



Bioinformatics analysis of *RNF20* gene

组长：张壮彪

主讲人：赵莹、张仁森

组员：5G07 、5G08、5G09全体成员

候占山 (5G07B) :

光频率和光强对杏鲍菇生长发育的影响及生理研究

- **研究内容:** 从菌丝成熟后刺激杏鲍菇原基形成开始, 研究不同的光频率、光强, 对原基形成、原基分化、子实体生长三个阶段的**形态变化**、**酶活性变化** (芳基醇脱氢酶、DNA拓扑异构酶、磷脂酰乙醇胺N-甲基转移酶)、**代谢组学变化**的影响。
- **研究方法:** 形态学测定采用观察与称测的方法; 酶活性与代谢组学测定采取仪器分析与试剂分析相结合的方法;
- **预期结果:** 明确不同光频率、光强下, 杏鲍菇在形态学、三种酶活性、代谢组学上的变化情况, 并在所测得的形态与生理变化情况的基础上, 找出生理变化与形态变化的规律。

韩建刚 (5G07C)

转录组学研究滩羊胚胎不同发育时期尾部脂肪沉积的分子机制

研究背景:

以尾型为标准，在形态学上可以将绵羊分为五个类型，分别是长脂尾、短脂尾、肥臀尾、长瘦尾、短瘦尾。目前研究表明沉积1kg脂肪所用的饲料能够形成2kg瘦肉，由此可见，不仅用于形成脂肪的饲料转化率低，而且形成的脂肪不符合人们的消费需求，经济价值。因此，小尾型绵羊是选育的一个重要方向。

技术方法:

RNA-seq、免疫组化、组织形态学分析。

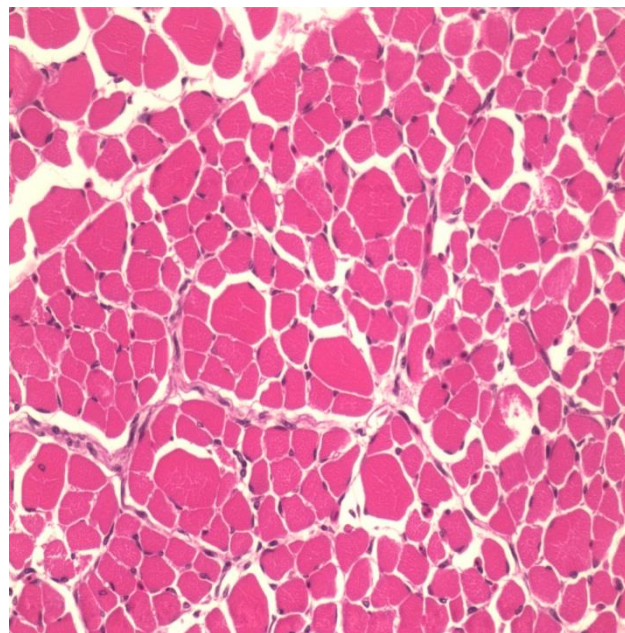


刘大鹏 (5G09A)

野鸭×北京鸭F2个体肌纤维定量分析与基因定位



北京鸭 (Z2)



野鸭

北京鸭和野鸭在肌纤维的数量和形态方面都有很大的差别。

通过500个F2个体的表型与GWAS分析，找出哪些基因调控北京鸭和野鸭在肌纤维的数量和形态方面差别。从而改进北京鸭的肌肉品质。

刘淑君 (5G09C)

一株拟杆菌转化谷氨酸的酶学机制研究

研究背景： 1.拟杆菌具有代谢谷氨酸盐的能力，谷氨酸盐是我们日常生活中味精的主要成分。拟杆菌产生的谷氨酸脱羧酶能够使谷氨酸脱羧产生 γ -氨基丁酸， γ -氨基丁酸在人体中具有重要的生理功能，被广泛应用于食品和制药工业中。

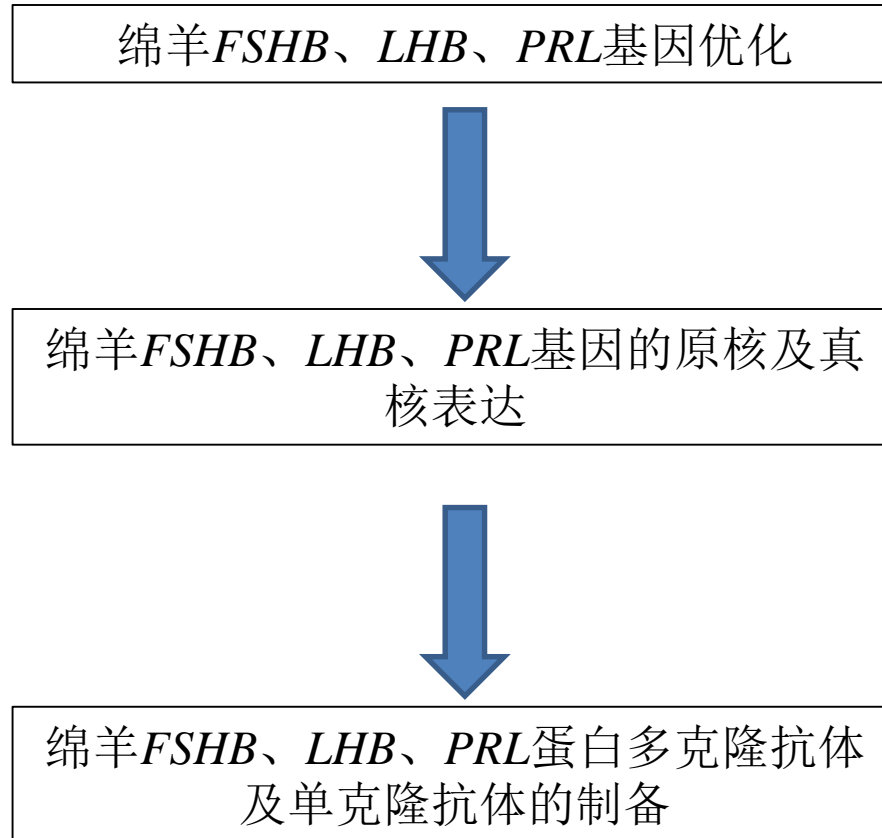
研究内容： 本研究拟从分子生物学和微生物学两个层面研究：**1.**利用体外重组相关的酶并对其结构和生化进行解析；**2.**多种底物培养，差异转录组分析。

研究目的： 解析酶的结构并明确其转化谷氨酸的酶学机制。

张仁森 (5G08C)

课题：《绵羊生殖激素 *FSH*、*LH*、*PRL* 检测抗体的制备》

技术路线：

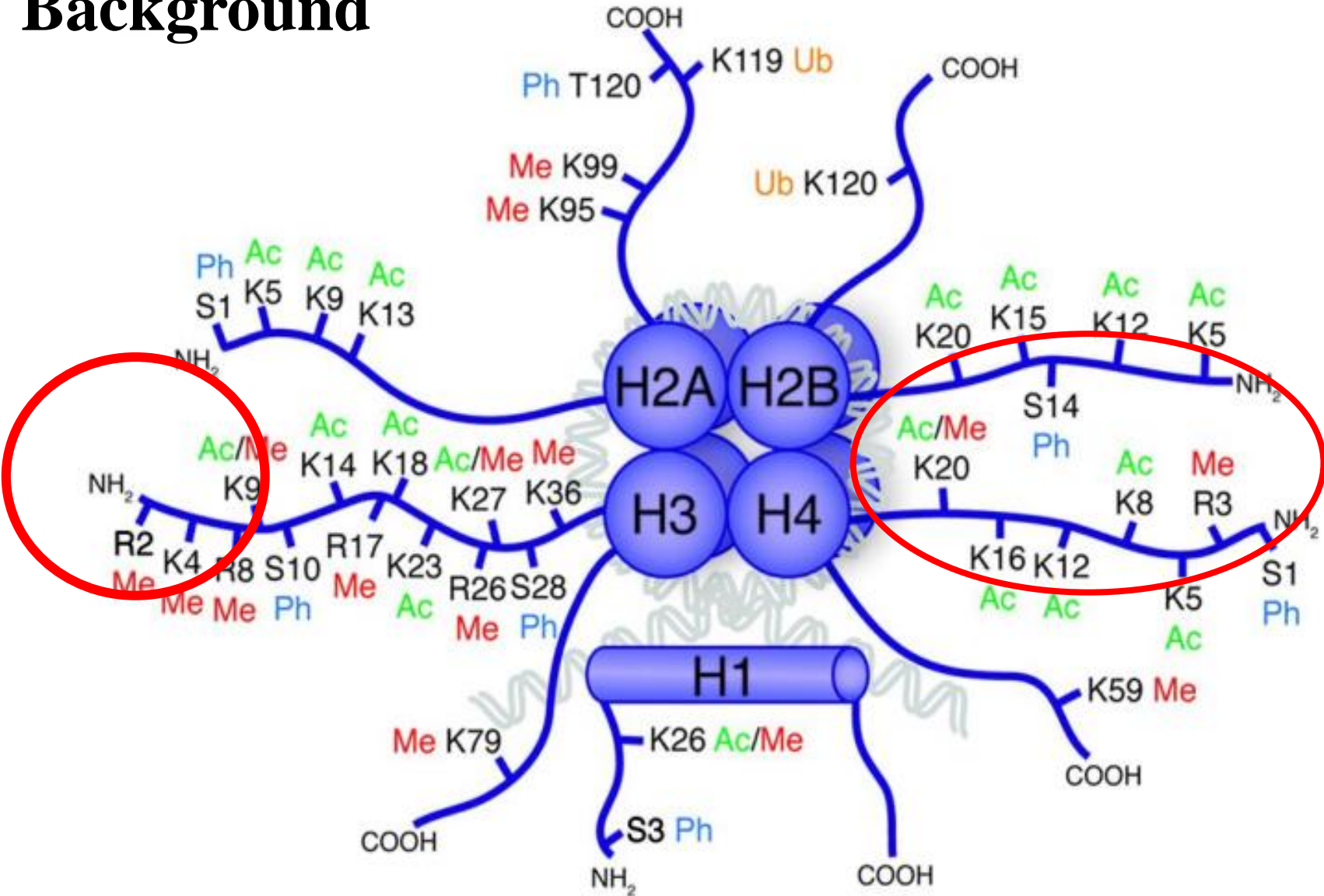




Contents

- **Background**
- Bioinformatics analysis
- Progress in RNF20

Background



核小体组蛋白的修饰成分



Background

组蛋白的翻译后修饰 (Post-Transcriptional Modifications, PTMs) 是表观遗传学领域的研究热点之一。目前已经有文献报道的组蛋白翻译后修饰包括甲基化、泛素化、乙酰化、磷酸化、类泛素化及 ADP 核糖基化等数十种发生在 60 多个不同的氨基酸位点的修饰类型。

(Kouzarides, 2007; Tan et al., 2011)



Background

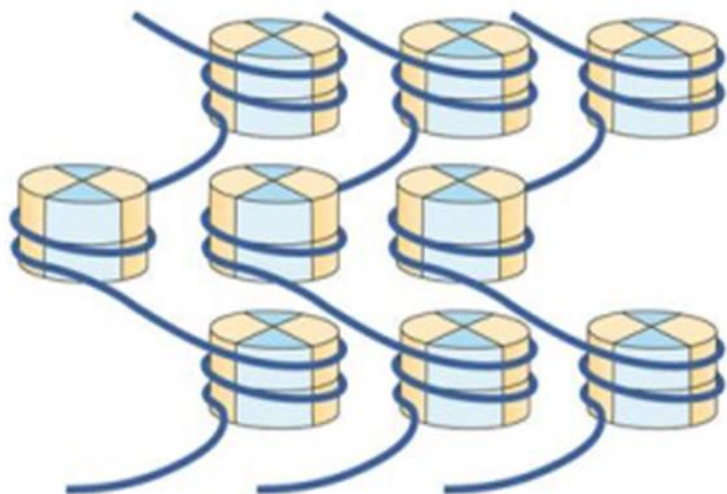
组蛋白修饰调控染色质相关细胞生物学过程一般有两种方式。

- **直接调节染色质的高级结构**，如组蛋白的超乙酰化（如组蛋白H4的第16位Lys的乙酰化，H4K16ac或者甲基化H4K20me3）等修饰可以阻抑高级复杂折叠状态的染色质向相对简单的30nm纤维甚至10nm纤维的疏松结构变化，从而抑制与DNA有关的转录、损伤修复等过程。
- **局部作用于染色体的特定区域疏松（H3K14ac）或者紧凑（H3K9me2）组蛋白和DNA之间的非共价结合**，以利于或阻止DNA结合蛋白的进入，进而调控诸如基因转录或者DNA损伤修复等相关过程。

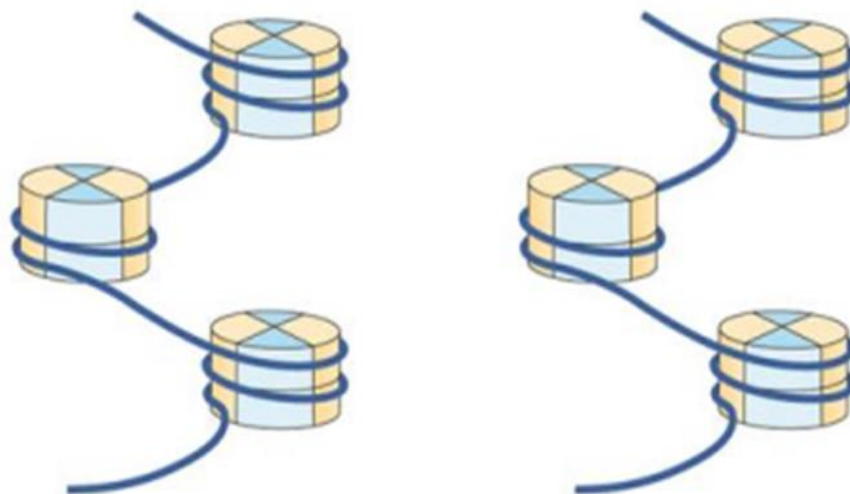


Background

b Inaccessible
Wild type

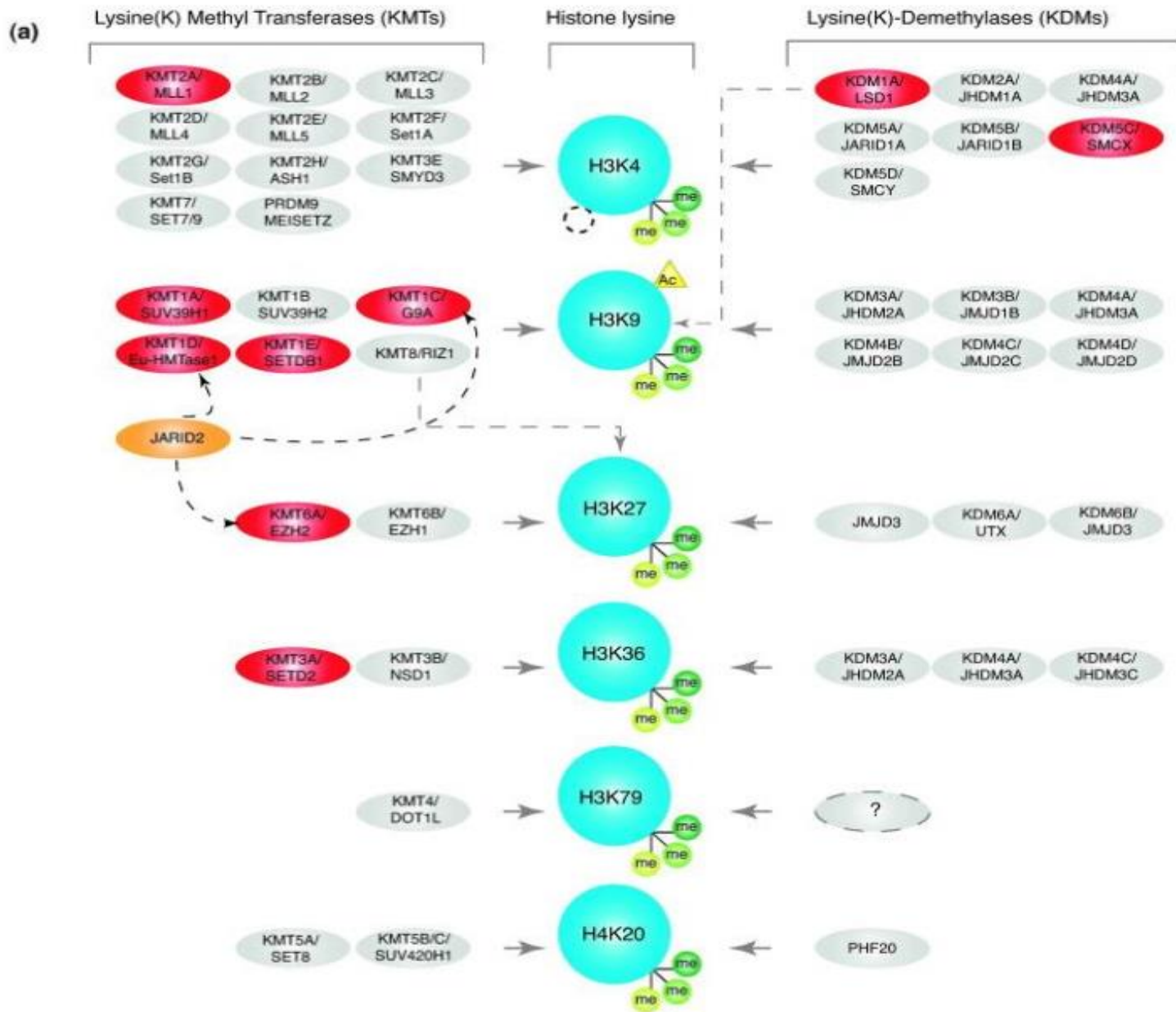


Accessible
H3K56ac

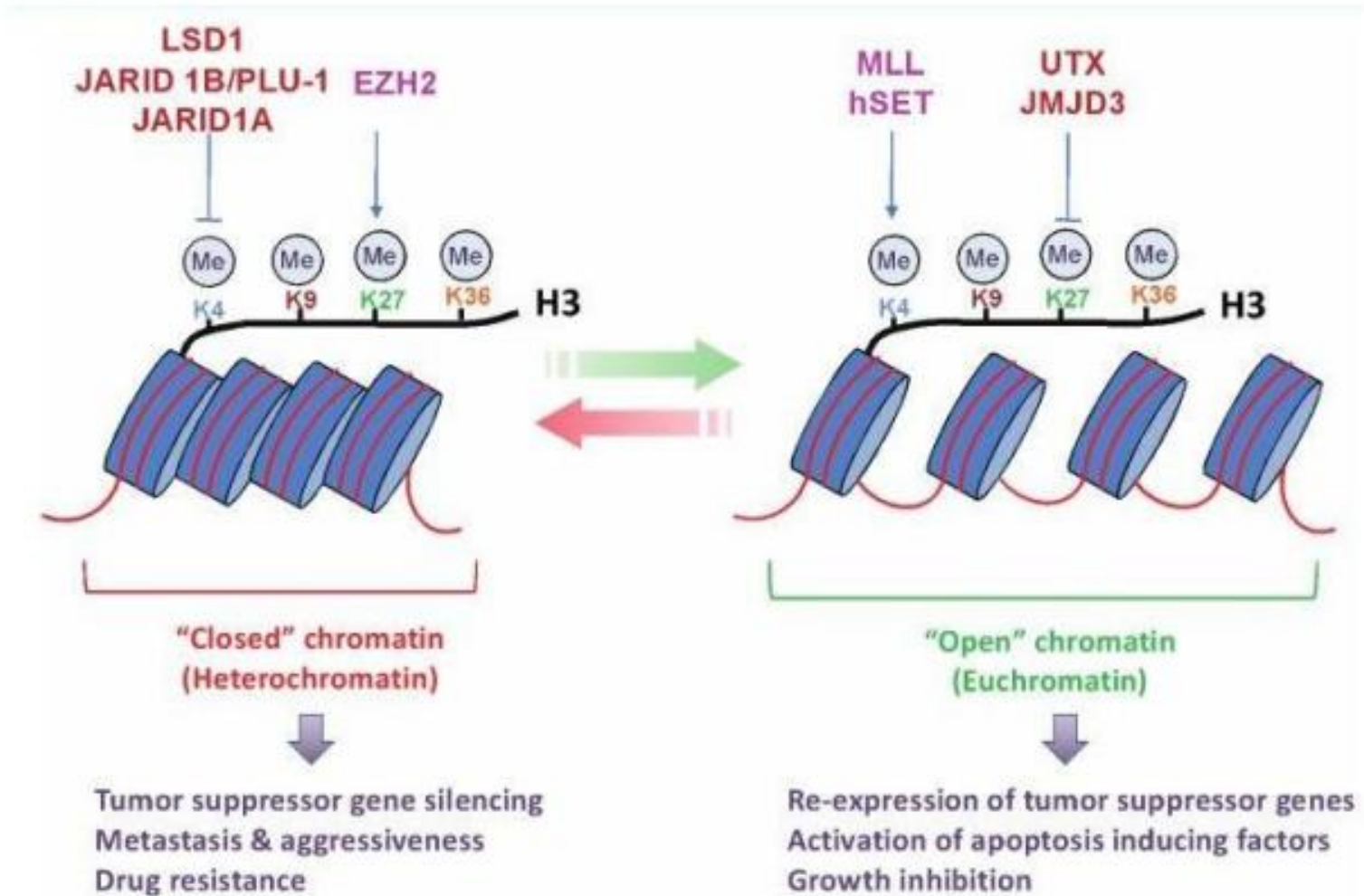


Nature Reviews | Molecular Cell Biology

组蛋白乙酰化修饰影响组蛋白与组蛋白及组蛋白与DNA之间的疏密关系



组蛋白甲基化主要集中在H3K4、H3K9、H3K27、H3K36、H3K79、H4K20等位点



组蛋白甲基化为染色体重塑因子提供一个平台，来调控染色体的松散结构。



Background

组蛋白泛素化

- **组蛋白泛素化**指由泛素激活酶E1 (ubiquitin activating enzyme) 泛素结合酶E2 (ubiquitin conjugating enzyme) 及特定的泛素连接酶E3 (ubiquitin-protein isopeptide ligase) 级联反应将76个氨基酸的泛素分子传递到底物上形成的反应。
- 泛素化修饰包括多泛素化修饰与单泛素化修饰，当底物蛋白被多泛素化修饰后会被转运到26S蛋白酶体中降解；而单泛素化修饰往往会被当做一种信号去发挥某种特定功能。
- 目前，单泛素化修饰就是组蛋白H2A的单泛素化 (H2Aub) 修饰和H2B的单泛素化 (H2Bub) 修饰。

哺乳动物体内，RNF20/RNF40复合体在H2Bub过程中，扮演着E3连接酶的角色。



Contents

- Background
- **Bioinformatics analysis**
- Progress in RNF20



Bioinformatics analysis

NCBI Information

在Gene下检索RNF20 PIG(*sus scrofa*), 包括RNF20基因的信息以及
在其他数据库中的相关链接。

RNF20 ring finger protein 20 [*Sus scrofa* (pig)]

Gene ID: 100154259, updated on 3-Sep-2017

Summary

Gene symbol	RNF20
Gene description	ring finger protein 20
See related	Ensembl:ENSSSCG00000005399
Gene type	protein coding
RefSeq status	MODEL
Organism	Sus scrofa
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae; Sus
Expression	Ubiquitous expression in ovary (RPKM 13.7), lung (RPKM 13.0) and 8 other tissues See more
Orthologs	human mouse all

同时，关联如下网站：

1、 https://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=100154259
(SNP位点数据库)

Gene Model (mRNA alignment) information from genome sequence ↑

Total gene model (contig mRNA transcript):

1

mrna	transcript	protein	mrna orientation	Contig	Contig Label	List SNP
XM_001926594.6	plus strand	XP_001926629.2	forward	NW_018084664.1	Sscrofa11.1	<- currently shown

in gene region
 cSNP
 has frequency
 double hit

gene model (contig mRNA transcript):	Contig Label	Contig	mrna	protein	mrna orientation	transcript	snp count
	Sscrofa11.1	NW_018084664.1	XM_001926594.6	XP_001926629.2	forward	plus strand	38, coding

Region	Chr. position	mRNA pos	dbSNP rs# cluster id	Heterozygosity	Validation	MAF	Allele origin	3D	Clinically Associated	Clinical Significance	Function	dbSNP allele	Protein residue	Codon pos	Amino acid pos	PubMed
	243232985	433	rs340471153	N.D.							synonymous	G	Arg [R]	3	82	
											contig reference	A	Arg [R]	3	82	
	243233165	493	rs319158519	N.D.							synonymous	A	Glu [E]	3	102	
											contig reference	G	Glu [E]	3	102	
	243233249	577	rs328823511	N.D.							synonymous	T	Pro [P]	3	130	

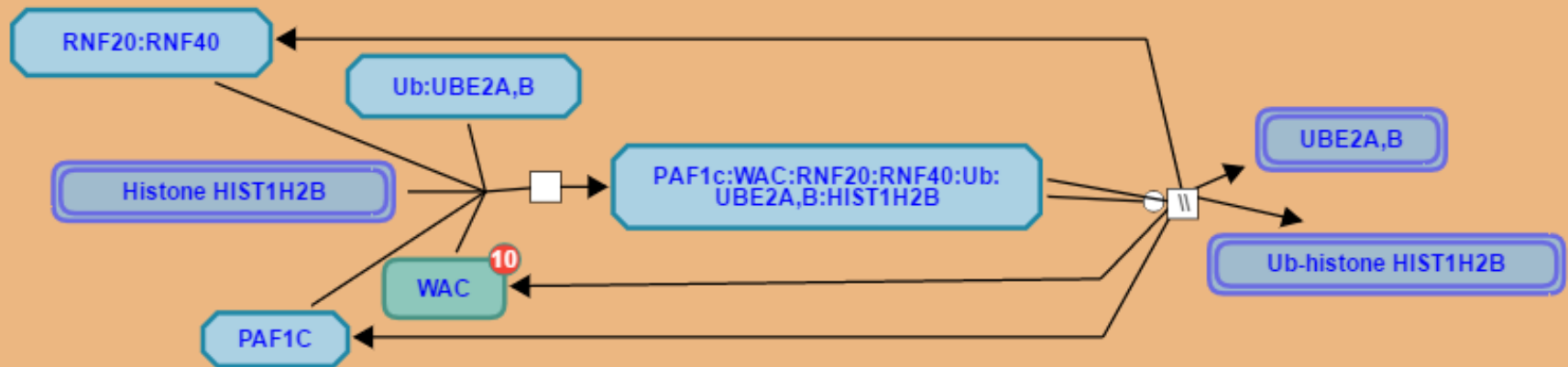
2、 <http://www.reactome.org/content/detail/R-SSC-8866654> (Reactome)

🔍 E3 ubiquitin ligases ubiquitinate target proteins

Stable Identifier	R-SSC-8866654
Type	Pathway
Species	Sus scrofa

Locations in the PathwayBrowser

- 🔍 Metabolism of proteins (Sus scrofa)
 - 🔍 Post-translational protein modification (Sus scrofa)
 - 🔍 Protein ubiquitination (Sus scrofa)
 - 🔍 E3 ubiquitin ligases ubiquitinate target proteins (Sus scrofa)



3. http://asia.ensembl.org/Sus_scrofa/Gene/Summary?g=ENSSSCG00000005399;r=1:243225719-243254557 (Ensembl)

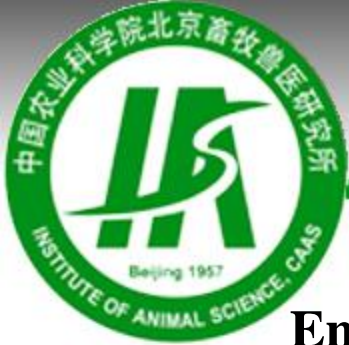
Description ring finger protein 20 [Source:HGNC Symbol;Acc:HGNC:10062]
Synonyms BRE1, BRE1A, FLJ11189, FLJ20382, KAIA2779, hBRE1
Location [Chromosome 1: 243,225,719-243,254,557](#) forward strand.
Sscrofa11.1:CM000812.5

About this gene This gene has 2 transcripts ([splice variants](#)), [100 orthologues](#), [1 paralogue](#) and is a member of [1 Ensembl protein family](#).

Transcripts [Hide transcript table](#)

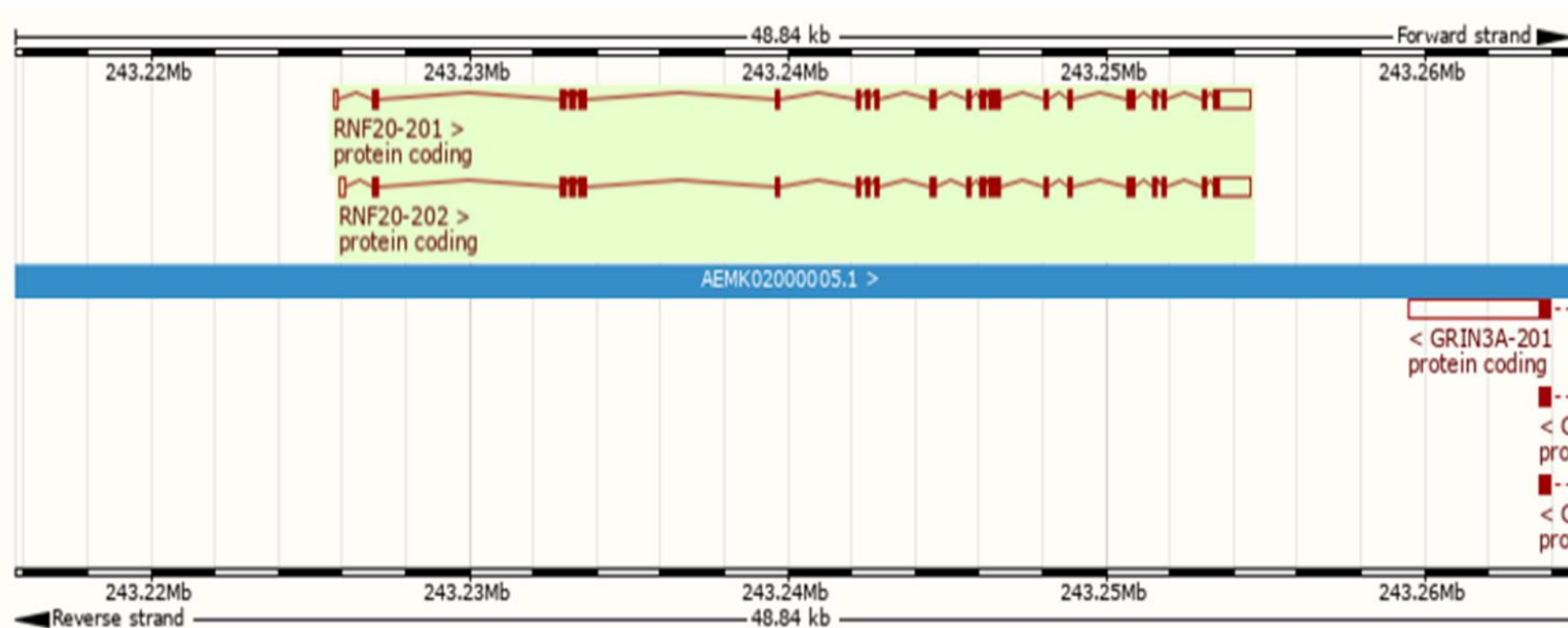
Show/hide columns (1 hidden)		Filter				
Name	Transcript ID	bp	Protein	Biotype	UniProt	Flags
RNF20-202	ENSSSCT000000024274.2	4150	975aa	Protein coding	F1SSB0	APPRIS P1
RNF20-201	ENSSSCT000000005942.3	4117	975aa	Protein coding	F1SSB0	APPRIS P1

猪RNF20基因位于1号染色体上。



Bioinformatics analysis

Ensembl 上寻找RNF20基因信息，发现猪 *RNF20* 基因具有两个转录本，长度分别为**4150bp**与**4117bp**，对应蛋白大小均为**975个氨基酸**。





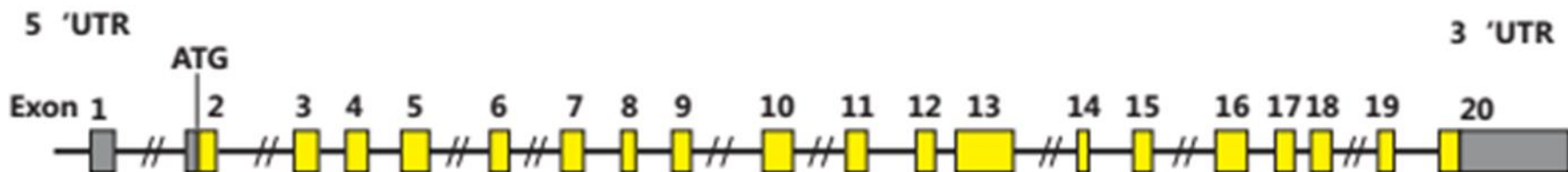
Bioinformatics analysis

启动子预测

<http://www.cbs.dtu.dk/services/Promoter/> Promoter 2.0 Prediction 进行启动子预测。

<http://www.softberry.com/berry.phtml?topic=index&group=programs&subgroup=promoter> Softberry预测

<http://bimas.dcr.t.nih.gov/molbio/proscan>. Promoter SCAN



RNF20 基因结构，20个外显子，其中起始密码子**ATG**位于第二外显子上

Ensembl上下游 启动子预测范围： ATG下游+200bp， 上游 -5kp。

Sequence, 7094 nucleotides		
Position	Score	Likelihood
200	1.106	Highly likely prediction
2800	1.131	Highly likely prediction
4500	1.088	Highly likely prediction
5900	0.650	Marginal prediction

Position Score Likelihood

where 'Position' is a position in the sequence,

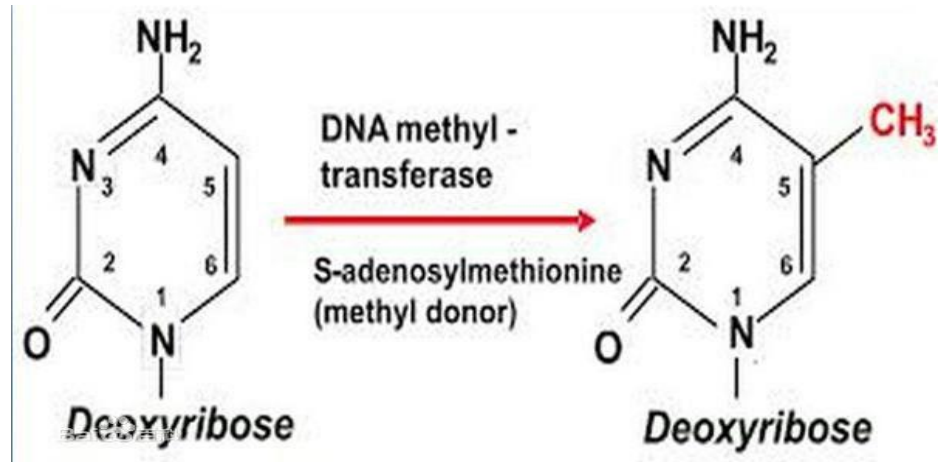
'Score' is the prediction score for a transcription start site occurring within 100 base pairs upstream from that position and

'Likelihood' is a descriptive label associated with that score. The scores are always positive numbers;

CpG岛预测

CpG岛主要位于基因的启动子（promotor）和外显子区域，是富含CpG二核苷酸的一些区域，长度为300—3000bp。

CpG位点经常发生甲基化修饰，而**CpG岛**往往是非甲基化。



<http://www.urogene.org/cgi-bin/methprimer2/MethPrimer.cgi> 预测CpG岛

https://www.ebi.ac.uk/Tools/seqstats/emboss_cpplot/

MP
ethrimer

MethPrimer - Input

MethPrimer allows you to design primers for most bisulfite conversion based PCR primers and to predict CpG islands on an input sequence. It also allow you to search for predesigned primers for human and mouse genes.

- **Design primers on your input sequence:** Paste an ORIGINAL DNA [sequence](#). Input sequence doesn't need virtual bisulfite conversion (e.g. convert 'C' to 'T'). Try this [Sample sequence](#).
- **Search for predesigned primers for human and mouse genes:** Type in the box below gene symbol, RefSeq ID, to search for pre-designed primers for protein coding genes, lncRNAs and miRNAs (e.g., ESR1, NM_000044, mir-1-1, PANDA).
- **CpG island prediction:** You can also use the program to predict CpG islands in a sequence.



Bioinformatics analysis

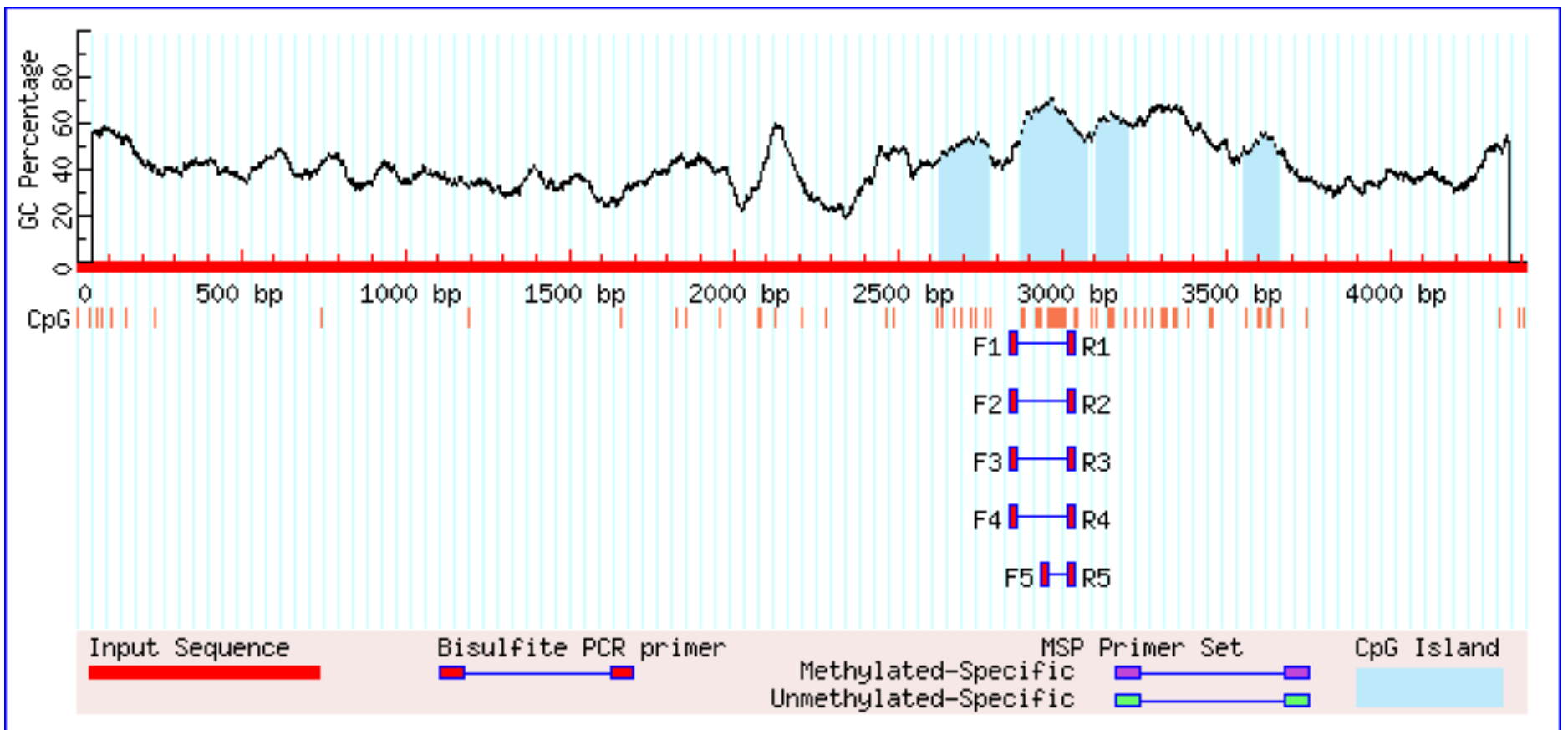
预测CpG岛的三个标准:

- ① GC最低含量
- ② CpG最低出现率
- ③ CpG岛最小长度 (一般 $\geq 200\text{bp}$)

其中, CpG最低出现率计算公式:

$$\text{CpG最低出现率} = \frac{\text{CpG碱基数量} \times \text{长度}}{\text{G碱基数量} \times \text{C碱基数量}}$$

将三种标准组合分析, 以期发现最多的CpG岛, 获得预测CpG岛最佳的参数组合。

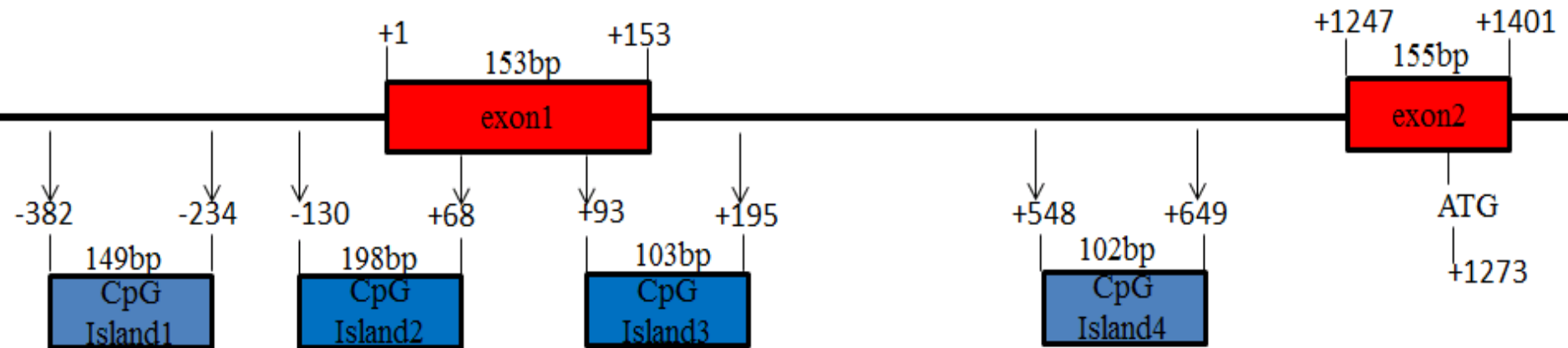


NAME	SIZE(bp)	Star-End
Island 1	149	2627-2775
Island2	198	2879-3076
Island3	103	3101-3203
Island4	102	3556-3657



Bioinformatics analysis

猪RNF20基因CpG岛位置示意图





Contents

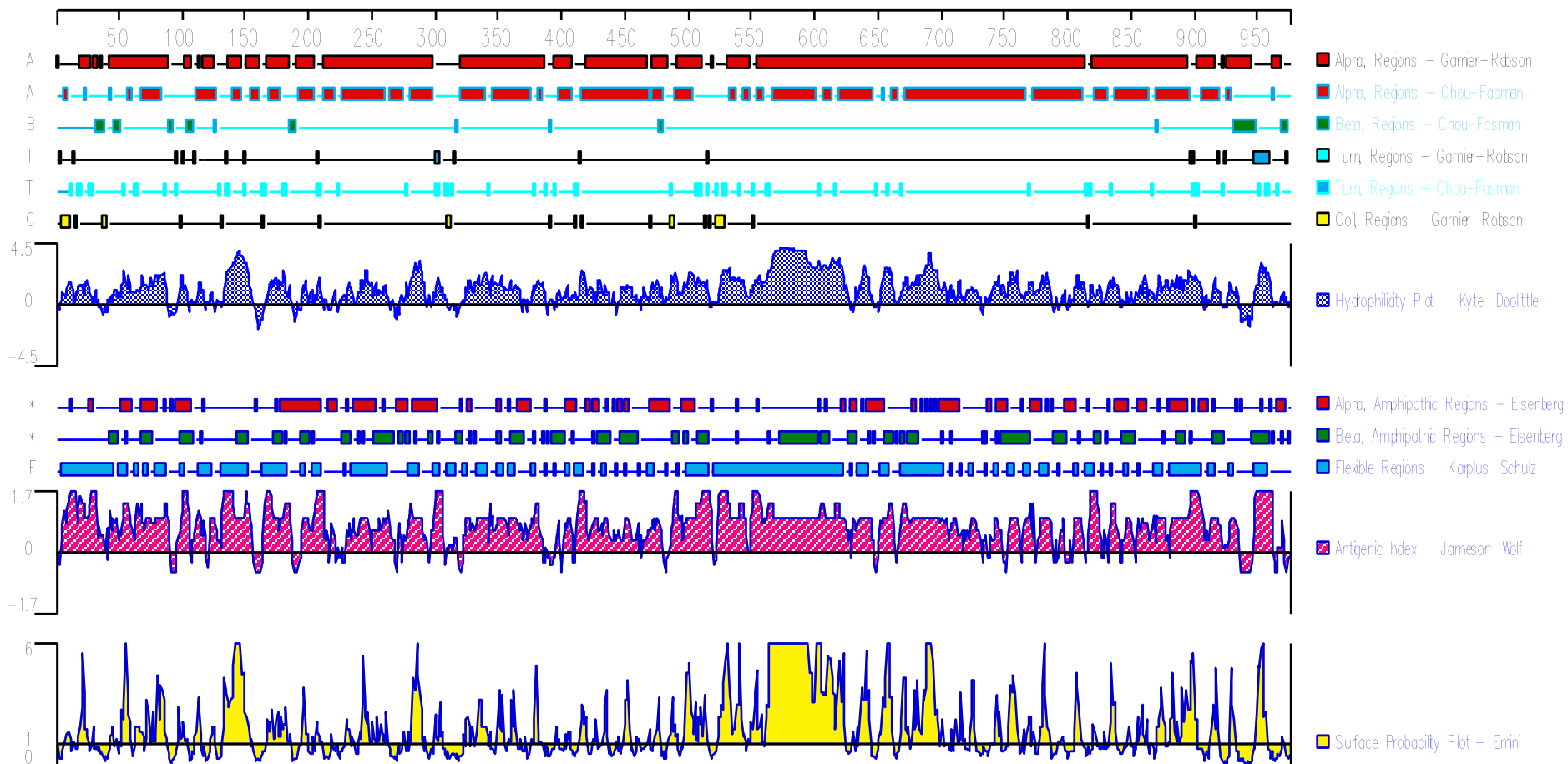
- Background
- Bioinformatics analysis
- **Progress in RNF20**

*RNF20*密码子偏好性分析

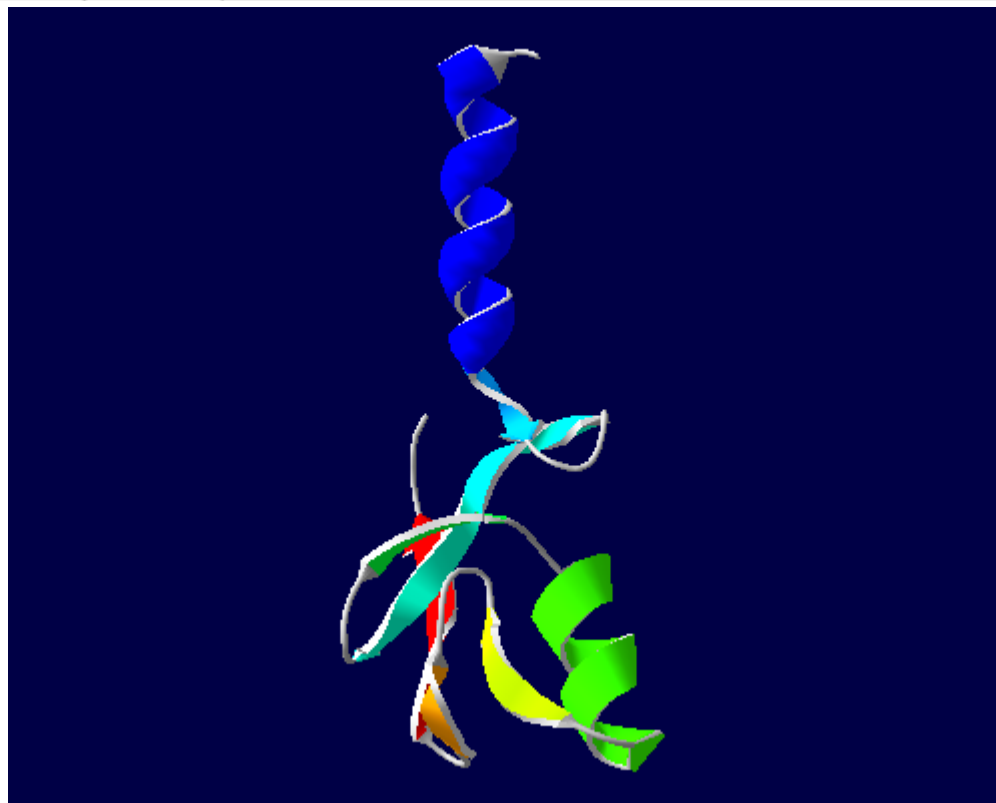
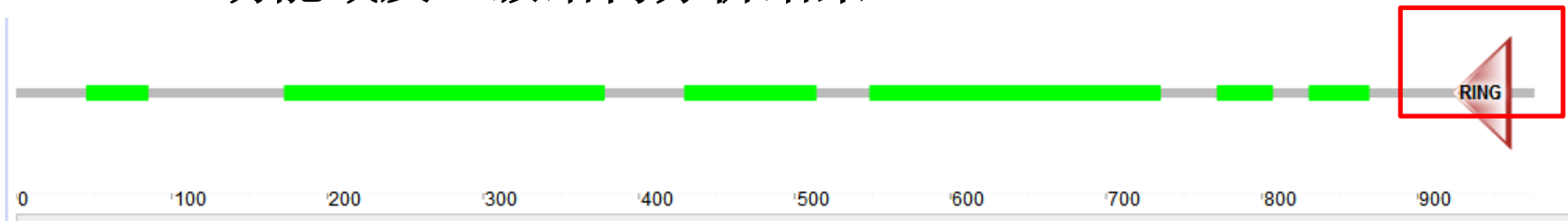
Coding GC	1st letter GC	2nd letter GC	3rd letter GC
47.85%	58.50%	31.15%	53.89%

#Codon	AA	Fraction	Frequency	Number
GCA	A	0.365	23.566	23
GCC	A	0.333	21.516	21
GCG	A	0.032	2.049	2
GCT	A	0.270	17.418	17
TGC	C	0.364	4.098	4
TGT	C	0.636	7.172	7
GAC	D	0.316	18.443	18
GAT	D	0.684	39.959	39
GAA	E	0.386	57.377	56
GAG	E	0.614	91.189	89
TTC	F	0.526	10.246	10
TTT	F	0.474	9.221	9
GGA	G	0.200	6.148	6
GGC	G	0.267	8.197	8
GGG	G	0.267	8.197	8
GGT	G	0.267	8.197	8
CAC	H	0.529	9.221	9
CAT	H	0.471	8.197	8
ATA	I	0.135	5.123	5
ATC	I	0.378	14.344	14
ATT	I	0.486	18.443	18
AAA	K	0.447	47.131	46

RNF20二级结构、亲疏水性、抗原性分析结果



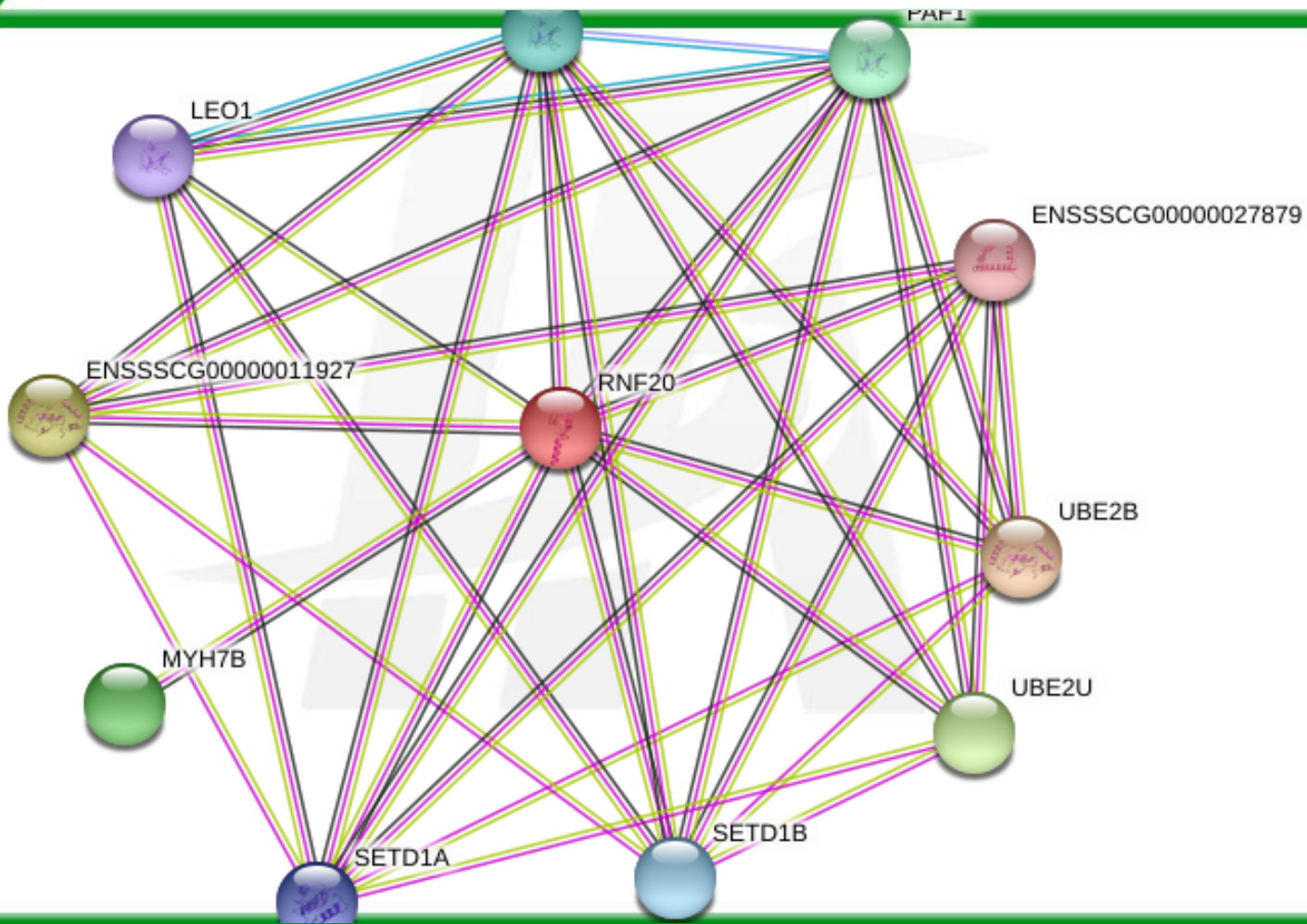
*RNF20*功能域及三级结构分析结果



注：因为没有合适的模板，所以以上预测结果仅预测了*RNF20*蛋白的907-975位蛋白的三级结构（在上图中以红框标出）。



RNF20蛋白互作图



Acknowledgements

◆ Prof. Luo Jingchu

◆ Group 7/8/9 members

◆ All the classmates

Thanks for your attention!