



# R语言零基础基因/数据差异分析（一）

已于 2025-11-12 13:43:32 修改

编辑



R语言零基础基因/... 专栏收录该内容

5 篇文章



本文档详细介绍了如何下载和处理基因数据，包括从GEO平台获取GSE137578基因数据，删除注释行，格式转换，以及利用GEO分析绘制拟火山图的步骤。同时强调了每一步的重要性，提供了数据处理和制图所需的数据结构及注意事项。

## 文章目录

[介绍](#)

[环境搭建](#)

[软件下载](#)

[结果展示](#)

[基因数据下载流程](#)

[基因数据处理](#)

[利用GEO分析绘制拟火山图](#)

本系列完全免费，目的在于发扬开源精神，**绝不会**收取任何费用！！

防止【CSDN】再次未经我允许，擅自列为Vip文章，请在[Github](#)获取免费全文。

地址：<https://github.com/FrankMilesFrms/> 下的 Zero-basis-genetic-data-difference-analysis-of-R-language

注意，本系列有连贯性，每一步都很详细，每一步都很重要，请耐心等待！！

如果你出现了其他错误，或有疑问，请在[评论区留言](#)。最好不要私信发送，消息经常被屏蔽。

## 介绍

内容来源：[csdn.net](#)

作者昵称：[Frms](#)

原文链接：[https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页：[https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

- |   |                            |
|---|----------------------------|
| 1 | 本系列文主要依据真实论文制图流程，详细说明制图过程， |
| 2 | 其中包括：                      |
| 3 | 1. 基因数据下载                  |
| 4 | 2. 制图所需数据格式                |
| 5 | 3. 火山图制作流程                 |
| 6 | 4. 聚类热图制作流程                |

## 环境搭建

### 软件下载 [🔗](#)

[移步至此学习](#)

## 结果展示

基因[数据处理](#) [🔗](#)

内容来源：[csdn.net](https://blog.csdn.net)

作者昵称：Frms

原文链接：[https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页：[https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

	A	B	C	D	E	F	G	H
1	ID_REF	GSM40819	GSM40819	GSM40819	GSM40819	GSM40819	GSM40819	22
2	(+)E1A_r6C	17.4871	17.49766	17.48005	17.47459	17.46447	17.45657	
3	(+)E1A_r6C	1.462313	1.368858	1.651992	1.343592	1.368677	1.408667	
4	(+)E1A_r6C	1.486171	1.602607	1.507144	1.549365	1.759575	1.544174	
5	(+)E1A_r6C	4.308478	4.552691	4.484554	5.18769	4.950299	4.904998	
6	(+)E1A_r6C	7.357802	7.503545	7.45115	8.053392	7.919011	7.804449	
7	(+)E1A_r6C	8.626238	8.769348	8.630123	9.238593	9.193987	9.104532	
8	(+)E1A_r6C	10.47383	10.54576	10.53931	11.05086	10.94846	10.86198	
9	(+)E1A_r6C	12.95811	12.96402	12.92151	13.29651	13.23619	13.17509	
10	(+)E1A_r6C	14.67739	14.68061	14.66084	15.08704	14.98829	14.91256	
11	(+)E1A_r6C	15.23048	15.24316	15.2295	15.51636	15.51329	15.47072	
12	3xSLv1	1.379378	1.483298	1.398159	1.344432	1.299346	1.366059	
13	A_19_P003	10.75624	10.59491	10.62647	8.930466	8.878832	8.956574	
14	A_19_P003	1.475371	1.485435	1.584281	1.462175	2.089482	2.306061	
15	A_19_P003	7.588089	7.196441	6.994468	6.336668	6.478453	7.146523	
16	A_19_P003	1.136111	1.121464	1.075821	1.216028	1.054102	4.35736	
17	A_19_P003	2.624631	2.835548	3.177159	2.002517	2.264473	2.836674	
18	A_19_P003	1.304456	1.310319	1.263381	1.20404	1.679507	1.265237	
19	A_19_P003	1.334372	1.342837	1.293907	1.217565	1.253325	2.25718	
20	A_19_P003	1.29993	1.357086	1.376344	1.344213	1.33693	1.318634	
21	A_19_P003	1.340204	1.208344	1.242823	1.284664	1.226446	1.22708	
22	A_19_P003	1.958668	3.067678	2.346347	3.596217	3.114963	1.968575	
23	A_19_P003	7.868755	8.052798	8.083573	7.835505	7.686385	7.800356	
24	A_19_P003	10.63384	10.6557	10.64452	11.81389	11.73952	11.57922	
25	A_19_P003	6.901836	6.56915	6.634874	6.561836	6.518923	6.394467	
26	A_19_P003	4.457062	4.30679	1.998701	2.807908	1.670587	3.051439	
27	A_19_P003	9.149735	8.420022	8.547316	8.279935	8.615163	8.790046	
28	A_19_P003	1.444675	1.428843	1.465946	1.449134	1.462573	1.488933	
29	A_19_P003	1.335129	2.657957	2.463007	2.947351	3.187096	2.792871	
		GSE137578_series_matrix			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/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

58329	ETG02_366	1.602682	1.801143	1.442416	1.273126	1.268069	1.289363				
58330	ETG04_277	1.281167	1.864192	1.488833	4.081985	1.405714	1.322708				
58331	ETG05_367	1.290035	1.541436	1.342019	1.587089	1.319309	1.440619				
58332	ETG05_660	1.610463	1.930714	1.579232	1.967686	1.39294	1.88451				
58333	ETG07_105	1.432225	1.339643	1.40671	1.457039	1.263779	1.283073				
58334	ETG08_142	1.345146	1.587567	1.397005	1.47015	1.558919	1.628247				
58335	ETG09_205	1.402411	1.301676	1.69388	1.510227	1.273777	1.275348				
58336	ETG09_354	1.4763	1.27138	1.375604	1.259499	1.247675	1.268259				
58337	ETG09_487	1.300019	1.261195	1.264987	1.242151	1.241167	1.25782				
58338	ETG10_134	1.670758	1.514649	1.293749	1.766785	1.349963	1.380169				
58339	ETG10_195	1.582988	1.451212	1.537539	1.513164	1.7045	1.41375				
58340	ETG10_234	1.295708	1.399567	1.381228	1.314529	1.343818	1.34781				
58341	ETG10_236	1.256427	1.581387	1.35334	1.235912	1.352213	1.453238				
58342	GE_BrightC	14.89571	14.95447	14.70748	14.37462	15.1048	15.10763				
58343	series_matrix_table_end										
58344											<a href="https://blog.csdn.net/qq_39751227/">https://blog.csdn.net/qq_39751227/</a>
58345											

注意删除末行注释

## 基因数据下载流程

以GSE137578基因为例，先下载 如图所示文件，并解压如图。

1.


内容来源: csdn.net

作者昵称: Frms

原文链接: [https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

平台 (1) [GPL21185](#) Agilent-072363 SurePrint G3 Human GE v3 8x60K 微阵列芯片 039494  
[探针名称版本]

样品 (6) [GSM4081917](#) HAEC\_Untreatment\_Rep1 (mRNA)  
 更多的... [GSM4081918](#) HAEC\_Untreatment\_Rep2 (mRNA)  
[GSM4081919](#) HAEC\_Untreatment\_Rep3 (mRNA)

关系

生物工程 [PRJNA565963](#)

使用 GEO2R 进行分析


下载家庭


[SOFT 格式的族文件](#)


[MINIML 格式的族文件](#)

[系列矩阵文件](#)

格式

[柔软的](#) 

[迷你机器语言](#) 

[文本文件](#) 

补充文件	尺寸	下载	文件类型/资源
<a href="#">GSE137578_RAW.tar</a>	73.4 兆字节	<a href="#">(http)</a> <a href="#">(自定义)</a>	TAR (TXT)

原始数据作为补充文件提供


样本表中包含的处理数据

[https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

2.

## Index of /geo/series/GSE137nnn/GSE137578/matrix

Name	Last modified	Size
<a href="#">Parent Directory</a>		-
<a href="#">GSE137578_series_matrix.txt.gz</a>	2021-06-27 00:33	2.0M

3. 新建一个excel  , 并将解压好的文本拖进去。

内容来源: [csdn.net](https://blog.csdn.net/qq_39751227)

作者昵称: [Frms](#)

原文链接: [https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

名称	修改日期	类型	大小
GSE137578_series_matrix	2021/7/15 14:02	文本文档	4,981 KB
GSE	2021/7/15 14:04	Microsoft Excel ...	7 KB

4. 注意，打开后，此时我们需要对数据进行处理。

GSE137578_series_matrix - Excel(产品激活失败)													登录	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	共享
A1	!Series_title																										
1	!Series_title	Novel mirnas participate in Atherosclerosis through inflammation and apoptosis pathways (HAEC mRNA)																									
2	!Series_geo_accession	GSE137578																									
3	!Series_status	Public on Sep 22 2020																									
4	!Series_submission_date	Sep 17 2019																									
5	!Series_last_update_date	Sep 27 2020																									
6	!Series_summary	Atherosclerosis (AS), which is a chronic inflammatory disease, is the leading cause of death worldwide. Although several miRNAs have been reported to participate in regulatin																									
7	!Series_overall_design	Totally, there were six mRNA samples divided into two equal-sized groups, AS group and normal control group.																									
8	!Series_type	Expression profiling by array																									
9	!Series_contributor	Jia-Yan,Guo																									
10	!Series_contributor	Han-Bing,Mei																									
11	!Series_contributor	Qing-Yuan,Meng																									
12	!Series_contributor	Zhen,Sheng																									
13	!Series_contributor	Murielle,M,茅niant																									
14	!Series_contributor	Hong,Yin																									
15	!Series_sample_id	GSM4081917 GSM4081918 GSM4081919 GSM4081920 GSM4081921 GSM4081922																									
16	!Series_contact_name	Zhen,Sheng																									
17	!Series_contact_email	zsheng@amgen.com																									
18	!Series_contact_institute	Amgen Biopharmaceutical Research & Development (Shanghai) Co., Ltd																									
19	!Series_contact_address	F13, Building 2, 4560 Jinke Road																									
20	!Series_contact_city	Shanghai																									
21	!Series_contact_zip/postal	201210																									
22	!Series_contact_country	China																									
23	!Series_supplementary_file	ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE137nnn/GSE137578/suppl/GSE137578_RAW.tar																									
24	!Series_platform_id	GPL21185																									
25	!Series_platform_taxid	9606																									
26	!Series_sample_taxid	9606																									
27	!Series_relation	BioProject: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA565963																									
28																											
29	!Sample title	HAEC Untreated Rep1 (mRNA) HAEC Untreated Rep2 (mRNA) HAEC Untreated Rep3 (mRNA) HAEC Untreated Rep4 (mRNA) HAEC Untreated Rep5 (mRNA)																									

1. 在以 ! 开头的行, 均是\*\*注释行\*\*, 要全部删除, 如图示,
2. 删除方法: 选择好区域后, 右键, 删除整行即可

GSE137578\_series\_matrix - 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 共享

	A	B	C	D	E	F	G	H	I	J	K	L
1	!Series_title	Novel mirnas participate in Atherosclerosis through inflammation and apoptosis pathways (HAEC mRNA)										
2	!Series_geo_accession	GSE137578										
3	!Series_status	Public on Sep 22 2020										
4	!Series_submission_date	Sep 17 2019										
5	!Series_last_update_date	Sep 27 2020										
6	!Series_summary	Atherosclerosis (AS), which is a chronic inflammatory disease, is the leading cause of death worldwide. Although several miRNAs have been reported to participate in regulati										
7	!Series_overall_design	Totally, there were six mRNA samples divided into two equal-sized groups, AS group and normal control group.										
8	!Series_type	Expression profiling by array										
9	!Series_contributor	Jia-Yan,,Guo										
10	!Series_contributor	Han-Bing,,Mei										
11	!Series_contributor	Qing-Yuan,,Meng										
12	!Series_contributor	Zhen,,Sheng										
13	!Series_contributor	Murielle,M,V茅niant										
14	!Series_contributor	Hong,,Yin										
15	!Series_sample_id	GSM4081917 GSM4081918 GSM4081919 GSM4081920 GSM4081921 GSM4081922										
16	!Series_contact_name	Zhen,,Sheng										
17	!Series_contact_email	zsheng@amgen.com										
18	!Series_contact_institute	Amgen Biopharmaceutical Research & Development (Shanghai) Co., Ltd										
19	!Series_contact_address	F13, Building 2, 4560 Jinke Road										
20	!Series_contact_city	Shanghai										
21	!Series_contact_zip/postal	201210										
22	!Series_contact_country	China										
23	!Series_supplementary_file	ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE137nnn/GSE137578/suppl/GSE137578_RAW.tar										
24	!Series_platform_id	GPL21185										
25	!Series_platform_taxid	9606										
26	!Series_sample_taxid	9606										
27	!Series_relation	BioProject: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA565963										
28												
29	!Sample title	HAEC Untreated Rep1 (mRNA)	HAEC Untreated Rep2 (mRNA)	HAEC Untreated Rep3 (mRNA)	HAEC Oxidized Rep1 (mRNA)	HAEC Oxidized Rep2 (mRNA)	HAEC Oxidized Rep3 (mRNA)					

GSE137578\_series\_matrix

删除之后:

内容来源: csdn.net

作者昵称: Frms

原文链接: [https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)



GSE137578\_series\_matrix - 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

E9 : 13.29650813

	A	B	C	D	E	F	G	H	I	J	K	L
1	ID_REF	GSM4081917	GSM4081918	GSM4081919	GSM4081920	GSM4081921	GSM4081922					
2	(+)E1A_r60.1	17.48710134	17.49766017	17.48005	17.47459	17.46447	17.45657					
3	(+)E1A_r60.3	1.462312578	1.368858407	1.651992	1.343592	1.368677	1.408667					
4	(+)E1A_r60_a104	1.486171298	1.602607403	1.507144	1.549365	1.759575	1.544174					
5	(+)E1A_r60_a107	4.308477578	4.552690618	4.484554	5.18769	4.950299	4.904998					
6	(+)E1A_r60_a135	7.357801775	7.503544758	7.45115	8.053392	7.919011	7.804449					
7	(+)E1A_r60_a20	8.626237791	8.769347592	8.630123	9.238593	9.193987	9.104532					
8	(+)E1A_r60_a22	10.47382667	10.54576398	10.53931	11.05086	10.94846	10.86198					
9	(+)E1A_r60_a97	12.95810701	12.9640176	12.92151	13.29651	13.23619	13.17509					
10	(+)E1A_r60_n11	14.6773913	14.68061269	14.66084	15.08704	14.98829	14.91256					
11	(+)E1A_r60_n9	15.23048111	15.24315894	15.2295	15.51636	15.51329	15.47072					
12	3xSLv1	1.379378437	1.483298174	1.398159	1.344432	1.299346	1.366059					
13	A_19_P00315452	10.75623522	10.59491456	10.62647	8.930466	8.878832	8.956574					
14	A_19_P00315492	1.475370885	1.485435071	1.584281	1.462175	2.089482	2.306061					
15	A_19_P00315493	7.588089457	7.196441477	6.994468	6.336668	6.478453	7.146523					
16	A_19_P00315502	1.136110924	1.121463728	1.075821	1.216028	1.054102	4.35736					
17	A_19_P00315506	2.624630955	2.835548399	3.177159	2.002517	2.264473	2.836674					
18	A_19_P00315518	1.30445565	1.310318887	1.263381	1.20404	1.679507	1.265237					
19	A_19_P00315519	1.334371574	1.342836885	1.293907	1.217565	1.253325	2.25718					
20	A_19_P00315529	1.299930376	1.357085951	1.376344	1.344213	1.33693	1.318634					
21	A_19_P00315541	1.340204091	1.208343806	1.242823	1.284664	1.226446	1.22708					
22	A_19_P00315543	1.958668428	3.067677893	2.346347	3.596217	3.114963	1.968575					
23	A_19_P00315551	7.868754867	8.052798248	8.083573	7.835505	7.686385	7.800356					
24	A_19_P00315581	10.63384409	10.65569558	10.64452	11.81389	11.73952	11.57922					
25	A_19_P00315584	6.901835865	6.569149831	6.634874	6.561836	6.518923	6.394467					
26	A_19_P00315593	4.457061905	4.306789979	1.998701	2.807908	1.670587	3.051439					
27	A_19_P00315603	9.149734504	8.420021956	8.547316	8.279935	8.615163	8.790046					
28	A_19_P00315625	1.444674752	1.42884266	1.465946	1.449134	1.462573	1.488933					
29	A_19_P00315627	1.335129437	2.657957241	2.463007	2.947351	3.187096	2.792871					

GSE137578\_series\_matrix

就绪

注意，尾部图片选择的整行都要删除：

内容来源：csdn.net

作者昵称：Frms

原文链接：[https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页：[https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

GSE137578\_series\_matrix - 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

A58343 : X ✓ fx lseries\_matrix\_table\_end

	A	B	C	D	E	F	G	H	I	J	K	L
58323	ERCC-00163_81	1.381120247	1.514078627	1.530836	1.607873	1.345794	1.673097					
58324	ERCC-00164_60	1.489989917	1.417526204	1.390162	1.560518	1.419947	1.615687					
58325	ERCC-00165_60	1.342669955	1.354468059	1.347193	1.381128	1.332374	1.359462					
58326	ERCC-00168_248	1.205076891	1.208801004	1.402477	1.194685	1.273737	1.318345					
58327	ERCC-00170_100	1.743079964	1.8489748	1.707604	1.64097	1.703133	1.413627					
58328	ERCC-00171_229	1.64281216	1.424622043	1.679792	1.402745	1.50248	1.609709					
58329	ETG02_36680	1.602681602	1.801143246	1.442416	1.273126	1.268069	1.289363					
58330	ETG04_27747	1.28116683	1.864191594	1.488833	4.081985	1.405714	1.322708					
58331	ETG05_36762	1.290034832	1.541435807	1.342019	1.587089	1.319309	1.440619					
58332	ETG05_66023	1.610462553	1.930713711	1.579232	1.967686	1.39294	1.88451					
58333	ETG07_105829	1.43222504	1.339643093	1.40671	1.457039	1.263779	1.283073					
58334	ETG08_142674	1.345145721	1.587566954	1.397005	1.47015	1.558919	1.628247					
58335	ETG09_205211	1.402410568	1.301675736	1.69388	1.510227	1.273777	1.275348					
58336	ETG09_35454	1.4762998	1.271379507	1.375604	1.259499	1.247675	1.268259					
58337	ETG09_48764	1.30001947	1.261194542	1.264987	1.242151	1.241167	1.25782					
58338	ETG10_13482	1.670757671	1.514649481	1.293749	1.766785	1.349963	1.380169					
58339	ETG10_195139	1.582988086	1.451211557	1.537539	1.513164	1.7045	1.41375					
58340	ETG10_234183	1.295708115	1.399566981	1.381228	1.314529	1.343818	1.34781					
58341	ETG10_236652	1.25642744	1.581386892	1.35334	1.235912	1.352213	1.453238					
58342	GE_BrightCorner	14.89570989	14.9544665	14.70748	14.37462	15.1048	15.10763					
58343	lseries_matrix_table_end											
58344												
58345												
58346												
58347												
58348												
58349												
58350												
58351												

GSE137578\_series\_matrix

就绪

删除尾部结果:

内容来源: csdn.net

作者昵称: Frms

原文链接: [https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

GSE137578\_series\_matrix - 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

A58343

	A	B	C	D	E	F	G	H	I	J	K	L
58323	ERCC-00163_81	1.381120247	1.514078627	1.530836	1.607873	1.345794	1.673097					
58324	ERCC-00164_60	1.489989917	1.417526204	1.390162	1.560518	1.419947	1.615687					
58325	ERCC-00165_60	1.342669955	1.354468059	1.347193	1.381128	1.332374	1.359462					
58326	ERCC-00168_248	1.205076891	1.208801004	1.402477	1.194685	1.273737	1.318345					
58327	ERCC-00170_100	1.743079964	1.8489748	1.707604	1.64097	1.703133	1.413627					
58328	ERCC-00171_229	1.64281216	1.424622043	1.679792	1.402745	1.50248	1.609709					
58329	ETG02_36680	1.602681602	1.801143246	1.442416	1.273126	1.268069	1.289363					
58330	ETG04_27747	1.28116683	1.864191594	1.488833	4.081985	1.405714	1.322708					
58331	ETG05_36762	1.290034832	1.541435807	1.342019	1.587089	1.319309	1.440619					
58332	ETG05_66023	1.610462553	1.930713711	1.579232	1.967686	1.39294	1.88451					
58333	ETG07_105829	1.43222504	1.339643093	1.40671	1.457039	1.263779	1.283073					
58334	ETG08_142674	1.345145721	1.587566954	1.397005	1.47015	1.558919	1.628247					
58335	ETG09_205211	1.402410568	1.301675736	1.69388	1.510227	1.273777	1.275348					
58336	ETG09_35454	1.4762998	1.271379507	1.375604	1.259499	1.247675	1.268259					
58337	ETG09_48764	1.30001947	1.261194542	1.264987	1.242151	1.241167	1.25782					
58338	ETG10_13482	1.670757671	1.514649481	1.293749	1.766785	1.349963	1.380169					
58339	ETG10_195139	1.582988086	1.451211557	1.537539	1.513164	1.7045	1.41375					
58340	ETG10_234183	1.295708115	1.399566981	1.381228	1.314529	1.343818	1.34781					
58341	ETG10_236652	1.25642744	1.581386892	1.35334	1.235912	1.352213	1.453238					
58342	GE_BrightCorner	14.89570989	14.9544665	14.70748	14.37462	15.1048	15.10763					
58343												
58344												
58345												
58346												
58347												
58348												
58349												
58350												
58351												

GSE137578\_series\_matrix

就绪

对于上述数据，我们所需要 **数据结构** 模样如上。

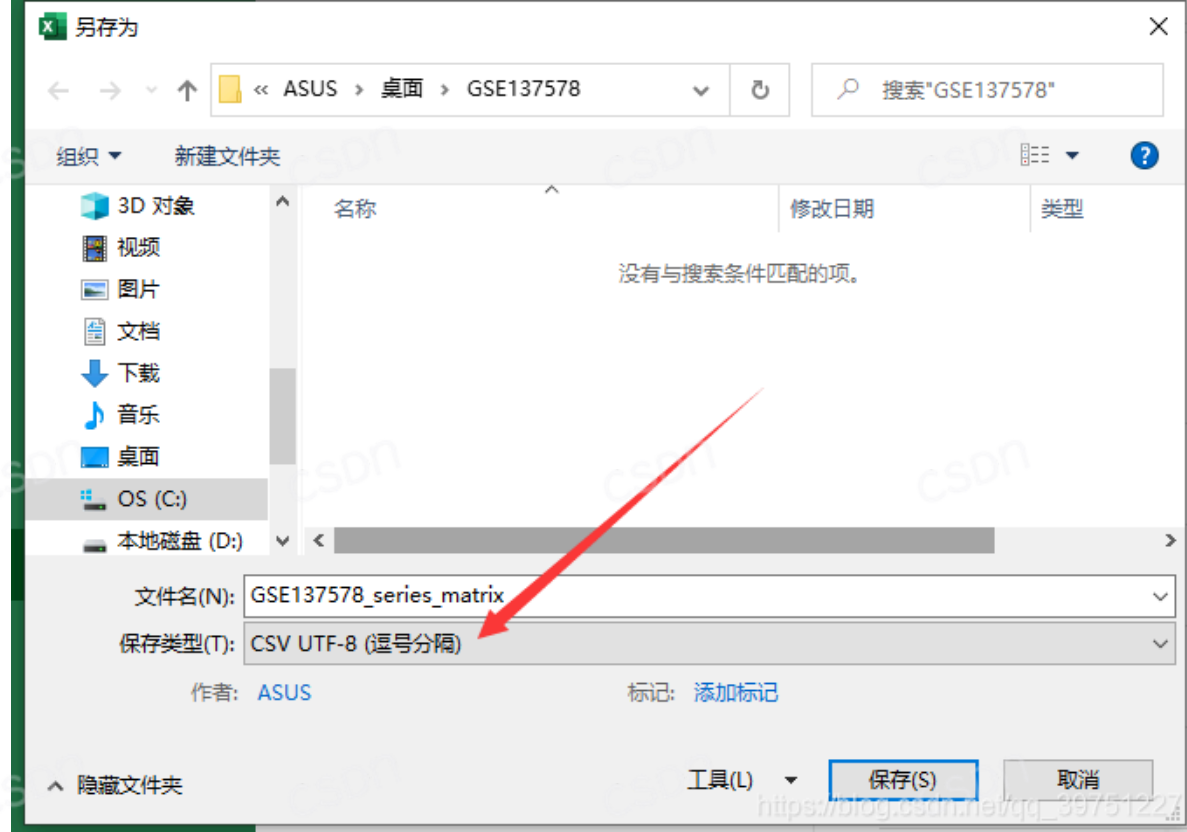
注意，我们上述下载的是基因实验数据，它用于绘制 **热图**，这在第三章会详细说明，这里点击 **另存CSV** 即可。

内容来源: csdn.net

作者昵称: Frms

原文链接: [https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)



## 利用GEO分析绘制拟火山图

### 1. 点击分析

内容来源: csdn.net

作者昵称: Frms

原文链接: [https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

样品 (6) GSM4081917 HAEC\_Untreatment\_Rep1 (mRNA)  
更多的... GSM4081918 HAEC\_Untreatment\_Rep2 (mRNA)  
GSM4081919 HAEC\_Untreatment\_Rep3 (mRNA)

关系

生物工程 PRJNA565963

使用 GEO2R 进行分析

下载家庭

[SOFT 格式的族文件](#)

[MINIML 格式的族文件](#)

[系列矩阵文件](#)

格式

[柔软的](#)

[迷你机器语言](#)

[文本文件](#)

补充文件	尺寸	下载	文件类型/资源
GSE137578_RAW.tar	73.4 兆字节	<a href="#">(http)</a> <a href="#">(自定义)</a>	TAR (TXT)

原始数据作为补充文件提供

样本表中包含的处理数据

[https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

2. 如图步骤,

1. 按住 Shift 选择组别,

2. 标记组别

3. 如此选择完你的组别

内容来源: csdn.net

作者昵称: Frms

原文链接: [https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

加入GEO

GSE137578

放

新型mirnas通过炎症和细胞凋亡途径 (HAEC mRNA) 参与动脉粥样硬化

▼ 样品

▼ 定义组

选择6出6个样品

输入组名: 列表

列

放

团体	加入	标	来源名称	细胞	治疗
一种	GSM4081917	HAEC_Untreatment_Rep1 (mRNA)	人主动脉内皮细胞, 未经处理, 复制 1	人主动脉内皮细胞	车辆
一种	GSM4081918	HAEC_Untreatment_Rep2 (mRNA)	人主动脉内皮细胞, 未经处理, 复制 2	人主动脉内皮细胞	车辆
一种	GSM4081919	HAEC_Untreatment_Rep3 (mRNA)	人主动脉内皮细胞, 未经处理, 复制 3	人主动脉内皮细胞	车辆
乙	GSM4081920	HAEC_oxLDL_Rep1 (mRNA)	人主动脉内皮细胞, 用 oxLDL 处理, 复制 1	人主动脉内皮细胞	氧化低密度脂蛋白
乙	GSM4081921	HAEC_oxLDL_Rep2 (mRNA)	人主动脉内皮细胞, 用 oxLDL 处理, 复制 2	人主动脉内皮细胞	氧化低密度脂蛋白
乙	GSM4081922	HAEC_oxLDL_Rep3 (mRNA)	人主动脉内皮细胞, 用 oxLDL 处理, 复制 3	人主动脉内皮细胞	氧化低密度脂蛋白

[https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

3. 进行分析

乙	GSM4081920	HAEC_oxLDL_Rep1 (mRNA)	人主动脉内皮细胞, 用 oxLDL 处理, 复制 1	人主动脉内皮细胞	氧化低密度脂蛋白
乙	GSM4081921	HAEC_oxLDL_Rep2 (mRNA)	人主动脉内皮细胞, 用 oxLDL 处理, 复制 2	人主动脉内皮细胞	氧化低密度脂蛋白
乙	GSM4081922	HAEC_oxLDL_Rep3 (mRNA)	人主动脉内皮细胞, 用 oxLDL 处理, 复制 3	人主动脉内皮细胞	氧化低密度脂蛋白

GEO2R

选项

剖面图

R 脚本

### 快速开始

- 如果出现提示, 请指定 GEO 系列登录名和平台。
- 单击“定义组”并输入您计划比较的样品组的名称, 例如测试和控制。
- 将样本分配给每个组。突出显示样本行, 然后单击组名称以将这些样本分配给该组。使用样本元数据 (标题、来源和特征) 列来帮助确定哪些样本属于哪个组。
- 单击“分析”以使用默认设置执行计算。
- 您可以在选项选项卡中更改设置。

如何使用

分析

[https://blog.csdn.net/qq\\_39751227/](https://blog.csdn.net/qq_39751227/)

4. 注意下面我们要绘制的火山图模样, 我们点击下载表格即可

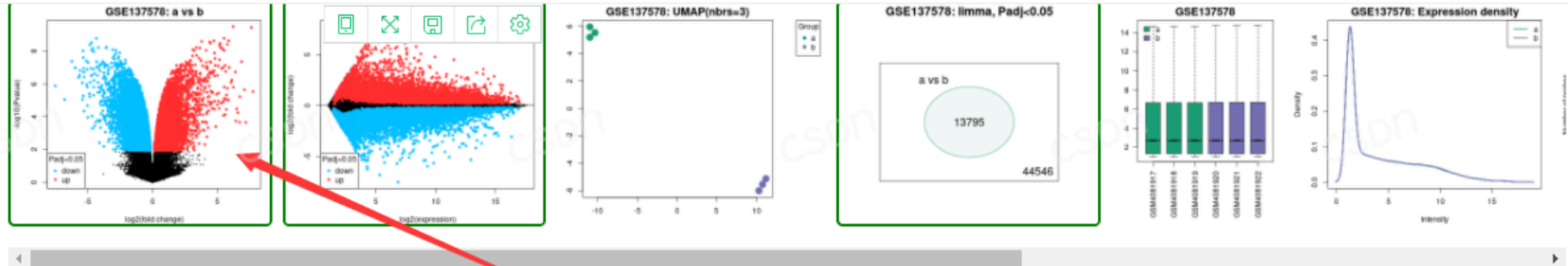
内容来源: csdn.net

作者昵称: Frms

原文链接: [https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227/](https://blog.csdn.net/qq_39751227/)





最高差异表达基因?

下载完整表格 选择列

ID	adj.P.Val的	P值	吨	乙	日志FC	GB_ACC	顺序	SPOT_ID
A_23_P10385	0.0000107	3.20e-10	145.4	12.24	6.22	NM_016448	GGAAGATGTTA	A_23_P10385
A_23_P256956	0.0000107	3.67e-10	141.4	12.21	7.63	NM_005733	TCAAGCCTTG	A_23_P256956
A_24_P322354	0.000018	1.70e-09	103.9	11.69	5.51	NM_001039535	ACCTGAAATG	A_24_P322354
A_23_P302672	0.000018	1.70e-09	-103.9	11.68	-4.34	BC013592	ATCAATGTTGT	A_23_P302672
A_33_P3374205	0.000018	2.37e-09	97.2	11.54	5.87	NM_002417	ATGTGCTCTG	A_33_P3374205
A_24_P346855	0.000018	2.39e-09	97	11.53	4.73	NM_002417	TCTCACCCTG	A_24_P346855
A_33_P3376116	0.000018	2.57e-09	95.6	11.5	3.78	NM_182513	TCCAGGAAATT	A_33_P3376116
A_23_P251421	0.000018	2.79e-09	94	11.46	4.42	NM_031942	ATTACTTGCA	A_23_P251421
A_22_P00006	0.000018	3.01e-09	-92.6	11.43	-4.73	XR_110175	GATTACAAGTT	A_22_P00006231
A_33_P3392187	0.000018	3.25e-09	91.2	11.39	4.47	NM_001080433	TCCAAATGCAA	A_33_P3392187

4. 文件格式是 tsv，我们利用excel打开，注意，在绘制热图时，我们需要它，所以，请进行备份！

内容来源: csdn.net

作者昵称: Frms

原文链接: [https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)



GSE137578 top.table - 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

A1 : x ✓ fx ID

	A	B	C	D	E	F	G	H	I	J
1	ID	adj.P.Val	P.Value	t	B	logFC	GB_ACC	SEQUENCE	SPOT_ID	
2	A_23_P10385	0.0000107	3.20E-10	1.45E+02	12.24225	6.22360265	NM_016448	GGAAGATGTTATTTATGACCA	A_23_P10385	
3	A_23_P256956	0.0000107	3.67E-10	1.41E+02	12.20565	7.62826318	NM_005733	TCAAGCCTTGACCACTTGTGAA	A_23_P256956	
4	A_24_P322354	0.000018	1.70E-09	1.04E+02	11.68504	5.51239755	NM_001039535	ACCTGAAATGCATTTAGTGTA	A_24_P322354	
5	A_23_P302672	0.000018	1.70E-09	-1.04E+02	11.68338	-4.34401927	BC013592	ATCAATGTTGTTTTGCATGTC	A_23_P302672	
6	A_33_P3374205	0.000018	2.37E-09	9.72E+01	11.53921	5.87395406	NM_002417	ATGTGCTCTGGGTTACCTGGT	A_33_P3374205	
7	A_24_P346855	0.000018	2.39E-09	9.70E+01	11.53435	4.73394617	NM_002417	TCTACCCTGGTAAGTACAGT	A_24_P346855	
8	A_33_P3376116	0.000018	2.57E-09	9.56E+01	11.50109	3.78222411	NM_182513	TCCAGGAAATTCATCAGCGA	A_33_P3376116	
9	A_23_P251421	0.000018	2.79E-09	9.40E+01	11.46287	4.41563086	NM_031942	ATTTACTTGCATATGTAAACC	A_23_P251421	
10	A_22_P00006231	0.000018	3.01E-09	-9.26E+01	11.42723	-4.7328389	XR_110175	GATTACAAGTTCTCTATAGGA	A_22_P00006231	
11	A_33_P3392187	0.000018	3.25E-09	9.12E+01	11.38962	4.46615468	NM_001080433	TCCAAATGCAATGATAGTTTC	A_33_P3392187	
12	A_23_P63789	0.000018	3.40E-09	9.04E+01	11.36709	5.19707619	NM_032997	TCAAAGATTCAGAGATTGGC	A_23_P63789	
13	A_24_P413884	0.0000186	3.83E-09	8.82E+01	11.30648	5.77657821	NM_001809	TAGTTTGTGAGTTACTCATGT	A_24_P413884	
14	A_21_P0003965	0.0000186	4.17E-09	-8.67E+01	11.2627	-4.84362635		CAATTTTAATGGAGCATGAA	A_21_P0003965	
15	A_23_P35219	0.0000186	4.48E-09	8.55E+01	11.22582	5.17592304	NM_002497	AGGAGTGAGAATTCTGAGAG	A_23_P35219	
16	A_33_P3316273	0.0000186	5.94E-09	-8.08E+01	11.0713	-3.55548526	NM_002983	TGCTTTTGTTCAGGGCTGTGA	A_33_P3316273	
17	A_22_P00018082	0.0000186	6.06E-09	-8.05E+01	11.05977	-4.19293576	BU154984	ATTCAGCTCTGATTGGAGAGA	A_22_P00018082	
18	A_22_P00001017	0.0000186	6.07E-09	8.04E+01	11.05889	5.25973938	NR_033957	CCCATATTCTTCATCTTGTTCC	A_22_P00001017	
19	A_24_P287941	0.0000186	6.14E-09	8.03E+01	11.05276	3.37759054	NM_013290	AAATTGCAGTAGCTTGAGGTT	A_24_P287941	
20	A_33_P3403576	0.0000186	6.19E-09	-8.01E+01	11.04834	-3.6073776	NM_201563	AAGAGCCCAATTACCAGAAC	A_33_P3403576	
21	A_23_P104651	0.0000186	6.56E-09	7.92E+01	11.01469	3.64260513	NM_080668	TCACCAGATGATGCAGAGTT	A_23_P104651	
22	A_23_P50508	0.0000186	6.85E-09	-7.85E+01	10.98981	-4.61913654	NM_003706	TCCAGATGGCCAGAATGAAT	A_23_P50508	
23	A_23_P57588	0.0000186	7.16E-09	7.78E+01	10.9641	5.40182345	NM_016426	CTCTGATCGACCTCATGACAA	A_23_P57588	
24	A_21_P0006266	0.0000186	8.17E-09	7.58E+01	10.88517	3.11078527		TCTACAAGCTCCCAAAGGAA	A_21_P0006266	
25	A_33_P3246833	0.0000186	8.22E-09	-7.57E+01	10.88196	-3.07124383	NM_173843	TGCAAGTTCCCTACTTCCTG	A_33_P3246833	
26	A_23_P92499	0.0000186	8.71E-09	-7.48E+01	10.84662	-4.70937093	NM_003264	AAGTCCTAGGTTCCCATATTT	A_23_P92499	
27	A_23_P58321	0.0000186	8.91E-09	7.45E+01	10.83295	4.13811878	NM_001237	AAGTTTGATAGATGCTGACCC	A_23_P58321	
28	A_23_P7827	0.0000186	9.06E-09	-7.42E+01	10.82247	-2.83272749	NM_001010919	CAAATTTTCATCACTGTATACT	A_23_P7827	
29	A_33_P3257678	0.0000186	9.14E-09	7.41E+01	10.81732	5.5173993	NM_001005464	AAGAAGTGGCGGTTGCGCCG	A_33_P3257678	

GSE137578 top.table

就绪 计数: 9

5. 在这里，我们只保留以下列，注意记住以下列的名字，这非常重要，它将决定代码是否成功运行！

保留的列头行名：

1. GB\_ACC（这是基因名，每个文件有所不同，按己索需）
2. P.Value（这里改名为FDR）
3. logFC
4. adj.P.Val

你只要记住你要保留的每行头个单元名称即可，至于是什么，自己可决定。

内容来源: csdn.net

作者昵称: Frms

原文链接: [https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)

我们如下整理，**请将非数据列放到开头。**

	A	B	C	D	E	F
1	GB_ACC	adj.P.Val	FDR	logFC		
2	NM_016448	0.0000107	3.20E-10	6.22360265		
3	NM_005733	0.0000107	3.67E-10	7.62826318		
4	NM_001039535	0.000018	1.70E-09	5.51239755		
5	BC013592	0.000018	1.70E-09	-4.34401927		
6	NM_002417	0.000018	2.37E-09	5.87395406		
7	NM_002417	0.000018	2.39E-09	4.73394617		
8	NM_182513	0.000018	2.57E-09	3.78222411		
9	NM_031942	0.000018	2.79E-09	4.41563086		
10	XR_110175	0.000018	3.01E-09	-4.7328389		
11	NM_001080433	0.000018	3.25E-09	4.46615468		
12	NM_032997	0.000018	3.40E-09	5.19707619		
13	NM_001809	0.0000186	3.83E-09	5.77657821		
14		0.0000186	4.17E-09	-4.84362635		
15	NM_002497	0.0000186	4.48E-09	5.17592304		
16	NM_002983	0.0000186	5.94E-09	-3.55548526		
17	BU154984	0.0000186	6.06E-09	-4.19293576		
18	NR_033957	0.0000186	6.07E-09	5.25973938		
19	NM_013290	0.0000186	6.14E-09	3.37759054		
20	NM_201563	0.0000186	6.19E-09	-3.6073776		
21	NM_080668	0.0000186	6.56E-09	3.64260513		
22	NM_003706	0.0000186	6.85E-09	-4.61913654		
23	NM_016426	0.0000186	7.16E-09	5.40182345		
24		0.0000186	8.17E-09	3.11078527		
25	NM_173843	0.0000186	8.22E-09	-3.07124383		

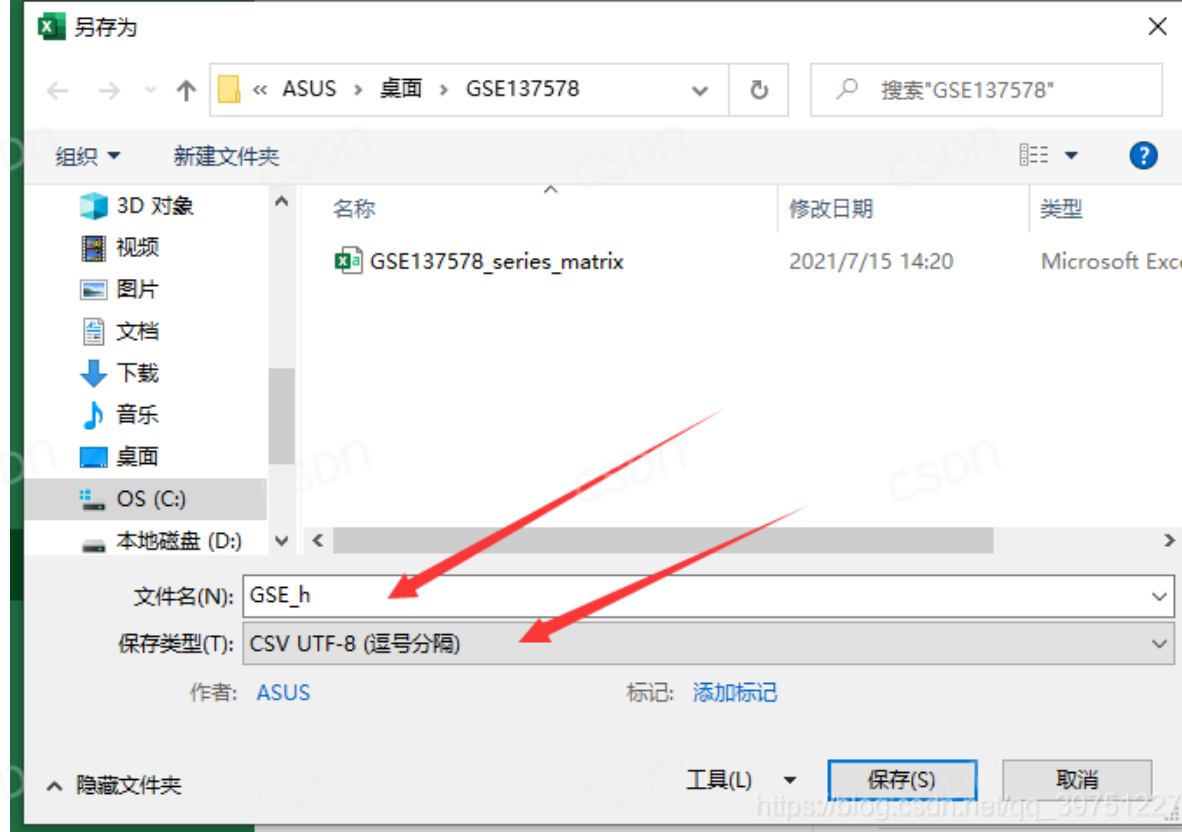
注意名称和另存格式

内容来源: csdn.net

作者昵称: Frms

原文链接: [https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)



之后，我们将利用此绘制火山图。

内容来源: csdn.net

作者昵称: Frms

原文链接: [https://blog.csdn.net/qq\\_39751227/article/details/118757653](https://blog.csdn.net/qq_39751227/article/details/118757653)

作者主页: [https://blog.csdn.net/qq\\_39751227](https://blog.csdn.net/qq_39751227)