

MITHrIL 2

Brief usage guide

The java executable of MITHrIL 2 is a container for several services related to pathway analysis and simulations. This document will describe the usage of the main enhancement of MITHrIL 2: the pathway analysis algorithm based on the meta-pathway environment.

To display the help screen of the program you can simply run:

```
java -jar MITHrIL2.jar merged-mithril
```

To use MITHrIL 2 with the default parameters (human organism, 2000 bootstrapping iterations, strongly validated miRNAs annotation), you can run the executable with the following command:

```
java -jar MITHrIL2.jar merged-mithril -verbose -i [INPUT FILE] -o  
[MAIN OUTPUT FILE] -p [PERTURBATIONS OUTPUT FILE]
```

All input and output files are tab-separated text files.

The main input file is a list of genes or microRNAs with their log-fold-change. Genes are identified using entrez ids, while miRNAs are identified with miRbase mature miRNA name. All genes/miRNAs in your platform must be added to the input file to compute pathway enrichment statistics. Non-statistically significant elements must be set to 0 in the input file. The following figure illustrate an example input file.

```
285550 0
285555 0
3640 0
286887 -0.997523211501566
hsa-miR-513b 5.0442798881368492
259266 3.35689130423164
2306 1.27485396534105
3638 0
4969 -3.11335084849119
```

Figure 1 - Example input file

The main output file contains all pathway statistics computed by MITHrIL. The file contains 9 fields: kegg pathway Id and name, raw accumulator (the sum of equation 3 and 4 of [1]), impact factor (see equation 2 in [1]), enrichment probability, total perturbation (the sum of all nodes perturbations), corrected accumulator (see equation 5 in [1]), p-value and adjusted p-value. The main values used for further analyses are impact factor (greater value implies greater impact of the deregulation in the pathway) and corrected accumulator. Figure 2 reports an example output.

# Pathway id	Pathway Name	Raw Accumulator	Impact Factor	Probability Pi	Total Perturbation	Corrected Accumulator	pValue	Adjusted pValue
path:hsa05034	Alcoholism - Enriched	-172,5098576	20,50160912	3,85075E-09	-126,370693	-118,4787805	6,27054E-13	1,74948E-10
path:hsa04022	cGMP-PKG signaling pathway - Enriched	-219,078777	19,91416489	2,27134E-08	-300,5818148	-141,9195726	3,49687E-12	3,25209E-10
path:hsa05206	MicroRNAs in cancer - Enriched	-85,0554624	19,3164622	1,74843E-08	-86,28160341	-62,58206413	2,71472E-12	3,25209E-10
path:hsa04510	Focal adhesion - Enriched	-190,983507	17,91754867	1,05182E-07	-271,8102718	-119,4208258	1,53879E-11	1,07331E-09
path:hsa04530	Tight junction - Enriched	-108,8461313	17,26638536	1,53877E-07	-135,533871	-76,76991382	2,22192E-11	1,23983E-09
path:hsa04917	Prolactin signaling pathway - Enriched	-122,4879603	15,35458307	1,50722E-06	-138,7041012	-71,7662054	2,00449E-10	9,32088E-09
path:hsa04151	PI3K-Akt signaling pathway - Enriched	-210,4807502	14,61258624	2,4131E-06	-309,7885363	-106,5228322	3,15248E-10	1,09943E-08
path:hsa04933	AGE-RAGE signaling pathway in diabetic complications - Enriched	-139,335694	15,09939023	2,12432E-06	-157,119936	-84,96881402	2,78875E-10	1,09943E-08
path:hsa04015	Rap1 signaling pathway - Enriched	-283,8037968	14,92233851	3,26945E-06	-372,8459078	-178,5240252	4,22161E-10	1,3087E-08
path:hsa01100	Metabolic pathways - Enriched	-400,466554	20,60197114	9,25004E-09	-594,075953	-154,1387215	8,13962E-10	1,62211E-08
path:hsa04072	Phospholipase D signaling pathway - Enriched	-220,5124098	14,81645826	4,61442E-06	-269,9778606	-133,7233485	5,8788E-10	1,62211E-08
path:hsa05220	Chronic myeloid leukemia - Enriched	-136,2920678	14,22074762	5,70861E-06	-129,4466994	-92,96616864	7,21211E-10	1,62211E-08

Figure 2 - Example output file

The perturbation output file contains all nodes statistics as computed by MITHrIL. The file contains 7 fields: pathway Id and name, gene/miRNA id and name, perturbation (see equation 1 in [1]), accumulator, and p-value (see section 4.1.1 in [2]).

# Pathway Id	Pathway Name	Gene Id	Gene Name	Perturbation	Accumulator	pValue
path:hsa00190	Oxidative phosphorylation - Enriched	64077	LHPP, HDHD2B	0	0	1
path:hsa00190	Oxidative phosphorylation - Enriched	5464	PPA1, HEL-S-66p, IOPPP, PP, PP1, SID6-8061	0	0	1
path:hsa00190	Oxidative phosphorylation - Enriched	hsa-miR-101-3p	hsa-miR-101-3p	-1,00943771	0	0,233883058
path:hsa00190	Oxidative phosphorylation - Enriched	245972	ATP6V0D2, ATP6D2, VMA6	1,579190022	0	0,056971514
path:hsa00190	Oxidative phosphorylation - Enriched	245973	ATP6V1C2, ATP6C2, VMA5	0	0	1
path:hsa00190	Oxidative phosphorylation - Enriched	479	ATP12A, ATP1A1	0	0	1
path:hsa00190	Oxidative phosphorylation - Enriched	513	ATP5D	0	0	1
path:hsa00190	Oxidative phosphorylation - Enriched	514	ATP5E, ATPE, MC5DN3	0,39003142	0	0,294852574
path:hsa00190	Oxidative phosphorylation - Enriched	515	ATP5F1, PIG47	0	0	1
path:hsa00190	Oxidative phosphorylation - Enriched	516	ATP5G1, ATP5A, ATP5G	0,45787118	0	0,297851074
path:hsa00190	Oxidative phosphorylation - Enriched	517	ATP5G2, ATP5A	0	0	1
path:hsa00190	Oxidative phosphorylation - Enriched	518	ATP5G3, P3	0	0	1
path:hsa00190	Oxidative phosphorylation - Enriched	cpd:C00008	ADP	9,524782206	9,524782206	0,021989005

Figure 3 - Example nodes perturbation output file

References

- [1] Alaimo, S., Giugno, R., Acunzo, M., Veneziano, D., Ferro, A., & Pulvirenti, A. (2016). Post-transcriptional knowledge in pathway analysis increases the accuracy of phenotypes classification. *Oncotarget*, 7(34), 54572.
- [2] Alaimo, S., Marceca, G. P., Ferro, A., & Pulvirenti, A. (2017). Detecting Disease Specific Pathway Substructures through an Integrated Systems Biology Approach. *Non-Coding RNA*, 3(2), 20.