

文本操作

- 点击“工具”菜单
- 点击“文本操作”



下载练习数据文件：[list1.xls](#), [list2.xls](#)

1、文本修剪

- 点击“文本修剪，导入制表符分隔的数据文件”
- 点击“文件” ——> “打开文本文件”
- 给出“list1.xls”
- 选择第 11 行，右击，选择“定义为变量名”

文件	分隔符	定义	帮助	保存
0:	[COMPARE_CRITERIA_V2]			
1:	\$NUM_OPTION_LINE=5			
2:	\$ARRAY_LIST_FILE=			
3:	\$COMPARE_ON_GENE_LIST=			
4:	\$COMPARE_ON_USE_LIST=1			
5:	\$AVERAGE_USING_STANDARD_ERROR=Yes			
6:	\$SOMIT_AFFY_CONTROL_GENE=Yes			
7:	\$NUM_CRITERION=1			
8:	\$Parenthesis	Combine	Baseline	Experiment
9:	\$No and	"16,17,18,19,20"	"21,22,23,24,25"	
10:	[COMPARE_RESULT]			
11:	probe set gene	Accession	EntrezGene	Description
12:	1417151_a_at	neurotensin receptor 2	NM_008747	18217
13:	1422790_at	natriuretic peptide precursor type C	NM_010933	18159
14:	1423396_at	"angiotensinogen (serpin peptidase inhibitor, clade A, member 8)"	AK018763	11606
15:	1450960_at	insulin-like growth factor binding protein-like 1	BM935068	75426
16:	1450611_at	orosomucoid 3	NM_013623	18407
17:	1423377_at	insulin-like growth factor binding protein-like 1	BM935068	75426
18:	1426225_at	"retinol binding protein 4, plasma"	U63146	19662
19:	1425116_a_at	spectrin beta 4	AY032655	80297
20:	1418743_a_at	tescalcin	NM_021344	57816
21:	1418739_at	serum/glucocorticoid regulated kinase 2	NM_013731	27219
22:	1418395_at	RIKEN cDNA 1300013J15 gene	NM_026183	67473
23:	1434596_at	"PREDICTED: Mus musculus hypothetical protein LOC625568 (LOC625568), mRNA"	BB194318	
24:	1420356_at	ninjurin 2	NM_016718	29862
25:	1422696_at	tweety homolog 1 (Drosophila)	NM_021324	57776
26:	1422586_at	endothelin converting enzyme-like 1	NM_021306	13599
27:	1424553_at	"Gup1, glycerol uptake/transporter homolog (yeast)"	BC024464	74770

- 定义为变量名
- 定义为第一行数据
- 删除被选的数据行
- 清除所有的数据行
- 定义数据字段分隔符：
 - 制表符分隔
 - 逗号分隔
 - 空格分隔
 - 其它分隔符分隔

• 右击第 12 行，选择“定义为第一行数据”，数据读入到右边框内，如下：

文件	分隔符	定义	帮助	保存
0:	[COMPARE_CRITERIA_V2]			
1:	\$NUM_OPTION_LINE=5			
2:	\$ARRAY_LIST_FILE=			
3:	\$COMPARE_ON_GENE_LIST=			
4:	\$COMPARE_ON_USE_LIST=1			
5:	\$AVERAGE_USING_STANDARD_ERROR=Yes			
6:	\$SOMIT_AFFY_CONTROL_GENE=Yes			
7:	\$NUM_CRITERION=1			
8:	\$Parenthesis	Combine	Baseline	Experiment
9:	\$No and	"16,17,18,19,20"	"21,22,23,24,25"	
10:	[COMPARE_RESULT]			
11:	probe set gene	Accession	EntrezGene	Description
12:	1417151_a_at	neurotensin receptor 2	NM_008747	18217
13:	1422790_at	natriuretic peptide precursor type C	NM_010933	18159
14:	1423396_at	"angiotensinogen (serpin peptidase inhibitor, clade A, member 8)"	AK018763	11606
15:	1450960_at	insulin-like growth factor binding protein-like 1	BM935068	75426
16:	1450611_at	orosomucoid 3	NM_013623	18407
17:	1423377_at	insulin-like growth factor binding protein-like 1	BM935068	75426
18:	1426225_at	"retinol binding protein 4, plasma"	U63146	19662
19:	1425116_a_at	spectrin beta 4	AY032655	80297
20:	1418743_a_at	tescalcin	NM_021344	57816
21:	1418739_at	serum/glucocorticoid regulated kinase 2	NM_013731	27219
22:	1418395_at	RIKEN cDNA 1300013J15 gene	NM_026183	67473
23:	1434596_at	"PREDICTED: Mus musculus hypothetical protein LOC625568 (LOC625568), mRNA"	BB194318	
24:	1420356_at	ninjurin 2	NM_016718	29862
25:	1422696_at	tweety homolog 1 (Drosophila)	NM_021324	57776
26:	1422586_at	endothelin converting enzyme-like 1	NM_021306	13599
27:	1424553_at	"Gup1, glycerol uptake/transporter homolog (yeast)"	BC024464	74770

• 点击“保存”，给出新文件名，如“list_new.xls”，生成新数据文件 list_new.xls。

2、合并两文本文件

根据字段 “probe.set” 匹配合并 list_new.xls 与 list2.xls

- 点击 “工具” ——> “文本操作” ——> “文件修剪、匹配合并”
- 点击 “文件一”，选择 “list1_new.xls”
- 点击 “文件二”，选择 “list2.xls”
- 点击文件一第一列 “probe.set”，选择 “将此列作为编号”，同样操作，选择 “probe.set” 为文件二的编号。

probe.set	gene	Accession	EntrezG	Description	fold change	lower bound	upper bound	difference	t statistic	P value	filtered	V13
1417151	neurot...	NM_00...	18217	"gb-NM...	-22.3	-18.01	-28.48	-1569.99	-15.032	0.000103	*	
1422790	natriuret...	NM_01...	18159	"gb-NM...	-13.11	-6.56	-34.34	-800.17	-3.776	0.018615	*	
1423396	"angiot...	AK0187...	11606	"gb-AK...	-7.82	-5.93	-10.67	-2253.08	-8.843	0.00065	*	
1450960	insulin-4...	BM935...	75426	gb-BM9...	-8.11	-5.91	-12.52	-1133.94	-13.248	0.000027	*	
1450611	orosom...	NM_01...	18407	"gb-NM...	-7.11	-5.34	-10.16	-500.3	-11.007	0.000129	*	
1423377	insulin-4...	BM935...	75426	"gb-AK...	-7.15	-5.13	-11.47	-1866.19	-13.179	0.000011	*	
1426225	"retinol ...	U63146	19662	"gb-U6...	-6.46	-4.78	-9.59	-910.24	-11.508	0.000049	*	
1425116	spectrin...	AY0326...	80297	"gb-AY...	-5.01	-3.77	-7.42	-1824.17	-18.295	0.000026	*	
1418743	tescalcin	NM_02...	57816	"gb-NM...	-4.71	-3.46	-7.2	-391.06	-12.07	0.000006	*	
1418745	osteom...	NM_01...	27047	"gb-NM...	-3.92	-3.42	-4.47	-1758.47	-11.746	0.000183	*	
1418739	serum/g...	NM_01...	27219	"gb-NM...	-4.32	-3.36	-5.68	-431.12	-8.001	0.000592	*	
1434596	"PREDI...	BB1943...	58003	"gb-BB...	-4.05	-3.31	-5.04	-1124.21	-9.487	0.000257	*	

probe.set	gene	Accession	EntrezG	Description	fold change	lower bound	upper bound	difference	t statistic	P value	filtered	V13
1417151	neurot...	NM_00...	18217	"gb-NM...	-22.3	-18.01	-28.48	-1569.99	-15.032	0.000103	*	
1422790	natriuret...	NM_01...	18159	"gb-NM...	-13.11	-6.56	-34.34	-800.17	-3.776	0.018615	*	
1423396	"angiot...	AK0187...	11606	"gb-AK...	-7.82	-5.93	-10.67	-2253.08	-8.843	0.00065	*	
1450960	insulin-4...	BM935...	75426	gb-BM9...	-8.11	-5.91	-12.52	-1133.94	-13.248	0.000027	*	
1450611	orosom...	NM_01...	18407	"gb-NM...	-7.11	-5.34	-10.16	-500.3	-11.007	0.000129	*	
1423377	insulin-4...	BM935...	75426	"gb-AK...	-7.15	-5.13	-11.47	-1866.19	-13.179	0.000011	*	
1426225	"retinol ...	U63146	19662	"gb-U6...	-6.46	-4.78	-9.59	-910.24	-11.508	0.000049	*	
1425116	spectrin...	AY0326...	80297	"gb-AY...	-5.01	-3.77	-7.42	-1824.17	-18.295	0.000026	*	
1418743	tescalcin	NM_02...	57816	"gb-NM...	-4.71	-3.46	-7.2	-391.06	-12.07	0.000006	*	
1418745	osteom...	NM_01...	27047	"gb-NM...	-3.92	-3.42	-4.47	-1758.47	-11.746	0.000183	*	
1418739	serum/g...	NM_01...	27219	"gb-NM...	-4.32	-3.36	-5.68	-431.12	-8.001	0.000592	*	
1434596	"PREDI...	BB1943...	58003	"gb-BB...	-4.05	-3.31	-5.04	-1124.21	-9.487	0.000257	*	

- 点击 “文件一” ——> “匹配文件二” ——> “只保留文件二中存在的行编号”

文件一 文件二 帮助 保存 输出到: C:\Users\Mark\Dropbox\EmpowerStats>manuals\demo\textTools\list1_new_new.xls											
打开	...	Descript...	fold cha...	lower b...	upper b...	differen...	t statistic	P value	filtered		
粘贴		gb:NM_...	-25.33	-19.19	-36.06	-1578.79	-15.082	0.000096	*		
删除		gb:NM_...	-13.27	-6.82	-30.62	-801.01	-3.785	0.018647	*		
添加		gb-AK0...	-8.32	-6.27	-11.5	-2273.01	-8.926	0.00063	*		
匹配文件二		gb-BM9...	-8.03	-5.98	-11.83	-1132.25	-13.39	0.000032	*		
输出到: list1_new_new.xls											
1426225...	retinol bi...	U63146	19662								
1425116...	spectrin...	AY0326...	80297								
1418743...	tescalcin	NM_02...	57816	gb:NM_...	-4.82	-3.68	-6.86	-393.51	-13.097	0.000002	*
1418739...	serum/g...	NM_01...	27219	gb:NM_...	-4.47	-3.59	-5.57	-435.44	-8.269	0.00074	*
1418395...	RIKEN ...	NM_02...	67473	gb:NM_...	-5.25	-3.39	-9.15	-452.83	-5.27	0.003677	*
1434596...	PREDI...	BB1943...		gb-BB1...	-4.08	-3.39	-4.91	-1126.15	-9.672	0.000322	*
1420356...	ninjurin 2	NM_01...	29862	gb:NM_...	-3.98	-3.32	-4.84	-437.32	-10.744	0.000121	*
1422696...	tweety ...	NM_02...	57776	gb:NM_...	-4.72	-3.32	-6.61	-562.52	-5.099	0.005783	*
1422586...	endothe...	NM_02...	13599	gb:NM_...	-4.63	-3.25	-7.67	-516.28	-8.862	0.00004	*
1424553...	Gup1, g...	BC0244...	74770	gb-BC0...	-4.45	-3.22	-7.12	-1807.94	-12.486	0.000004	*
C:\Users\Mark\Dropbox\EmpowerStats>manuals\demo\textTools\list1_new.xls 290 行											
probe set	gene	Accessi...	EntrezG...	Descript...	fold cha...	lower b...	upper b...	differen...	t statistic	P value	filtered
1417151...	neurote...	NM_00...	18217	"gb:NM...	-22.3	-18.01	-28.48	-1569.99	-15.032	0.000103	*
1422790...	natriuret...	NM_01...	18159	"gb:NM...	-13.11	-6.56	-34.34	-800.17	-3.776	0.018615	*
1423396...	"angiot...	AK0187...	11606	"gb-AK...	-7.82	-5.93	-10.67	-2253.08	-8.843	0.00065	*
1450960...	insulin-ii...	BM935...	75426	gb-BM9...	-8.11	-5.91	-12.52	-1133.94	-13.248	0.000027	*
1450611...	orosom...	NM_01...	18407	"gb:NM...	-7.11	-5.34	-10.16	-500.3	-11.007	0.000129	*
1423377...	insulin-ii...	BM935...	75426	"gb-AK...	-7.15	-5.13	-11.47	-1866.19	-13.179	0.000011	*
1426225...	"retinol ...	U63146	19662	"gb:U6...	-6.46	-4.78	-9.59	-910.24	-11.508	0.000049	*
1425116...	spectrin...	AY0326...	80297	"gb:AY...	-5.01	-3.77	-7.42	-1824.17	-18.295	0.000026	*

- 点击“输出到 C:\xxxx\list1_new_new.xls”，给出输出文件名，如“list1_new_matched.xls”

新数据文件“list1_new_matched.xls”生成，该文件是从原“list1_new.xls”中挑出来部分记录，其“probe.set”包含在list2.xls中。原list1_new.xls中有290条计量，新list1_new_matched.xls中只有259条记录。

同样方法，可以从“list1_new.xls”剔除“probe.set”存在于list2.xls中的记录，或匹配“probe.set”合并两文件的列。