r winrequirements.txt --user` on Windows

:warning:Depending on your distribution, installing **tkinter** might be necessary to generate graphs.
It can be called `python3-tk` for debian/ubuntu/fedora or just `tk` on arch-based linux. # `db` and ANNOVAR Database Management

db is the option to manage your ANNOVAR databases. First you need to download ANNOVAR.

First use

If you use TAPES to manage ANNOVAR databases for the first time you need to provide the location of your ANNOVAR folder using:

```
python3 tapes.py db -s -A /path/to/annovar/
```

This will create 2 useful files in the src folder : **db_config.json** and **db_vcf.json** in the src folder.

Simplified database download

You can download only the databases needed by TAPES using:

```
python3 tapes.py db -b --acmg -a hg19
```

Full database management

Using the file db_config.json, you can see which databases are downloaded and which are missing. To flag a database for download, open **db_config.json**

using any editor and replace “**MISSING**” by “**DOWNLOAD**” or “**DOWN**”.
then run:

```
python3 tapes.py db -b -a hg38
```

The databases flagged for download will all be downloaded

Options

- `-s` `--see_db`|*flag*|See which databases are present in the ANNOVAR folder

- `-b` `--build_db`|*flag*| Download flagged databases
- `--acmg` |*flag*| Bypasses the db_config.json file and download only necessary databases for TAPES
- `-a` `--assembly`|*string*| Either hg19 or hg38 - default is hg19

`annotate`

annotate will use ANNOVAR to annotate a vcf file using a simplified command line.

If you have not already, you should read the entry about database management first.
TAPES will accept **vcf**, **bcf**, **gzipped bcf** and **vcf** as well as **bgzipped vcf** as input for annotation. All formats will be **converted to decomposed vcf** prior to annotation.

Input file

`-i` `--input` TAPES will accept vcf, bcf, gzipped bcf and bgzipped vcf as input files for the `annotate` option ### Output file `-o` `--output` TAPES allow to output either annotated csv (comma separated), txt (tab separated) or vcf files.

WARNING If your vcf is multi-sample and your output is csv or txt, two files will be created. One file without the sample genotyping and one file with the sample genotyping data, which will have the suffix: ”_with_samples”

Simplified annotation

To annotate a vcf files using just the necessary annotation for TAPES, then run:

```
python3 tapes.py annotate -i /path/to/variant_file.bcf -o /path/to/annotated_file.vcf
--amcg
```

The `--acmg` tag will only use the necessary annotations for TAPES.

Full annotations and db_vcf.json

If you want to annotate your files with other annotations, open the file **db_vcf.json** with any editor. It will show every databases detected in your

annovar folder. Replace “**NO**” by “**YES**” for the databases you wish to use for annotation.
Then run:

```
python3 tapes.py annotate -i /path/to/variant_file.vcf -o /path/to/annotated_file.vcf
--ref_anno knownGene
```

OPTIONS

- `-a` `--assembly`|*string*| The assembly used. Either hg19 or hg38 - default = hg38
- `--acmg`|*flag*| Bypassed the **db__vcf.json** file and only annotate the variant file with the necessary annotations for TAPES sorting.
- `--ref_anno`|*string*| The reference annotation system used. Either ‘refGene’ for RefSeq, ‘ensGene’ for ENSEMBL or ‘knownGene’ for UCSC annotation - default = refGene

`sort`

sort is the main function of TAPES. It will prioritise variants and generate reports.

INPUT

`-i` or `--input` The input file will always be an ANNOVAR annotated file. It can be either: - A vcf file (multi or single sample) - A txt (tab separated) or csv (comma separated) file (ANNOVAR does not keep genotyping data in those format but you can use TAPES to annotate you original vcf files to the format of your liking)

OUTPUT

`-o` or `--output` The output file can be: - a txt (tab separated) or csv (comma separated) accompanied by an xlsx file - a folder containing either multiple txt (tab separated) or csv (comma separated) files.

EXAMPLES

Simple examples

```
python3 tapes.py sort -i /path/to/annotated_file.vcf -o /path/to/output.txt
```
for txt+xlsx output
```
python3 tapes.py sort -i /path/to/annotated_file.vcf -o /path/to/output.csv
```
for csv+xlsx output
```
python3 tapes.py sort -i /path/to/annotated_file.vcf -o /path/to/output/
```
for folder csv output
```
python3 tapes.py sort -i /path/to/annotated_file.vcf -o /path/to/output/
```

`--tab` for folder txt output ### Example of full command `python3 tapes.py sort -i /path/to/annotated_file.vcf -o /path/to/output/ --tab --acmg --by_sample --by_gene --enrichr --disease "autosomal dominant" --kegg "Pathways in cancer" --list "ATM APC MUTYH AKT1 CASP8 DMD MSH3 MSH6 MLH1 MSH2 POLE POLD POLQ KRAS" -a hg19 --BIG --trio ./trio.txt`

Reports

The different reports generated (if using the FOLDER mode) will have different suffixes appended to them. For example:

- report_**by_sample**.txt for the `--by_sample` option
- report_**GO_Biological_Process_2018**.txt for the `--enrichr` option using the GO_Biological_Process_2018 library

OPTIONS

- `-a --assembly` |*string*| The assembly used. Either hg19 or hg38 - default is hg19

- `--acmg`|*flag*| Will check for all necessary annotations before proceeding. Will exit the program if alle the necessary annotations were not found. Do not use this unless your annotation file is fully “TAPES compliant”

- `--BIG`|*0 to 1 float*|If your file is particularly big, remove variants that are not likely to be pathogenic from the output.Take the percentage of chance to be pathogenic as an argument - default = 0.35

- `--by_gene`|*flag*| Will generate a report of the mutational burden for each gene with associated samples for each variant.
- `--by_sample`|*flag*|Will generate a report grouping all pathogenic variants per sample

- `--cutoff`|*0 to 1 float*| Select the cutoff for “rare” variant. Used in ACMG BS1 - default = 0.005

- `-d --disease` |*string*| Will generate a report similar to the main report including only the variants involved in a particular disease. Based on the Disease column from ANNOVAR - default = “cancer”

- `-e --enrichr`|*string*| Implementation of the EnrichR API. Analyse top mutations (>0.85 in Pathology prediction) with EnrichR API to determine the pathways disrupted by pathogenic variants - default = GO_Biological_Process_2018 - See the EnrichR librairies page for the full list of available EnrichR Libraries

- `--kegg` |*string*|Similar to list but if you do not know all the genes involved in a given pathway. Will generate a report similar to the main report including only variants found in the genes related to a given pathway. See the KEGG pathways page for the full list of available pathways
- `--list` |*string* or *path*| Will generate a report similar to the main report including only variants found in genes list. Using gene symbols. Can use either a string containing gene symbols separated by a space and in quotes(eg. "MLH1 APC MLH2 BRCA2 PTEN") or a txt file with one gene symbol per line

- `-t --threads` |int| Number of threads to use for sorting | Still not optimised and require a large amount of ram (around 2Gb per thread + 2*size of the input file)
- `--tab` |*flag*| If the output mode is folder, outputs txt (tab separated) files instead of csv (comma separated)
- `--trio`|*path*| Path to a tab separated txt file. Use if you have trio data with unnaffected parents to detect de-novo mutations. Necessary for ACMG PS2 criteria. See the trio wiki page for more information.
- `--pp2_percent`|*0 to 100 int*|Threshold for PP2, considering PP2 positive if variant is missense in a gene where more than the threshold are pathogenic missense - default = 80

- `--pp2_min`|*int*| Number of minimum pathogenic missense variants in gene to consider PP2

- `--bp1_percent` |*0 to 100 int*|Threshold for BP1, considering BP1 positive if variant is missense in a gene where less than the threshold are pathogenic missense - default = 15

`analyse` or `analyze`

analyse allow users to generate secondary reports without having to generate the main report again, saving time.

Input

`-i --input`
Any kind of main report from TAPES either txt (tab separated) or csv (comma separated) file.

Output

`-o --output` Will output a txt (tab separated) or csv (comma separated) file containing the desired report

Examples

```
python tapes.py analyse -i /path/to/main_report.csv -o /path/to/new_report.csv --by_gene
python tapes.py analyze -i /path/to/main_report.csv -o /path/to/new_report.txt --by_sample
python tapes.py analyse -i /path/to/main_report.txt -o /path/to/new_report.txt --enrichr GO_Molecular_Function_2018
```

Warning

analyse will only work one report at a time. For example:
`python tapes.py analyse -i /path/to/main_report.csv -o /path/to/new_report.csv --by_gene --by_sample` will **not** work. ## Available Report Options See the entire details of options at the sort function page.
* `--by_gene` * `--by_sample` * `--list` and `--kegg` * `--enrichr` * `--disease`

Analysis strategies

Options such as `--disease`, `--list` and `--kegg` will output a file with a similar format to the main report. The output file will only keep genes passing a filter.

This means that those can be used with other options such as `--by_gene` or `--by_sample`.

For example:

```
python tapes.py analyse -i ./main_report.txt -o ./only_cancer_genes.txt --disease cancer
```

then

```
python tapes.py analyse -i ./only_cancer_genes.txt -o ./cancer_sample.txt --by_sample
```

Will give, for each sample, the 5 most pathogenic variant only in genes that are involved in cancer pathways. # `decompose`

decompose allow users to manually decompose their vcf files prior to annotation. Decomposing vcf files allow remove multi-alleles (eg. REF-A ALT-T,C,GT) loci from vcf files to have a 'one variant per line' file.

Usage

```
python tapes.py decompose -i ./input.vcf -o ./output_decomposed.vcf
```

Credits

The `decompose` function is using the vcf_parser module. # Setting up TAPES in a python Virtual Environment

Using python3 on Linux or windows:

- `cd` to the TAPES directory
- `python3 -m venv tapes_env` creating a virtual environment called 'tapes_env'
- `source env/bin/activate` Activate the virtual environment (`.\env\Scripts\activate` on windows)
- `pip3 install -r requirements.txt --user` to install dependencies (winrequirements.txt on windows)
- Use TAPES
- `deactivate` to leave the virtual environment # Quick Start (Choose a starting point)

Unannotated VCF file

I just want to use the necessary ANNOVAR annotations for TAPES sorting

- Download ANNOVAR
- `python3 tapes.py db -s -A /path/to/annovar/`

- `python3 tapes.py db -b --acmg --assembly hg19`
- `python3 tapes.py annotate -i /path/to/file.vcf -o /path/to/output.vcf --acmg -a hg19` ### I want to customise the ANNOVAR annotations used
- See detailed instructions on Database Management and Custom annotations

VEP or ANNOVAR annotated file

I want to sort and prioritise my annotated VCF file - `python3 tapes.py sort -i /to/annotated/file.vcf -o /to/output/folder/ --tab` ### I want to re-analyse TAPES main report - `python3 tapes.py analyse -i /to/main/report.txt -o /to/output/report.txt --single_option`

Using the toy dataset

try:

```
python3 tapes.py sort -i ./toy_dataset/toy_annovar_multi.vcf -o
./Reports/ --tab --by_gene --by_sample --enrichr --list "MLH1
MSH6 MSH2" --disease "autosomal dominant" --kegg "pathways in
cancer"
```

TAPES Workflow using the provided toy dataset

Here you can find an example for workflow starting from a ***multi-sample VCF*** file and a fresh ANNOVAR download

Using ANNOVAR wrapping

First, `cd` to the main TAPES folder.

First use:

```
python3 tapes.py db -s -A /path/to/annovar/
```

To determine the ANNOVAR folder.

Then:

```
python3 tapes.py db -b --acmg
```

To download necessary databases (this might take some time).

Then:

```
python3 tapes.py annotate -i ./toy_dataset/toy_multi.vcf -o
./annotated_multi.vcf --acmg
```

To annotate the original vcf.

Then either: *

```
python3 tapes.py sort -i ./annotated_multi.vcf -o
./toy_report/ --tab
```

To prioritise the annotated vcf and generate the main output.

or

-
  ```
  python3 tapes.py sort -i ./annotated_multi.vcf -o ./toy_reports/
  --tab --by_gene --by_sample --disease cancer --enrichr
  --list "TP53 BRCA1 BRCA2 PTEN KRAS" --trio ./toy_dataset/trio.txt
  --kegg "pathways in cancer"
  ```

To prioritise the annotated vcf, generate the main output and other few secondary reports

Without ANNOVAR WRAPPING

Try

- `python3 tapes.py sort -i ./toy_dataset/toy_annovar_multi.vcf -o ./toy_report/ --tab`

To prioritise the annotated vcf and generate the main output.

or

- `python3 tapes.py sort -i ./toy_dataset/toy_vep_multi.vcf -o ./toy_reports/ --tab --by_gene --by_sample --disease cancer --enrichr --list "TP53 BRCA1 BRCA2 PTEN KRAS" --trio ./toy_dataset/trio.txt   --kegg "pathways in cancer"`

To prioritise the annotated vcf, generate the main output and other few secondary reports

Main report

This is the report generated by the `sort` option. TAPES main report contains all annotations present in the input annotated VCF file. Several other columns will be added depending on the options used with the `sort` option.

Added columns

- **One column per sample containing the variant genotype (0/0, 0/1 or 1/1)**
- **WT count**: The number of homozygous wild type (Ref/Ref) individuals
- **Het count**: The number of heterozygous (Ref/Alt) individuals
- **Hom count**: The number of homozygous variant (Alt/Alt) individuals
- **PS2 column for each Trio** (if `--trio` was used): the de-novo status of the specific variant
- **Odds ratio**: the odds ratio calculation (only if at least 2 individuals in the cohorts are affected by a variant)
- **CI** : confidence interval associated with the odds ratio (95%)
- **p-value** : p-value associated with the Odds ratio calculation
- **PSV1_contrib / PS2_contrib / ... / BP7_contrib**: the status of the associated ACMG criteria for this variant (0 or 1)
- **Probability_Path**: The probability that this variant is pathogenic
- **Prediction_ACMG_tapes** : the ACMG category assigned by TAPES # –by_gene report

By using the tag `--by_gene` with the `analyse` or `sort` option, you will generate a report that will group variants by gene.

Each gene will be assigned a "mutational burden" score that is useful to assess how much this gene is affected by variants in a cohort (if the input file is a multi-sample vcf).

The gene burden score is the sum of the probability of pathogenicity (**0.8 and above** otherwise excluded) or variant **i** multiplied by the number of individuals affected by variant **i**.

A warning will be associated if the genes is a FLAG genes (frequently mutated in exome studies), if the number of sample affected seems to be to high (half of the variants in a gene with more than half of the individuals affected) or if the gene is especially long (more than 250,000bp, which are expected to harbour more variants).

This report is useful to detect genes frequently mutated in a cohort sharing a same phenotype (but where several different variants are present).

Example

ERC2 502726

Chr	Start	End	Ref	Alt	Allele	Exonic_function	Probability	Pathogenic_ACMG	Samples
19	45867550	45867550	C	T	T	missense_variant	0.9993	Pathogenic	Sample_12, Sample_15, Sample_18, Sample_21, Sample_24, Sample_27, Sample_28, Sample_34, Sample_35, Sample_37, Sample_38, Sample_40, Sample_41, Sample_42, Sample_45, Sample_47
19	45856371	45856371	G	A	A	missense_variant	0.9993	Pathogenic	Sample_12, Sample_15, Sample_18, Sample_21, Sample_24, Sample_27, Sample_28, Sample_34, Sample_35, Sample_37, Sample_38, Sample_40, Sample_41, Sample_42, Sample_45, Sample_47
19	45860626	45860626	G	C	C	missense_variant	0.9941	Pathogenic	Sample_12
19	45867532	45867532	C	T	T	missense_variant	0.9941	Pathogenic	Sample_16, Sample_36
19	45855808	45855807	CTG	G	G	frameshift_variant	0.9941	Likely Pathogenic	Sample_27
19	45855574	45855574	G	A	A	missense_variant	0.9878	Likely Pathogenic	Sample_16, Sample_36
19	45860760	45860760	C	T	T	missense_variant	0.9878	Likely Pathogenic	Sample_3, Sample_23, Sample_31
19	45856059	45856059	C	G	G	missense_variant	0.8999	Likely Pathogenic	
19	45860760	45860760	C	T	T	missense_variant	0.8121	VUS	Sample_21
19	45858047	45858047	C	T	T	missense_variant	0.8121	VUS	Sample_26

ERCC2	50726						
19	45868168	45868168	C	C	missense_variant	0.8121	VUS	Sample_3, Sample_21, Sample_23, Sample_27, Sample_31, Sample_32, Sample_45, Sample_46
19	45867139	45867139	C	C	missense_variant	0.8121	VUS	Sample_21

– by_sample report

By using the tag `--by_sample` with the `analyse` or `sort` option, you will generate a report that will show for each sample the 5 most probably pathogenic variants.

This obviously requires multi-sample vcf.

Example

Sample_9									
Chr	Start	End	Ref	Alt	Allele	ExonicFunc.refGene	Gene.refGene	Probability_Pathogenic	Prediction_ACMG_tapes
17	7574017	7574017	C	T	T	missense_variant	TP53	0.9999	Pathogenic
17	7574017	7574017	C	T	T	missense_variant	TP53	0.9999	Pathogenic
13	32953932	32953932	T	A	A	stop_gained	BRCA2	0.9999	Pathogenic
17	7574017	7574017	C	T	T	missense_variant	TP53	0.9999	Pathogenic
14	45606305	45606305	G	A	A	stop_gained	FANCM	0.9998	Pathogenic
Sample_12									
Chr	Start	End	Ref	Alt	Allele	ExonicFunc.refGene	Gene.refGene	Probability_Pathogenic	Prediction_ACMG_tapes
13	32968863	32968863	C	G	G	stop_gained	BRCA2	1	Pathogenic
17	59793412	59793412	G	A	A	stop_gained	BRIP1	1	Pathogenic
17	7574017	7574017	C	T	T	missense_variant	TP53	0.9999	Pathogenic
17	7574017	7574017	C	T	T	missense_variant	TP53	0.9999	Pathogenic
17	7574017	7574017	C	T	T	missense_variant	TP53	0.9999	Pathogenic

–enrichr

(default is: GO_Biological_Process_2018)

usage :

```
python3 tapes.py sort -i ./input.vcf -o ./output_folder/ --enrichr
GO_Molecular_Function_2018
python3 tapes.py sort -i ./input.vcf -o ./output_folder/ -e
GO_Molecular_Function_2018
```

By using the option `-e` or `--enrichr` (see here for possible libraries) with the `analyse` or `sort` option, you will generate a report containing a pathway analysis of the genes containing probably pathogenic variants.

This generates a list of genes containing at least one probably pathogenic variant (probability of pathogenicity >= 0.85) and perform a pathway analysis using this gene-list.

Return the 11 most relevant pathways.

Example

Name	P-value	Z-score	Combined score	Genes	Adjusted p-value
DNA repair (GO:0006281)	8.6317[illegible] 36	[illegible]	[illegible]	[[illegible] 'XPC', 'BRCA1', 'BRCA2', 'BRIP1', 'WRN', 'PMS2', 'MUTYH', 'MEN1', 'FANCI', 'FANCM', 'FANCL', 'FANCA', 'FANCC', 'MLH1', 'FANCE', 'PALB2', 'FANCG', 'FANCF', 'DDB2', 'MSH6', 'RAD51C', 'MSH2', 'ERCC3', 'FANCD2', 'ERCC4', 'ERCC2', 'ATM', 'TP53']	4.40475870923533E-32
DNA metabolic process (GO:0006259)	6.5041[illegible] 25	[illegible]	[illegible]	[[illegible] 'FANCL', 'FANCA', 'FANCC', 'XPC', 'BRCA1', 'BRCA2', 'PALB2', 'FANCG', 'DDB2', 'MSH6', 'BRIP1', 'WRN', 'RAD51C', 'MSH2', 'TERT', 'ERCC3', 'FANCD2', 'ERCC4', 'ATM', 'TP53', 'MUTYH', 'MEN1']	1.65954273197225E-21
cellular response to DNA damage stimulus (GO:0006974)	1.8143[illegible] 21	[illegible]	[illegible]	[[illegible] 'FANCL', 'FANCA', 'FANCC', 'XPC', 'BRCA1', 'FANCG', 'DDB2', 'MSH6', 'BRIP1', 'WRN', 'RAD51C', 'MSH2', 'APC', 'ERCC3', 'FANCD2', 'ERCC4', 'ATM', 'TP53', 'MUTYH', 'MEN1']	3.08615764825346E-18
interstrand cross-link repair (GO:0036297)	7.6509[illegible] 16	[illegible]	[illegible]	[[illegible] 'FANCM', 'FANCD2', 'FANCL', 'ERCC4', 'FANCA', 'FANCC', 'FANCE', 'FANCG', 'FANCF']	9.76066382065821E-13
DNA biosynthetic process (GO:0071897)	8.8058[illegible] 15	[illegible]	[illegible]	[[illegible] 'BRIP1', 'WRN', 'RAD51C', 'TERT', 'ATM', 'BRCA1', 'BRCA2', 'PALB2']	8.98723664024476E-12

Name	P-value	Z-score	Combined score	Genes	Adjusted p-value
strand displacement (GO:0000732)	2.45906[illegible]-14	[illegible]	[illegible]	[illegible]RIP1’, ‘WRN’, ‘RAD51C’, ‘ATM’, ‘BRCA1’, ‘BRCA2’, ‘PALB2’]	2.09143563280303E-11
response to UV (GO:0009411)	1.27104[illegible]-12	[illegible]	[illegible]	[illegible]WRN’, ‘MSH2’, ‘ERCC3’, ‘ERCC4’, ‘ERCC2’, ‘ERCC5’, ‘TP53’, ‘DDB2’, ‘MEN1’]	9.26593034091148E-10
response to ionizing radiation (GO:0010212)	5.3087[illegible]-12	[illegible]	[illegible]	[illegible]AD51C’, ‘MSH2’, ‘FANCD2’, ‘ERCC4’, ‘ATM’, ‘BRCA1’, ‘TP53’, ‘MEN1’]	3.38633399416261E-09
DNA recombination (GO:0006310)	1.45632[illegible]-11	[illegible]	[illegible]	[illegible]RIP1’, ‘WRN’, ‘RAD51C’, ‘ATM’, ‘BRCA1’, ‘BRCA2’, ‘PALB2’]	8.25738942037681E-09
negative regulation of cell proliferation (GO:0008285)	1.72767[illegible]-11	[illegible]	[illegible]	[illegible]TSC2’, ‘TSC1’, ‘GATA2’, ‘APC’, ‘WT1’, ‘NF1’, ‘NF2’, ‘VHL’, ‘RAF1’, ‘TP53’, ‘HRAS’, ‘BAP1’, ‘MEN1’]	8.81634387742703E-09
double-strand break repair (GO:0006302)	4.83636[illegible]-11	[illegible]	[illegible]	[illegible]RIP1’, ‘WRN’, ‘RAD51C’, ‘MSH2’, ‘ERCC4’, ‘ATM’, ‘BRCA1’, ‘BRCA2’, ‘PALB2’]	2.24363406392391E-08

–list

Used to filter the main report based on user-supplied genes. usage:

```
python3 tapes.py sort -i ./input.vcf -o ./output_folder/ --list
"MLH1 PMS2 MSH6 MSH2 APC"
```

```
python3 tapes.py sort -i ./input.vcf -o ./output_folder/ --list
./gene_list.txt
```

By using the option `--list` with the `sort` or `analyse` option, a report containing only variants located in the user-supplied list of genes.

`--list` requires gene SYMBOLS, either **in quotes separated by a space** or as a **.txt file with one gene symbol per lines**.

The format of this report will be identical to the main report. # –disease #### Used to filter the main report based on the associated disease description.

usage:

```
python3 tapes.py sort -i ./input.vcf -o ./output_folder/ --disease
"autosomal recessive"
python3 tapes.py sort -i ./input.vcf -o ./output_folder/ -d
"folate metabolism"
```

By using the option `--disease` or `-d` with the `sort` or `analyse` option, a report will be created containing only variant with the search term in their “disease” column.

User can use a variety of terms such as “autosomal dominant”, “multiple sclerosis”, “cancer”, etc.

Quotes are expected if the search term contains multiple words.

The format of this report will be identical to the main report. # –kegg

Used to filter the main report based on functional kegg (Kyoto Encyclopedia of Genes and Genome) pathways. usage:

```
python3 tapes.py sort -i ./input.vcf -o ./output_folder/ --kegg
"fc epsilon ri signaling pathway"
```

By using the option `--kegg` with the `sort` or `analyse` option, a report containing only variants located in the genes contained in the designated pathway will be created.

A list of KEGG patwhays can be found here.
The format of this report will be identical to the main report. # PS4/Enrichment calculation

TAPES can calculates the PS4 ACMG criteria (enrichment of a variant in a diseased population vs general population) without the need of sequenced controls.

Calculating Odds Ratio using Fisher’s exact test require integers while most of the databases only gives minor allele frequencies. ## OLD METHOD By assuming that most very rare variants are heterozygous, we extrapolate the number of individuals with the variant and the number of individuals without the variant in public databases (ExAC or gNomad).

If MAF in the controls population
MAF = y^-x
then the number of individuals with the variant is
n = ceil(y)
and the number of individuals without the variant is **N = (x/2) - n**

If MAF = 3.23^-5 then n = 4 and N = 49996.

Then Odds ratios are calculated using Fisher's exact test (one sided to test for enrichment only).

NEW METHOD

To compensate for the difference between extrapolating from MAF and using exact number of individuals from public databases, TAPES now first calculate OR the old method.

Then another OR is calculated using **n = ceil(y/3)*****10 and N = (((x/2) - n)*****10) - n**

The final OR is a mean between the OLD and NEW OR calculation.

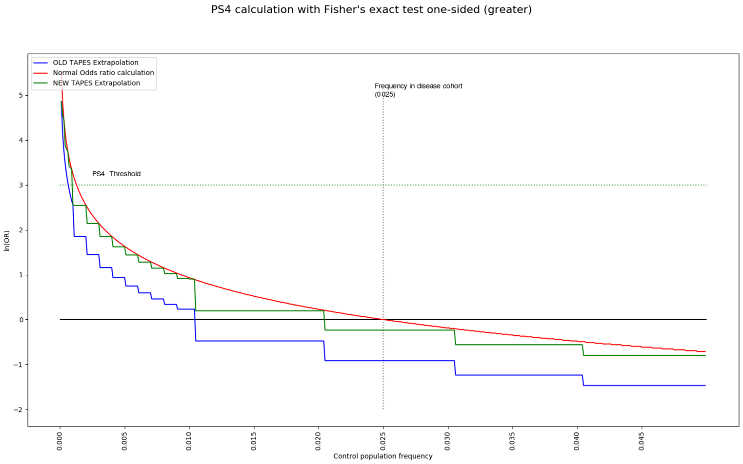
Graphs

Odds Ratio (OR) calculation (for PS4 criteria) was recently changed to be closer to the reality. In short the old extrapolation from MAF is first calculated and then another OR calculation is made using a smaller frequency in the control population. Then a mean between the 2 results is calculated. This represents (around) 30% less difference between the extrapolated OR and the reality. Keep in mind that TAPES OR calculation will always be more stringent than the normal calculation to avoid excessive false positive, meaning OR will (nearly) always be lower (only enrichment is tested). (see graph and table below)

Affected in general population	Unaffected in general population	Frequency	Normal OR calculation	Old TAPES extrapolation	New TAPES extrapolation
1	9999	1.00E-04	256.38	128.18	128.165
2	9998	2.00E-04	128.18	64.08	96.105
3	9997	3.00E-04	85.44	42.71	85.405
4	9996	4.00E-04	64.08	32.03	48.03
5	9995	5.00E-04	51.26	25.62	44.815
6	9994	6.00E-04	42.71	21.34	42.67
7	9993	7.00E-04	36.6	18.29	30.47
8	9992	8.01E-04	32.03	16	29.32
9	9991	9.01E-04	28.46	14.22	28.425
99	9901	1.00E-02	2.56	1.26	2.5
100	9900	1.01E-02	2.54	1.26	2.465
101	9899	1.02E-02	2.51	1.26	2.465
102	9898	1.03E-02	2.49	1.26	2.465
103	9897	1.04E-02	2.46	1.26	2.465
104	9896	1.05E-02	2.44	1.26	2.465
105	9895	1.06E-02	2.42	0.62	1.215

Affected in general population	Unaffected in general population	Frequency	Normal OR calculation	Old TAPES extrapolation	New TAPES extrapolation
106	9894	1.07E-02	2.39	0.62	1.215
107	9893	1.08E-02	2.37	0.62	1.215

Assuming a frequency of 0.025 in the case cohort



# ACMG Criteria Assignment

Pathogenic Criteria

PVS1

Will be assigned to a variant if it is a stopgain or frameshift deletion/insertion located 50 bp further than the end of the final exon. (Based on the ExonicFunc column anf the REK_canon library) Will be assigned to a splicing variant with a dbscSNV score of more than 0.6 (ADA or RF) (Based on the Func column and the dbscSNV score annotation)

PS1

Will be assigned if a variant have the same AA ref and AA alt as a known pathogenic variant. Using all known pathogenic variants from clinvar

PS2

Will be assigned if a variant is assumed de novo and parents are disease free. This requires trio data. See the dedicated wiki page for more informations on trio data.

PS3

Will be assigned if clinvar classifies the variant as Pathogenic or drug reponse and the level of evidence is either ‘practice guideline’ or ‘reviewed by expert panel’

PS4

Will be assigned if a variant is enriched in the samples provided. Requires either ‘output_with_samples.csv’ from the annotation to keep sample genotyping data or an annotated multi-sample vcf. PS4 will take the affected individuals with the mutations and the total number of individuals in the disease cohort and compare it to the data from gnomad_genome and gnomad_exome. The number of individuals with and without variants in public data is extrapolated with the following formula: Minor allele frequency in control population (MAF) = MAF_c=y× 10 ^(-x) Number of individuals with the variant in control population = n_c= y Total number of individuals in control population = N_c= 10 ^x/2-n_c Then a fisher’s exact test is performed to calculate the odd ratios, the confidence interval and the p value. PS4 will only be considered if at least 2 samples are affected by a variant. Otherwise, Intervar PS4 database, based on GWAS database will be used. PS4 will be assigned if the Odd Ratio is superior to 20, the confidence interval does not cross one and the p value is under 0.01

PM1

Will be assigned if the variant is a Missense variant (nonsynonymous SNV) and is located in a in a domain without benign variants (Using Intervar db) for benign domains

PM2

Will be assigned if the variant is in a recessive gene and has a frequency under 0.005 or is in a dominant gene and has no frequency data available. Recessive and Dominant/Haploinsufficient genes were infered using Pli and Prec scores computed by Lek et al, 2016. A gene is considered dominant dominant with a pli >0.85 and recessive if prec >0.85

PM4

Will be assigned if the variant is an in-frame deletion/insertion in a non-repeat region of the gene. Using the repeat_dict database.

PM5

Will be assigned if a variant have the same AA ref and a different AA alt as a known pathogenic variant. Using all known pathogenic variants from clinvar

PP2

Will be assigned if the variant is Missense (nonsynonymous SNV) in a gene where missense variants represents at least 80 percent of all known pathogenic variants (using PP2_BP1 database)

PP3

Will be assigned if the variant is predicted to be pathogenic using various in-silico prediction tools (sift, lrt, mutationtaster, mutation assessor, fathmm, provean, meta svm, meta lr, mcap, mkl, genocanyon, gerp). Each prediction tool will add +1 or -1. A score over 3 will assign PP3 to a variant.

PP5

Will be assigned the variant is classified as pathogenic or likely pathogenic by clinvar but the evidence is limited.

Benign criteria

BA1

Will be assigned to a variant if its frequency in gnomad_exome/exac or gnomad_genome is superior to 0.05

BS1

Will be assigned to a variant if its frequency is superior to a cutoff (0.005) for a rare disease.

BS2

Will be assigned if the variant was observed in a healthy individual as homozygous for a recessive disease and heterozygous for a dominant disease. (Using Intervar db BS2_hom_het)

BS3

Will be assigned if clinvar classifies the variant as Benign or likely benign and the level of evidence is either ‘practice guideline’ or ‘reviewed by expert panel’

BP1

Will be assigned if the variant is Missense (nonsynonymous SNV) in a gene where missense variants represents at most 10 percent of all known pathogenic variants (using PP2_BP1 database).

BP3

Will be assigned if the variant is an in-frame deletion/insertion in a repeat region of the gene. (Using the repeat_dict database).

BP4

Will be assigned if the variant is predicted to be benign using various in-silico prediction tools (sift, lrt, mutationtaster, mutation assessor, fathmm, provean, meta svm, meta lr, mcap, mkl, genocanyon, gerp). Each prediction tool will add +1 or -1. A score of 0 or under will assign BP4 to a variant.

BP6

Will be assigned the variant is classified as Benign or likely benign by clinvar but the evidence is limited.

BP7

Will be assigned if a variant if synonymous and no splicing impact is predicted by dbscSNV (score under 0.6)

Implementation

TAPES was written in python3.

It was thoroughly tested on both Linux (Ubutun, Manjaro and Fedora) and Windows. Due to the lack of access to it, TAPES was not tested on macOS. However, Travis CI shows that the main sort function does work on it.

TAPES used Travis CI for continuous integration. Please refer to this page for build history.

You can check TAPES major releases here. # `--trio`

The `--trio` option allows TAPES to assign the PS2 criteria, detecting de-novo mutations from offspring in trios with healthy parents.

A few clarifications on `--trio` :
- `--trio` will remove from the final report the "healthy" parents in trios - You **can** use the same sample in different trios (for example in a case of multiple siblings) but you probably should not use an individual as case offspring and also control parent, as it would mean the individual in not a "healthy" parent. - In the main report, a new column for each trio will be created using the family ID. **The PS2 criteria will be assigned globally if one of the variant is de-novo in any trio in the cohort**. ## Warnings

Here are a few things to keep in mind when using TAPES

ACMG classification and Pathogenicity prediction

The ACMG criteria and classification (from Benign to Pathogenic or class 1 to 5) reflect the **PROBABILITY** of a variant to be pathogenic. A variant of class 4 (Likely Pathogenic) is **NOT MORE** pathogenic than a variant of class 5 (Pathogenic). Instead, it is just **MORE LIKELY** to be pathogenic.

The same goes with the ACMG criteria modeling from Tavtigian et al that TAPES uses.

This has implications in the `--by_gene` score used in TAPES. This score is the sum of all probabilities multiplied by the number of sample affected. It reflect the probability of the genes to be affected by a pathogenic variant in the cohort.

ANNOVAR Necessary annotations

TAPES need a number of annotations to give the best possible estimation of a variant pathogenicity. You can use the `--acmg` flag when annotating with TAPES to directly use them.

The necessary/recommended databases are the following:

Annotation Required	ACMG Classification Criteria
Gene reference annotation (Refseq, ENSEMBL or UCSC)	All
dbscSNV	PVS1
gnomad_genome and either gnomad_exome or exac	PS1
Clinvar (20151201 or above)	PS3 / PP5 / BP6
DbNSFP (30a or above)	PP3 /BP4
dbsnp or avsnp	PS4

If one or more these annotations are missing. Do not use the `--acmg` flag while sorting.
TAPES will prioritise the variant using as many annotation as possible but will obviously be less powerfull.

VEP Necessary annotations

TAPES processing of VEP annotated vcf relies heavily on dbNSFP annotations. RefSeq reference annotations are recommended for VEP annotated vcf processing.

Necessary dbNSFP annotations : - Interpro (for domain annotation) - for in-silico prediction: - SIFT - LRT - MutationTaster - MutationAssessor - FATHMM - fathmm-MKL - PROVEAN - MetaSVM and MetaLR - M-CAP - GenoCanyon - GERP++ - Polyphen-2 - clinvar - gnomad_exome **or** ExAC - gnomad_genome - dbscSNV

Kegg pathways keys

- 2-oxocarboxylic acid metabolism
- abc transporters
- acute myeloid leukemia
- adherens junction
- adipocytokine signaling pathway
- adrenergic signaling in cardiomyocytes
- african trypanosomiasis
- age-rage signaling pathway in diabetic complications
- alanine, aspartate and glutamate metabolism
- alcoholism
- aldosterone synthesis and secretion
- aldosterone-regulated sodium reabsorption
- allograft rejection
- alpha-linolenic acid metabolism
- alzheimer disease
- amino sugar and nucleotide sugar metabolism
- aminoacyl-trna biosynthesis
- amoebiasis
- amphetamine addiction
- ampk signaling pathway
- amyotrophic lateral sclerosis
- antifolate resistance
- antigen processing and presentation
- apelin signaling pathway
- apoptosis
- apoptosis - multiple species
- arachidonic acid metabolism
- arginine and proline metabolism
- arginine biosynthesis

- arrhythmogenic right ventricular cardiomyopathy
- ascorbate and aldarate metabolism
- asthma
- autoimmune thyroid disease
- autophagy - animal
- autophagy - other
- axon guidance
- b cell receptor signaling pathway
- bacterial invasion of epithelial cells
- basal cell carcinoma
- basal transcription factors
- base excision repair
- beta-alanine metabolism
- bile secretion
- biosynthesis of amino acids
- biosynthesis of unsaturated fatty acids
- biotin metabolism
- bladder cancer
- breast cancer
- butanoate metabolism
- c-type lectin receptor signaling pathway
- caffeine metabolism
- calcium signaling pathway
- camp signaling pathway
- carbohydrate digestion and absorption
- carbon metabolism
- cardiac muscle contraction
- cell adhesion molecules
- cell cycle
- cellular senescence
- central carbon metabolism in cancer

- cgmp-pkg signaling pathway
- chagas disease
- chemical carcinogenesis
- chemokine signaling pathway
- cholesterol metabolism
- choline metabolism in cancer
- cholinergic synapse
- chronic myeloid leukemia
- circadian entrainment
- circadian rhythm
- citrate cycle
- cocaine addiction
- collecting duct acid secretion
- colorectal cancer
- complement and coagulation cascades
- cortisol synthesis and secretion
- cushing syndrome
- cysteine and methionine metabolism
- cytokine-cytokine receptor interaction
- cytosolic dna-sensing pathway
- d-arginine and d-ornithine metabolism
- d-glutamine and d-glutamate metabolism
- dilated cardiomyopathy
- dna replication
- dopaminergic synapse
- drug metabolism - cytochrome p450
- drug metabolism - other enzymes
- ecm-receptor interaction
- egfr tyrosine kinase inhibitor resistance
- endocrine and other factor-regulated calcium reabsorption
- endocrine resistance

- endocytosis
- endometrial cancer
- epithelial cell signaling in helicobacter pylori infection
- epstein-barr virus infection
- erbb signaling pathway
- estrogen signaling pathway
- ether lipid metabolism
- fanconi anemia pathway
- fat digestion and absorption
- fatty acid biosynthesis
- fatty acid degradation
- fatty acid elongation
- fatty acid metabolism
- fc epsilon ri signaling pathway
- fc gamma r-mediated phagocytosis
- ferroptosis
- fluid shear stress and atherosclerosis
- focal adhesion
- folate biosynthesis
- foxo signaling pathway
- fructose and mannose metabolism
- gabaergic synapse
- galactose metabolism
- gap junction
- gastric acid secretion
- gastric cancer
- glioma
- glucagon signaling pathway
- glutamatergic synapse
- glutathione metabolism
- glycerolipid metabolism

- glycerophospholipid metabolism
- glycine, serine and threonine metabolism
- glycolysis / gluconeogenesis
- glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate
- glycosaminoglycan biosynthesis - heparan sulfate / heparin
- glycosaminoglycan biosynthesis - keratan sulfate
- glycosaminoglycan degradation
- glycosphingolipid biosynthesis - ganglio series
- glycosphingolipid biosynthesis - globo and isoglobo series
- glycosphingolipid biosynthesis - lacto and neolacto series
- glycosylphosphatidylinositol
- glyoxylate and dicarboxylate metabolism
- gnrh signaling pathway
- graft-versus-host disease
- hedgehog signaling pathway
- hematopoietic cell lineage
- hepatitis b
- hepatitis c
- hepatocellular carcinoma
- herpes simplex infection
- hif-1 signaling pathway
- hippo signaling pathway
- hippo signaling pathway - multiple species
- histidine metabolism
- homologous recombination
- human cytomegalovirus infection
- human immunodeficiency virus 1 infection
- human papillomavirus infection
- human t-cell leukemia virus 1 infection
- huntington disease
- hypertrophic cardiomyopathy

- il-17 signaling pathway
- inflammatory bowel disease
- inflammatory mediator regulation of trp channels
- influenza a
- inositol phosphate metabolism
- insulin resistance
- insulin secretion
- insulin signaling pathway
- intestinal immune network for iga production
- jak-stat signaling pathway
- kaposi sarcoma-associated herpesvirus infection
- legionellosis
- leishmaniasis
- leukocyte transendothelial migration
- linoleic acid metabolism
- lipoic acid metabolism
- long-term depression
- long-term potentiation
- longevity regulating pathway
- longevity regulating pathway - multiple species
- lysine degradation
- lysosome
- malaria
- mannose type o-glycan biosynthesis
- mapk signaling pathway
- maturity onset diabetes of the young
- measles
- melanogenesis
- melanoma
- metabolic pathways
- metabolism of xenobiotics by cytochrome p450

- micrornas in cancer
- mineral absorption
- mismatch repair
- mitophagy - animal
- morphine addiction
- mrna surveillance pathway
- mtor signaling pathway
- mucin type o-glycan biosynthesis
- n-glycan biosynthesis
- natural killer cell mediated cytotoxicity
- necroptosis
- neomycin, kanamycin and gentamicin biosynthesis
- neuroactive ligand-receptor interaction
- neurotrophin signaling pathway
- nf-kappa b signaling pathway
- nicotinate and nicotinamide metabolism
- nicotine addiction
- nitrogen metabolism
- nod-like receptor signaling pathway
- non-alcoholic fatty liver disease
- non-homologous end-joining
- non-small cell lung cancer
- notch signaling pathway
- nucleotide excision repair
- lfactory transduction
- ne carbon pool by folate
- cyte meiosis
- steoclast differentiation
- ther glycan degradation
- ther types of o-glycan biosynthesis
- varian steroidogenesis

- xidative phosphorylation
- xytocin signaling pathway
- p53 signaling pathway
- pancreatic cancer
- pancreatic secretion
- pantothenate and coa biosynthesis
- parathyroid hormone synthesis, secretion and action
- parkinson disease
- pathogenic escherichia coli infection
- pathways in cancer
- pentose and glucuronate interconversions
- pentose phosphate pathway
- peroxisome
- pertussis
- phagosome
- phenylalanine metabolism
- phenylalanine, tyrosine and tryptophan biosynthesis
- phosphatidylinositol signaling system
- phospholipase d signaling pathway
- phosphonate and phosphinate metabolism
- phototransduction
- pi3k-akt signaling pathway
- platelet activation
- platinum drug resistance
- porphyrin and chlorophyll metabolism
- ppar signaling pathway
- primary bile acid biosynthesis
- primary immunodeficiency
- prion diseases
- progesterone-mediated oocyte maturation
- prolactin signaling pathway

- propanoate metabolism
- prostate cancer
- proteasome
- protein digestion and absorption
- protein export
- protein processing in endoplasmic reticulum
- proteoglycans in cancer
- proximal tubule bicarbonate reclamation
- purine metabolism
- pyrimidine metabolism
- pyruvate metabolism
- rap1 signaling pathway
- ras signaling pathway
- regulation of actin cytoskeleton
- regulation of lipolysis in adipocytes
- relaxin signaling pathway
- renal cell carcinoma
- renin secretion
- renin-angiotensin system
- retinol metabolism
- retrograde endocannabinoid signaling
- rheumatoid arthritis
- riboflavin metabolism
- ribosome
- ribosome biogenesis in eukaryotes
- rig-i-like receptor signaling pathway
- rna degradation
- rna polymerase
- rna transport
- salivary secretion
- salmonella infection

- selenocompound metabolism
- serotonergic synapse
- shigellosis
- signaling pathways regulating pluripotency of stem cells
- small cell lung cancer
- snare interactions in vesicular transport
- sphingolipid metabolism
- sphingolipid signaling pathway
- spliceosome
- staphylococcus aureus infection
- starch and sucrose metabolism
- steroid biosynthesis
- steroid hormone biosynthesis
- sulfur metabolism
- sulfur relay system
- synaptic vesicle cycle
- synthesis and degradation of ketone bodies
- systemic lupus erythematosus
- t cell receptor signaling pathway
- taste transduction
- taurine and hypotaurine metabolism
- terpenoid backbone biosynthesis
- tgf-beta signaling pathway
- th1 and th2 cell differentiation
- th17 cell differentiation
- thermogenesis
- thiamine metabolism
- thyroid cancer
- thyroid hormone signaling pathway
- thyroid hormone synthesis
- tight junction

- tnf signaling pathway
- toll-like receptor signaling pathway
- toxoplasmosis
- transcriptional misregulation in cancer
- tryptophan metabolism
- tuberculosis
- type i diabetes mellitus
- type ii diabetes mellitus
- tyrosine metabolism
- ubiquinone and other terpenoid-quinone biosynthesis
- ubiquitin mediated proteolysis
- valine, leucine and isoleucine biosynthesis
- valine, leucine and isoleucine degradation
- vascular smooth muscle contraction
- vasopressin-regulated water reabsorption
- vegf signaling pathway
- vibrio cholerae infection
- viral carcinogenesis
- viral myocarditis
- vitamin b6 metabolism
- vitamin digestion and absorption
- wnt signaling pathway # EnrichR Libraries
- Genes_Associated_with_NIH_Grants
- Cancer_Cell_Line_Encyclopedia
- Achilles_fitness_decrease
- Achilles_fitness_increase
- Aging_Perturbations_from_GEO_down
- Aging_Perturbations_from_GEO_up
- Allen_Brain_Atlas_down
- Allen_Brain_Atlas_up
- ARCHS4_Cell-lines

- ARCHS4_IDG_Coexp
- ARCHS4_Kinases_Coexp
- ARCHS4_TFs_Coexp
- ARCHS4_Tissues
- BioCarta_2013
- BioCarta_2015
- BioCarta_2016
- BioPlex_2017
- ChEA_2013
- ChEA_2015
- ChEA_2016
- Chromosome_Location
- Chromosome_Location_hg19
- CORUM
- Data_Acquisition_Method_Most_Popular_Genes
- dbGaP
- Disease_Perturbations_from_GEO_down
- Disease_Perturbations_from_GEO_up
- Disease_Signatures_from_GEO_down_2014
- Disease_Signatures_from_GEO_up_2014
- Drug_Perturbations_from_GEO_2014
- Drug_Perturbations_from_GEO_down
- Drug_Perturbations_from_GEO_up
- DrugMatrix
- DSigDB
- ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X
- ENCODE_Histone_Modifications_2013
- ENCODE_Histone_Modifications_2015
- ENCODE_TF_ChIP-seq_2014
- ENCODE_TF_ChIP-seq_2015
- Enrichr_Libraries_Most_Popular_Genes

- Enrichr_Submissions_TF-Gene_Coocurrence
- Epigenomics_Roadmap_HM_ChIP-seq
- ESCAPE
- GeneSigDB
- Genome_Browser_PWMs
- GO_Biological_Process_2013
- GO_Biological_Process_2015
- GO_Biological_Process_2017
- GO_Biological_Process_2017b
- GO_Biological_Process_2018
- GO_Cellular_Component_2013
- GO_Cellular_Component_2015
- GO_Cellular_Component_2017
- GO_Cellular_Component_2017b
- GO_Cellular_Component_2018
- GO_Molecular_Function_2013
- GO_Molecular_Function_2015
- GO_Molecular_Function_2017
- GO_Molecular_Function_2017b
- GO_Molecular_Function_2018
- GTEx_Tissue_Sample_Gene_Expression_Profiles_down
- GTEx_Tissue_Sample_Gene_Expression_Profiles_up
- HMDB_Metabolites
- HomoloGene
- Human_Gene_Atlas
- Human_Phenotype_Ontology
- HumanCyc_2015
- HumanCyc_2016
- huMAP
- Jensen_COMPARTMENTS
- Jensen_DISEASES

- Jensen_TISSUES
- KEA_2013
- KEA_2015
- KEGG_2013
- KEGG_2015
- KEGG_2016
- Kinase_Perturbations_from_GEO_down
- Kinase_Perturbations_from_GEO_up
- Ligand_Perturbations_from_GEO_down
- Ligand_Perturbations_from_GEO_up
- LINCS_L1000_Chem_Pert_down
- LINCS_L1000_Chem_Pert_up
- LINCS_L1000_Kinase_Perturbations_down
- LINCS_L1000_Kinase_Perturbations_up
- LINCS_L1000_Ligand_Perturbations_down
- LINCS_L1000_Ligand_Perturbations_up
- MCF7_Perturbations_from_GEO_down
- MCF7_Perturbations_from_GEO_up
- MGI_Mammalian_Phenotype_2013
- MGI_Mammalian_Phenotype_2017
- MGI_Mammalian_Phenotype_Level_3
- MGI_Mammalian_Phenotype_Level_4
- Microbe_Perturbations_from_GEO_down
- Microbe_Perturbations_from_GEO_up
- miRTarBase_2017
- Mouse_Gene_Atlas
- MSigDB_Computational
- MSigDB_Oncogenic_Signatures
- NCI-60_Cancer_Cell_Lines
- NCI-Nature_2015
- NCI-Nature_2016

- NURSA_Human_Endogenous_Complexome
- Old_CMAP_down
- Old_CMAP_up
- OMIM_Disease
- OMIM_Expanded
- Panther_2015
- Panther_2016
- Pfam_InterPro_Domains
- Phosphatase_Substrates_from_DEPOD
- PPI_Hub_Proteins
- Reactome_2013
- Reactome_2015
- Reactome_2016
- RNA-Seq_Disease_Gene_and_Drug_Signatures_from_GEO
- SILAC_Phosphoproteomics
- Single_Gene_Perturbations_from_GEO_down
- Single_Gene_Perturbations_from_GEO_up
- SysMyo_Muscle_Gene_Sets
- TargetScan_microRNA
- TargetScan_microRNA_2017
- TF-LOF_Expression_from_GEO
- TF_Perturbations_Followed_by_Expression
- Tissue_Protein_Expression_from_Human_Proteome_Map
- Tissue_Protein_Expression_from_ProteomicsDB
- Transcription_Factor_PPIs
- TRANSFAC_and_JASPAR_PWMs
- Virus_Perturbations_from_GEO_down
- Virus_Perturbations_from_GEO_up
- VirusMINT
- WikiPathways_2013
- WikiPathways_2015

- WikiPathways_2016