

其它 - 小工具 - 简易 ID 转换[人]

id	official_name	gene_id_ensembl109	gene_biotype_ensembl109	entrez_id	other1	other2
ENSG00000223972	DDX11L1	ENSG00000223972	transcribed_unprocessed_pseudo	100287029	3	6
ENSG00000227232	WASH7P	ENSG00000227232	unprocessed_pseudogene	653635	2	5
ENSG00000278267	MIR6859-1	ENSG00000278267	miRNA	102466751	9	0
ENSG00000243485	MIR1302-2HG	ENSG00000243485	IncRNA	107985730	9	0
ENSG00000284332	MIR1302-2	ENSG00000284332	miRNA	100302278	7	6
ENSG00000237613	FAM138A	ENSG00000237613	IncRNA	645520	5	9
ENSG00000268020	OR4G4P	ENSG00000268020	unprocessed_pseudogene	79504	7	5
ENSG00000240361	OR4G11P	ENSG00000240361	transcribed_unprocessed_pseudo	403263	7	7
ENSG00000186092	OR4F5	ENSG00000186092	protein_coding	79501	4	6
ENSG00000238009	NA	NA	NA	NA	7	7
ENSG00000239945	NA	NA	NA	NA	8	4
ENSG00000233750	CICP27	ENSG00000233750	processed_pseudogene	100420257	5	5
ENSG00000268903	NA	NA	NA	NA	9	2

网址: https://www.xiantao.love



更新时间: 2023.05.15



目录

基本概念 3
应用场景 3
分析流程
结果解读
数据格式
参数说明 7
物种 7
结果说明 8
主要结果 8
方法学 9
如何引用
常见问题





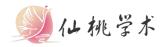
基本概念

- ▶ ID 转换:将输入的分子 ID 信息转换成常见的分子信息,比如基因名、Ensembl编号等
 - 目前支持:基因名、基因别名、Ensembl 编号、Entrez 编号的 ID 转换。 (其他 ID 格式,比如 IncRNA 数据库的编号、芯片探针号,目前还不 支持)



应用场景

简易 ID 转换通常:通过数据下载或整理数据得到相关分子 ID 信息相关数据转换成分析流程中常用的分析数据(信息),主要用于转换一些常见的 ID



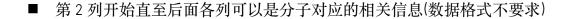
分析流程

上传数据 → 数据处理(清洗) → ID 转换

- ▶ 数据格式: xlsx / csv / txt 文件格式:
 - 第1列表示分子列表,可以上传(或支持)基因名、基因别名、Ensembl编号、Entrez编号的ID
 - 不能上传无法识别的特殊字符或者非法的字符
 - 不能一整列数据都是空的内容



.4	A	В	C
1	id	other1	other2
2	ENSG00000223972	3	6
3	ENSG00000227232	2	5
4	ENSG00000278267	9	0
5	ENSG00000243485	9	0
6	ENSG00000284332	7	6
7	ENSG00000237613	5	9
8	ENSG00000268020	7	5
9	ENSG00000240361	7	7
10	ENSG00000186092	4	6
11	ENSG00000238009	7	7
12	ENSG00000239945	8	4
13	ENSG00000233750	5	5
14	ENSG00000268903	9	2
15	ENSG00000269981	10	0
16	ENSG00000239906	4	10



- ▶ 数据处理: <u>只对第1列(分子列表)数据进行相关处理</u>,处理掉无法识别的 特殊字符等
- ▶ ID 转换:
 - 将处理过后的分子 ID 与分子注释数据库中的分子 ID 进行匹配,并得到最终匹配成功的分子 ID 对应到注释库中的内容



结果解读

id	official_name	gene_id_ensembl109	gene_biotype_ensembl109	entrez_id	other1	other2
ENSG00000223972	DDX11L1	ENSG00000223972	transcribed_unprocessed_pseudo	100287029	3	6
ENSG00000227232	WASH7P	ENSG00000227232	unprocessed_pseudogene	653635	2	5
ENSG00000278267	MIR6859-1	ENSG00000278267	miRNA	102466751	9	0
ENSG00000243485	MIR1302-2HG	ENSG00000243485	IncRNA	107985730	9	0
ENSG00000284332	MIR1302-2	ENSG00000284332	miRNA	100302278	7	6
ENSG00000237613	FAM138A	ENSG00000237613	IncRNA	645520	5	9
ENSG00000268020	OR4G4P	ENSG00000268020	unprocessed_pseudogene	79504	7	5
ENSG00000240361	OR4G11P	ENSG00000240361	transcribed_unprocessed_pseudo	403263	7	7
ENSG00000186092	OR4F5	ENSG00000186092	protein_coding	79501	4	6
ENSG00000238009	NA	NA	NA	NA	7	7
ENSG00000239945	NA	NA	NA	NA	8	4
ENSG00000233750	CICP27	ENSG00000233750	processed_pseudogene	100420257	5	5
ENSG00000268903	NA	NA	NA	NA	9	2

▶ id: 表示上传分子列表 ID

▶ official_name: 表示分子列表通用名字(分子名)

▶ gene_id_ensemblxxx: 表示 Ensembl 数据库中提供的 gene_id

▶ gene_biotyope_ensemblxxx: 表示 Ensembl 数据库中提供的 gene_biotype

▶ entrez_id: 表示 EntrezID 编号

▶其它的表示上传数据各列(与分子列表对应)



数据格式

1	Α	В	С
1	id	other1	other2
2	ENSG00000223972	3	6
3	ENSG00000227232	2	5
4	ENSG00000278267	9	0
5	ENSG00000243485	9	0
6	ENSG00000284332	7	6
7	ENSG00000237613	5	9
8	ENSG00000268020	7	5
9	ENSG00000240361	7	7
10	ENSG00000186092	4	6
11	ENSG00000238009	7	7
12	ENSG00000239945	8	4
13	ENSG00000233750	5	5
14	ENSG00000268903	9	2
15	ENSG00000269981	10	0
16	ENSG00000239906	4	10

数据要求:

- ▶ 列数: 至少需要 1 列(分子列表)以上的数据,最多 10 列数据,至少需要 10 个以上的样本(10 行),暂时支持最多 70000(行)个以上的样本
- ▶ 第1列表示分子列表,可以上传(或支持)基因名、基因别名、Ensembl编号、 Entrez编号的ID
 - 不能上传无法识别的特殊字符或者非法的字符
 - 不能一整列数据都是空的内容
- ▶ 第2列开始直至后面各列可以是分子对应的相关信息(数据格式不要求)
- ▶ 上传数据中列明不能重复,也不能上传空的列名



参数说明

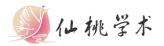
(说明:标注了颜色的为常用参数。)

物种



▶ 物种:可以选择不同的物种对分子列表进行 ID 转换,目前只支持人源





结果说明

主要结果

ID转换

简易ID转换: 对提供的数据进行ID转换(目前仅支持gene_name|ensembl_id|entrezid三种类型)

id	official name	gene id ensembl109	gene biotype ensembl109	entrez id	other1	other
IU.	Official_flatfie	gene_id_ensembilios		entrez_lu	Otheri	Other
ENSG00000223972	DDX11L1	ENSG00000223972	transcribed_unprocessed_pseudo	100287029	3	6
ENSG00000227232	WASH7P	ENSG00000227232	unprocessed_pseudogene	653635	2	5
ENSG00000278267	MIR6859-1	ENSG00000278267	miRNA	102466751	9	0
ENSG00000243485	MIR1302-2HG	ENSG00000243485	IncRNA	107985730	9	0
ENSG00000284332	MIR1302-2	ENSG00000284332	miRNA	100302278	7	6
ENSG00000237613	FAM138A	ENSG00000237613	IncRNA	645520	5	9
ENSG00000268020	OR4G4P	ENSG00000268020	unprocessed_pseudogene	79504	7	5
ENSG00000240361	OR4G11P	ENSG00000240361	transcribed unprocessed pseudo	403263	7	7
ENSG00000186092	OR4F5	ENSG00000186092	protein_coding	79501	4	6
ENSG00000238009	NA	NA	NA	NA	7	7
ENSG00000239945	NA	NA	NA	NA	8	4
ENSG00000233750	CICP27	ENSG00000233750	processed_pseudogene	100420257	5	5
ENSG00000268903	NA	NA	NA	NA	9	2

D转换xls

- •提供下载的数据的行的顺序是和原数据对齐的,可以打开后直接复制到原来的数据中
- · official_name: 通用名字



方法学

统计分析和可视化均在 R 4.2.1 版本中进行

处理过程:

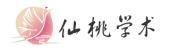
(1) 利用对应的分子注释库和 Ensembl 注释库对分子进行 ID 转换

补充说明:

(1) 物种: 当前只支持人源(Homo sapiens)

(2) ID 转换库(包): org.Hs.eg.db





如何引用

生信工具分析和可视化用的是 R 语言,<mark>可以直接写自己用 R 来进行分析和可视化即可</mark>,可以无需引用仙桃,如果想要引用仙桃,可以在致谢部分 (Acknowledge) 致谢仙桃学术(www.xiantao.love)。

方法学部分可以参考对应说明文本中的内容以及一些文献中的描述。





常见问题

1. 页面出来的结果为什么这么少?

答:页面仅仅为预览结果,完全的结果请下载表格数据进行查看

2. 下载的结果为什么没有变化

答:如果下载的结果中没有新增加数据列,说明上传的数据的第一列无法匹配到 注释库中的内容

3. 下载的结果中为什么注释的列不全

答: 如果数值的列**为全**空,就会去掉,所以无这些列说明这些列中没有注释内容