

# 表达谱芯片验证探针选择的注意事项

尊敬的客户:

首先十分感谢您选择了上海伯豪作为您的芯片服务商,我们一直致力于向客户提供稳定、可靠的数据.为客户的科研工作保驾护航。

当您拿到芯片数据后,下一步的工作内容将是对芯片的数据结果进行验证。而验证的工作可以分为以下的四个步骤:

- 1. 利用RT-qPCR等方法,在相同的样本内对芯片本身的数据质量进行评估;
- 2. 利用PCA或聚类分析等方法,对实验设计中的样本分组质量进行评估;
- 3. 根据文献报道或已有知识判断芯片数据中找到的差异基因是否符合预期;
- 4. 扩大样本量, 检验在新的个体或处理中, 差异基因是否依然存在。

而在以上的几步验证过程中,第一步的验证要注意的事项却往往被忽略而出现问题。为了让您能够准确有效地选取合适的探针,我们在下文中列举了表达谱芯片的验证中涉及到的探针筛选和引物设计的原则及注意事项供您参考。此外还通过一个真实案例的分析,加深理解,希望让您在后续的实验及研究中少走弯路。

#### 靶基因(探针)的选择及PCR验证引物设计原则:

1. 选出的目的探针的荧光原始信号值应该较高。

首先应该排除Flag为A的探针,而选择P探针(如果组间差异大,可以选择一组为P,另一组为A),其次对于P探针的信号值最好有一个筛选:如当选择的探针信号值在7(取了以2为底的对数)附近时,原始的荧光信号值约为2<sup>7</sup>=128,信号值偏低。在信号较低的情况下,线性关系较差,为了提高PCR结果和芯片结果趋势一致的比例,建议选择的原始荧光信号值>300,

#### 即从数据表中选择探针信号值为8以上的探针数据更容易得到验证。

2. 针对同一个基因且荧光值都较高的几条探针,尽量选择**靶序列在基因3'端的探针**。由于mRNA在反转录的时候用的oligod(T),从基因的3'向5'端的方向进行反转录,由于受酶的活性或者mRNA的高级机构的影响,反转录的时候,mRNA的3'端被优先反转录,有的基因的5'端反转录效率低,在基因5'端的探针一般检测不到好的信号,而且进行定量验证时也难以得到有效的扩增。



注: 芯片上可能设计有多个探针用于检测相同的基因, 所以选定好某个基因后, 将此基因对应的所有探针信号都筛选出来, 会有助于进一步选择合适的探针以及方便后续的验证服务工作。

3. PCR扩增产物的序列最好包含芯片上探针的序列。

如上文所述,最好选择位于基因3'端的探针作为候选靶标,而设计引物也应该位于探针序列处。因为如果设计在基因5'端或基因中部时,不但会由于反转效率低而无法有效扩增外,还可能由于存在着差异剪接,只有部分剪接体能被扩增出来,而得到与芯片不一致的结果,所以引物设计在包含探针序列的3'UTR区最好。

4. 引物设计出来后要对引物的扩增效果进行验证

并不是每个基因都能设计出理想的引物,引物质量也影响着定量检测的效果,如果引物的熔解曲线不好,说明有非特异性扩增,须重新设计引物;引物的扩增Ct值最好在20-25之间,如果引物的Ct值过高,说明引物扩增效率不好或模板降解。

5. 对于芯片结果的验证, 最好用与芯片实验相同的样本

定量验证所用的模板最好要与芯片实验的同样的样本,如果是生物学重复或者相似的样本,由于个体差异或制备动物模型条件不同等原因,验证的结果可能会与芯片结果不一致。比较极端的情况下,即使是同一块组织切两半,一半样本用于做芯片,一半样本用于验证,得到的结果也不一定完全一致,这可能是由于样本的异质性造成的。以肿瘤组织为例:不同位置的肿瘤样本,肿瘤细胞的比例可能不一样,造成结果的不一致。



附: 探针选择的真实案例解析:

案例中所用的芯片为Affymetrix U133 plus 2.0表达谱芯片,客户所提供的待验证的探针ID号为: 211506\_s\_at, 240052\_at两探针,由此可知目的基因为: IL8和ITPR1这两个基因,分别对应的转录本为NM 000584和NM 001099952。

首先对探针的原始信号及位置分析,发现探针所处的位置都位于基因的5°端或中部,杂交信号值偏低,不适合作为验证。具体原因如下:

## 一、 对探针的原始信号值进行分析:

对在归一化后的数据中查找211506\_s\_at, 240052\_at两探针的原始荧光信号值:

表 1 归一化后的数据中探针的信号值:

Probe\_Set\_ID 样本 样本 样本 样本 样本 样本 211506 s at 10.79701 8.934139 8.847228 8.5538 6. 309962 8.604987 240052 at 6. 380231 6. 945526 6. 541215 8.041 7. 46361 7. 234888 根据以上的信号值计算得到的探针原始荧光荧光信号值:

Probe\_Set\_ID A\_NS  $C_NS$ D NS E NS F NS B\_NS 79.34 211506\_s\_at 1779. 19 489. 15 460. 55 375. 79 389.37 240052 at 83. 30 123. 26 93. 13 263.38 176. 51 150.63

计算实际信号值后发现:除了一个探针的数值高于1000外,平均值在200多,探针信号值都偏低;如果探针信号值如此低,对其做位置的分析基本上会发现探针序列位于基因的5°端或中间。

所以根据以上的这两个探针设计引物进行验证,很可能得不到好的结果。所以我们建议根据上述五原则另选探针进行验证。

地址:上海市张江高科技园区李冰路 151 号(201203) 技术服务热线: 800-820-5086 / 400-880-5086 电话: 021-51320288 传真: 021-51320266 3/12



## 二、 进一步对探针的位置进行分析:

在 NCBI 中将探针对应的序列与目标基因 mRNA 序列的比对:



probe ID	gene ID	mRNA	probe location	probe location	consesus
		length	(satrt)	(end)	ratio
211506_S_AT	NM_000584	1718bp	210	425	100%
240052_AT	NM_001099952	10098bp	6384	6497	22. 14%

如表2所示,两探针分别在基因的5'端及中部,此外有一个探针的匹配还有问题,都不适合于进行验证。具体信息如下:

1. 211506\_S\_AT探针, 检测转录本为NM\_000584(全长1718bp)。

电话: 021-51320288 传真: 021-51320266 4/12



### >HG-U133 PLUS 2:211506 S AT (探针序列所在区域)

211506\_S\_AT与NM\_000584(全长1718bp)比对结果如下在基因的 210~425bp处,靠近基因的5'区,不适合于检测:

Query	1	GTGTGAAGGTGCAGTTTTGCCAAGGAGTGCTAAAGAACTTAGATGTCAGTGCATAAAGAC 60
Sbjct	210	GTGTGAAGGTGCAGTTTTGCCAAGGAGTGCTAAAGAACTTAGATGTCAGTGCATAAAGAC 269
Query	61	ATACTCCAAACCTTTCCACCCCAAATTTATCAAAGAACTGAGAGTGATTGAGAGTGGACC 120
Sbjct	270	ATACTCCAAACCTTTCCACCCCAAATTTATCAAAGAACTGAGAGTGATTGAGAGTGGACC 329
Query	121	ACACTGCGCCAACACAGAAATTATTGTAAAGCTTTCTGATGGAAGAGAGCTCTGTCTG
Sbjct	330	ACACTGCGCCAACACAGAAATTATTGTAAAGCTTTCTGATGGAAGAGAGCTCTGTCTG
Query	181	CCCCAAGGAAAACTGGGTGCAGAGGGTTGTGGAGAA 216

2. 240052\_AT探针, 检测转录本为NM\_001099952 (Length=10098)。

>HG-U133\_PLUS\_2: 240052\_AT (探针序列所在区域)

Sbjct 390 CCCCAAGGAAAACTGGGTGCAGAGGGTTGTGGAGAA 425

地址:上海市张江高科技园区李冰路 151 号(201203) 电话: 021-51320288 传真: 021-51320266 技术服务热线: 800-820-5086 / 400-880-5086 技术服务网站: http://www.shbio.cn

5 / 12

acccaggttgctagtgctgactacgatgcagaactccaacctttctcttccacgtaaggt atctcatgaggggataggagcttctgaactgttaggtcacacagaatgaagattccttgc tccaagctaagaagtgaatcccaacagcatcgagtactaaataccttggttctgaaagtcacaaagaggagaaacagtggaaacgccattccctct

240052\_AT与NM\_001099952 (Length=10098) 比对结果如下,匹配差:

Query 62 CAGAACTGCATAGCCACCCATGAATCCAATGGCATTGACATCACAGCCCTGATCCTC 121

Sbjct 6384 CAGAACTGCATAGCCACCCATGAATCCAATGGCATTGACATCACAGCCCTGATCCTC 6443

Query 122 AATGATATCAATCCTTTGGGAAAGAAGAGGATGGACCTTGTGTTAGAACTGAAG 175

Sbjct 6444 AATGATATCAATCCTTTGGGAAAGAAGAGGATGGACCTTGTGTTAGAACTGAAG 6497

探针区域 515bp, 其中只有一小半与转录本序列匹配, 且在转录本的中部, 不适于验证。

三、 进一步对目的基因的的其他探针进行分析:

有时选到的目的基因,的一个探针不适合而放弃此基因的验证很可惜,此时就可以分析对于目的基因是否还有其他可以选择的探针。比如对于目的基因 NM\_000584 及 NM\_001099952 分析后,可发现还有其他的探针检测同一基因:

Probe_Set_ID	样本	样本	样本	样本	样本	样本	Gene Symbol	Entrez Gene	RefSeq Transcript ID
211506_s_at	10.80	8. 93	8. 85	8. 55	6. 31	8. 60	IL8	3576	NM_000584
202859_x_at	14. 32	13. 10	13. 12	13. 27	11. 46	13. 39	IL8	3576	NM_000584
Probe_Set_ID	样本	样本	样本	样本	样本	样本	Gene Symbol	Entrez Gene	RefSeq Transcript ID
240052_at	6. 38	6. 95	6. 54	8. 04	7. 46	7. 23	I TPR1	3708	NM_001099952 /// NM_002222
211323_s_at	7. 00	7. 55	7. 18	7. 10	7. 21	6. 63	I TPR1	3708	NM_001099952 /// NM_002222
203710_at	10. 75	10. 91	10. 94	10.83	11. 11	10. 60	I TPR1	3708	NM_001099952 /// NM_002222
216944_s_at	9. 84	9. 67	10. 14	9. 81	10. 04	9. 68	I TPR1	3708	NM_001099952 /// NM_002222

地址:上海市张江高科技园区李冰路 151 号 (201203) 技术服务热线: 800-820-5086 / 400-880-5086

技术服务网站: http://www.shbio.cn

传真: 021-51320266

6/12

电话: 021-51320288



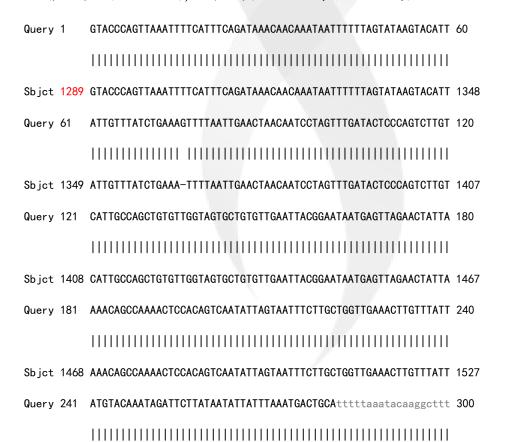
对于目的基因的其他的探针的信号值及探针位置进行上述分析,判断这些探针是否可以用于验证设计引物:

1. 对于IL8基因在芯片中的另一探针:

202859 X AT探针, 其所在区域序列如下:

gtacccagttaaattttcatttcagataaacaacaaataatttttagtataagtacatt attgtttatctgaaagttttaattgaactaacaatcctagtttgatactcccagtcttgt cattgccagctgttggtagtgctgtgttgaattacggaataatgagttagaactatta aaacagccaaaactccacagtcaatattagtaatttcttgctggttgaaacttgtttatt atgtacaaatagattcttataatattattaaatgactgcattttaaatacaaggcttt atattttaactttaagtgttttattgtgctctccaaattttttactgtttctgatt gtat

202859\_X\_AT与NM\_000584(全长1718bp)比对结果如下(在1289-1651 处靠近基因的3'端,所以其荧光信号值也较高):



地址:上海市张江高科技园区李冰路 151 号(201203) 电话: 021-51320288 传真: 021-51320266 **7/12** 技术服务热线: 800-820-5086 / 400-880-5086 技术服务网站: http://www.shbio.cn



Sbjct 1528 ATGTACAAATAGATTCTTATAATATTATTTAAATGACTGCATTTTTAAATACAAGGCTTT 1587

Query 361 GTAT 364

Sbjct 1648 GTAT 1651

从实验结果来看,202859探针也存在着在不同样本间差异的现象,可 作为候选探针。

注:探针信号值差异数据在一倍左右,验证中也可能差异不明显,需要考虑后选择。

Probe_Set_ID	样本	样本	样本	样本	样本	样本	Gene Symbol	Entrez Gene	RefSeq Transcript ID
211506_s_at	10.80	8. 93	8. 85	8. 55	6. 31	8. 60	IL8	3576	NM_000584
202859_x_at	14. 32	13. 10	13. 12	13. 27	11. 46	13. 39	IL8	3576	NM_000584

## 2. 而对于ITPR1基因还设计以下的探针区域:

>HG-U133\_PLUS\_2:211323\_S\_AT

211323 S AT与NM 001099952 (Length=10098) 比对结果如下:

#### 

电话: 021-51320288 传真: 021-51320266

技术服务网站: http://www.shbio.cn

8/12

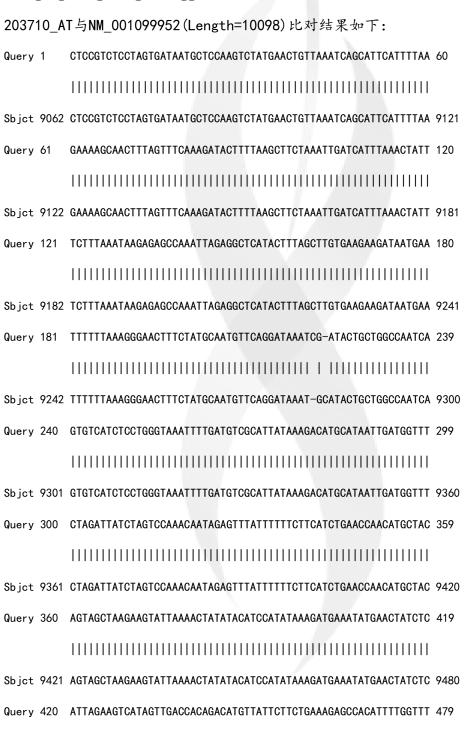




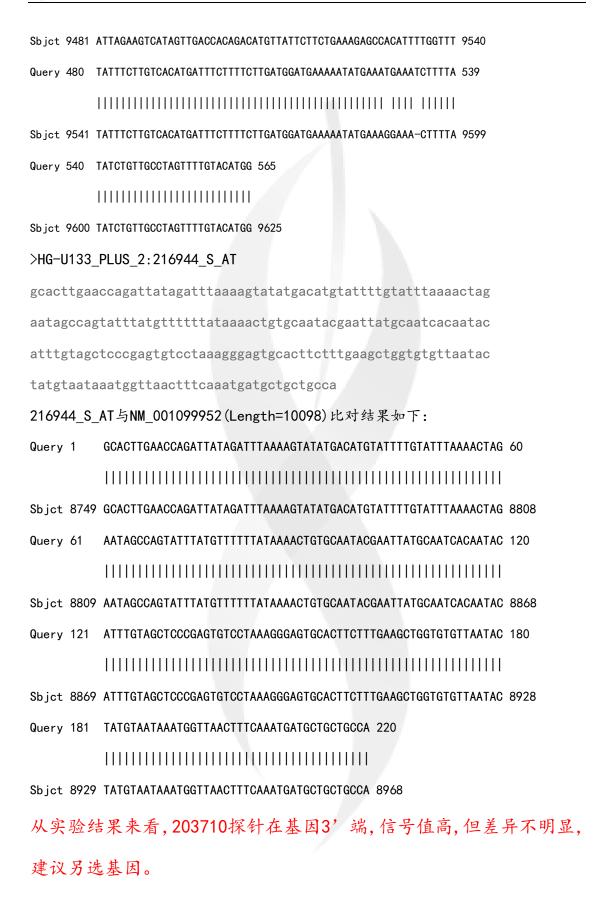
地址:上海市张江高科技园区李冰路 151 号 (201203) 技术服务热线: 800-820-5086 / 400-880-5086 电话: 021-51320288 传真: 021-51320266

9 / 12





地址:上海市张江高科技园区李冰路 151 号(201203) 电话: 021-51320288 传真: 021-51320266 **10/12** 技术服务热线: 800-820-5086 / 400-880-5086 技术服务网站: http://www.shbio.cn



地址:上海市张江高科技园区李冰路 151 号(201203) 电话: 021-51320288 传真: 021-51320266 **11/12** 技术服务热线: 800-820-5086 / 400-880-5086 技术服务网站: http://www.shbio.cn



Probe_Set_ID	样本	样本	样本	样本	样本	样本	Gene Symbol	Entrez Gene	RefSeq Transcript ID
240052_at	6. 38	6. 95	6. 54	8. 04	7. 46	7. 23	I TPR1	3708	NM_001099952 /// NM_002222
211323_s_at	7. 00	7. 55	7. 18	7. 10	7. 21	6. 63	I TPR1	3708	NM_001099952 /// NM_002222
203710_at	10. 75	10. 91	10.94	10. 83	11.11	10. 60	I TPR1	3708	NM_001099952 /// NM_002222
216944_s_at	9.84	9. 67	10.14	9. 81	10.04	9. 68	I TPR1	3708	NM_001099952 /// NM_002222

上海伯豪生物技术有限公司 2016-1-20