

## 易侏统计：从文本文件读入数据

在创建新项目窗口，如给出的文本文件有中文或不规则的符号与格式，易侏自动打开此模块。如点击“>> 定义文本文件格式”或“通过复制粘贴创建数据文件”，均打开此模块。

如需粘贴文本，右击左窗口，然后选“粘贴文本”到左窗口。

读入文本文件操作步骤：

### (1) 定义变量名：

右击左窗口文本文件变量名行，在弹出菜单里选“定义为变量名”，如图示：

The screenshot shows a window titled "读入文本文件并生成制表符分隔的数据文件". The main area displays a text file with various parameters and a table of data. A context menu is open over the first row of the table, with the option "定义为变量名" (Define as variable name) selected.

probe_set	gene	Accession	EntrezGene	Description	fold change	lower bound of FC	upper bound of FC	difference of means t	t	P value	filtered	V13
1417151_a_at	neurotensin receptor 2	NM_008747	18217	"gb:NM_008747.1 /DB_XREF=gi:6116601.1	18217	-19.19	-36.06	-1578.79	-15.082	0.000096	*	
1422790_at	natriuretic peptide precursor type C	NM_010933	18159	"gb:NM_010933.1 /DB_XREF=gi:116601.1	18159	-6.82	-30.62	-801.01	-3.785	0.018647	*	
1450960_at	"angiotensinogen (serpin peptidase inhibitor, clade A, member 8)"	AK018763	11660	"gb:AK018763.1 /DB_XREF=gi:116601.1	11660	-8.32	-6.27	-2273.01	-8.926	0.00063	*	
1450960_at	insulin-like growth factor binding protein-like 1	BM935068	75426	"gb:BM935068.1 /DB_XREF=gi:7305358.1	75426	-8.03	-5.98	-1132.25	-13.39	0.000032	*	
1450611_at	orosomucoid 3	NM_013623	18407	"gb:NM_013623.1 /DB_XREF=gi:7305358.1	18407	-8.45	-5.75	-513.3	-10.967	0.00007	*	
1423377_at	insulin-like growth factor binding protein-like 1	BM935068	75426	"gb:AK012705.1 /DB_XREF=gi:151544.1	75426	-8.12	-5.62	-1902.51	-13.428	0.000008	*	
1426225_at	"retinol binding protein 4, plasma"	U63146	19662	"gb:U63146.1 /DB_XREF=gi:151544.1	19662	-6.62	-5.04	-914.34	-11.827	0.000063	*	
1425116_a_at	spectrin beta 4	AY032655	80297	"gb:AY032655.1 /DB_XREF=gi:16117404.1	80297	-5.51	-4.14	-1865.57	-20.045	0.000001	*	
1418743_a_at	tescalcin	NM_021344	57816	"gb:NM_021344.1 /DB_XREF=gi:10946661.1 /GEN=Testes	57816	-4.82	-3.68	-393.51	-13.097	0.000002	*	
1418739_at	serum/glucocorticoid regulated kinase 2	NM_013731	27219	"gb:NM_013731.1 /DB_XREF=gi:10946631.1 /GEN=Ttyh1 /FEA=FmRNA /CNT=1	27219	-4.47	-3.59	-8269	0.00074	*		
1418395_at	RIKEN cDNA 1300013J15 gene	NM_026183	67473	"gb:NM_026183.1 /DB_XREF=gi:1935355.1 /FEA=FmRNA	67473	-5.25	-3.39	-452.83	-5.27	0.003677	*	
1434596_at	"PREDICTED: Mus musculus hypothetical protein LOC625568 (LOC625568), mRNA"	BB194226	BB194226	"gb:BB194226.1 /DB_XREF=gi:16117404.1	BB194226	-3.28	-4.91	-1136.15	0.000222	*		
1420356_at	ninjurin 2	NM_016718	29862	"gb:NM_016718.1 /DB_XREF=gi:7710069.1 /GEN=Ninj2 /FEA=FmRNA /CNT=8 /TID=Mm.103668.1	29862	-2.28	-1.91	-157.89	-1.91	0.056137	*	
1422696_at	tweety homolog 1 (Drosophila)	NM_021324	57776	"gb:NM_021324.1 /DB_XREF=gi:10946631.1 /GEN=Ttyh1 /FEA=FmRNA /CNT=1	57776	-4.7	-3.59	-8269	0.00074	*		
1422586_at	endothelin converting enzyme-like 1	NM_021306	13599	"gb:NM_021306.1 /DB_XREF=gi:10946605.1 /GEN=Ecel1 /FEA=FmRNA	13599	-5.25	-3.39	-452.83	-5.27	0.003677	*	
1424553_at	"Gup 1, glycerol uptake/transporter homolog (yeast)"	BC024464	74770	"gb:BC024464.1 /DB_XREF=gi:1935355.1 /FEA=FL	74770	-3.28	-4.91	-1136.15	0.000222	*		
1418745_at	osteomodulin	NM_012050	27047	"gb:NM_012050.1 /DB_XREF=gi:6754933.1 /GEN=Omd /FEA=FmRNA /CNT=25 /TID=Mm	27047	-3.28	-4.91	-1136.15	0.000222	*		

右窗口将显示一行变量名：

probe_set	gene	Accessi...	EntrezG...	Descript...	fold_ch...	lower_b...	upper_b...	differen...	t_statistic	P_value	filtered	V13
-----------	------	------------	------------	-------------	------------	------------	------------	-------------	-------------	---------	----------	-----

必要时定义分隔符：易侏软件自动识别制表符、空格、逗号分隔符，如果自动识别失败，可以点击菜单“分隔符”选择相应的分隔符。

### (2) 定义第一行数据：

点击左窗口文本文件第一行数据，在弹出菜单里选“定义为第一行数据”，如上图所示。数据将从此行开始逐行读入到右窗口，如下图所示：

	probe_set	gene	Accessi...	EntrezG...	Descript...	fold_ch...	lower_b...	upper_b...	differen...	t_statistic	P_value	filtered	V13
1	1417151...	neurote...	NM_00...	18217	gb:NM_...	-25.33	-19.19	-36.06	-1578.79	-15.082	0.000096	*	
2	1422790...	natriuret...	NM_01...	18159	gb:NM_...	-13.27	-6.82	-30.62	-801.01	-3.785	0.018647	*	
3	142339...	angioten...	AK0187...	11660	gb:AK0...	-8.32	-6.27	-11.5	-2273.01	-8.926	0.00063	*	
4	145096...	insulin-li...	BM935...	75426	gb:BM9...	-8.03	-5.98	-11.83	-1132.25	-13.39	0.000032	*	
5	145061...	orosom...	NM_01...	18407	gb:NM_...	-8.45	-5.75	-15.07	-513.3	-10.967	0.00007	*	
6	142337...	insulin-li...	BM935...	75426	gb:AK0...	-8.12	-5.62	-14.17	-1902.51	-13.428	0.000008	*	
7	142622...	retinol bi...	U63146	19662	gb:U63...	-6.62	-5.04	-9.32	-914.34	-11.827	0.000063	*	
8	142511...	spectrin...	AY0326...	80297	gb:AY0...	-5.51	-4.14	-8.2	-1865.57	-20.045	0.000001	*	
9	141874...	tescalcin	NM_02...	57816	gb:NM_...	-4.82	-3.68	-6.86	-393.51	-13.097	0.000002	*	
10	141873...	serum/g...	NM_01...	27219	gb:NM_...	-4.47	-3.59	-5.57	-435.44	-8.269	0.00074	*	
11	141839...	RIKEN ...	NM_02...	67473	gb:NM_...	-5.25	-3.39	-9.15	-452.83	-5.27	0.003677	*	
12	142455...	"Gup 1, g...	BC0244...	74770	gb:BC0...	-3.28	-4.91	-1136.15	0.000222	*			

### (3) 定义最后一行数据（可选）：

点击左窗口文本文件最后一行数据，在弹出菜单里选“定义为最后一行数据”，如图：



### (4) 修改变量名，删除不必要的列（变量）：

点击右窗口第一行相应的变量名，可选择“删除本列”或“修改本列标题”，如图：

	probe_set	gene	Accessi...	EntrezG...	Descript...	fold_ch...	lower_b...	upper_b...	differen...	t_statistic	P_value	filtered	V13
216	144990...	selectin...	M72332	20344	gb:M72...	11.51	7.3	19.63	389.81	5.408			
217	142499...	thymidyl...	BC0201...	22171 /...	gb:BC0...	9.82	7.33	12.46	1781.32	6.018			
218	142754...	hyaluro...	X64550	15366	gb:X64...	10.67	7.38	16.72	670.56	7.35	0.001505		
219	142285...	high mo...	X58380	15364	gb:NM...	13.89	7.4	23.52	814.37	3.503	0.024608	*	
220	141828...	RAD51 ...	NM_01...	19361	gb:NM...	9.44	7.47	12.19	842.3	10.65	0.000343	*	
221	144889...	RAD51 ...	BC0037...	19362	gb:BC0...	10.85	7.54	17.43	247.32	8.529	0.000747	*	

### (5) 重新排列变量顺序（可选）：

如要重新排列输出文件的列顺序，点击右窗口第一行相应的变量名，按下鼠标左键不放，然后移动该列到相应的位置即可。

点击“保存到: XXXX.xls”菜单，可修改输出文件名。

点击“运行”易俚将读取右边（定义好的）数据。