

踩坑记录-GB_ACC转换基因-不允许有重复的'row.names' -R语言零基础基因/数据差异分析 (三)

已于 2022-06-22 15:53:40 修改

编辑



R语言零基础基因/数... 专栏收录该内容

5 篇文章

摘要 本文指导如何将GB_ACC转换为基因名，提供两种方法，推荐使用简单的方法避免繁琐操作。涉及排序、去重、正则表达式处理及CSV格式校验，确保数据无重复'row.names'问题。

文章目录

[GB_ACC转换成基因](#)

[直接利用GEO分析](#)

[**GB_ACC的第二种方法（繁琐）**](#)

[不允许有重复的'row.names'解决问题](#)

GB_ACC转换成基因

直接利用GEO分析

在上个系列中，我们记住了 GB_ACC，但制作热图，需要我们将其转换成基因名，如果你已经转换好了或者不需要制作热图，前参考后续系列。

方法1 (推荐)

如下图示意，将GB_ACC转换成基因名。

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227



Top differentially expressed genes ?

[Download full table](#) [Select columns](#)



ID	adj.P.Val	P.Value	t	B	logFC	GENE_SYMBOL	SEQUENCE	SPOT_ID
A_23_P10385	0.0000107	3.20e-10	145.4	12.24	6.22	DTL	GGAAGATGTTA...	A_23_P10385
A_23_P256956	0.0000107	3.67e-10	141.4	12.21	7.63	KIF20A	TCAAGCCTTG...	A_23_P256956
A_24_P322354	0.000018	1.70e-09	103.9	11.69	5.51	SKA1	ACCTGAAATG...	A_24_P322354
A_23_P302672	0.000018	1.70e-09	-103.9	11.68	-4.34	DDIT4L	ATCAATGTTGT...	A_23_P302672
A_33_P3374205	0.000018	2.37e-09	97.2	11.54	5.87	MKI67	ATGTGCTCTG...	A_33_P3374205
A_24_P346855	0.000018	2.39e-09	97	11.53	4.73	MKI67	TCTCACCCCTG...	A_24_P346855
A_33_P3376116	0.000018	2.57e-09	95.6	11.5	3.78	SPC24	TCCAGGAAATT...	A_33_P3376116
A_23_P251421	0.000018	2.79e-09	94	11.46	4.42	CDCA7	ATTTACTTGCA...	A_23_P251421
A_22_P00006...	0.000018	3.01e-09	-92.6	11.43	-4.73	LOC644135	GATTACAAGTT...	A_22_P00006231
A_33_P3392187	0.000018	3.25e-09	91.2	11.39	4.47	CCDC85A	TCCAATGCAA...	A_33_P3392187
A_23_P63780	0.000018	3.40e-09	90.4	11.37	5.2	ZIMINT	TCAAAAGATTC...	A_23_P63780

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

Check all

Data columns

- Adj P-value
- P-value
- t-statistic
- B-value
- logFC
- F-statistic

Annotation columns

- | | | |
|---------------------------------------------|-----------------------------------------------|-------------------------------------------------|
| <input checked="" type="checkbox"/> ID | <input type="checkbox"/> CONTROL_TYPE | <input type="checkbox"/> REFSEQ |
| <input type="checkbox"/> GB_ACC | <input type="checkbox"/> LOCUSLINK_ID | <input checked="" type="checkbox"/> GENE_SYMBOL |
| <input type="checkbox"/> GENE_NAME | <input type="checkbox"/> UNIGENE_ID | <input type="checkbox"/> ENSEMBL_ID |
| <input type="checkbox"/> ACCESSION_STRING | <input type="checkbox"/> CHROMOSOMAL_LOCATION | <input type="checkbox"/> CYTOBAND |
| <input type="checkbox"/> DESCRIPTION | <input type="checkbox"/> GO_ID | <input checked="" type="checkbox"/> SEQUENCE |
| <input checked="" type="checkbox"/> SPOT_ID | | |

Set **Cancel** **Restore defaults**

2

1

3

https://blog.csdn.net/qq_39751227

上图选项框勾选后，就会出现对应的列在列表中，如下图勾选 GENE_SYMBOL后的效果。

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

Top differentially expressed genes

[Download full table](#) [Select columns](#)

ID	adj.P.Val	P.Value	t	B	logFC	GENE_SYMBOL	SEQUENCE	SPOT_ID
A_23_P10385	0.0000107	3.20e-10	145.4	12.24	6.22	DTL	GGAAGATGTTA...	A_23_P10385
A_23_P256956	0.0000107	3.67e-10	141.4	12.21	7.63	KIF20A	TCAAGCCTG...	A_23_P256956
A_24_P322354	0.000018	1.70e-09	103.9	11.69	5.51	SKA1	ACCTGAAATG...	A_24_P322354
A_23_P302672	0.000018	1.70e-09	-103.9	11.68	-4.34	DDIT4L	ATCAATGTTGT...	A_23_P302672
A_33_P3374205	0.000018	2.37e-09	97.2	11.54	5.87	MKI67	ATGTGCTCTG...	A_33_P3374205
A_24_P346855	0.000018	2.39e-09	97	11.53	4.73	MKI67	TCTCACCCCTG...	A_24_P346855
A_33_P3376116	0.000018	2.57e-09	95.6	11.5	3.78	SPC24	TCCAGGAAATT...	A_33_P3376116
A_23_P251421	0.000018	2.79e-09	94	11.46	4.42	CDCA7	ATTTACTTGCA...	A_23_P251421
A_22_P00006...	0.000018	3.01e-09	-92.6	11.43	-4.73	LOC644135	GATTACAAGTT...	A_22_P00006231
A_33_P3392187	0.000018	3.25e-09	91.2	11.39	4.47	CCDC85A	TCCAAATGCAA...	A_33_P3392187
A_23_P63789	0.000018	3.40e-09	90.4	11.37	5.2	ZWINT	TCAAAGATTCA...	A_23_P63789
A_24_P413884	0.0000186	3.83e-09	88.2	11.31	5.78	CENPA	TAGTTTGTGAG...	A_24_P413884
A_21_P0003965	0.0000186	4.17e-09	-86.7	11.26	-4.84		CAATTTAACG...	A_21_P0003965
A_23_P35219	0.0000186	4.48e-09	85.5	11.23	5.18	NEK2	AGGAGTGAGA...	A_23_P35219
A_33_P3316273	0.0000186	5.94e-09	-80.8	11.07	-3.56	CCL3	TGCTTTGTC...	A_33_P3316273
A_22_P00018...	0.0000186	6.06e-09	-80.5	11.06	-4.19		ATTCAGCTCTG...	A_22_P00018082
A_22_P00001	0.0000186	6.07e-09	80.4	11.06	5.26	LINC00842	CCCATATTCTT	A_22_P00001017

如此就可以实现GB_ACC与基因名的对应。

GB_ACC的第二种方法（繁琐）

我并不推荐这种方法，它十分的繁琐且对新手不友好，如果你是新手，请忽视掉它。

在后面的文章中，如果特别指明（请忽视）说明是不重要且无伤大雅的。

要复制的列：

内容来源：csdn.net

作者昵称：Frms

原文地址：https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页：https://blog.csdn.net/qq_39751227

GSE137578.top.table (1) - Excel(产品激活失败)

文件 开始 插入 页面布局 公式 数据 审阅 视图 帮助 特色功能 团队 百度网盘 搜索

G1 GB_ACC

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	ID	adj.P.Val	P.Value	t	B	logFC	GB_ACC	SEQUENCE	SPOT_ID						
2	A_23_P103	1.07E-05	3.20E-10	#####	12.24225	-6.223603	NM_016448	GGAAGATGTTATTATGACCAATATA_23_P10385							
3	A_23_P256	1.07E-05	3.67E-10	#####	12.20565	-7.628263	NM_005733	TCAAGCCTTGACCACTTGTGATGACA_23_P256956							
4	A_24_P322	0.000018	1.70E-09	#####	11.68504	-5.512398	NM_001039535	ACCTGAAATGCATTAGTGTACACCA_24_P322354							
5	A_23_P302	0.000018	1.70E-09	1.04E+02	11.68338	4.3440193	BC013592	ATCAATGTTGTCATGTCTATTGA_23_P302672							
6	A_33_P337	0.000018	2.37E-09	#####	11.53921	-5.873954	NM_002417	ATGTGCTCTGGTACCCGGTCTTA_33_P3374205							
7	A_24_P346	0.000018	2.39E-09	#####	11.53435	-4.733946	NM_002417	TCTCACCCCTGGTAAGTACAGTATTTA_24_P346855							
8	A_33_P337	0.000018	2.57E-09	#####	11.50109	-3.782224	NM_182513	TCCAGGAAATTATCAGCGACTACCA_33_P3376116							
9	A_23_P251	0.000018	2.79E-09	#####	11.46287	-4.415631	NM_031942	ATTACTTGCATATGTAACCATTG(A_23_P251421							
10	A_22_P000	0.000018	3.01E-09	9.26E+01	11.42723	4.7328389	XR_110175	GATTACAAGTTCTATAGGAAGACA_22_P00006231							
11	A_33_P339	0.000018	3.25E-09	#####	11.38962	-4.466155	NM_001080433	TCCAAATGCAATGATAGTTCTTGT(A_33_P3392187							
12	A_23_P637	0.000018	3.40E-09	#####	11.36709	-5.197076	NM_032997	TCAAAGATTAGAGATTGGCTTTG(A_23_P63789							
13	A_24_P413	1.86E-05	3.83E-09	#####	11.30648	-5.776578	NM_001809	TAGTTTGTGAGTTACTCATGTGACT(A_24_P413884							
14	A_21_P000	1.86E-05	4.17E-09	8.67E+01	11.2627	4.8436264		CAATTAAATGGAGCATGAAGATT(A_21_P0003965							
15	A_23_P352	1.86E-05	4.48E-09	#####	11.22582	-5.175923	NM_002497	AGGAGTGAGAATTCTGAGAGTCAG(A_23_P35219							
16	A_33_P331	1.86E-05	5.94E-09	8.08E+01	11.0713	3.5554853	NM_002983	TGCTTTGTTCAAGGGCTGTGATCGG(A_33_P3316273							
17	A_22_P000	1.86E-05	6.06E-09	8.05E+01	11.05977	4.1929358	BU154984	ATTAGCTCTGATTGGAGAGATAAA_22_P00018082							
18	A_22_P000	1.86E-05	6.07E-09	#####	11.05889	-5.259739	NR_033957	CCCATATTCTCATCTTGTCCCTGG(A_22_P00001017							
19	A_24_P287	1.86E-05	6.14E-09	#####	11.05276	-3.377591	NM_013290	AAATTGCGAGTAGCTTGAGGTTAAC(A_24_P287941							
20	A_33_P340	1.86E-05	6.19E-09	8.01E+01	11.04834	3.6073776	NM_201563	AAGAGCCCCAATTACCGAACACAA_33_P3403576							
21	A_23_P104	1.86E-05	6.56E-09	#####	11.01469	-3.642605	NM_080668	TCACCAAGATGATGCAGAGTTGAGA(A_23_P104651							
22	A_23_P505	1.86E-05	6.85E-09	7.85E+01	10.98981	4.6191365	NM_003706	TCCAGATGGGAGAACATGAATGTGA(A_23_P50508							
23	A_23_P575	1.86E-05	7.16E-09	#####	10.9641	-5.401823	NM_016426	CTCTGATCGACCTCATGACAAACACA_23_P57588							
24	A_21_P000	1.86E-05	8.17E-09	#####	10.88517	-3.110785		TCTACAAGCTCCCAAAAGAAAAACA(A_21_P0006266							
25	A_33_P324	1.86E-05	8.22E-09	7.57E+01	10.88196	3.0712438	NM_173843	TGCAAAGTTCCCTACTTCTGTGAC(A_33_P3246833							
26	A_23_P924	1.86E-05	8.71E-09	7.48E+01	10.84662	4.7093709	NM_003264	AAGTCCTAGGTTCCCATAATTAAAGA(A_23_P92499							
27	A_23_P583	1.86E-05	8.91E-09	#####	10.83295	-4.138119	NM_001237	AAAGTTGATAGATGCTGACCCATAC(A_23_P58321							
28	A_23_P782	1.86E-05	9.06E-09	7.42E+01	10.82247	2.8327275	NM_001010919	CAAATTCATCACTGTATACTTCA(A_23_P7827							
29	A_33_P325	1.86E-05	9.14F-09	#####	10.81732	-5.517399	NM_001005464	AAGAAGTGGCGGTTGGCCGGAGGA_33_P3257678							

但在复制选择的列之前，我们需要对它进行排序，这个非常重要

内容来源: csdn.net

作者昵称: Frms

原文地址: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

GSE137578.top.table (1) - Excel(产品激活失败)

文件 开始 插入 页面布局 公式 数据 审阅 视图 帮助 特色功能 团队 百度网盘 操作说明 搜索

等线 11 A A % ,

G1 GB_ACC

ID	adj.P.Val	P.Value	t	B	logFC	GB_ACC	SEQUENCE	SPOT_ID
A_23_P10385	1.07E-05	3.20E-10	#####	12.24225	-6.223603	NM_0164	TGACCAATAT	A_23_P10385
A_23_P256956	1.07E-05	3.67E-10	#####	12.20565	-7.628263	NM_0057	TTGTGATGAC	A_23_P256956
A_24_P322354	0.000018	1.70E-09	#####	11.68504	-5.512398	NM_0010	AGTGTACACC	A_24_P322354
A_23_P302672	0.000018	1.70E-09	1.04E+02	11.68338	4.3440193	BC013592	CATGTCTATTG	A_23_P302672
A_33_P3374205	0.000018	2.37E-09	#####	11.53921	-5.873954	NM_0024	CCTGGCTTTA	A_33_P3374205
A_24_P346855	0.000018	2.39E-09	#####	11.53435	-4.733946	NM_1825	TACAGTATTT	(A_24_P346855
A_33_P3376116	0.000018	2.57E-09	#####	11.50109	-3.782224	NM_0319	AGCGACTACCA	A_33_P3376116
A_23_P251421	0.000018	2.79E-09	#####	11.46287	-4.415631	XR_11017	AAACCATTTG	(A_23_P251421
A_22_P00006231	0.000018	3.01E-09	9.26E+01	11.42723	4.7328389	NM_0010	ATAGGAAGACA	A_22_P00006231
A_33_P3392187	0.000018	3.25E-09	#####	11.38962	-4.466155	NM_0329	AGTTTCTTGT	(A_33_P3392187
A_23_P63789	0.000018	3.40E-09	#####	11.36709	-5.197076	NM_0018	TTGGCTTTG	(A_23_P63789
A_21_P0003965	1.86E-05	4.17E-09	8.67E+01	11.2627	4.8436264	NM_00249	ATGAAGATT	(A_21_P0003965
A_23_P35219	1.86E-05	4.48E-09	#####	11.22582	-5.175923	NM_0029	GAGAGTCAG	(A_23_P35219
A_33_P3316273	1.86E-05	5.94E-09	8.08E+01	11.0713	3.5554853	BU154984	CTGTGATCGC	(A_33_P3316273
A_22_P00018082	1.86E-05	6.06E-09	8.05E+01	11.05977	4.1929358	NR_03395	CAACACAAA	(A_22_P00018082
A_22_P0001017	1.86E-05	6.07E-09	#####	11.05889	-5.259739	NM_01329	升序(S)	① 这里一定要进行排序
A_24_P287941	1.86E-05	6.14E-09	#####	11.05276	-3.377591	NM_2015	Z↓ 降序(O)	
A_33_P3403576	1.86E-05	6.19E-09	8.01E+01	11.04834	3.6073776	NM_0806	将所选单元格颜色放在最前面(C)	
A_23_P104651	1.86E-05	6.56E-09	#####	11.01469	-3.642605	NM_00370	将所选字体颜色放在最前面(E)	
A_23_P50508	1.86E-05	6.85E-09	7.85E+01	10.98981	4.6191365	NM_01642	将选定格式的图标置顶(F)	
A_23_P57588	1.86E-05	7.16E-09	#####	10.9641	-5.401823	NM_17384	自定义排序(U)...	
A_21_P0006266	1.86E-05	8.17E-09	#####	10.88517	-3.110785	NM_00326	ATATTTAAGA	(A_23_P92499
A_33_P3246833	1.86E-05	8.22E-09	7.57E+01	10.88196	3.0712438	NM_00123	TGACCCATACA	(A_23_P58321
A_23_P92499	1.86E-05	8.71E-09	7.48E+01	10.84662	4.7093709	NM_0010	CAAAATTCACTACG	TATACTTCAA
A_23_P58321	1.86E-05	8.91E-09	#####	10.83295	-4.138119	NM_00100	TGAGAAGTGGCGGTT	CGGCCGGAGG
A_23_P7827	1.86E-05	9.06E-09	7.42E+01	10.82247	2.8327275	NM_001069	A_33_P3257678	ACTTCTCAGATCAATCGTCATCCT
A_33_P3257678	1.86E-05	9.14E-09	#####	10.81732	-5.517399	NM_015714	A_23_P19291	(A_23_P19291
A_23_P19291	1.86E-05	9.26E-09	#####	10.80919	-3.213546	NM_001244008		

GSE137578.top.table (1)

就绪

再删除重复，如图

内容来源: csdn.net

作者昵称: Frms

原文地址: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

文件 开始 插入 页面布局 公式 数据 审阅 帮助 特色功能 团队 百度网盘 搜索

登录 退出 分享

自 Access 自 Web 自其他来源 现有连接 显示查询 新建查询 从表格 最近使用的源 全部刷新 属性 连接 编辑链接 排序 A↓ Z↓ A↑ Z↑ 筛选 清除 重新应用 高级 分列 快速填充 删除 数据验 合并计算 关系 模拟分析 预测 工作表 预测 组合 取消组合 分类汇总 分级显示

1 在这里，我们删除重复值，并剔除空格

	A_23_P103	1.07E-05	3.20E-10	#####	12.24E20	-0.223003	NM_U10448	GGAAGGATGTTAATGACCAATAA	A_23_P103
2	A_23_P256	1.07E-05	3.67E-10	#####	12.20565	-7.628263	NM_005733	TCAAGCTTGACCCTGTGATGACA	A_23_P256_56
3	A_24_P322	0.000018	1.70E-09	#####	11.68504	-5.512398	NM_001039535	ACCTGAAATGCATTAGTGTACCCA	A_24_P322334
4	A_23_P302	0.000018	1.70E-09	1.04E+02	11.68338	4.3440193	BC013592	ATCAATGTTGATGTCTATTGA	A_23_P302672
5	A_33_P337	0.000018	2.37E-09	#####	11.53921	-5.873954	NM_002417	ATGTGCTCTGGTACCGTGGTCTTA	(A_33_P3374205
6	A_24_P346	0.000018	2.39E-09	#####	11.53435	-4.733946	NM_002417	TCTCACCCCTGGTAAGTACAGTATTT	(A_24_P346855
7	A_33_P337	0.000018	2.57E-09	#####	11.50109	-3.782224	NM_182513	TCCAGGAAATTATCAGCGACTACCA	A_33_P3376116
8	A_23_P251	0.000018	2.79E-09	#####	11.46287	-4.415631	NM_031942	ATTACTTGCATATGTAACCATTG	(A_23_P251421
9	A_22_P000	0.000018	3.01E-09	9.26E+01	11.42723	4.7328389	XR_110175	GATTACAAGTCTCATAGGAAGACA	A_22_P0006231
10	A_33_P339	0.000018	3.25E-09	#####	11.38962	-4.466155	NM_001080433	TCCAAATGCAATGATAGTTCTTGT	(A_33_P3392187
11	A_23_P637	0.000018	3.40E-09	#####	11.36709	-5.197076	NM_032997	TCAAAGATTAGAGATTGGCTTTG	A_23_P63789
12	A_24_P413	1.86E-05	3.83E-09	#####	11.30648	-5.776578	NM_001809	TAGTTTGTGAGTTACTCATGTGACT	A_24_P413884
13	A_21_P000	1.86E-05	4.17E-09	8.67E+01	11.2627	4.8436264		CAATTAAATGGAGCATGAAGATT	(A_21_P0003965
14	A_23_P352	1.86E-05	4.48E-09	#####	11.22582	-5.175923	NM_002497	AGGAGTGAGAATTCTGAGAGTCAG	(A_23_P35219
15	A_33_P331	1.86E-05	5.94E-09	8.08E+01	11.0713	3.5554853	NM_002983	TGCTTTGTTCAAGGCTGTGATCGG	(A_33_P3316273
16	A_22_P000	1.86E-05	6.06E-09	8.05E+01	11.05977	4.1929358	BU154984	ATTAGCTCTGATTGGAGAGATAA	A_22_P00018082
17	A_22_P000	1.86E-05	6.07E-09	#####	11.05889	-5.259739	NR_033957	CCCATATTCTCATCTTGTCCCTGG	(A_22_P00001017
18	A_24_P287	1.86E-05	6.14E-09	#####	11.05276	-3.377591	NM_013290	AAATTGCGAGTAGCTTGAGGTTAAC	A_24_P287941
19	A_33_P340	1.86E-05	6.19E-09	8.01E+01	11.04834	3.6073776	NM_201563	AAGAGCCCCAATTACCGAACACAA	(A_33_P3403576
20	A_23_P104	1.86E-05	6.56E-09	#####	11.01469	-3.642605	NM_080668	TCACCAAGATGATGCAGAGTTGAGA	(A_23_P104651
21	A_23_P505	1.86E-05	6.85E-09	7.85E+01	10.98981	4.6191365	NM_003706	TCCAGATGGGAGAACATGAATGTG	(A_23_P50508
22	A_23_P575	1.86E-05	7.16E-09	#####	10.9641	-5.401823	NM_016426	CTCTGATCGACCTCATGACAAACACA	(A_23_P57588
23	A_21_P000	1.86E-05	8.17E-09	#####	10.88517	-3.110785		TCTACAAGCTCCCAAGGGAAAACA	(A_21_P0006266
24	A_33_P324	1.86E-05	8.22E-09	7.57E+01	10.88196	3.0712438	NM_173843	TGCAAAGTCCCCTACTTCTGTGAC	(A_33_P3246833
25	A_23_P924	1.86E-05	8.71E-09	7.48E+01	10.84662	4.7093709	NM_003264	AAAGTCTAGGTTCCATATTAAGA	(A_23_P92499
26	A_23_P583	1.86E-05	8.91E-09	#####	10.83295	-4.138119	NM_001237	AAAGTTGATAGATGCTGACCCATAC	(A_23_P58321
27	A_23_P782	1.86E-05	9.06E-09	7.42E+01	10.82247	2.8327275	NM_001010919	CAAATTTCATCACTGTATACTTTCA	(A_23_P7827
28	A_33_P325	1.86E-05	9.14F-09	#####	10.81732	-5.517399	NM_001005464	AAGAAGTGGCGGTCGGCCGGAGG	(A_33_P3257678

之后打开David,

地址: <https://david.ncifcrf.gov/>

内容来源: csdn.net

作者昵称: Frms

原文地址: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

go富集分析 - 国内版 简 GO, KEGG, DO富 DAVID Functional A x https://david.ncifcrf.gov/ https://david.ncifcrf.gov/ 历史记录 + 6 热搜 译器 翻译 游戏 搜索 光照

登录账号 收藏 手机收藏夹 谷歌 网址大全 ASUS E- Links Tripadvi 医学书籍 中国知网 Sci-Hub iData-知 爱爱医学 必应 百度云搜 Me GeekM https://

DAVID Bioinformatics Resources 6.8

Laboratory of Human Retrovirology and Immunoinformatics (LHRI)

Home Start Analysis Shortcut to DAVID Tools Technical Center Downloads & APIs Term of Service About DAVID About LHRI

Overview

The Database for Integrated Discovery (DAVID) v6.8 comprises a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

- Identify enriched biological themes, particularly GO terms
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathway maps
- Display related many-genes-to-many-terms on 2-D view.
- Search for other functionally related genes not in the list
- List interacting proteins
- Explore gene names in batch
- Link gene-disease associations
- Highlight protein functional domains and motifs
- Redirect to related literatures

Functional Annotation
- Functional Annotation Clustering
- Functional Annotation Chart
- Functional Annotation Table

Gene Functional Classification

Gene ID Conversion

Gene Name Batch Viewer

Hot Links

Postdoctoral Fellow position available in LHRI

The Laboratory of Human Retrovirology and Immunoinformatics (LHRI) has collaborated with the National Institute of Allergy and Infectious Diseases (NIAID) and supported NIAID clinical trials for patients infected with HIV mutants resisting anti-retroviral therapy. LHRI has isolated the multiple-class drug-resistant (MDR) variants from patients and characterized each variant's drug sensitivity and infectivity. The study aims to define salvage therapy and develop novel therapy (chemotherapy and immunotherapy). During the investigation, LHRI has characterized the emergence of novel mutations on drug susceptibility and viral replication. LHRI is a pioneer in researching the anti-viral cytokine, Interleukin-27, DNA-repair protein (Ku70)-mediated innate immune response against HIV and other virus co-infection, and novel subsets of immune cells. LHRI maintains the Database for Annotation, Visualization and Integrated Discovery (DAVID). Postdoctoral Fellow position available to perform Microbiology/Cellular Immunology research in our Basic Research Section.

今日优选 <https://david.ncifcrf.gov/conversion.jsp>

我的视频 每日关注 网站信用 图片 热门 下载

由于我做的是人类基因，所以下图步骤3选择可能不同，请自己选择

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

Gene ID Conversion Tool

Help and Tool Manual

Upload List Background

Upload Gene List 1

Demolist 1 Demolist 2
Upload Help

Step 1: Enter Gene List

A: Paste a list
 Clear

Or

B: Choose From a File
选择文件 未选择任何文件
Multi-List File ?

Step 2: Select Identifier
GENBANK_ACCESSION

Option 1:
Convert the gene list being selected in left panel to ENTREZ_GENE_ID (Default)

For species: Type your species name or id (e.g. Homo sapiens; 9606)

Submit to Conversion Tool

Option 2: Go Back to Submission Form

Step 3: List Type
Gene List (selected)
Background

Step 4: Submit List
Submit List

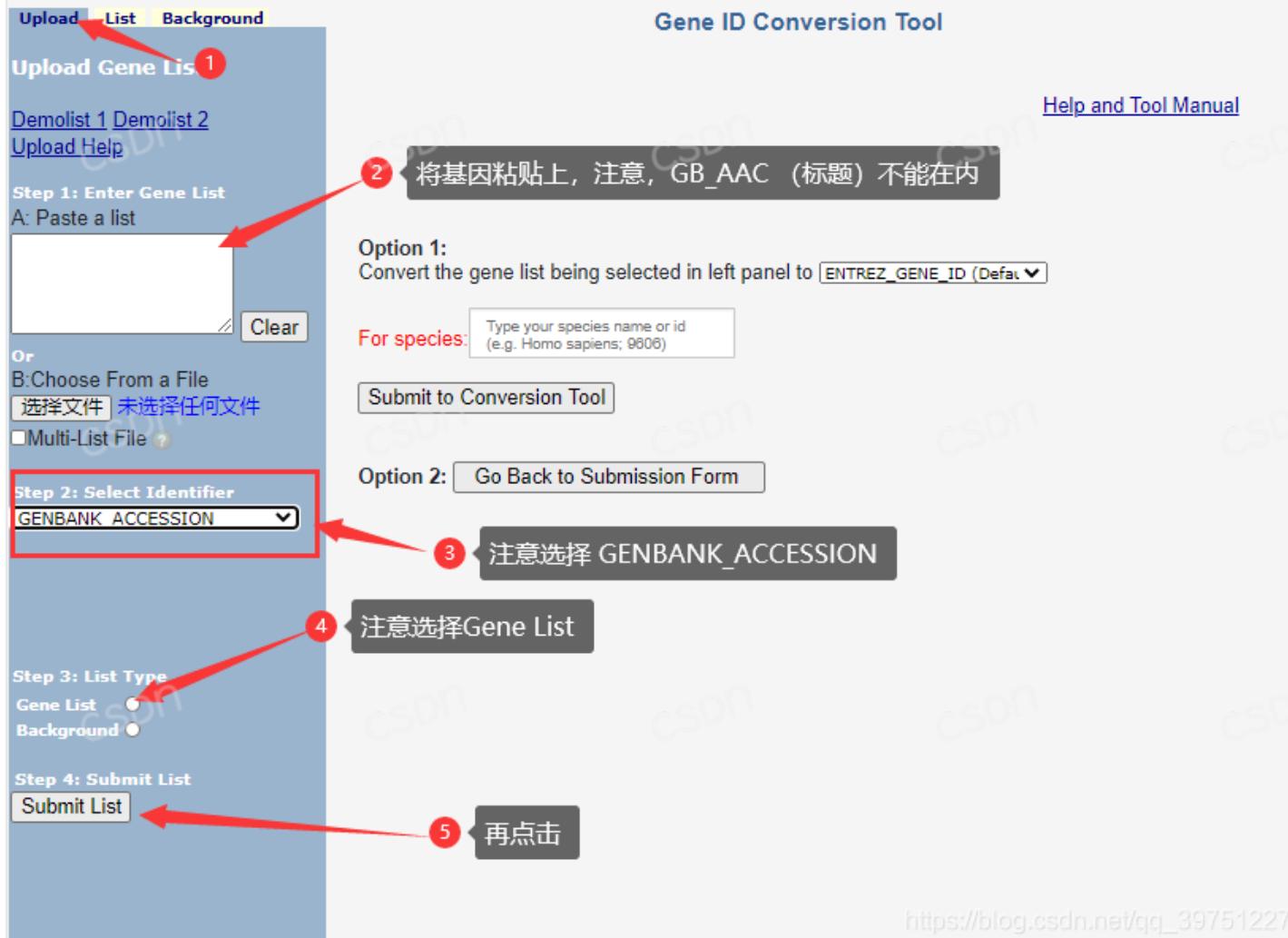
1

2 将基因粘贴上，注意，GB_AAC（标题）不能在内

3 注意选择 GENBANK_ACCESSION

4 注意选择 Gene List

5 再点击



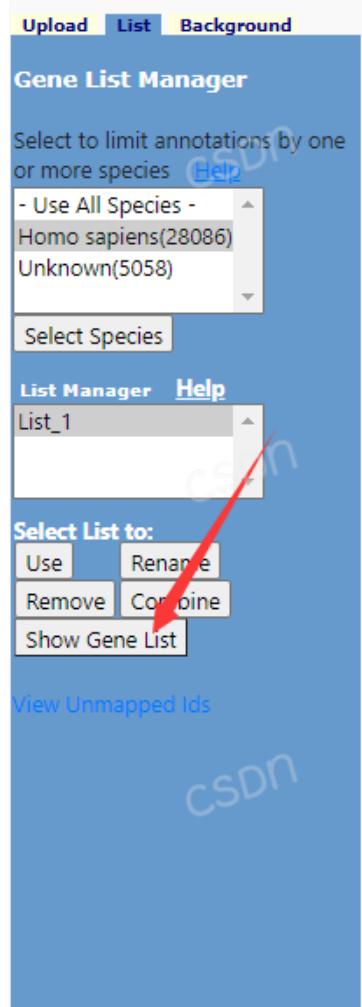
https://blog.csdn.net/qq_39751227

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227



之后我们可点击下载文件

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

Gene List Report

[Help and Manual](#)

Current Gene List: List_1

Current Background: Homo sapiens

24300 DAVID IDs

[Download File](#)

GENBANK_ACCESSION	Gene Name	Related Genes	Species
AA215598	long intergenic non-protein coding RNA 1170(LINC01170)	RG	Homo sapiens
AA293893	MAGI2 antisense RNA 1(MAGI2-AS1)	RG	Homo sapiens
AA393706	uncharacterized LOC105369443(LOC105369443)	RG	Homo sapiens
AA861243	long intergenic non-protein coding RNA 1617(LINC01617)	RG	Homo sapiens
AA936120	MAS related GPR family member F(MRGPRF)	RG	Homo sapiens
AB000465	NOP14 antisense RNA 1(NOP14-AS1)	RG	Homo sapiens
AB000466	NOP14 antisense RNA 1(NOP14-AS1)	RG	Homo sapiens
AB001328	solute carrier family 15 member 1(SLC15A1)	RG	Homo sapiens
AB002058	purinergic receptor P2X 6(P2RX6)	RG	Homo sapiens
AB002446	uncharacterized LOC728254(LOC728254)	RG	Homo sapiens
AB006651	mediator complex subunit 14(MED14)	RG	Homo sapiens
AB007870	nucleoporin 58(NUP58)	RG	Homo sapiens
AB007921	tetrastricopeptide repeat domain 39A(TTC39A)	RG	Homo sapiens
AB007954	uncharacterized LOC57235(KIAA0485)	RG	Homo sapiens
AB007978	uncharacterized LOC57242(KIAA0509)	RG	Homo sapiens
AB011145	endoplasmic reticulum protein 44(ERP44)	RG	Homo sapiens

注意，有些GB_ACC是没有基因的，所以我们需要对下载的文件进行进一步排序。

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791作者主页: https://blog.csdn.net/qq_39751227

XR_430626	uncharacterized LOC102723847 (LOC102723847)	RG	Homo sapiens
XR_430813	uncharacterized LOC100134423 (LOC100134423)	RG	Homo sapiens
XR_430888	uncharacterized LOC107984092 (LOC107984092)	RG	Homo sapiens
XR_431115	uncharacterized LOC102725193 (LOC102725193)	RG	Homo sapiens
XR_431974	uncharacterized LOC102724520 (LOC102724520)	RG	Homo sapiens
XR_432395	uncharacterized LOC101928906 (LOC101928906)	RG	Homo sapiens
XR_432469	uncharacterized LOC101927993 (LOC101927993)	RG	Homo sapiens
XR_433481	uncharacterized LOC101927943 (LOC101927943)	RG	Homo sapiens
XR_433584	long intergenic non-protein coding RNA 1595 (LINC01595)	RG	Homo sapiens
XR_433775	uncharacterized LOC102724491 (LOC102724491)	RG	Homo sapiens
XR_252233			
CD674917			
CU675766			
DB090165			
XR_246724			
XR_245666			
XR_250430			
XM_006710077			
AK094244			
X51791			
CA412304			
AW770842			
XR_426704			
BF570763			
BX108699			
BC068076			
XM_006710171			
BG218723			
XM_006718601			
DB026626			
XR_249390			
DA959652			
XR_425878			
XR_241657			
AK130723			
AF258585			
BC073932			

基因的排序 (请忽视)

准备工具

Sublime Text, 请自行下载

除如图所示外，其他全复制。

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

ENBANK_ACCESSION	Name	Species
EU794585	alpha-1-B glycoprotein(A1BG)	Homo sapiens
NM_130786	alpha-1-B glycoprotein(A1BG)	Homo sapiens
NM_000014	alpha-2-macroglobulin (A2M)	Homo sapiens
NR_040112	alpha-2-macroglobulin pseudogene 1(A2MP1)	Homo sapiens
NM_001017928	coiled-coil domain containing 58(CCDC58)	Homo sapiens
NM_000662	N-acetyltransferase 1(NAT1)	Homo sapiens
NM_000015	N-acetyltransferase 2(NAT2)	Homo sapiens
NM_001085	serpin family A member 3(SERPINA3)	Homo sapiens
NM_001086	arylacetamide deacetylase(AADAC)	Homo sapiens
NM_001087	angio associated migratory cell protein(AAMP)	Homo sapiens
NM_001166579	aralkylamine N-acetyltransferase(AANAT)	Homo sapiens
NM_001605	alanyl-tRNA synthetase(AARS)	Homo sapiens
NM_152608	SDE2 telomere maintenance homolog (SDE2)	Homo sapiens
NR_104662	uncharacterized LOC102467081(LOC102467081)	Homo sapiens
NM_000663	4-aminobutyrate aminotransferase(ABAT)	Homo sapiens
NR_104659	uncharacterized LOC102467080(LOC102467080)	Homo sapiens
NM_005502	ATP binding cassette subfamily A member 1(ABCA1)	Homo sapiens
NM_001606	ATP binding cassette subfamily A member 2(ABCA2)	Homo sapiens
NM_001089	ATP binding cassette subfamily A member 3(ABCA3)	Homo sapiens
NM_004299	ATP binding cassette subfamily B member 7(ABCB7)	Homo sapiens
NM_001025091	ATP binding cassette subfamily F member 1(ABCF1)	Homo sapiens
NR_104618	long intergenic non-protein coding RNA 1017(LINCO1017)	Homo sapiens
NM_000380	ATP binding cassette subfamily A member 4(ABCA4)	Homo sapiens
NR_104617	long intergenic non-protein coding RNA 1377(LINCO1377)	Homo sapiens
NM_005157	ABL proto-oncogene 1, non-receptor tyrosine kinase(ABL1)	Homo sapiens
NM_007313	ABL proto-oncogene 1, non-receptor tyrosine kinase(ABL1)	Homo sapiens
NR_104613	heart tissue-associated transcript 5(HRAT5)	Homo sapiens
NM_144633	potassium voltage-gated channel subfamily H member 8(KCNH8)	Homo sapiens
NM_001091	amine oxidase, copper containing 1(AOC1)	Homo sapiens
NM_001136001	ABL proto-oncogene 2, non-receptor tyrosine kinase(ABL2)	Homo sapiens
NM_007314	ABL proto-oncogene 2, non-receptor tyrosine kinase(ABL2)	Homo sapiens
NR_104655	uncharacterized LOC102467079(LOC102467079)	Homo sapiens
NM_020469	ABO blood group (transferase A, alpha 1-3-N-acetylgalactosaminyltransferase; transf	
NM_001159746	active BCR-related(ABR)	Homo sapiens
NW_021962	active BCR-related(ABR)	Homo sapiens
NR_104633	long intergenic non-protein coding RNA 603(LINCO0603)	Homo sapiens
NM_001607	acetyl-CoA acyltransferase 1(ACAA1)	Homo sapiens
NR_047462	long intergenic non-protein coding RNA 492(LINCO0492)	Homo sapiens
NM_198839	acetyl-CoA carboxylase alpha(ACACA)	Homo sapiens
NM_001093	acetyl-CoA carboxylase beta(ACACB)	Homo sapiens
AT701053	acetyl-CoA carboxylase beta(ACACB)	Homo sapiens
NM_001608	acyl-CoA dehydrogenase, long chain(ACADL)	Homo sapiens
NW_000016	acyl-CoA dehydrogenase, C-4 to C-12 straight chain(ACADM)	Homo sapiens
NM_000017	acyl-CoA dehydrogenase, C-2 to C-3 short chain(ACADS)	Homo sapiens
NM_001609	acyl-CoA dehydrogenase, short/branched chain(ACADSB)	Homo sapiens
NM_000018	acyl-CoA dehydrogenase, very long chain(ACADVL)	Homo sapiens
NM_000019	acetyl-CoA acetyltransferase 1(ACAT1)	Homo sapiens
NW_005891	acetyl-CoA acetyltransferase 2(ACAT2)	Homo sapiens
NM_001139459	consortin, connexin sorting protein(CNST)	Homo sapiens
NM_182609	consortin, connexin sorting protein(CNST)	Homo sapiens
NM_183377	acid sensing ion channel subunit 2(ASIC2)	Homo sapiens
NM_020039	acid sensing ion channel subunit 1(ASIC1)	Homo sapiens
NM_000665	acetylcholinesterase (cerebral blood vessel)(ACUB)	Homo sapiens

不要复制

先将文件贴到新的excel 中，进行排序

内容来源: csdn.net

作者昵称: Frms

原文地址: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

文件 开始 插入 页面布局 公式 数据 审阅 视图 帮助 特色功能 团队 百度网盘 操作说明搜索

GSE137578.top.table (1) - Excel

A1 alpha-1-B glycoprotein(A1BG) Homo sapiens

1 EU794525 alpha-1-B glycoprotein(A1BG) Homo sapiens

2 NM_134025 alpha-1-B glycoprotein(A1BG) Homo sapiens

3 NM_0012222M Homo sapiens

4 NR_047262 pseudogene 1(A2MP1) Homo sapiens

5 NM_0012222M training 58(CCDC58) Homo sapiens

6 NM_0012222M AT1) Homo sapiens

7 NM_0012222M AT2) Homo sapiens

8 NM_0012222M 3(SERPINA3) Homo sapiens

9 NM_0012222M base(AADAC) Homo sapiens

10 NM_0012222M ery cell protein(AAMP) Homo sapiens

11 NM_0012222M transferase(AANAT) Homo sapiens

12 NM_0012222M (AARS) Homo sapiens

13 NM_134025 lance homolog(SDE2) Homo sapiens

14 NR_104613 2467081(LOC102467081) Homo sapiens

15 NM_0012222M transferase(ABAT) Homo sapiens

16 NR_104613

17 NM_0012222M

18 NM_0012222M

19 NM_0012222M

20 NM_0012222M

21 NM_0012222M

22 NR_104613

23 NM_0012222M

24 NR_104613

25 NM_0012222M

26 NM_007313 ABL proto-oncogene 1, non-receptor tyrosine kinase(ABL1) Homo sapiens

27 NR_104613 heart tissue-associated transcript 5(HRAT5) Homo sapiens

排序(O)

- 升序(S)
- 降序(O)
- 将所选单元格颜色放在最前面(C)
- 将所选字体颜色放在最前面(F)
- 将选定格式的图标置顶(E)
- 自定义排序(U)...

Sheet1

https://blog.csdn.net/qq_39751227

排序完成之后，再复制全部 (ctrl + A全选, ctrl + C 复制) 新建一个文本文档并用 sublime text 打开，粘贴基因。

正则表达式

```
java
1 | 查询多个括号:
2 | \(\+[^\\n]+\+\)
3 |
```

内容来源: csdn.net 键获取完整项目代码 | 复制
作者昵称: Frms
原文地址: https://blog.csdn.net/qq_39751227/article/details/118791791
作者主页: https://blog.csdn.net/qq_39751227

```
4 | 删除括号左右侧:  
5 | \)+[^\n]+\n  
   | \n+[^\n]+\(\
```

点击查询，找到查找和替换，然后逐个输入上方正则表达式（就是上面每行表达式，注意中文是注解行，就别输入了）替换成换行符（也就是写"\n",注意不带引号）。

一般来说括号内是基因，有的 **基因说明** 中，括号内不一定是基因，请在进行上面正则表达式操作之前，**手动删除用括号括起来但里面又不是基因的文本（连带它的括号一起删了）**。

注意输入正则表达式顺序已排好，依次操作就好，别整错顺序了。

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

C:\Users\ASUS\Desktop\GSE137578\新建文本文档.txt - Sublime Text

文件(F) 编辑(E) 选择(S) 查找(I) 查看(V) 转到(G) 工具(T) 项目(P) 首选项(N) 帮助(H)

OPEN FILES

- 新建文本文档.txt
- 基因查找正则表达式.txt

基因查找正则表达式.txt

```
18 NM_001606 ATP binding cassette subfamily A member 2(ABCA2) Homo sapiens
19 NM_001089 ATP binding cassette subfamily A member 3(ABCA3) Homo sapiens
20 NM_004299 ATP binding cassette subfamily B member 7(ABC7) Homo sapiens
21 NM_001025091 ATP binding cassette subfamily F member 1(ABCF1) Homo sapiens
22 NR_104618 long intergenic non-protein coding RNA 1017(LINC01017) Homo sapiens
23 NM_000350 ATP binding cassette subfamily A member 4(ABCA4) Homo sapiens
24 NR_104617 long intergenic non-protein coding RNA 1377(LINC01377) Homo sapiens
25 NM_005157 ABL proto-oncogene 1, non-receptor tyrosine kinase(ABL1) Homo sapiens
26 NM_007313 ABL proto-oncogene 1, non-receptor tyrosine kinase(ABL1) Homo sapiens
27 NR_104613 heart tissue-associated transcript 5(HRAT5) Homo sapiens
28 NM_144633 potassium voltage-gated channel subfamily H member 8(KCNH8) Homo sapiens
29 NM_001091 amine oxidase, copper containing 1(AOC1) Homo sapiens
30 NM_001136001 ABL proto-oncogene 2, non-receptor tyrosine kinase(ABL2) Homo sapiens
31 NM_007314 ABL proto-oncogene 2, non-receptor tyrosine kinase(ABL2) Homo sapiens
32 NR_104655 uncharacterized LOC102467079(LOC102467079) Homo sapiens
33 NM_020469 ABO blood group (transferase A, alpha-1-3-N-acetylgalactosaminyltransferase; transferase B, alpha-1-3-galactosyltransferase)(ABO) Homo sapiens
34 NM_001159746 active BCR-related(ABR) Homo sapiens
35 NM_021962 active BCR-related(ABR) Homo sapiens
36 NR_104633 long intergenic non-protein coding RNA 603(LINC00603) Homo sapiens
37 NM_001607 acetyl-CoA acyltransferase 1(ACAA1) Homo sapiens
38 NR_047462 long intergenic non-protein coding RNA 492(LINC00492) Homo sapiens
39 NM_198839 acetyl-CoA carboxylase alpha(ACACA) Homo sapiens
40 NM_001093 acetyl-CoA carboxylase beta(ACACB) Homo sapiens
41 AY701053 acetyl-CoA carboxylase beta(ACACB) Homo sapiens
42 NM_001608 acyl-CoA dehydrogenase, long chain(ACADL) Homo sapiens
43 NM_000016 acyl-CoA dehydrogenase, C-4 to C-12 straight chain(ACADM) Homo sapiens
44 NM_000017 acyl-CoA dehydrogenase, C-2 to C-10 straight chain(ACADS) Homo sapiens
45 NM_001609 acyl-CoA dehydrogenase, short/branched chain(ACADSB) Homo sapiens
46 NM_000018 acyl-CoA dehydrogenase, very long chain(ACADVL) Homo sapiens
47 NM_000019 acetyl-CoA acetyltransferase 1(ACAT1) Homo sapiens
```

* Aa ⌂ Find What: \(\+[^\\n]\+\(\ ① 点击这里选择则查找替换

Replace With: ② 点击它

③ 输入此, 手动删除括号

Find Replace F3 Find All Replace All

1 of 23 matches; Wrapped past end of file

https://blog.csdn.net/qq_39751227

上图注意按顺序点击。

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

C:\Users\ASUS\Desktop\GSE137578\新建文本文档.txt - Sublime Text

文件(F) 编辑(E) 选择(S) 查找(I) 查看(V) 转到(G) 工具(T) 项目(P) 首选项(N) 帮助(H)

OPEN FILES

- 新建文本文档.txt
- 基因查找正则表达式.txt

基因查找正则表达式.txt

```
96 NM_207334 family with sequence similarity 43 member B(FAM43B) Homo sapiens
97 NM_000666 aminoacylase 1(ACY1) Homo sapiens
98 NM_001302617 acylphosphatase 1(ACYP1) Homo sapiens
99 NM_138448 acylphosphatase 2(ACYP2) Homo sapiens
100 AF339825 long intergenic non-protein coding RNA 555(LINC00555) Homo sapiens
101 NM_000022 adenosine deaminase(ADA) Homo sapiens
102 NR_047483 long intergenic non-protein coding RNA 554(LINC00554) Homo sapiens
103 NM_001164489 ADAM metallopeptidase domain 8(ADAM8) Homo sapiens
104 NM_001109 ADAM metallopeptidase domain 8(ADAM8) Homo sapiens
105 NR_047482 DOCK9 antisense RNA 2 (DOCK9-AS2) Homo sapiens
106 NM_001110 ADAM metallopeptidase domain 10(ADAM10) Homo sapiens
107 NM_001111 adenosine deaminase, RNA specific(ADAR) Homo sapiens
108 NM_001112 adenosine deaminase, RNA specific B1(ADARB1) Homo sapiens
109 NR_027673 adenosine deaminase, RNA specific B1(ADARB1) Homo sapiens
110 BC039360 long intergenic non-protein coding RNA 564(LINC00564) Homo sapiens
111 AK289895 adenosine deaminase, RNA specific B2 (ADARB2) Homo sapiens
112 NM_018702 adenosine deaminase, RNA specific B2 (ADARB2) Homo sapiens
113 BC034981 long intergenic non-protein coding RNA 561(LINC00561) Homo sapiens
114 NM_138805 family with sequence similarity 3 member D(FAM3D) Homo sapiens
115 AF339811 long intergenic non-protein coding RNA 557(LINC00557) Homo sapiens
116 NM_021116 adenylate cyclase 1(ADCY1) Homo sapiens
117 AF086361 long intergenic non-protein coding RNA 556(LINC00556) Homo sapiens
118 NM_020546 adenylate cyclase 2(ADCY2) Homo sapiens
119 NR_047484 PDX1 antisense RNA 1(PDX1-AS1) Homo sapiens
120 NM_004036 adenylate cyclase 3(ADCY3) Homo sapiens
121 NR_047492 long intergenic non-protein coding RNA 562(LINC00562) Homo sapiens
122 NM_183357 adenylate cyclase 5(ADCY5) Homo sapiens
123 NR_047495 long intergenic non-protein coding RNA 565(LINC00565) Homo sapiens
124 NM_015270 adenylate cyclase 6(ADCY6) Homo sapiens
125 NR_047493 long intergenic non-protein coding RNA 563(LINC00563) Homo sapiens
126 NM_001114 adenylate cyclase 7(ADCY7) Homo sapiens
```

* Aa ⌂ Find What: `\)+[^\n]+\n` 3

Replace With: `\n`

Find Replace All

111 of 288 matches

https://blog.csdn.net/qq_39751227

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

C:\Users\ASUS\Desktop\GSE137578\新建文本文档.txt - Sublime Text

文件(F) 编辑(E) 选择(S) 查找(I) 查看(V) 转到(G) 工具(T) 项目(P) 首选项(N) 帮助(H)

OPEN FILES

- 新建文本文档.txt
- 基因查找正则表达式.txt

基因查找正则表达式.txt

```
95 NM_000020 activin A receptor like type 1(ACVR1)
96 NM_207334 family with sequence similarity 43 member B(FAM43B)
97 NM_000666 aminoacylase 1(ACY1)
98 NM_001302617 acylphosphatase 1(ACYP1)
99 NM_138448 acylphosphatase 2(ACYP2)
100 AF339825 long intergenic non-protein coding RNA 555(LINC00555)
101 NM_000022 adenosine deaminase(ADA)
102 NR_047483 long intergenic non-protein coding RNA 554(LINC00554)
103 NM_001164489 ADAM metallopeptidase domain 8(ADAM8)
104 NM_001109 ADAM metallopeptidase domain 8(ADAM8)
105 NR_047482 DOCK9 antisense RNA 2 (DOCK9-AS2)
106 NM_001110 ADAM metallopeptidase domain 10(ADAM10)
107 NM_001111 adenosine deaminase, RNA specific(ADAR)
108 NM_001112 adenosine deaminase, RNA specific B1(ADARB1)
109 NR_027673 adenosine deaminase, RNA specific B1(ADARB1)
110 BC039360 long intergenic non-protein coding RNA 564(LINC00564)
111 AK289895 adenosine deaminase, RNA specific B2(ADARB2)
112 NM_018702 adenosine deaminase, RNA specific B2(ADARB2)
113 BC034981 long intergenic non-protein coding RNA 561(LINC00561)
114 NM_138805 family with sequence similarity 3 member D(FAM3D)
115 AF339811 long intergenic non-protein coding RNA 557(LINC00557)
116 NM_021116 adenylate cyclase 1(ADCY1)
117 AF086361 long intergenic non-protein coding RNA 556(LINC00556)
118 NM_020546 adenylate cyclase 2(ADCY2)
119 NR_047484 PDX1 antisense RNA 1(PDX1-AS1)
120 NM_004036 adenylate cyclase 3(ADCY3)
121 NR_047492 long intergenic non-protein coding RNA 562(LINC00562)
122 NM_183357 adenylate cyclase 5(ADCY5)
123 NR_047495 long intergenic non-protein coding RNA 565(LINC00565)
124 NM_015270 adenylate cyclase 6(ADCY6)
125 NR_047493 long intergenic non-protein coding RNA 563(LINC00563)
126 NM_001114 adenylate cyclase 7(ADCY7)
```

* Aa ⌂ Find What: \n+[^\\n]+\\(
Replace With: \\n

Find All Replace All

111 of 288 matches

https://blog.csdn.net/qq_39751227

上两个图示范了一个正则表达式的用法，请自行完成后续正则表达式的操作。

删除末尾多余文字

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

```
281 BIN1  
282 AMT  
283 TTC24  
284 AMY1C  
285 ANG  
286 ANGPT1  
287 CCDC185  
288 ANGPT2  
289 ANK1) Homo sapiens
```

https://blog.csdn.net/qq_39751227

删除开头多余文字

```
(S) 查找(I) 查看(V) 转到(G) 工具(T) 项目(P) 首选项(N) 帮助(H)  
新建文本文档.txt 基因查找正则表达式.txt  
1 EU794585 alpha-1-B glycoprotein(A1BG)  
2 A1BG  
3 A2M  
4 A2MP1  
5 CCDC58  
6 NAT1  
7 NAT2  
8 SERPINA3  
9 AADAC  
10 AAMP  
11 AANAT  
12 AARS  
13 SDE2  
14 LOC102467081  
15 ABAT  
16 LOC102467080  
17 ABCA1  
18 ABCA2  
19 ABCA3  
20 ABCB7  
21 ABCF1  
22 LINC01017  
23 ABCA4  
24 LINC01377  
25 ABL1  
26 ABL1  
27 HRAT5
```

https://blog.csdn.net/qq_39751227
内容来源: csdn.net

之后全部复制贴回对应行即可，另存为**csv文件**，我们这里命名为**GSE_HM**

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	adj.P.Val	P.Value	t	B	logFC	GB_ACC	SEQUENCE	SPOT_ID					
2	A_22_P00018823	0.00399	4.15E-04	8.34	-0.07346	1.0199385	A23747	TCTCTTGATGCCGTATCCAGAGGA	A_22_P00018823	ALOX5				
3	A_33_P3329958	0.311093	1.40E-01	1.75	-6.6896	0.1160892	A25493	CTGAGCCTGGCGCGCAGCACCCA	A_33_P3329958	LINC00556				
4	A_23_P12620	0.20024	6.86E-02	2.32	-5.93983	0.2657097	A25966	CATTATGGCTTGCACATCTCTTTAT	A_23_P12620	ADH1B				
5	A_22_P00005154	0.292034	1.26E-01	-1.84	-6.57888	-0.067717	A25968	GTTCTTAAAGGGGATGGAAATCTA	A_22_P00005154	LINC00557				
6	A_21_P0009693	0.314165	1.43E-01	-1.74	-6.70575	-0.087999	A25969	GTAAGAGAACATGCTGAGTAATGCAG	A_21_P0009693	LINC00558				
7	A_33_P3213522	0.395638	2.09E-01	1.44	-7.07894	0.0966604	A25970	TTTGAATACATATGTATAACCTTGG	A_33_P3213522	ALPP				
8	A_23_P121898	0.020295	3.70E-03	5.14	-2.63459	0.3740517	AA045595	TGCAAAGCCGTAACCTAAATGCATA	A_23_P121898	APBB1				
9	A_22_P00000227	0.422043	2.32E-01	-1.36	-7.17748	-0.05073	AA062935	CTACGTAGACCATTTCATGTGTATA	A_22_P00000227	APBA2				
10	A_22_P00001509	0.249287	9.66E-02	-2.04	-6.30365	-0.13359	AA211868	AATCAACACCCACTGTGATAAACCA	A_22_P00001509	ANK3				
11	A_23_P73526	0.149517	4.66E-02	2.63	-5.51916	1.0963425	AA213559	CAAGGCTCTGAAATGCCAACACG	A_23_P73526	ACTL6A				
12	A_21_P0012651	0.133547	4.03E-02	-2.75	-5.36014	-0.500611	AA215598	ACTCATGAAAAAGCCCTCGGGAGG	A_21_P0012651	ADARB2				
13	A_21_P0011325	0.289145	1.24E-01	1.85	-6.56168	0.0809466	AA218968	CTAAAGCATTTCTTGGATGGTGATT	A_21_P0011325	ACTG2				
14	A_23_P312174	0.104784	2.97E-02	-3.02	-5.02074	-0.689723	AA234626	TTCCTCAGAGCCTTGGAAATTCTATTA	A_23_P312174	ACACB				
15	A_33_P3316313	0.178623	5.89E-02	-2.44	-5.77557	-0.180787	AA262526	CAAGGAAGCTTTCATCTCTTAAA	A_33_P3316313	LINC00561				
16	A_33_P3274560	0.093757	2.58E-02	-3.14	-4.86506	-0.529154	AA292921	ATGTCTAAAGCTTGCATTCAAGCA	A_33_P3274560	LINC00562				
17	A_22_P00003103	0.389492	2.03E-01	1.46	-7.05474	0.7031414	AA293893	AGCAGGCTGGTGGAGGCAGAGCAGA	A_22_P00003103	LINC01599				
18	A_21_P0006164	0.427581	2.37E-01	1.35	-7.19782	0.2594651	AA311247	AAAGGTAAAGATTGCATAGAGAA	A_21_P0006164	UFL1-AS1				
19	A_22_P00022385	0.0243	4.70E-03	-4.86	-2.91162	-1.012119	AA311918	GACACTGCTACATGCCAGAACAA	A_22_P00022385	A1BG				
20	A_23_P63379	0.361161	1.80E-01	-1.56	-6.93405	-0.299064	AA315543	ATGGCCCTTCCCTAGATATACTGC	A_23_P63379	ALDOAP2				
21	A_33_P3299285	0.263839	1.06E-01	1.97	-6.40034	0.2907585	AA333164	AAATGCAGCTGCTCAAAGCCTCTA	A_33_P3299285	A2M				
22	A_33_P3253234	0.33639	1.60E-01	1.65	-6.81972	0.1976279	AA335815	TCAGAGTGCCTGTTGCTGTACAA	A_33_P3253234	NAT2				
23	A_21_P0001058	0.401331	2.14E-01	1.43	-7.1021	0.0689164	AA345637	AAACAATTAAATTAGGGACAAAGA	A_21_P0001058	ACADM				
24	A_33_P3232655	0.379949	1.95E-01	-1.5	-7.01344	-0.057785	AA354059	TTTCGCCACCATGACCAACTTCAGC	A_33_P3232655	ACADS				
25	A_21_P0004544	0.20202	6.94E-02	-2.31	-5.95351	-0.08567	AA371472	GCTAGCATATTAGAGCGCTAACAA	A_21_P0004544	ACADV1				
26	A_33_P3224095	0.231762	8.58E-02	2.14	-6.17869	2.6222833	AA378382	TTGTTTTCATGCTTGCAGGATTGG	A_33_P3224095	ACAT1				
27	A_24_P254506	0.276336	1.14E-01	-1.91	-6.47878	-0.069612	AA382521	ATGAAAAGAGAGAACACTCCACCTA	A_24_P254506	ACVRL1				
28	A_24_P759477	0.239354	9.02E-02	-2.1	-6.2323	-0.373204	AA382715	AGAATATGGTACTTCACATAATAT	A_24_P759477	ADA				
29	A_23_P18267	0.208021	7.25F-02	2.27	-5.99949	0.1469642	AA383451	CCACCCCCAGCAGCTGCAGCATGTT	A_23_P18267	ADRB2				

总之，一定要注意排序，排序是都有的数据（这里是GB_ACC）

不允许有重复的' row.names' 解决问题

在R studio中，我们发现输入以下代码可能报错：

```
r
1 | df = read.csv("文件目录", header = T, row.names = 1)
```

```
1 | 不允许有重复的' row.names' 解决问题
```

出错原因是**数据格式不对**，但这在网上很少解释清楚，常错的原因有以下几点：

内容来源: csdn.net

作者昵称: Frms

原文地址: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

- | | |
|---|--------------|
| 1 | 1. 第一行有重复名 |
| 2 | 2. CSV文件格式错误 |

原因1用Excel修改第一行重复名称即可。

我们介绍2如何解决，这里我们先用 **sublime text** （软件，自行下载）打开文件修改即可，（事实上，任何文本编辑器都可以）。

我们需要分析的csv的数据格式如下：

一键获取完整项目代码 | 复制

1	1. 每行末尾必须有英文符号的逗号
2	2. 最后一行需要换行（即保留一行空白）
3	3. 如果存在末行连续逗号，要删除，如图。
4	4. 注意，开头没有空格

```

33 M13712,9.1342645577,3.788884458,3.589298889,3.793233419,8.688488888,8.982721999,
40 SNORA6,6.916103562,6.927701925,7.171010622,7.956925056,8.539428128,8.356588493,
41 FTL,2.542559725,2.723108464,1.915541806,1.023080373,1.031172611,1.028791315,
42 STXBP5-AS1,8.888303373,8.559949593,8.6302204,9.930272329,10.32627517,10.07183756,
43 PARP9,7.856418473,7.471736308,7.515550174,5.873730039,5.944156249,6.112442391,
44 CASC15,6.307522121,6.546775439,6.548798923,7.695901972,7.674456297,7.598914013,
45 SPATA22,2.488128533,3.380152957,3.02900219,1.684208773,1.746724302,1.382890118,
46 VLDLR-AS1,10.63384409,10.65569558,10.64451802,11.81389286,11.73951679,11.57921787,
47 LMC1-AS1,8.277915105,7.689313839,7.698282271,8.900205371,9.13790645,9.169347015,
48 MEG3,8.091852134,8.362672903,8.24939414,6.599731899,6.37464649,6.292378941,
49 GS1-259H13.2,10.38270147,10.22298906,10.31473995,12.37796504,12.50552936,12.26959956,
50 FAM105A,6.030540902,5.930833908,6.015771589,3.666757349,4.130155652,3.509733207,
51 LINC0113,3.171773672,2.842703202,3.175306171,4.275265731,4.576879039,3.992524185,
52 ,,,,
53 ,,,,
54 ,,,,
55

```

https://blog.csdn.net/qq_39751227



内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227

```
33 11.2172,31.3428,19977,31.73833,11.38,31.5032,31.8833,31.7382,31.713,31.883,18.88888,31.5327,21.9933,  
40 SNORA6,6.916103562,6.927701925,7.171010622,7.956925056,8.539428128,8.356588493,  
41 FTL,2.542559725,2.723108464,1.915541806,1.023080373,1.031172611,1.028791315,  
42 STXBP5-AS1,8.888303373,8.559949593,8.6302204,9.930272329,10.32627517,10.07183756,  
43 PARP9,7.856418473,7.471736308,7.515550174,5.873730039,5.944156249,6.112442391,  
44 CASC15,6.307522121,6.546775439,6.548798923,7.695901972,7.674456297,7.598914013,  
45 SPATA22,2.488128533,3.380152957,3.02900219,1.684208773,1.746724302,1.382890118,  
46 VLDLR-AS1,10.63384409,10.65569558,10.64451802,11.81389286,11.73951679,11.57921787,  
47 LMCD1-AS1,8.277915105,7.689313839,7.698282271,8.900205371,9.13790645,9.169347015,  
48 MEG3,8.091852134,8.362672903,8.24939414,6.599731899,6.37464649,6.292378941,  
49 GS1-259H13.2,10.38270147,10.22298906,10.31473995,12.37796504,12.50552936,12.26959956,  
50 FAM105A,6.030540902,5.930833908,6.015771589,3.666757349,4.130155652,3.509733207,  
51 LINC01133,3.171773672,2.842703202,3.175306171,4.275265731,4.576879039,3.992524185,  
52
```

https://blog.csdn.net/qq_39751227

这样问题就基本解决了。

内容来源: csdn.net

作者昵称: Frms

原文链接: https://blog.csdn.net/qq_39751227/article/details/118791791

作者主页: https://blog.csdn.net/qq_39751227