

表达谱芯片验证基因挑选注意事项

尊敬的客户：您好！

首先非常感谢您选择欧易生物作为研究合作伙伴，我们将一如既往地致力于为科研工作者提供真实可靠的数据，加速您的实验进程。当您对表达谱芯片的数据分析之后，接下来可能需要对芯片结果进行荧光定量 PCR 验证。在进行验证的时候，有些客户反映出一些问题，比如基因扩增不出，Ct 值过高（超过 30），与芯片结果不吻合，非特异性扩增等等。针对这样的问题，根据欧易生物实验人员和数据分析人员丰富的经验，我们总结了以下注意事项，希望对您的实验有所帮助。

1、选择信号值比较高的探针。

首先看基因对应探针的检出情况，选择检出的探针用于后续验证(Agilent 标记为“Detected”，Affymetrix 芯片标记为“P”)。对于检出的探针，尽量选择标准化信号值大于 8 以上的探针。标准化信号值以 2 为底进行 log 转化，如当标准化信号值为 7，其真实信号值为 $2^7=128$ ，这个值是偏低，可能会因为 Ct 值过高而导致结果不可信。建议选择真实信号 200-400 及以上(根据样本类型不同，可稍作变化)，也就是选择标准化信号值 8 以上的探针用于后续验证更容易得到比较准确和有效的结果。

2、选择位置靠近 3'UTR 的探针。

对于 Agilent 表达谱芯片和 Affymetrix 3'IVT 芯片而言，探针一般设计在 3'UTR 区。表达谱芯片样品需要将 RNA 先反转录成 cDNA 后再进行后续实验。mRNA 反转录引物是 oligo d(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 的高级结构的影响，5'端反转录效率降低，在基因 5'端的探针检测信号值会偏低，在进行定量验证时也难以得到有效的扩增，因此在有几条探针对应同一个基因或者转录本的时候，优先选择位置靠近 3'UTR 区的探针进行验证实验。

3、引物设计的位置靠近探针，PCR 产物最好包含探针序列。

如上所述，设计引物时尽量选择靠近 3'UTR 区的探针附近。由于选择性剪切或选择性转录终止等生理因素会造成的异构转录本的形成，因此在芯片探针设计的时候，就可能出现几个探针检测同一个转录本的情况，且这些探针检测的信

号值会存在差异。在这种情况下，我们建议根据您的研究目的留下相应的探针进行验证，(也可全部留下，针对每个探针分别进行验证)。在进行引物设计的时候，建议扩增产物包含探针序列。

4、选择在各个样品间稳定表达的内参基因。

实时荧光定量 PCR 通常会选择一定的内参基因进行校正和标准化。选择合适的内参基因是保证试验成功的基础。物种或者样本类型不同，选择的内参基因略有差异。选取的内参基因往往是一些管家基因，常用的包括 β -actin, GAPDH, 18S rRNA 等。内参基因需要选择在各个样品间稳定表达的基因。如果在实验前不知道选择哪个管家基因作为内参基因，可以先选择几个候选的内参基因做预实验，最终确定稳定表达的基因用于后续定量实验的校正。

5、设计出来的引物最好对扩增效率进行验证

软件设计出来的引物我们一般优先选择打分高的引物，但是并不是所有的引物都可以有好的定量效果。如果引物设计不好，可能会出现非特异性扩增，熔解曲线会出现多个峰值，导致扩增效果不好，Ct 值过高。好的扩增引物 Ct 值一般最好是在 20-25 之间，超过 30 则数据可信度低，可能是因为该基因丰度确实很低，也可能是引物扩增效率低或模板降解。

6、如果仅是针对芯片结果进行验证，最好使用与芯片实验相同的样本，建议在芯片实验完成后半年内进行。

我们一般建议用于芯片的样本进行备份，一方面防止实验中出现意外，另一方面可以用于后续的验证实验。用相同的样本进行验证一般符合程度会比较高，但是也可能出现由于样本保存时间过长导致降解或者样本异质性等各方面的原因出现符合程度较差的情况。

探针选择案例解析：

案例中所用的芯片为 Agilent SurePrint G3 Human Gene Expression 8x60K v2 Microarray, 客户想要验证基因 ANKRD36(一般公司需要提供探针号和 Genebank 号，不能仅指定基因，在这里指定基因是做示例展示探针挑选的过程)。

1、对探针的信号值进行分析。

先从差异基因中将对应的探针全部筛选出来，包括如下三条。

| ProbeName | FC (abs) | Regulation | control.tx | case.tx | GeneSymbol | Description | EntrezGene | GenbankAccession |
|---------------|----------|------------|------------|---------|------------|-------------------|------------|------------------|
| A_33_P3335030 | 2.12662 | down | 8.31606 | 7.2275 | ANKRD36 | Homo sapi | 375248 | NM_001164315 |
| A_21_P0002474 | 2.4359 | down | 8.597503 | 7.31305 | ANKRD36 | Homo sapiens anky | | NM_001164315 |
| A_21_P0002475 | 2.48332 | down | 5.05152 | 3.73925 | ANKRD36 | Homo sapiens anky | | NM_001164315 |

从结果来看，三条探针检测的信号值都表现为下调约 2 倍，其中前两条信号值都在 7 以上，作为备选的探针。

2、对探针的位置进行分析

打开网页(NCBI blast):

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=megaBlast&SHOW_DEFAULTS=on&BLAST_SPEC=blast2seq&LINK_LOC=align2seq，输入第一条探针的序列(探针序列列表见报告中 GEO 文件(Agilent)或者 Target sequence 文件(Affymetrix))和对应转录本的 Genebank 号，如下图，点击“BLAST”。

blastn blasto blastx tblastn tblastx

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

CGAGGAAGATTCTGTTTGTAGTATAGCCAGAGAAGAAAAGGATGGAGAAAAATCTAGGAC

Clear Query subrange

From To

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

☒ Align two or more sequences

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence

NM_001164315

Clear Subject subrange

From To

Or, upload file

Program Selection

Optimize for

☒ Highly similar sequences (megablast)

☐ More dissimilar sequences (discontiguous megablast)

☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

出现下图:

Download GenBank Graphics

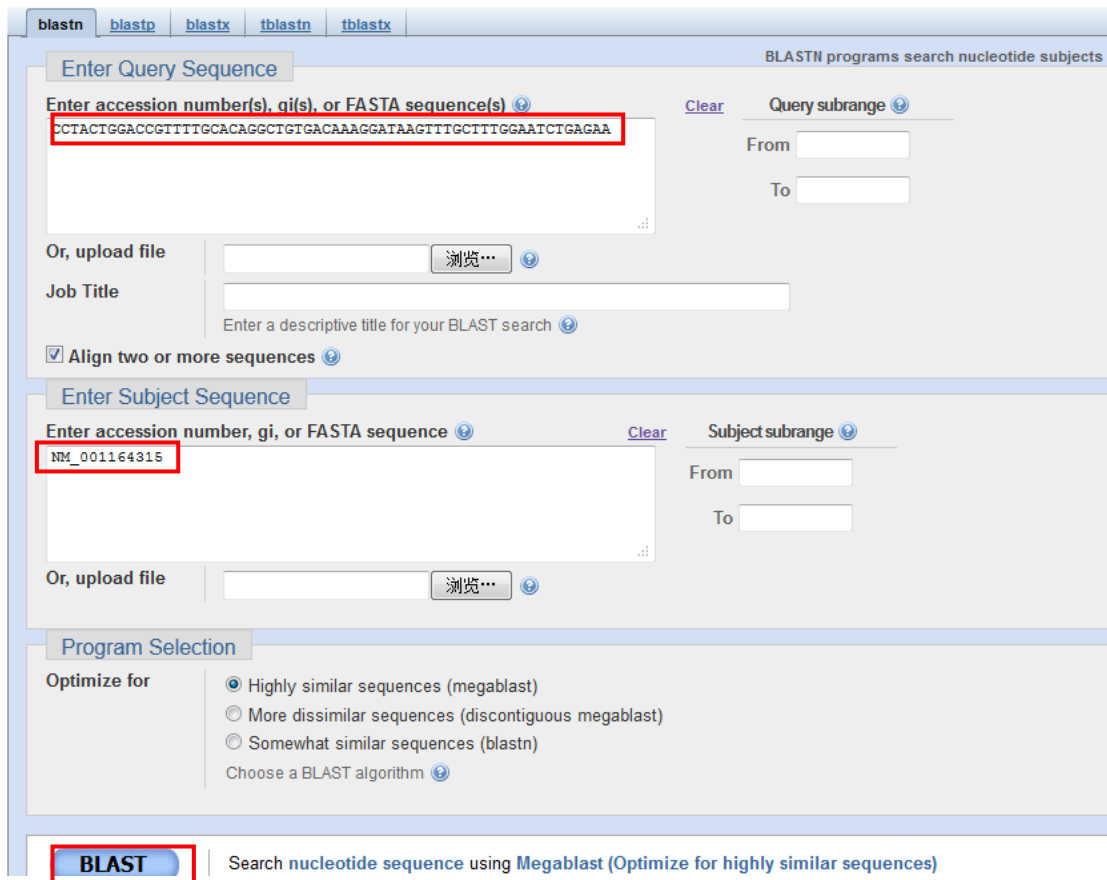
Homo sapiens ankyrin repeat domain 36 (ANKRD36), mRNA

Sequence ID: [ref|NM_001164315.1|](#) Length: 6269 Number of Matches: 1

Range 1: 3347 to 3405 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

| Score | Expect | Identities | Gaps | Strand |
|---------------|---|------------|----------|-----------|
| 99.0 bits(53) | 5e-25 | 57/59(97%) | 0/59(0%) | Plus/Plus |
| Query 2 | CAGGAAGATTCTGTTTGTAGTATAGCCAGAGAAGAAAAGGATGGAGAAAAATCTAGGAC | 60 | | |
| Sbjct 3347 | CAGGAAGATTCTGTTTGTAGTATAGCCAGAGAAAACAAGGATGGAGAAAAATCTAGGAC | 3405 | | |

输入第二条探针的序列和对应转录本的 Genebank 号, 如下图, 点击“BLAST”。



blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide subjects

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [From](#) [To](#)

CCTACTGGACCGTTTTCACAGGCTGTGACAAAGGATAAGTTTGCTTTGGAATCTGAGAA

Or, upload file [浏览...](#)

Job Title [Enter a descriptive title for your BLAST search](#)

☒ Align two or more sequences

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence [Clear](#) Subject subrange [From](#) [To](#)

NM_001164315

Or, upload file [浏览...](#)

Program Selection

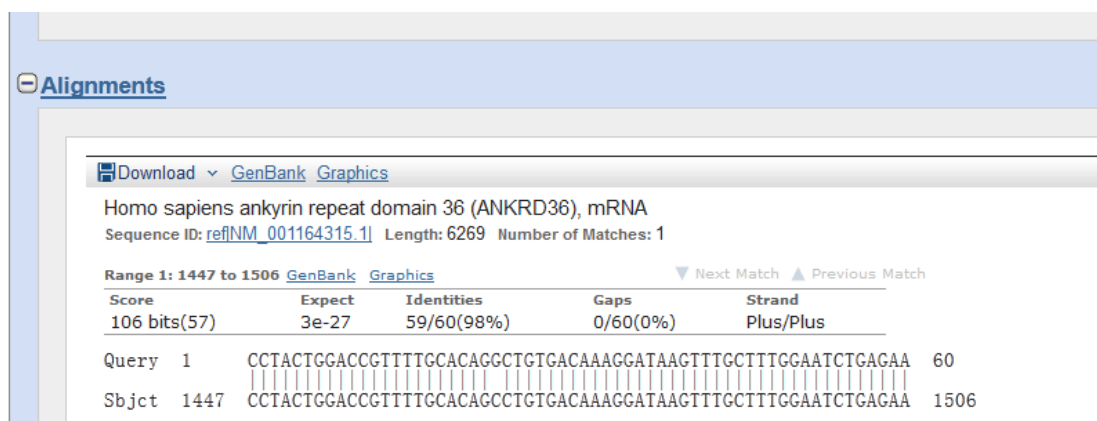
Optimize for

- ☒ Highly similar sequences (megablast)
- ☐ More dissimilar sequences (discontiguous megablast)
- ☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

出现下图:



Alignments

Download [GenBank](#) [Graphics](#)

Homo sapiens ankyrin repeat domain 36 (ANKRD36), mRNA
Sequence ID: [ref|NM_001164315.1|](#) Length: 6269 Number of Matches: 1

Range 1: 1447 to 1506 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

| Score | Expect | Identities | Gaps | Strand |
|--------------|--------|------------|----------|-----------|
| 106 bits(57) | 3e-27 | 59/60(98%) | 0/60(0%) | Plus/Plus |

Query 1 CCTACTGGACCGTTTTCACAGGCTGTGACAAAGGATAAGTTTGCTTTGGAATCTGAGAA 60

Sbjct 1447 CCTACTGGACCGTTTTCACAGGCTGTGACAAAGGATAAGTTTGCTTTGGAATCTGAGAA 1506

第一条序列更接近 3'UTR 区, 所以优先选择第一条序列。设计引物时, 扩增出来的片段尽量包含探针序列。