

猪核糖体蛋白S10的生物信息学 分析



组别：G08 组

组员：蒲 蕾 刘文舒

崔小燕 高 倩

2014年1月25日

报告提纲

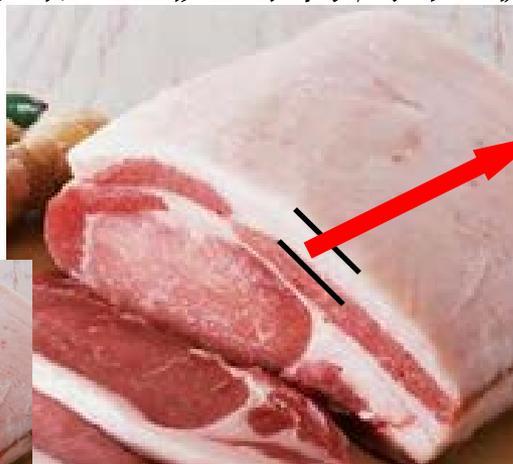
- 基因**RPS10**的研究背景
- 基因**RPS10**核酸及蛋白结构分析
- 课题的进一步研究计划

研究背景

猪肉品质

猪肉品质由20多种的指标构成，如肉色、pH、系水力、**皮下脂肪**、滴水损失、多汁性、嫩度、大理石纹、肌肉脂肪、瘦肉率各部分胴体的产肉率等等。

猪的皮下脂肪：俗称背膘，是指猪的真皮层以下肌肉组织以上的脂质部分。它可以直接反应胴体品质和瘦肉率的高低，所以在实际育种生产中很重视对背膘性状的选育。



皮下脂肪

半个世纪以来我国猪肉消费的趋势



50-80年代



90年代到现在



将来



东北民猪



东北民猪肉

东北民猪是华北猪种，在世界地方猪品种排行第四，它肉质坚实，大理石纹分布均匀，肉色鲜红，给人口感细腻多汁，色香味俱全，它以产仔量高，抗病强、耐粗饲、杂交效果显著。

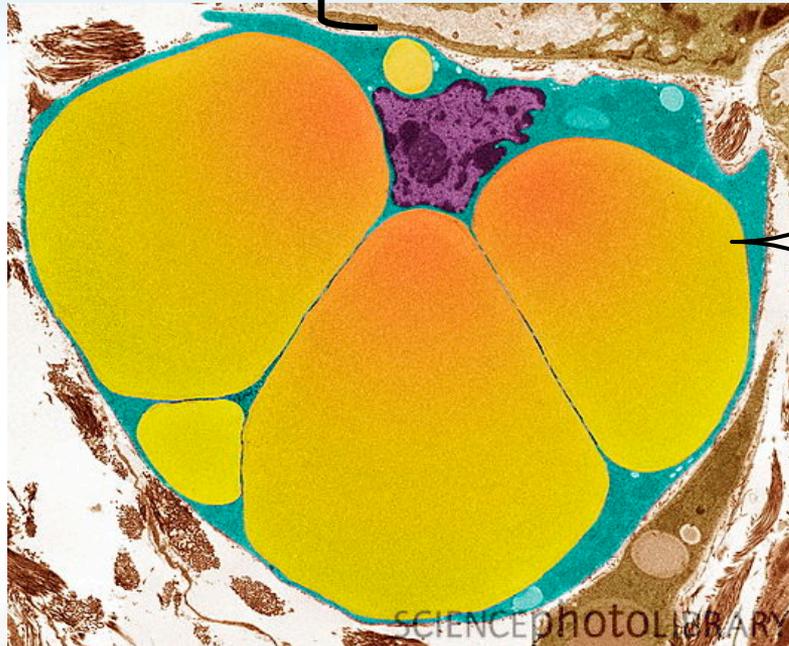
背膘厚！！

本实验室通过将大白猪和民猪进行杂交，并构建了 F_2 代资源群，对此资源群进行了全基因组关联分析，发现了与猪背膘厚显著相关的候选基因RPS10。

脂肪细胞

数目——增殖

充脂——分化



代谢途径:

脂肪酸 \rightarrow 脂酰CoA



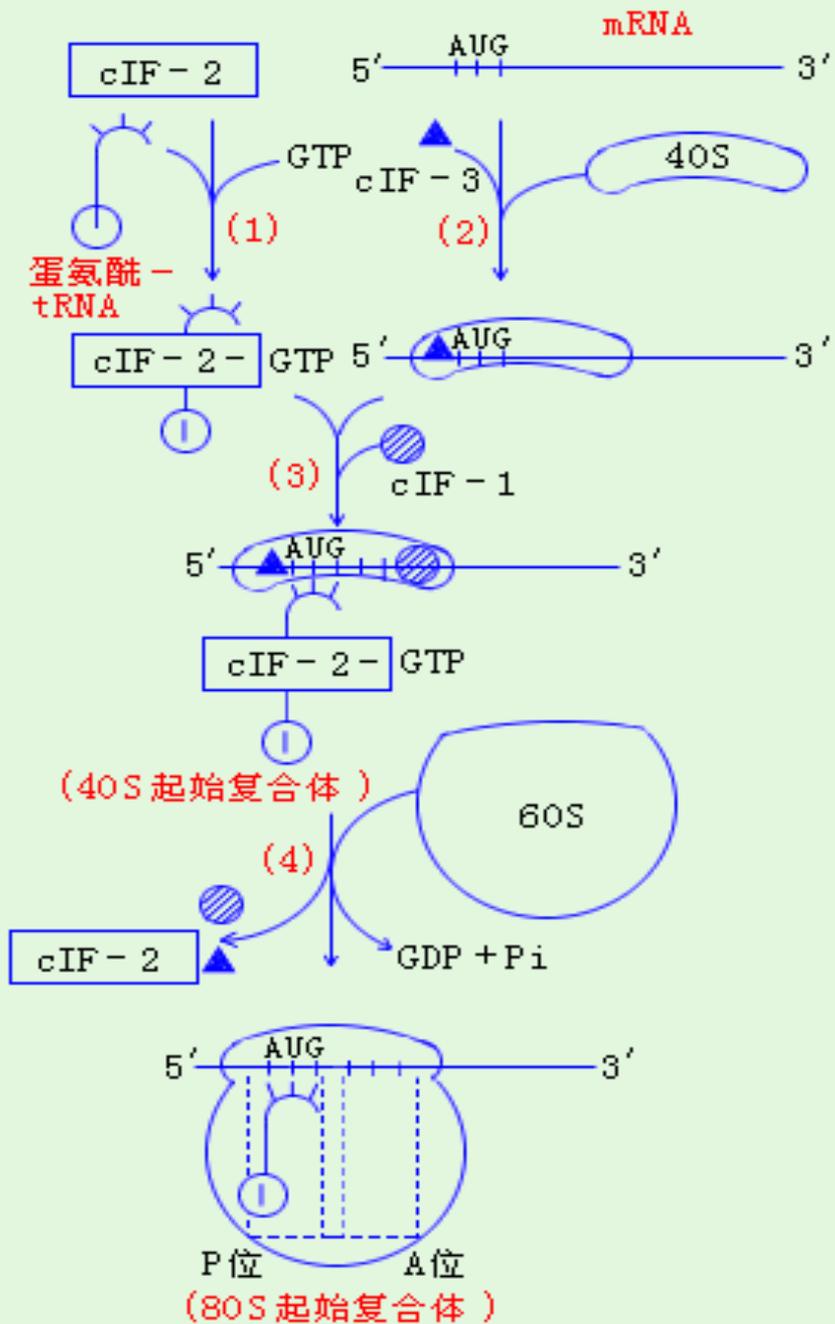
进入线粒体代谢利用

合成途径:

脂肪酸从头合成途径

甘油三酯合成途径

相关信号通路 **JNK/MAPK**、**PI3K/Akt-mTOR**、**Wnt/ β -Catein**



- **RPS10: Ribosomal protein S10, 核糖体蛋白S10, 构成40S核糖体小亚基。**
- **翻译起始阶段, 40S核糖体小亚基与起始因子3和mRNA结合成一个三元复合物。起始因子2与甲硫氨酰-tRNA和GTP结合成另一个三元复合物。两个复合物在起始因子1的作用下形成了起始复合体。**
- **猪40S核糖体小亚基含有30多个成员, 负责RNA模板的固定, 和起始密码子的识别。**

- RPS10: mRNA登陆号: NM_001244106, 1406bp, 位于第7染色体上。
蛋白的登陆号: NP_001231035.1, 165AA。
- 可能的功能: 参与细胞蛋白的生成, 影响细胞增殖过程。

本课题拟通过对RPS10基因进行生物信息学方面的预测，从而为此基因成脂方面功能的验证奠定基础。

报告提纲

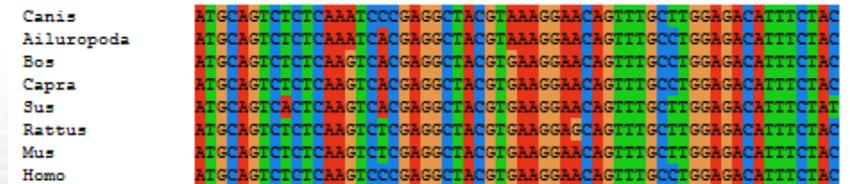
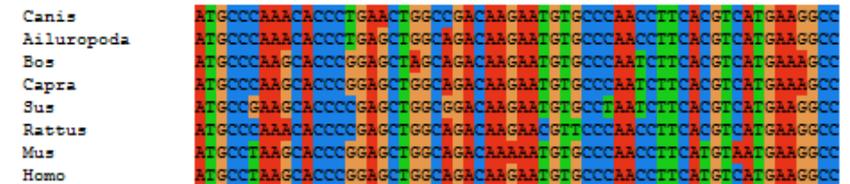
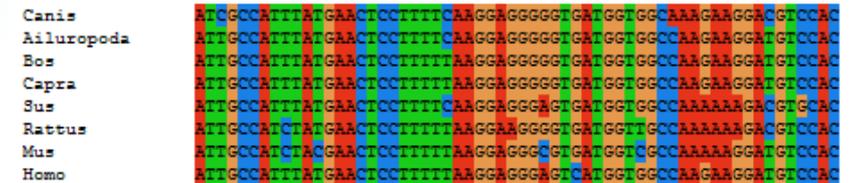
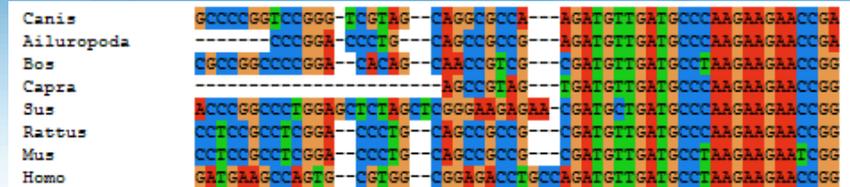
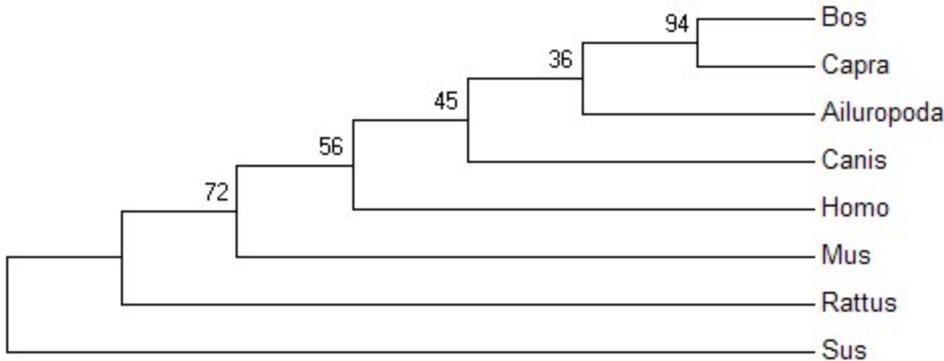
- 基因RPS10的研究背景
- 基因**RPS10**核酸及蛋白结构分析
- 课题的进一步研究计划

基因RPS10的生物信息学分析

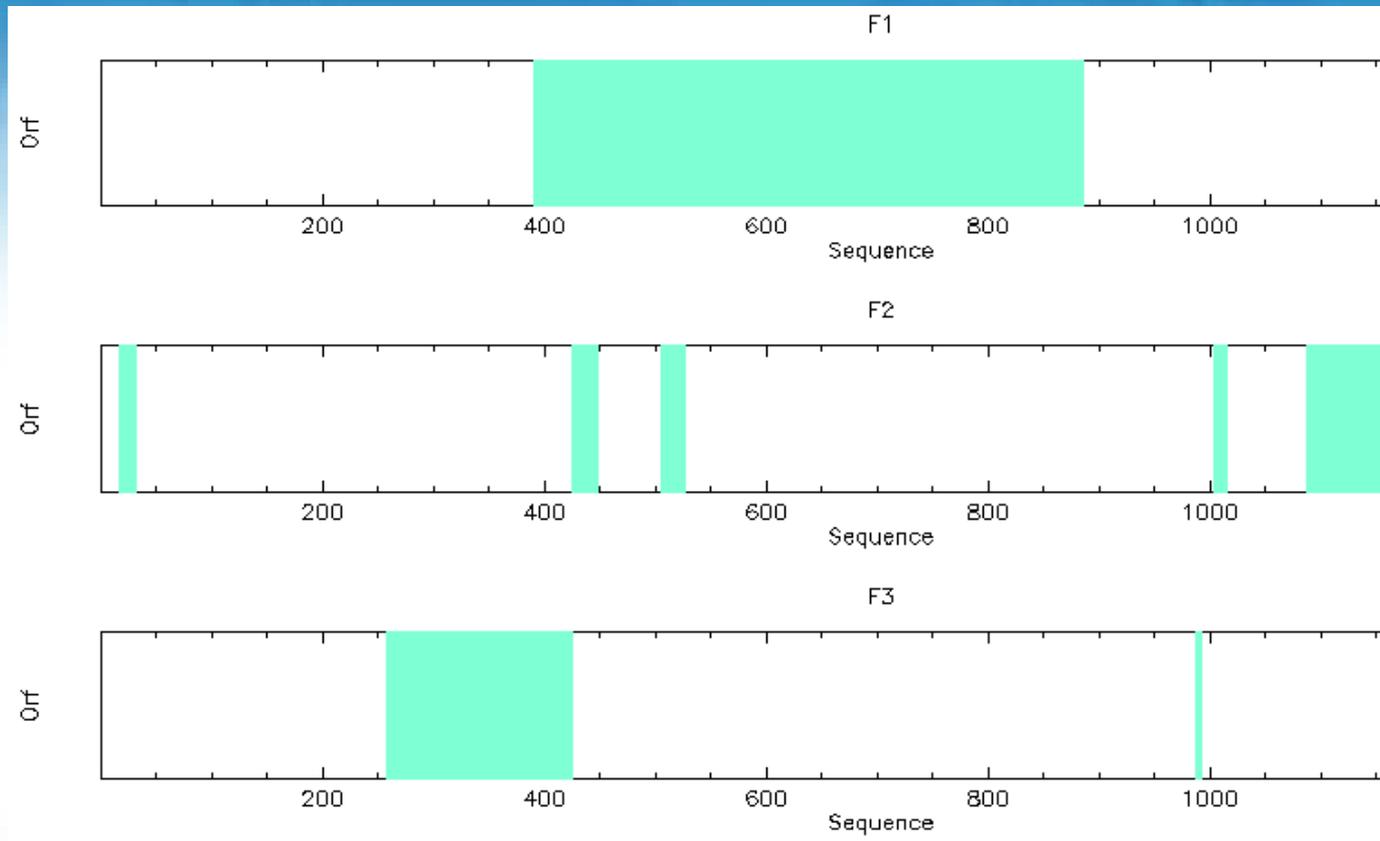
核酸方面分析

提取大熊猫 (Ailuropoda melanoleuca)、牛 (Bos Taurus)、狼 (Canis lupus)、山羊 (Capra hircus)、智人 (Homo sapiens)、大鼠 (Rattus norvegicus) 小鼠 (musculus) 和猪 (Sus scrofa) 此基因进行多序列比对, 发现此基因mRNA序列的相似性较高。

构建引导树如下: 猪的RPS10基因没有旁系同源性较高的基因。



ORF区的预测



从此图中可看出本基因的开放式阅读框位于F1区域内。在开放式阅读框内设计干扰RNA，干扰效果会更好一些。

基因RPS10的生物信息学分析

蛋白质方面的预测

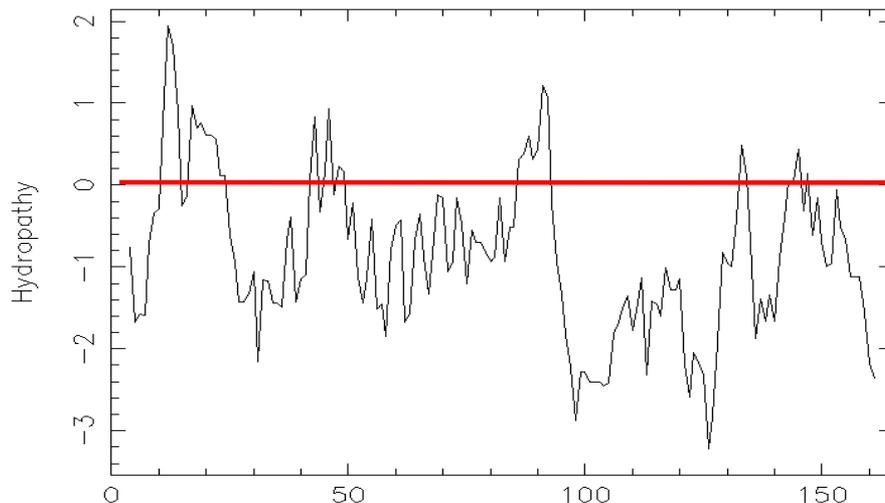
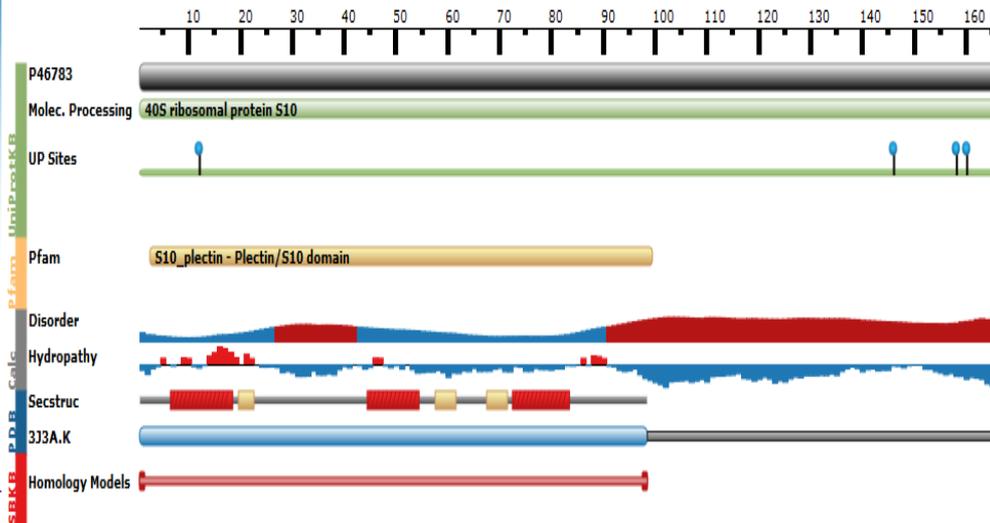
PEPSTATS of from 1 to 165

Molecular weight = 18931.74 Residues = 165
Average Residue Weight = 114.738 Charge = 15.5
Isoelectric Point = 10.7355
A280 Molar Extinction Coefficient = 20340
A280 Extinction Coefficient 1mg/ml = 1.07
Improbability of expression in inclusion bodies = 0.620

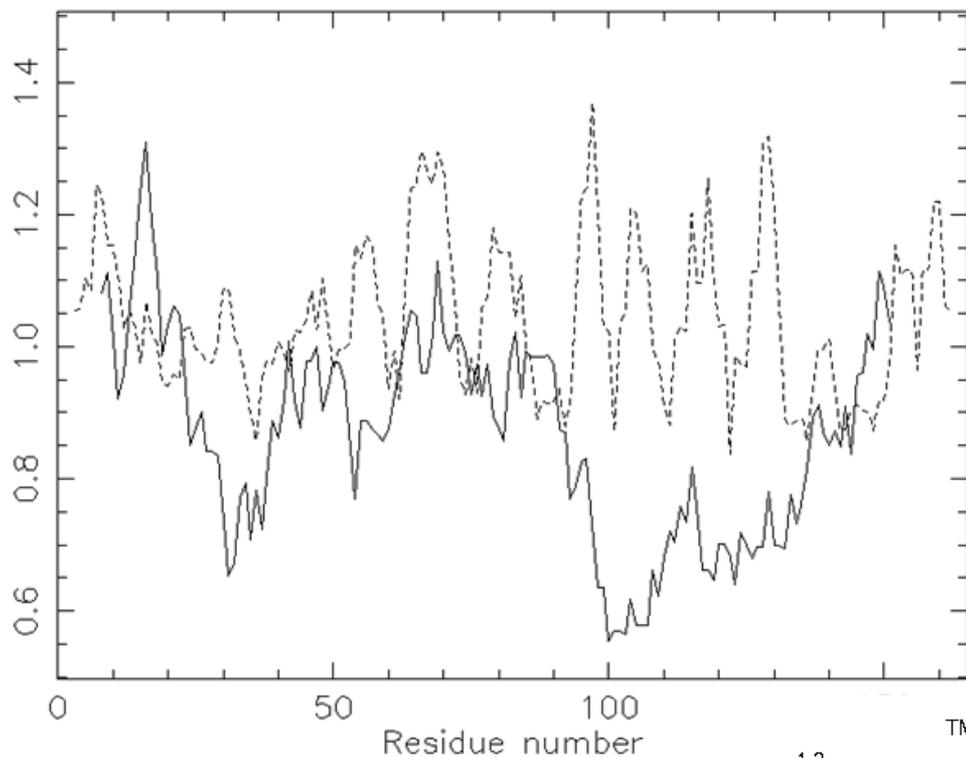
Residue	Number	Mole%	DayhoffStat
A = Ala	14	8.485	0.987
B = Asx	0	0.000	0.000
C = Cys	0	0.000	0.000
D = Asp	6	3.636	0.661
E = Glu	12	7.273	1.212
F = Phe	7	4.242	1.178
G = Gly	15	9.091	1.082
H = His	5	3.030	1.515
I = Ile	4	2.424	0.539
J = ---	0	0.000	0.000
K = Lys	13	7.879	1.194
L = Leu	12	7.273	0.983
M = Met	6	3.636	2.139
N = Asn	4	2.424	0.564
O = ---	0	0.000	0.000
P = Pro	15	9.091	1.748
Q = Gln	6	3.636	0.932
R = Arg	18	10.909	2.226
S = Ser	5	3.030	0.433
T = Thr	6	3.636	0.596
U = ---	0	0.000	0.000
V = Val	8	4.848	0.735
W = Trp	2	1.212	0.932
X = Xaa	0	0.000	0.000
Y = Tyr	7	4.242	1.248
Z = Glx	0	0.000	0.000

Gene name: RPS10

Length:165

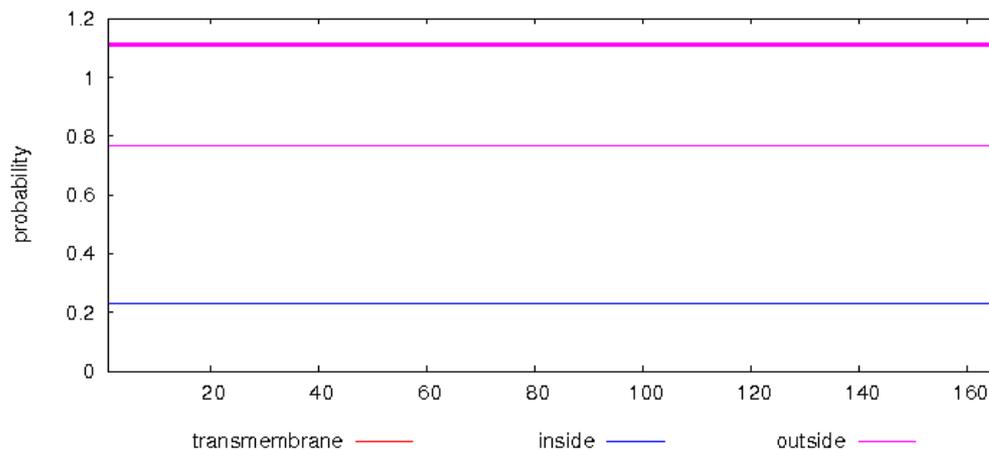


Tmap



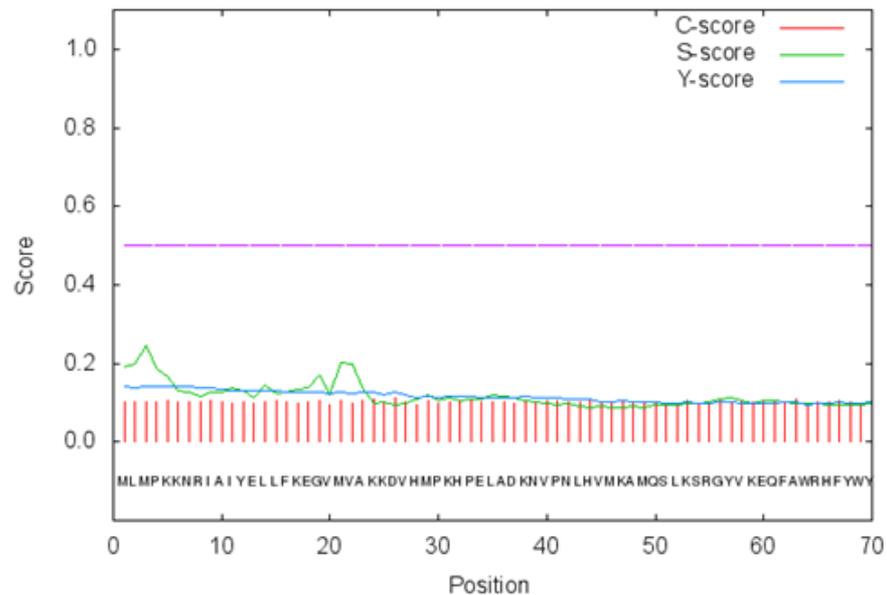
非跨膜蛋白

TMHMM posterior probabilities for gj|346421343|ref|NP_001231035.1|



```
# SignalP-4.1 euk predictions
>gi_346421343_ref_NP_001231035.1_ ribosomal protein S10 _Sus scrofa_
```

SignalP-4.1 prediction (euk networks): gi_346421343_ref_NP_001231035.1_



```
# Measure Position Value Cutoff signal peptide?
max. C 26 0.113
max. Y 12 0.132
max. S 3 0.244
mean S 1-11 0.159
D 1-11 0.147 0.450 NO
```

Name=gi_346421343_ref_NP_001231035.1_ SP='NO' D=0.147 D-cutoff=0.450 Networks=SignalP-noTM

不包含信号肽

SMART分析

Length 165 aa

Source database UniProt

Identifiers UPI00017EF9EC

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than the same piece of sequence; the priority for display is given by SMART > PFAM > PROSPERO repeats > Signal peptide. Domains above diagram are marked as 'overlap' in the right side table below.

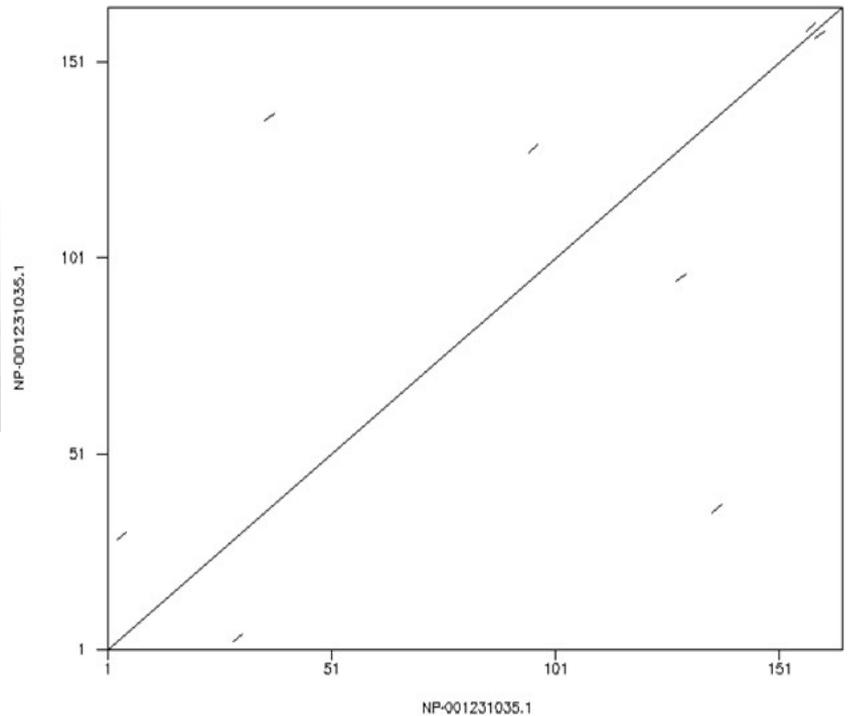
Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
Pfam:S10_pectin	3	101	1.6e-49
low complexity	150	165	N/A

Dottup

Dottup: fasta::683052:NP-001231035.1 vs fasta::683053:NP...

Sun 1 Dec 2013 19:30:08



Filter

Dataset

UniProtKB/Swiss-Prot (11) ▾

Taxonomy

Metazoa (250) ▾

Filter

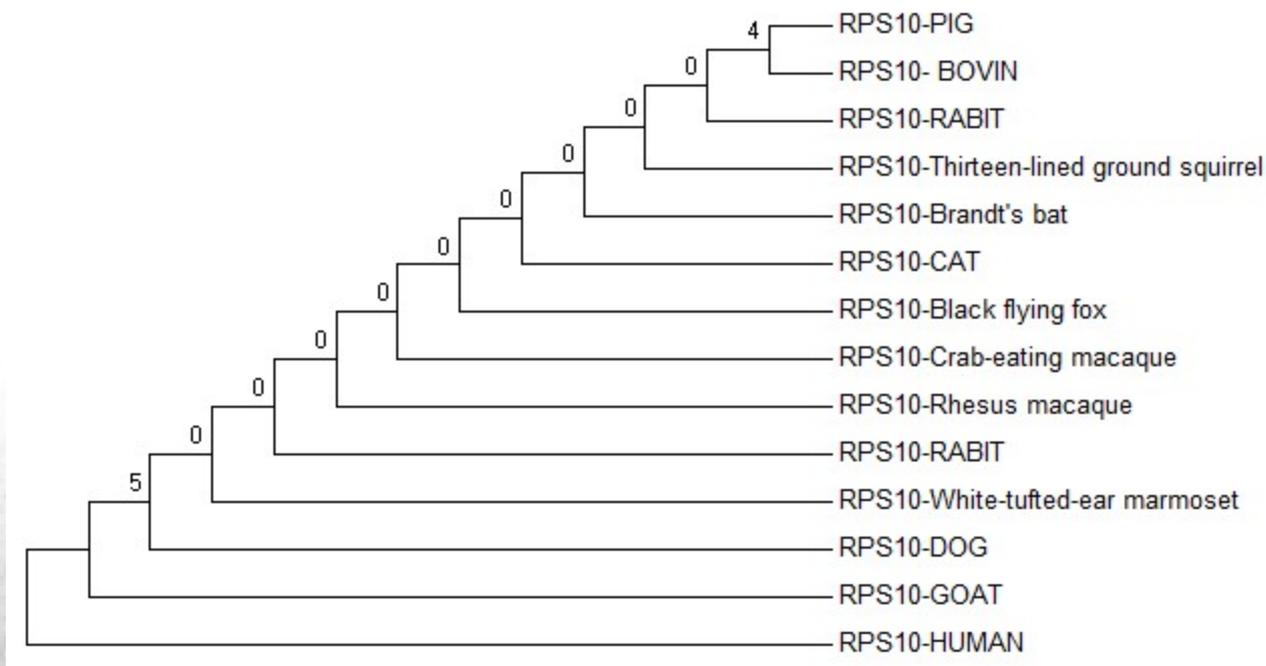
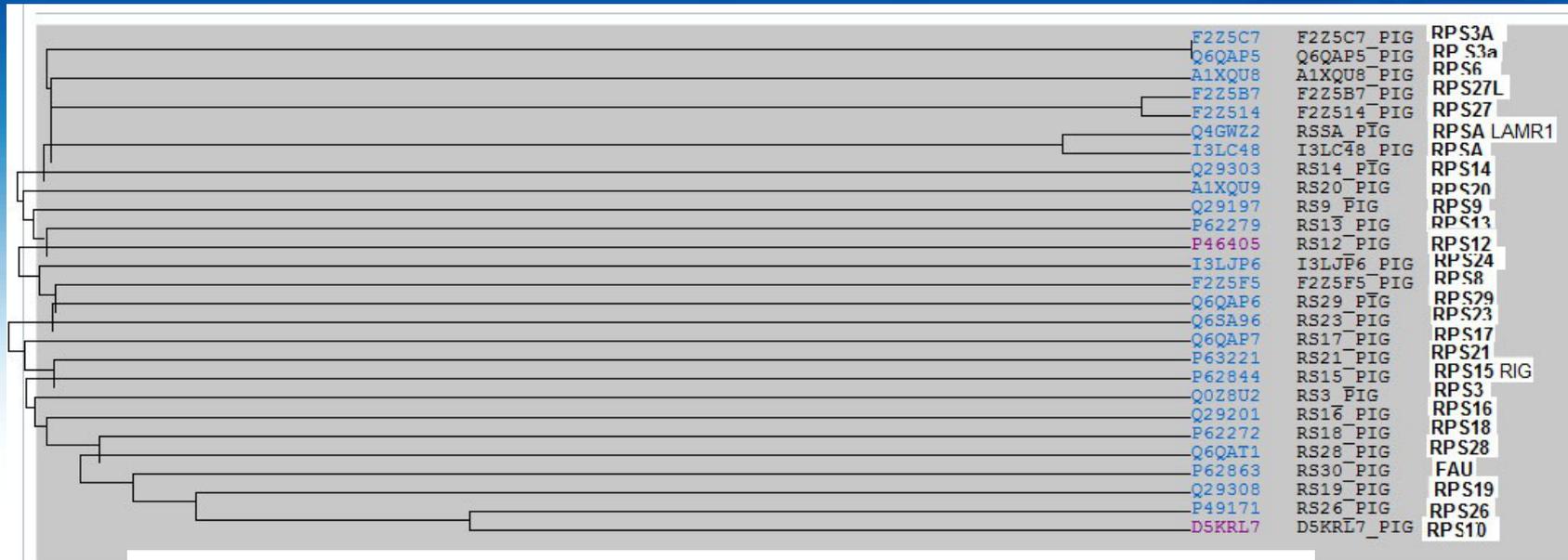
Graphical overview

Color code for identity 0-100% =



Accession	Entry name	0Query hit165	0Match hit (sqrt scale)442	Name (Organism)
<input checked="" type="checkbox"/> Query 2013113061US2BLVGI				
<input checked="" type="checkbox"/> P46783	RS10_HUMAN			40S ribosomal protein S10 (Homo sapiens)
<input checked="" type="checkbox"/> Q3T0F4	RS10_BOVIN			40S ribosomal protein S10 (Bos taurus)
<input checked="" type="checkbox"/> P63326	RS10_RAT			40S ribosomal protein S10 (Rattus norvegicus)
<input checked="" type="checkbox"/> P63325	RS10_MOUSE			40S ribosomal protein S10 (Mus musculus)
<input checked="" type="checkbox"/> Q07254	RS10_XENLA			40S ribosomal protein S10 (Xenopus laevis)
<input checked="" type="checkbox"/> Q90YR4	RS10 ICTPU			40S ribosomal protein S10 (Ictalurus punctatus)
<input checked="" type="checkbox"/> Q9NQ39	RS10L_HUMAN			Putative 40S ribosomal protein S10-like (Homo sapiens)
<input type="checkbox"/> O77302	RS10_LUMRU			40S ribosomal protein S10 (Lumbricus rubellus)
<input type="checkbox"/> Q962R9	RS10_SPOFR			40S ribosomal protein S10 (Spodoptera frugiperda)
<input type="checkbox"/> Q9VWG3	RS10B_DROME			40S ribosomal protein S10b (Drosophila melanogaster)
<input type="checkbox"/> Q9VB14	RS10A_DROME			40S ribosomal protein S10a (Drosophila melanogaster)

和BLAST中的比对结果类似，可看出猪的RPS10蛋白与人、牛、大鼠、小鼠、非洲爪蟾及鲶鱼的相似度达到90%以上。



Mus musculus ribosomal protein S10 (Rps10), mRNA
Sequence ID: [reflNM_025963.3](#) Length: 573 Number of Matches: 1

Range 1: 38 to 566 [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
641 bits(347)	0.0	471/532(89%)	3/532(0%)	Plus/Plus
Query 389	CGATGCTGATGCCCAAGAAGAACCGGATTGCCATTTATGAACTCCTTTTCAAGGAGGGAG	448		
Sbjct 38	CGATGTTGATGCCTAAGAAGAATCGGATTGCCATCTACGAACTCCTTTTAAGGAGGGCG	97		
Query 449	TGATGGTGGCCAAAAAAGACGTGCACATGCCGAAGCACCOCGAGCTGGCGGACAAGAATG	508		
Sbjct 98	TGATGGTCGCCAAAAAAGGATGTCCACATGCCTAAGCACCOCGAGCTGGCAGACAAAAATG	157		
Query 509	TGCCTAATCTTTCACGTCAATGAAGGCCATGCAGTCACTCAAGTCACGAGGCTACGTGAAGG	568		
Sbjct 158	TGCCCAACCTTCATGTAATGAAGGCCATGCAGTCTCTCAAGTCTCGAGGCTACGTGAAGG	217		
Query 569	AACAGTTTGCTTGGAGACATTTCTATTGGTACCTTACCAACGAGGGTATCCAGTATCTCC	628		
Sbjct 218	AACAGTTTGCTTGGAGACATTTCTACTGGTACCTTACGAACGAGGGCATCCAGTATCTCC	277		
Query 629	GTGATTACCTCCACCTCCCCCTGAGATTGTGCTGCTACCTGCGCCGACGCGCTCCTG	688		
Sbjct 278	GAGACTACCTGCACCTACCCCGGAGATCGTGCCCGCCACCCTGCGTGCAGCCGTCCCG	337		
Query 689	AGACTGGCAGGCCACGGCCAAAAGTCTGGAGGGAGAGCGACCTGCACGATTACACGAG	748		
Sbjct 338	AGACCGGCAGGCCTCGGCCAAAAGTCCAGAGGGTGAGCGACCTGCAAGATTACAAGAG	397		
Query 749	GGGAAGCTGACAGAGACACCTACAGACGAAGCGCTGTGCCCCCTGGTCTGACAAGAAA	808		
Sbjct 398	GGGAGGCTGACAGAGACACCTACAGAAGGAGCGCTGTGCCCCCTGGAGCTGACAAGAAA	457		
Query 809	CCGAGGCCGGAGCTGGGTACGCAACAGAGTTCCAATTTAGAGGCGGCTTTGGTCTGGAC	868		
Sbjct 458	CTGAGGCTGGGGCTGGCTCAGCCACTGAGTCCAGTTTATAGAGGCGGCTTTGGTCTGGAC	517		
Query 869	GTGGTCAGCCTCCTCAGTAAAGTTGGAAGAGATTGTTTTGTGTTGAATAAAC	920		
Sbjct 518	GTGGTCAGCCACCTCAGTGAAGTTGGA-GT--TTATGTTGATTGAATAAAC	566		

Result: Untitled.markx0.2013-11-27 19:56 PM

[Friendly Print](#)

```
#=====  
#  
# Aligned_sequences: 2  
# 1: NP_080239.1  
# 2: NP_001231035.1  
# Matrix: EPAM70  
# Gap_penalty: 14  
# Extend_penalty: 4  
#  
# Length: 165  
# Identity: 164/165 (99.4%)  
# Similarity: 164/165 (99.4%)  
# Gaps: 0/165 (0.0%)  
# Score: 1104  
#  
#=====  
  
      10      20      30      40      50  
NP_080 MLMPKKNRIAIYELLFKEGVMVAKKDVHMPKGPPELADKNVPLHVMKAMQ  
      * * * * *  
NP_001 MLMPKKNRIAIYELLFKEGVMVAKKDVHMPKGPPELADKNVPLHVMKAMQ  
      * * * * *  
      10      20      30      40      50  
  
      60      70      80      90     100  
NP_080 SLKSRGYVKEQFAWRHFYWYLTNEGQYLRDYLHLPPEIVPATLRRSRPE  
      * * * * *  
NP_001 SLKSRGYVKEQFAWRHFYWYLTNEGQYLRDYLHLPPEIVPATLRRSRPE  
      * * * * *  
      60      70      80      90     100  
  
      110     120     130     140     150  
NP_080 TGRPRPKGTEGERPARFTRGEADRDTYRRSAVPPGADKKAEAGAGSATEF  
      * * * * *  
NP_001 TGRPRPKGLEGERPARFTRGEADRDTYRRSAVPPGADKKAEAGAGSATEF  
      * * * * *  
      110     120     130     140     150  
  
      160  
NP_080 QFRGGFGRGRGPPQ  
      * * * * *  
NP_001 QFRGGFGRGRGPPQ  
      * * * * *  
      160
```

由于实验材料的限制，所以成脂相关实验需在细胞系中先进行，功能较为明确后才会利用猪的脂肪细胞进行实验，现在国内外利用的较多的为小鼠的3T3-L1细胞系，所以在设计引物或者小干扰RNA时会利用保守的区段进行设计。蛋白序列比对发现小鼠RPS10与猪的RPS10蛋白只有一个氨基酸的差异，故在选择抗体的时候可以共用。

基因功能研究

人 ✓
猪 ✗

```
Result: Untitled.markx0.2013-11-27 20:10 PM View
Friendly Print

# Aligned_sequences. 2
# 1: NM_001014.4
# 2: NM_001244106.1
# Matrix: EDNAMAT
# Gap_penalty: 14
# Extend_penalty: 4
#
# Length: 614
# Identity: 527/614 (85.8%)
# Similarity: 527/614 (85.8%)
# Gaps: 6/614 (1.0%)
# Score: 2247
#
#
#=====
          30      40      50      60      70
NM_001  GGAAGAGACGCAGCACCGCGCATGCTCCTTCCCTTCCAGCCCCGGTACCG
NM_001  GGCAGA-ACTCCGCACTCCG--TGCTGCGTCTTACC CGGCCCTGGAGCT
          330      340      350      360      370

          80      90      100     110     120
NM_001  GACCCTGCAGCCGCAGA--GATGTTGATGCCTAAGAAGAACCGGATTGCC
NM_001  CTAGCT-CGGGAAGAGAACGATGCTGATGCCAAGAAGAACCGGATTGCC
          380      390      400     410     420

          130     140     150     160     170
NM_001  ATTTATGAACTCCTTTTTAAGGAGGGAGTCATGGTGGCCAAGAAGGATGT
NM_001  ATTTATGAACTCCTTTTTCAAGGAGGGAGTGATGGTGGCCAAAAAAGACGT
          430     440     450     460     470

          180     190     200     210     220
NM_001  CCACATGCCTAAGCACCCGGAGCTGGCAGACAAGAATGTGCCAACCTTC
NM_001  GCACATGCCGAAGCACCCCGAGCTGGCGGACAAGAATGTGCTAATCTTC
          480     490     500     510     520

          230     240     250     260     270
NM_001  ATGTCATGAAGGCCATGCAGTCTCTCAAGTCCCGAGGCTACGTGAAGGAA
```

mRNA 同源性 85.8%

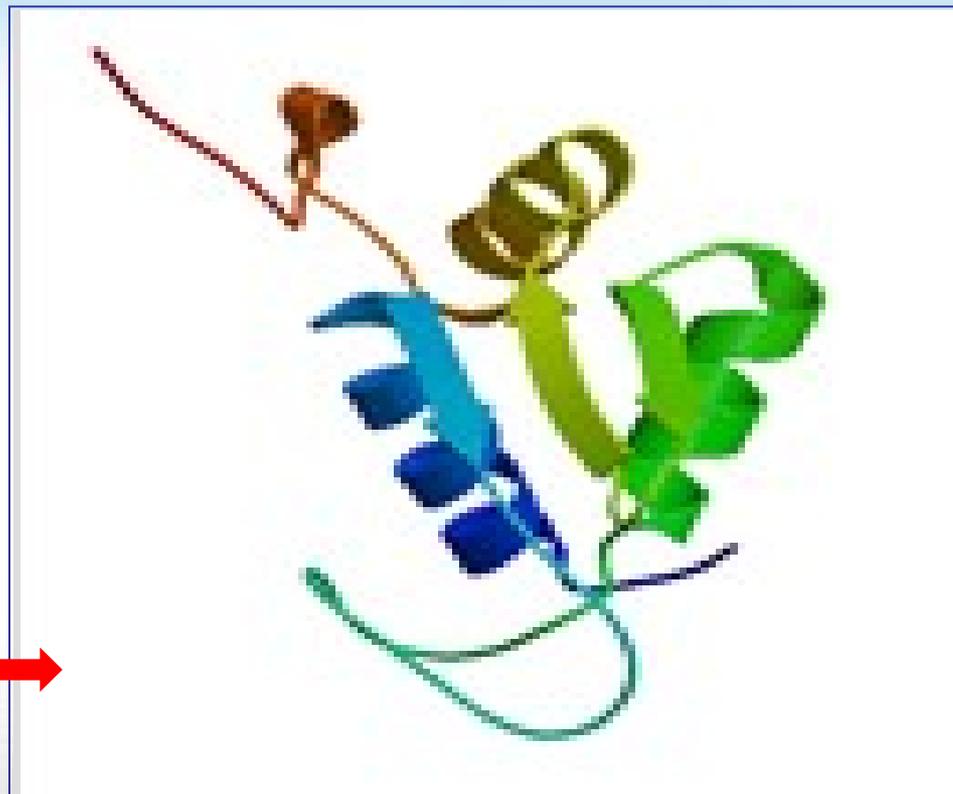
蛋白99.4%的相似，只有一个氨基酸的差异。

```
Result: Untitled.markx0.2013-11-27 19:54 PM View
Friendly Print

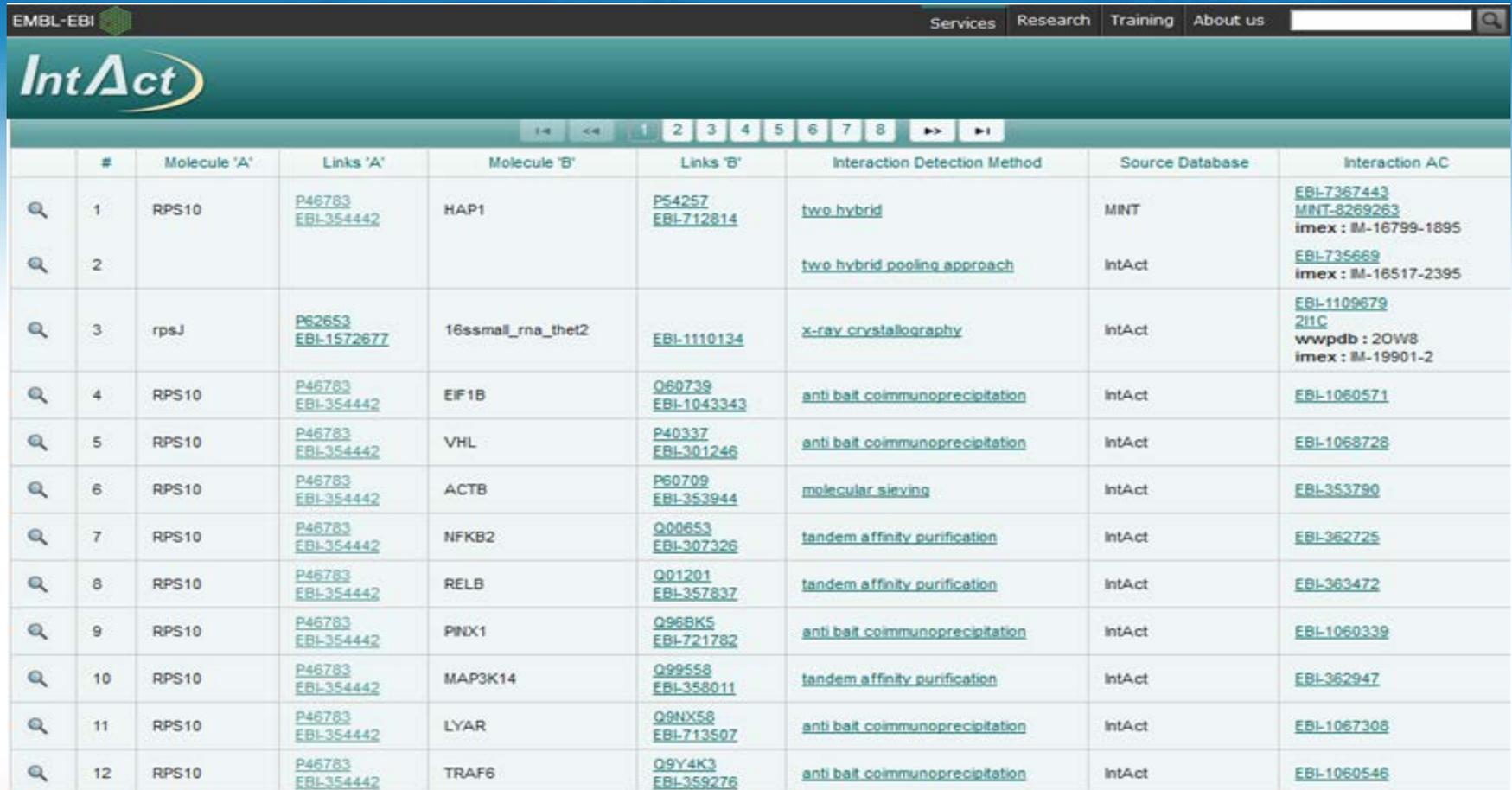
# Aligned_sequences. 2
# 1: AAH73799.1
# 2: NP_001231035.1
# Matrix: EPAM70
# Gap_penalty: 14
# Extend_penalty: 4
#
# Length: 165
# Identity: 164/165 (99.4%)
# Similarity: 164/165 (99.4%)
# Gaps: 0/165 (0.0%)
# Score: 1106
#
#
#=====
          10      20      30      40      50
AAH737  MLMPEKKNRIAIYELLFKEGVMVAKEKVHMPKHPPELADKNVPLHVMKAMQ
NP_001  MLMPEKKNRIAIYELLFKEGVMVAKEKVHMPKHPPELADKNVPLHVMKAMQ
          60      70      80      90     100
AAH737  SLKSRGYVKEQFAWRHPFYWYLTNEGIQYLRDYLHLPPPEIVPATLRRSRPE
NP_001  SLKSRGYVKEQFAWRHPFYWYLTNEGIQYLRDYLHLPPPEIVPATLRRSRPE
          110     120     130     140     150
AAH737  TGRPRPKGLEGERPARLRTGEADRDTYRRSAVPPGADKKAEGAGSATEF
NP_001  TGRPRPKGLEGERPARFRTGEADRDTYRRSAVPPGADKKAEGAGSATEF
          160
AAH737  QFRGGFGRGRGPPQ
NP_001  QFRGGFGRGRGPPQ
          160
```

以人的40S核糖体小亚基的三维结构，以此结构为模板建模。这个模型只有0-99位的氨基酸模型。

SWISS MODEL SUS →



蛋白质互作



The screenshot shows the IntAct database interface. At the top, there is a navigation bar with 'Services', 'Research', 'Training', and 'About us' links, and a search box. Below the navigation bar is the IntAct logo. The main content area displays a table of protein interactions. The table has columns for '#', 'Molecule 'A'', 'Links 'A'', 'Molecule 'B'', 'Links 'B'', 'Interaction Detection Method', 'Source Database', and 'Interaction AC'. The table lists 12 interactions, each with a magnifying glass icon in the first column. The interactions are as follows:

#	Molecule 'A'	Links 'A'	Molecule 'B'	Links 'B'	Interaction Detection Method	Source Database	Interaction AC
1	RPS10	P46783 EBL-354442	HAP1	P54257 EBL-712814	two hybrid	MINT	EBL-7387443 MINT-8269263 imex : IM-16799-1895
2					two hybrid pooling approach	IntAct	EBL-735669 imex : IM-16517-2395
3	rpsJ	P62653 EBL-1572677	16ssmall_rna_the2	EBL-1110134	x-ray crystallography	IntAct	EBL-1109679 ZINC wwpdb : 2OW8 imex : IM-19901-2
4	RPS10	P46783 EBL-354442	EIF1B	Q60739 EBL-1043343	anti bait coimmunoprecipitation	IntAct	EBL-1060571
5	RPS10	P46783 EBL-354442	VHL	P40337 EBL-301246	anti bait coimmunoprecipitation	IntAct	EBL-1068726
6	RPS10	P46783 EBL-354442	ACTB	P60709 EBL-353944	molecular sieving	IntAct	EBL-353790
7	RPS10	P46783 EBL-354442	NFKB2	Q00653 EBL-307326	tandem affinity purification	IntAct	EBL-362725
8	RPS10	P46783 EBL-354442	RELB	Q01201 EBL-357837	tandem affinity purification	IntAct	EBL-363472
9	RPS10	P46783 EBL-354442	PINX1	Q968K5 EBL-721782	anti bait coimmunoprecipitation	IntAct	EBL-1060339
10	RPS10	P46783 EBL-354442	MAP3K14	Q99558 EBL-358011	tandem affinity purification	IntAct	EBL-362947
11	RPS10	P46783 EBL-354442	LYAR	Q9NX58 EBL-713507	anti bait coimmunoprecipitation	IntAct	EBL-1067308
12	RPS10	P46783 EBL-354442	TRAF6	Q9Y4K3 EBL-359276	anti bait coimmunoprecipitation	IntAct	EBL-1060546

HAP1: ASY基因的同系物，跟细胞凋亡相关；

VHL: Von hippel lindau (人名命名)，与肿瘤的发生有关；

NFKB2: 细胞核因子K基因结合因子，调控细胞凋亡；

PINX1: 蛋白酶抑制基因，抑制癌症的发生；

EIF1B: 真核基因翻译起始因子1，控制蛋白的翻译；

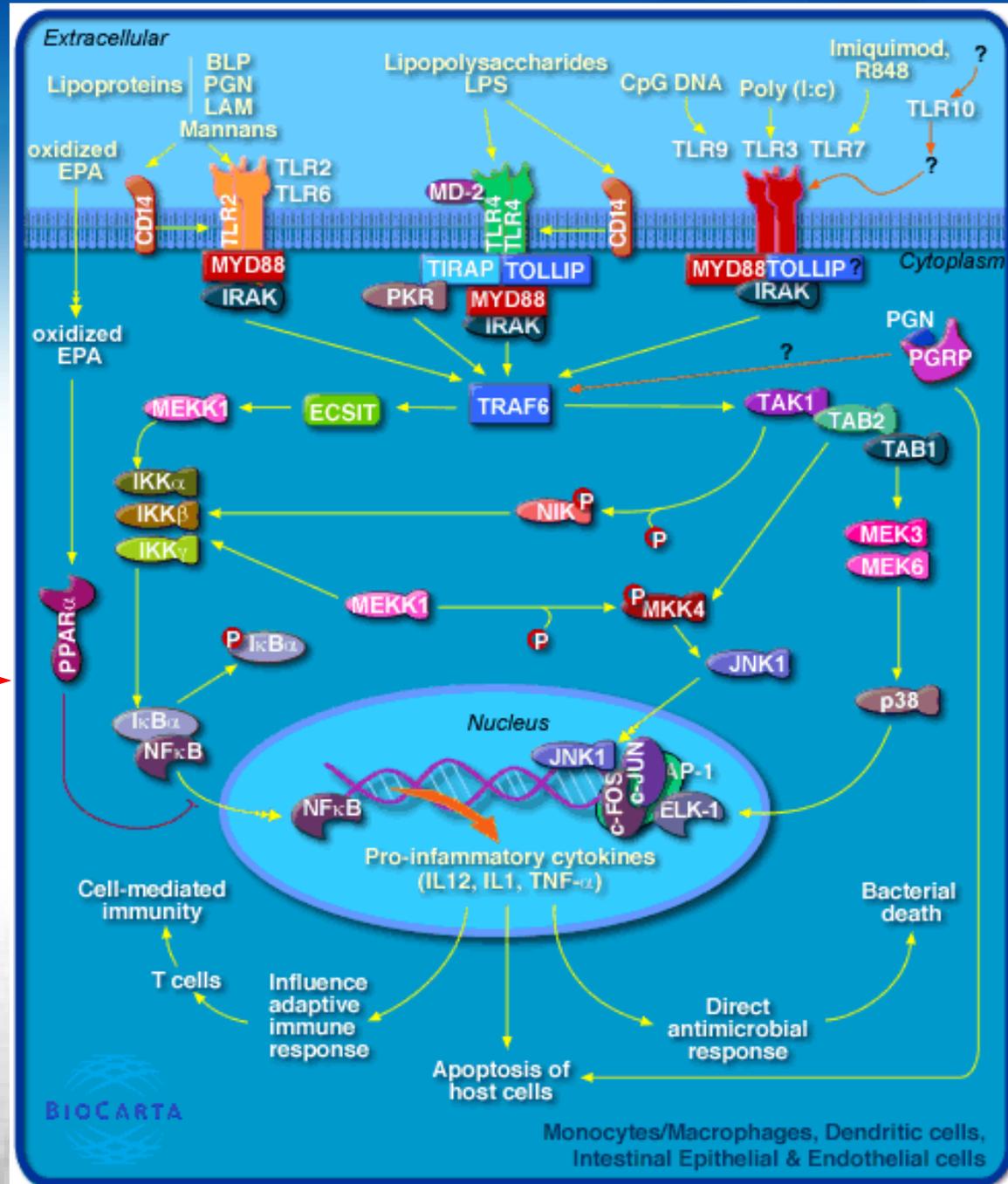
ACTB: 细胞质肌动蛋白，曾经被用于对照marker；

RELB: 可以与NFKB结合，共同发挥功能；

MAP3K14: 促有丝分裂蛋白3激酶14，调控细胞增殖。

RELB可以与NFκB结合，MAP3K14为NFκB的诱导激酶。所以我们将问题转移为RPS10对NFκB的功能影响。

NFκB基因的功能 →



RPS10



NFkB

怎么结合？
结合先后？
如何调控？



报告提纲

- 基因RPS10的研究背景
- 基因RPS10核酸及蛋白结构分析
- 课题的进一步研究计划

课题进一步研究计划

1. 根据猪和鼠mRNA的保守序列，设计干扰shRNA，进行基因干扰实验。
2. 干扰后检测成脂相关信号通路及NF κ B信号通路关键基因的表达。
3. 确定干扰作用后，进行更细致研究RPS10与NF κ B作用的机理研究。

感悟： 生物信息学分析贯穿我们实验的各个环节，实验前的预测只是一小部分，大部分分析还需要在一定实验结果基础上进行分析，指导下一步实验的进行。

通过以上生物信息学分析，使得本基因的研究思路更加明朗，为**RPS10**基因成脂方面的实验验证奠定了基础。

致 谢

首先感谢罗老师认真负责的授课和细心的指导；感谢本组成员对此次报告所付出的大量时间、努力和做出的积极准备！最后希望罗老师身体健康，工作顺利；各位同学回所后，多发好文章，实验顺利~

欢迎大家批评指正！