

甲基化芯片数据分析

主要内容

- 从GEO下载数据
- 质控
- 归一化
- 差异甲基化位点
- 差异甲基化区域
- 注释

甲基化芯片

Illumina Human Methylation 450K BeadChip芯片可检测人全基因组近450,000个甲基化位点，具有单碱基的分辨率。全面覆盖了96%的CpG岛，并根据需求加入了CpG岛以外的CpG位点、人类干细胞非CpG甲基化位点、正常组织与肿瘤（多种癌症）组织差异甲基化位点、编码区以外的CpG岛、miRNA启动子区域和GWAS疾病相关区域的位点，同时覆盖了Human Methylation27 BeadChip的90%的位点。

芯片的值

- A探针（非甲基化）的数目U
- B探针（甲基化）的数目M
- β 值或者m值

$$\beta = \frac{M}{(M + U + 100)} \quad m = \log \left(\frac{\beta}{1 - \beta} \right)$$

- β 值反映了能够和给定被甲基化的序列匹配的寡核苷酸的比率，序列中的甲基化率
- M值可以消除探针不同而造成的影响

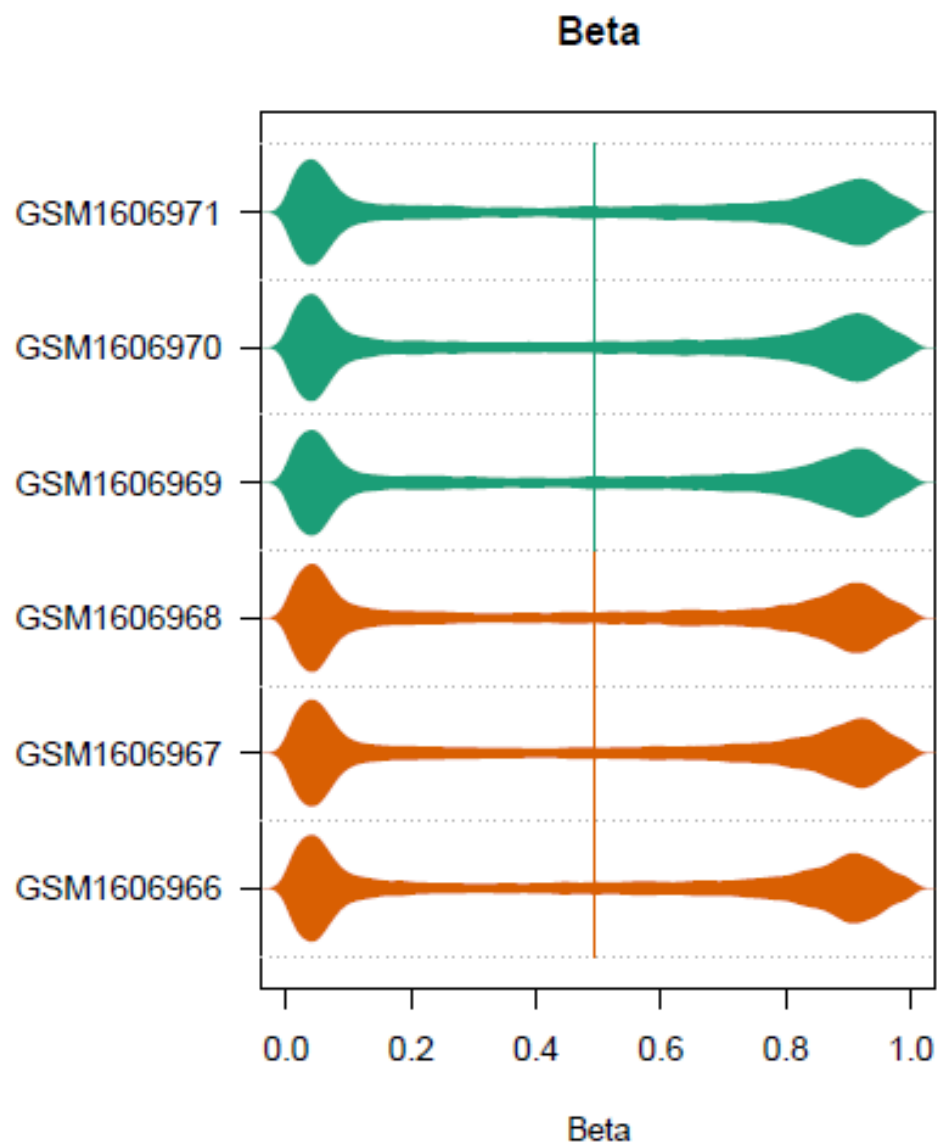
数据下载

1. GEO

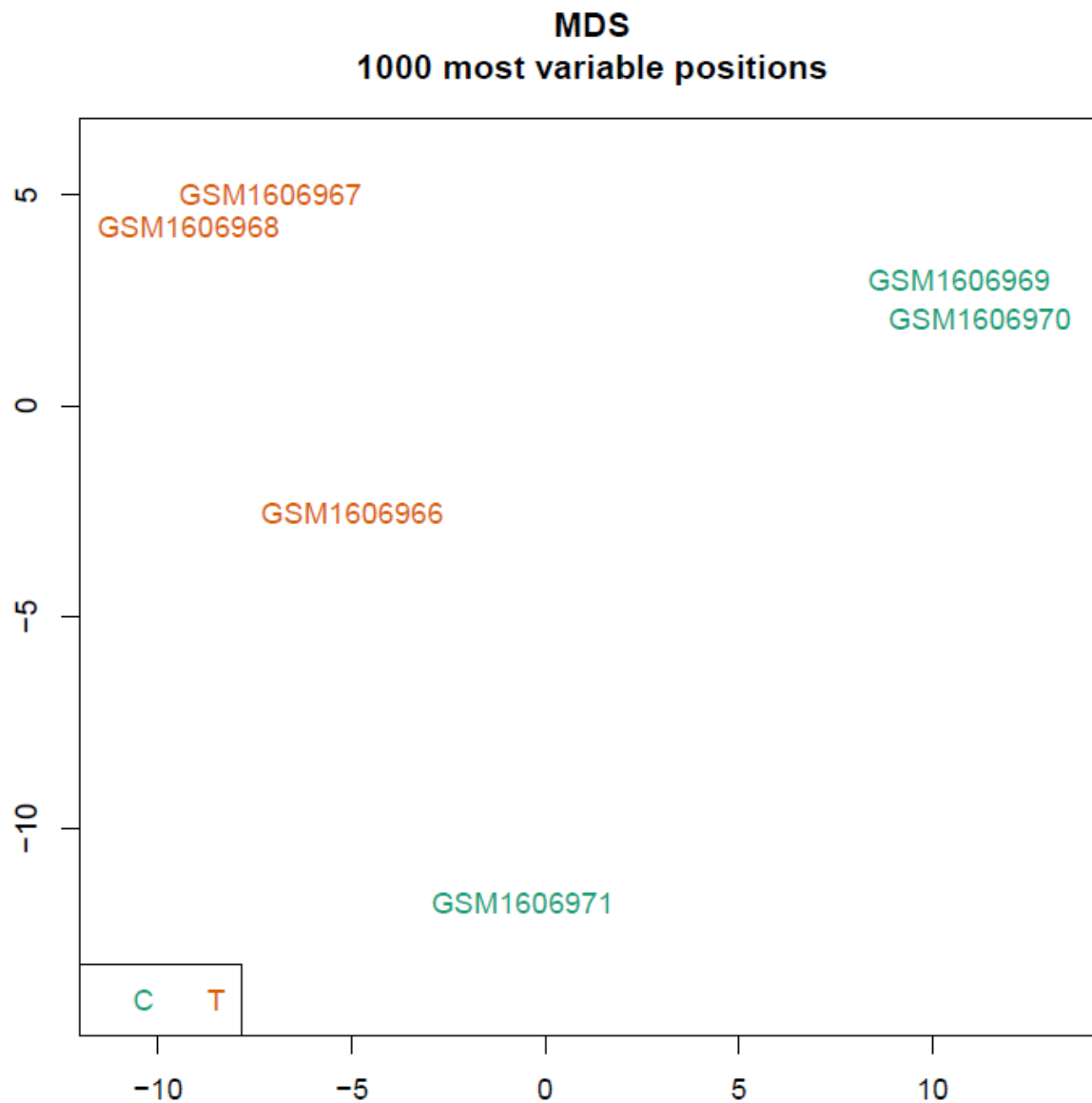
<http://www.ncbi.nlm.nih.gov/geo/>

2. TCGA、EBI等其它网站

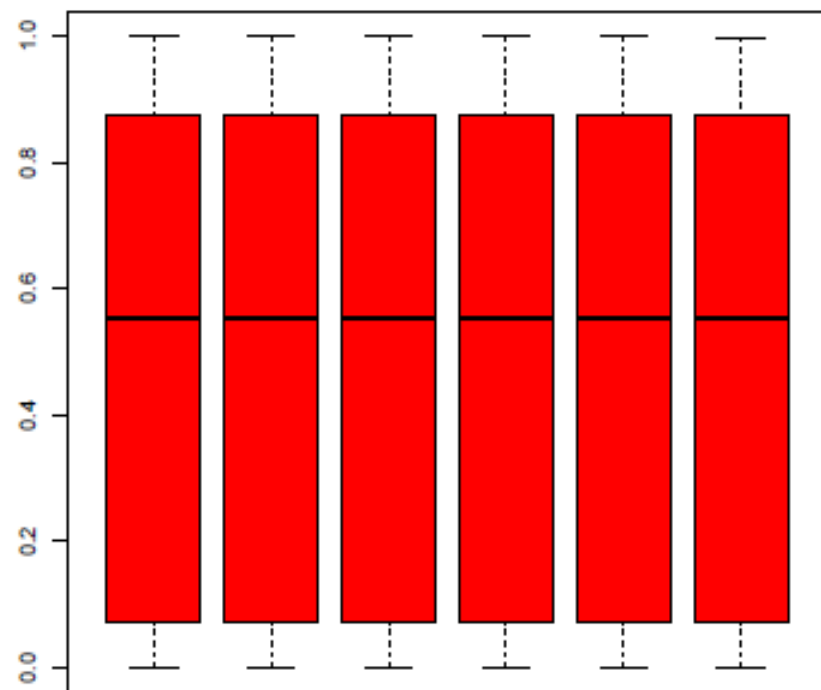
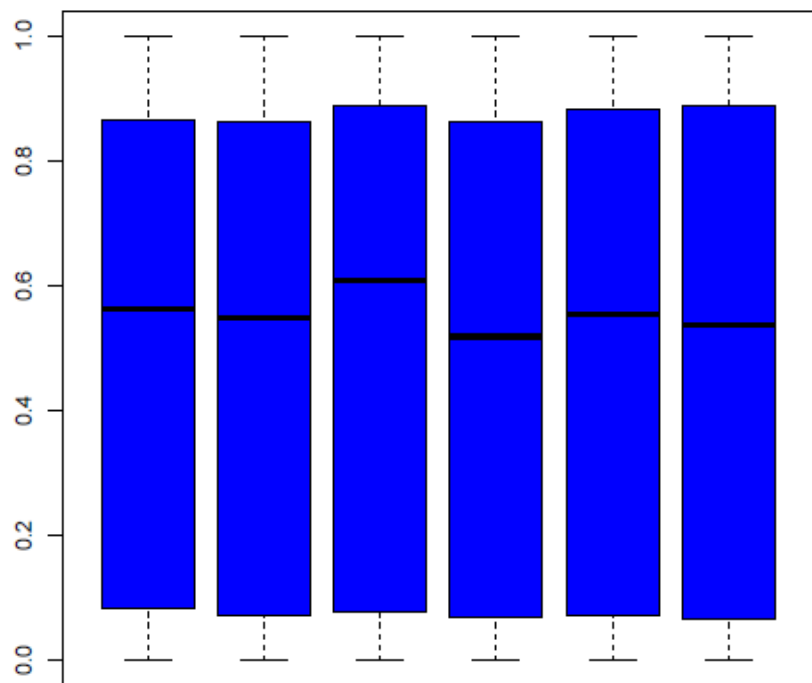
质控



质控



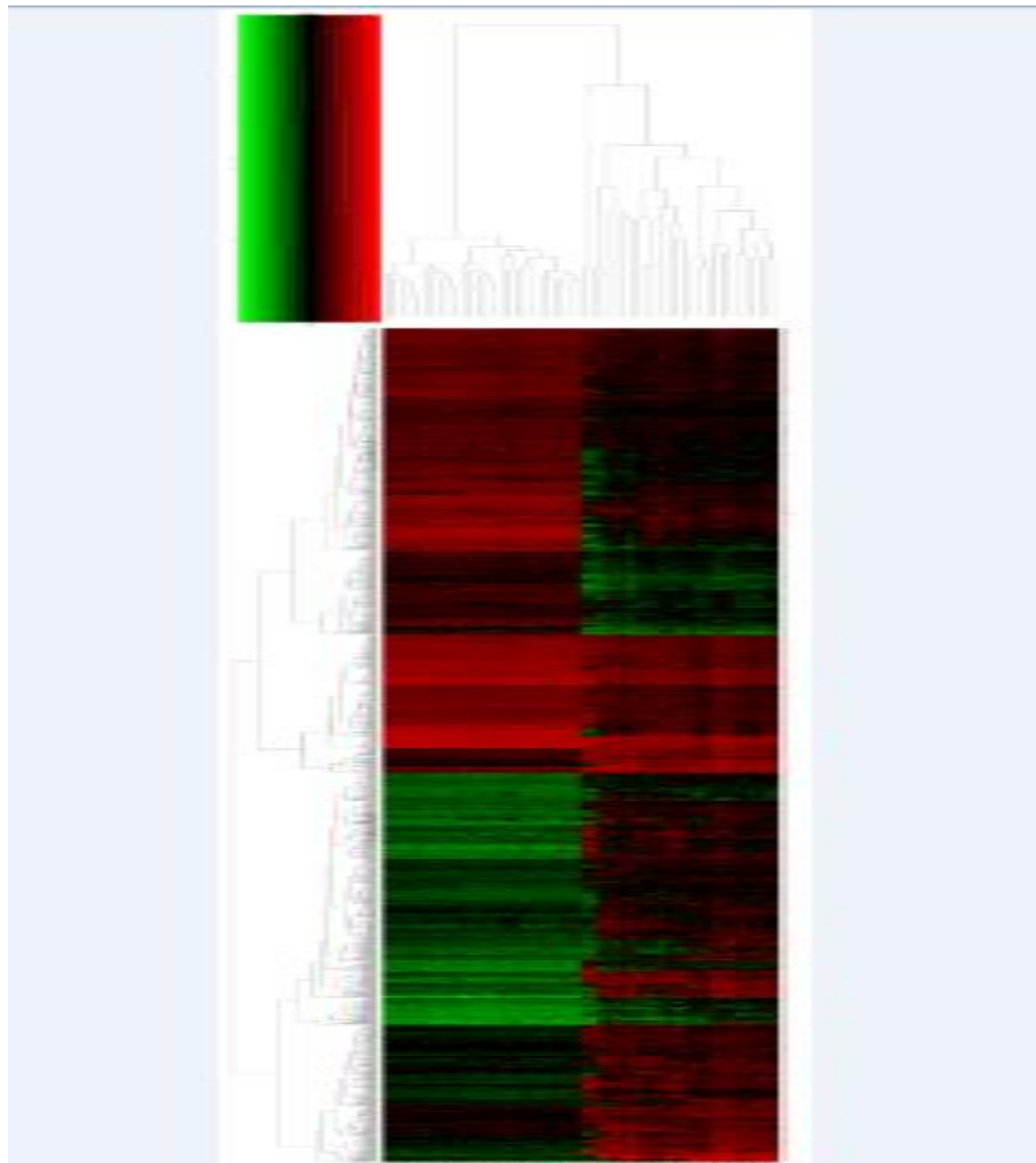
归一化



差异甲基化位点

Id	intercept	f	pval	qval
cg19229544	-5.819817249	9137.68304	7.18E-08	0.013840388
cg13706769	3.192061949	7427.708505	1.09E-07	0.013840388
ch.1.1262625R	-16.60962605	6836.276705	1.28E-07	0.013840388
cg01972335	2.144506398	5554.829221	1.94E-07	0.01571845
cg19506433	-0.65297139	4201.708582	3.39E-07	0.02196957
cg05945291	-16.60962605	2652.551359	8.51E-07	0.033324073
cg21081097	-6.283964656	2621.428714	8.71E-07	0.033324073
cg08094135	2.349302795	2547.055433	9.22E-07	0.033324073
cg25010146	-5.218557921	2449.915749	9.97E-07	0.033324073

热图



差异甲基化区域

chr	start	end	value	area	cluster	indexSt art	indexE nd	L	cluster L	p.value	fwer
chr1	2525719 1	2525823 6	0.43417 0033	8.68340 0661	4866	11732	11751	20	33	0	0
chr11	2397201	2397977	0.42381 6783	8.05251 8884	31573	274030	274048	19	37	0	0
chr6	3193837 2	3193911 2	- 0.43463 8412	7.38885 3009	163981	163875	163891	17	109	0	0
chr6	1498059 95	1498067 32	- 0.55850 1181	6.14351 2994	170492	181364	181374	11	11	0	0
chr2	2092242 25	2092251 83	- 0.61795 2848	3.70771 7086	115900	73375	73380	6	8	0	0

差异甲基化区域

value--average smooth methylation difference within block

area--block area estimate ($\text{abs}(\text{value}) * \text{length}$)

cluster--id of cluster blockgroup within which block occurs

indexStart--index of first cluster in block

indexEnd--index of last cluster in block

L--number of clusters in block

clusterL--number of probes in block

p.value--permutation p.value based on difference conditioned on length

fwer--family-wise error rate estimate based on difference conditioned on length

p.valueArea--permutation p.value based on area

fwerArea--family-wise error rate estimate based on area

差异甲基化区域注释

Chr	Start	End	Ref	Alt	Func	Gene	
chr1	25257191	25258236	0	0	intronic	RUNX3	
chr11	2397201	2397977	0	0	ncRNA_intronic	CD81-AS1	
chr6	31938372	31939112	0	0	exonic	DXO	
chr6	149805995	149806732	0	0	UTR5	ZC3H12D	
chr2	209224225	209225183	0	0	downstream	PIKFYVE	
chrY	6778641	6778641	0	0	upstream	TBL1Y	

注释网站: <http://wannovar.usc.edu/>