哺乳动物RNA聚合酶复合物的功能和结构分析

小组编号：G06　　报告人：牛迪　　小组成员：李姝含 刘雅迪 张毅

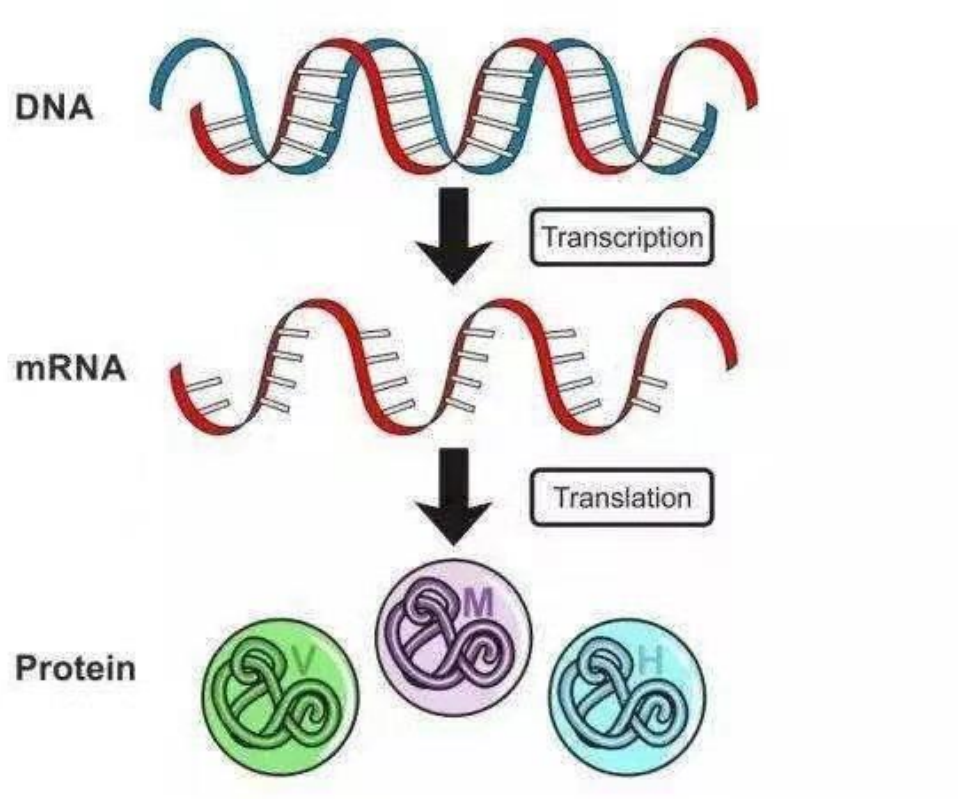
ABC课程期末报告

PART 01

背景介绍

Background

背景介绍

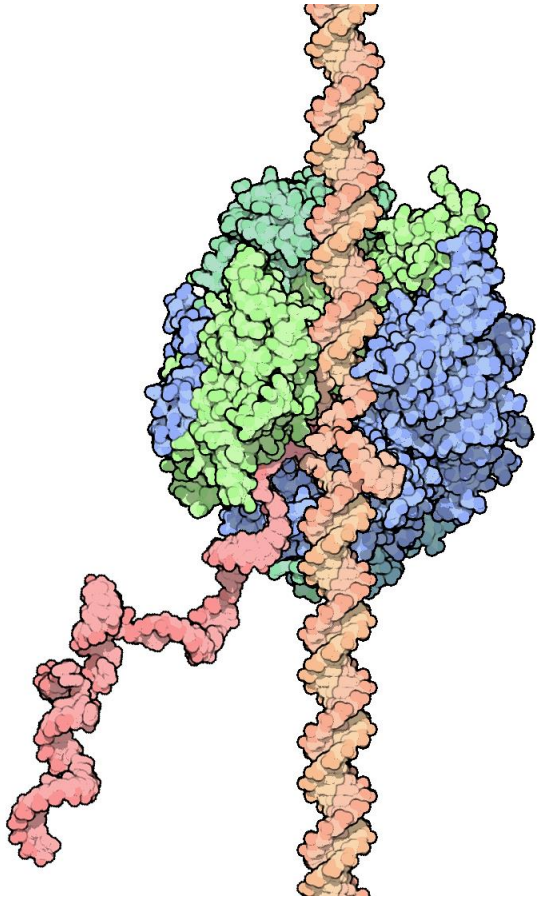


中心法则示意图

生物体遗传信息的传递遵循中心法则，是遗传信息在细胞内的生物大分子间转移的基本法则。遗传信息从DNA传递给RNA，再从RNA传递给蛋白质，即完成遗传信息的转录和翻译的过程。也可以从DNA传递给DNA，即完成DNA的复制过程。这是所有有细胞结构的生物所遵循的法则。

在基因转录过程中，RNA聚合酶起到关键作用。

背景介绍



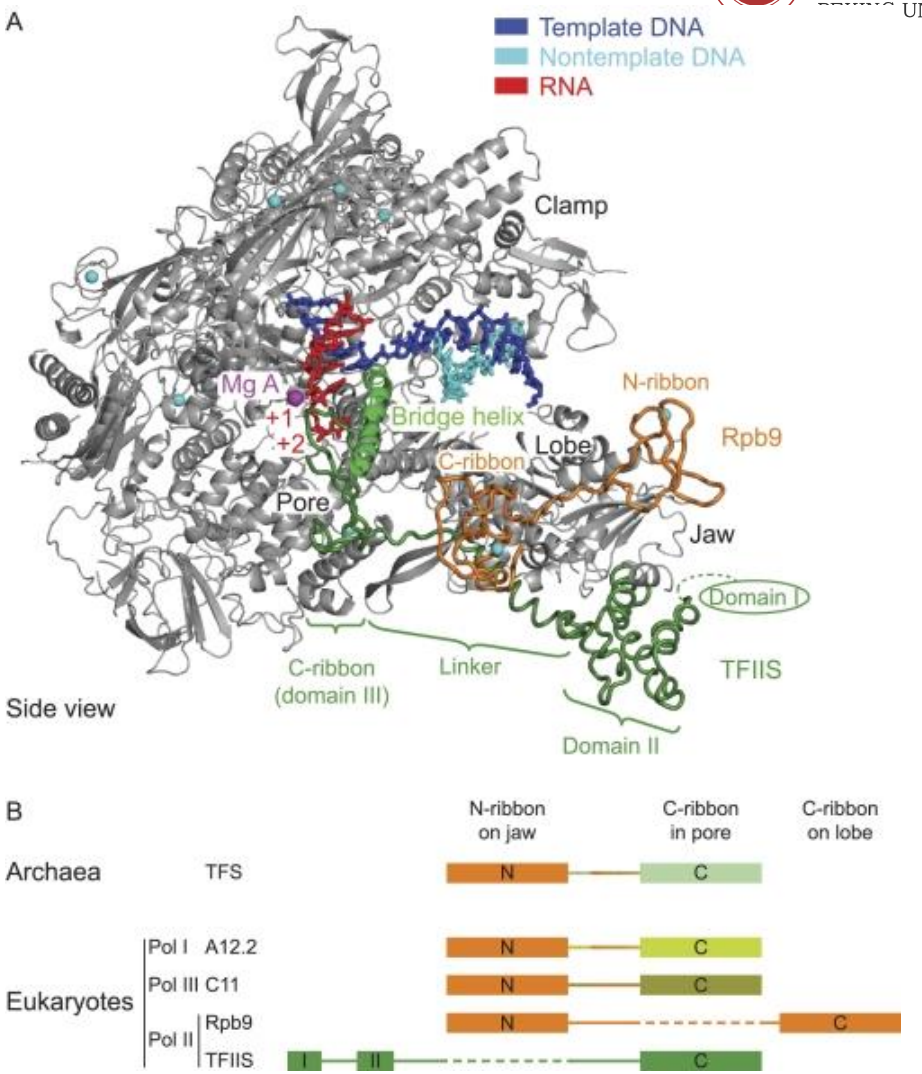
工作中的RNA聚合酶模式图

RNA聚合酶负责生物体RNA合成，是决定细胞命运、维持生命活动的关键复合酶之一。真核生物具有三种 RNA 聚合酶，即 RNA 聚合酶Ⅰ、Ⅱ和Ⅲ。它们分别参与不同类型基因的转录，对于基因的表达起着重要的作用。

背景介绍

三种RNA聚合酶在结构上存在一定的相似性：

- **核心组分**为10个蛋白质亚基形成的活性位点。10 个核心组分构成了类似“蟹钳”的结构，中间的裂缝可以在转录时装载 DNA 分子；另外还形成了两个通道，用于底物核糖核苷酸的进入以及RNA 产物的输出; 活性位点中的高度保守区域使RNA聚合酶能够维持转录泡结构，促进核苷酸的增添及随着模板链移动，起到稳定核苷酸链的作用，从而保证转录的顺利进行。
- **外周组分**为RNA聚合酶的柄部，在转录的起始和终止以及RNA的剪切中发挥重要作用。三种 R NA 聚合酶分别含有两个外周组分，此外， R NA 聚合酶 I 还含有两个外周组分A49 /A34.5， R NA 聚合酶III还有五个外周组分，分别为 C37 /C53 二聚体和 C82 /C34 /C31 三聚体。



Pol II-TFIIS核酸复合物的结构

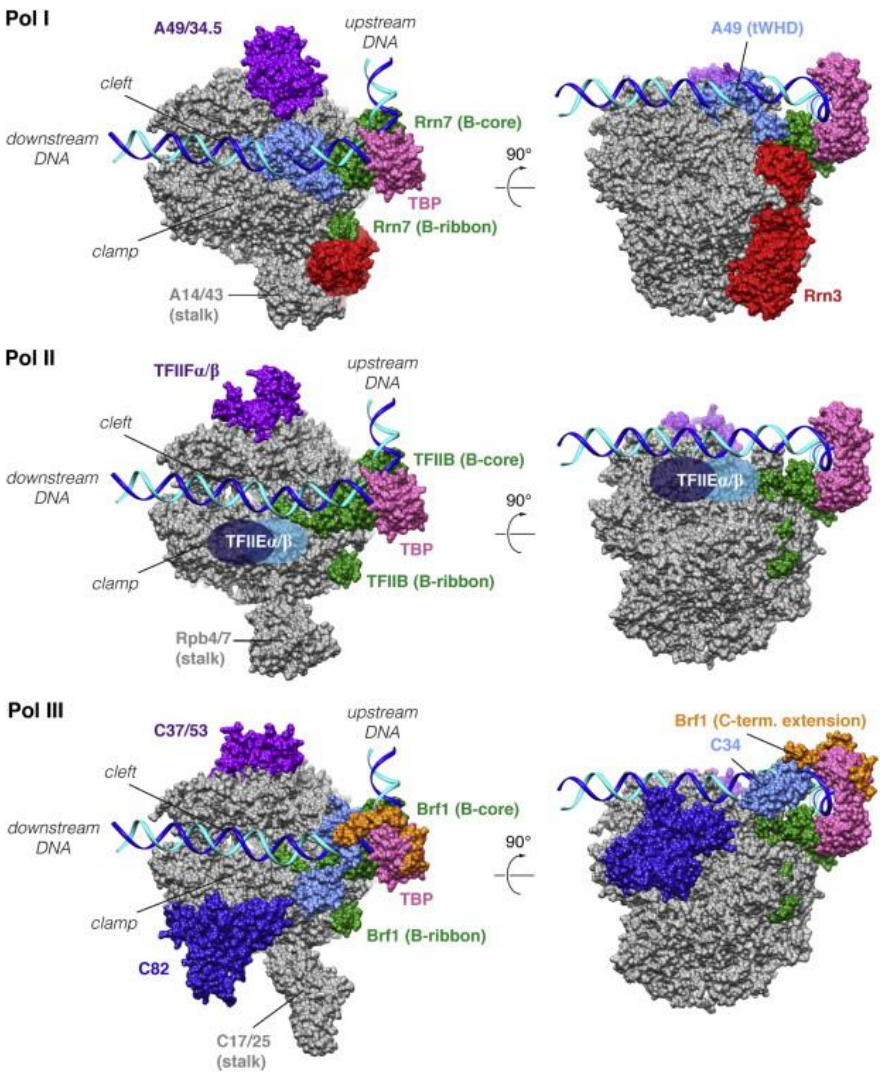
背景介绍

RNA 聚合酶在基因的转录和表达过程中起重要作用，但不同转录的基因有所不同:

- RNA 聚合酶 I 专门转录 rRNA;
- RNA 聚合酶II转录 mRNA 以及一些小核RNA;
- RNA 聚合酶III则负责转录 5SrRNA、tRNA 以及 U6 小核RNA 等

RNA 聚合酶介导的转录过程包括前起始复合物的形成、起始、延伸和终止。真核 RNA 聚合酶不能直接识别启动子，需要形成前起始复合物才能启动转录。

RNA 聚合酶 I 转录因子	TBP、TFIIB 、UAF、CF
RNA 聚合酶II 转录因子	TBP、TFIIB、TFIID、TFIIE、TFIIF 、TFIIH
RNA 聚合酶III 转录因子	TBP、TFIIB、TFIIIB、TFIIIC



转录起始复合物的保守核心区域

背景介绍

研究RNA聚合酶的结构和功能对于揭示真核生物复杂的转录机制具有重大的实际意义。美国斯坦福大学医学院RogerD.Kornberg教授利用酵母作为模型，通过结构生物学、生物化学及酵母遗传学技术阐明了RNAPⅡ复合体结构，揭示了真核生物转录的分子机制，被授予2006年度诺贝尔化学奖。

RNA聚合酶Ⅱ能够合成mRNA及部分 nocoding RNA，是决定基因表达、细胞命运及器官发育等生命活动的关键调控机器，是RNA聚合酶研究的关注热点，故本次报告希望通过生物信息技术的手段，探讨RNA聚合酶Ⅱ的功能和结构。

PART 01

高通量测序

High-throughput sequencing

高通量测序

北京大学 PEKING UNIVERSITY

数据获取来源：
GEO网站

下载了小鼠胚胎干细胞中RNA Pol1 Pol2 Pol3的ChIP-seq测序数据结构

Series GSE145791 Query DataSets for GSE145791

Status	Public on Jun 02, 2020
Title	Genome-wide analyses of chromatin interactions after the loss of Pol I, Pol II and Pol III [ChIP-seq]
Organism	Mus musculus
Experiment type	Genome binding/occupancy profiling by high throughput sequencing
Summary	Genome binding/occupancy profiling by high throughput sequencing | Expression profiling by high throughput sequencing | Other We analyzed Hi-C, HiChIP, Ocean-C, RNA-seq, ATAC-seq, ChIP-seq and 4C-seq data to provide the most comprehensive evidence to date to demonstrate that transcription plays a marginal role in organizing the 3D genome in mammalian cells: 1) degraded Pol I, Pol II and Pol III proteins in mESCs, and showed their loss results in little or no changes of global 3D chromatin structures for the first time; 2) selected RNA polymerases high abundance binding sites-associated interactions and found they still persist after the degradation; 3) generated higher resolution chromatin interaction maps and revealed that transcription inhibition mildly alters small loop domains; 4) identified Pol II bound but CTCF and Cohesin unbound loops and disclosed that they are largely resistant to transcription inhibition Interestingly, we found that Pol II depletion for a longer time significantly affects the chromatin accessibility and Cohesin occupancy, suggesting RNA polymerases are capable of affecting the 3D genome indirectly. So, the direct and indirect effects of transcription inhibition explain the previous confusing effects on the 3D genome. We conclude that Pol I, Pol II, and Pol III loss only mildly alter chromatin interactions in mammalian cells, suggesting the 3D chromatin structures are preestablished and relatively stable.

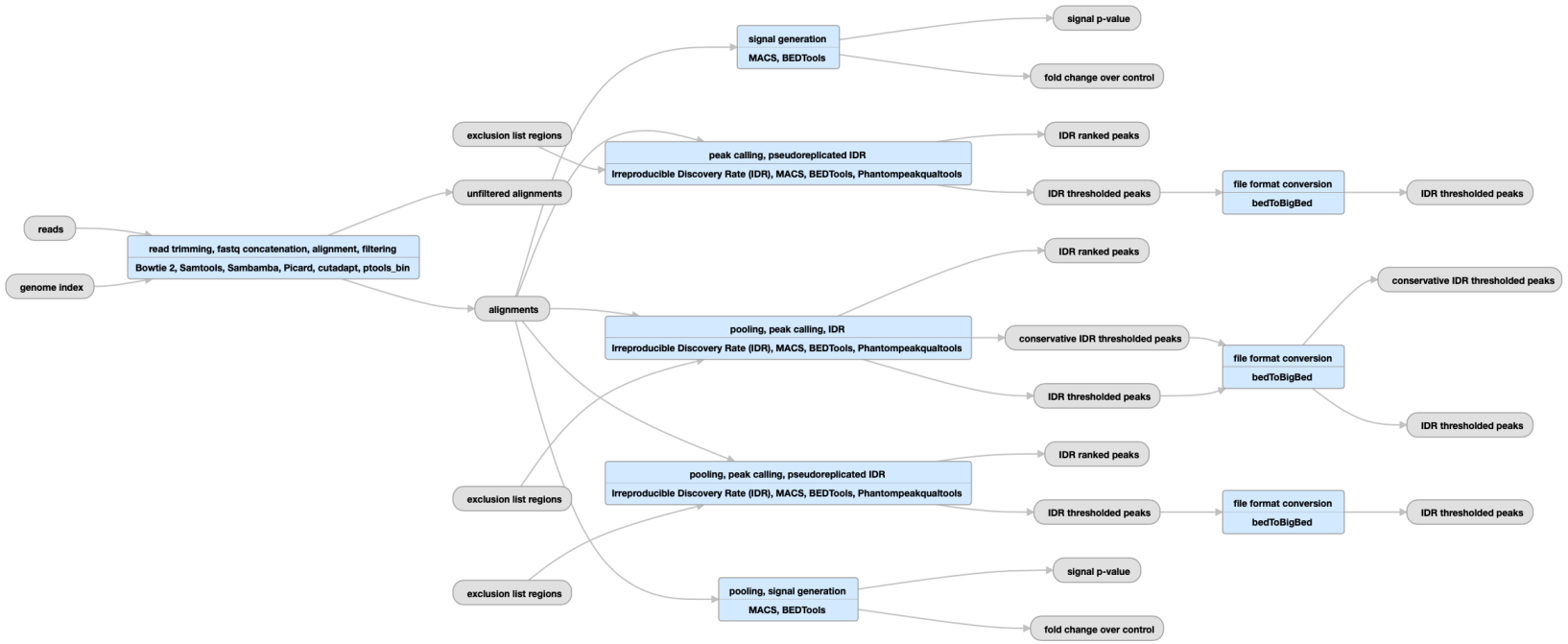
RESEARCH Open Access

Genome-wide analyses of chromatin interactions after the loss of Pol I, Pol II, and Pol III

Yongpeng Jiang[1†], Jie Huang[1†], Kehuan Lun[1†], Boyuan Li[1†], Haonan Zheng[1], Yuanjun Li[1], Rong Zhou[1], Wenjia Duan[1], Chenlu Wang[1], Yuanqing Feng[1], Hong Yao[1], Cheng Li[2] and Xiong Ji[1*]

北京大学 PEKING UNIVERSITY

ChIP-seq分析流程



GSM4333012: GFP_Pol1_ChIP_rep1

GSM4333014: GFP_Pol2_ChIP_rep1

GSM4333016: GFP_Pol3_ChIP_rep1

GSM4333018: GFP_Input_ChIP_rep1

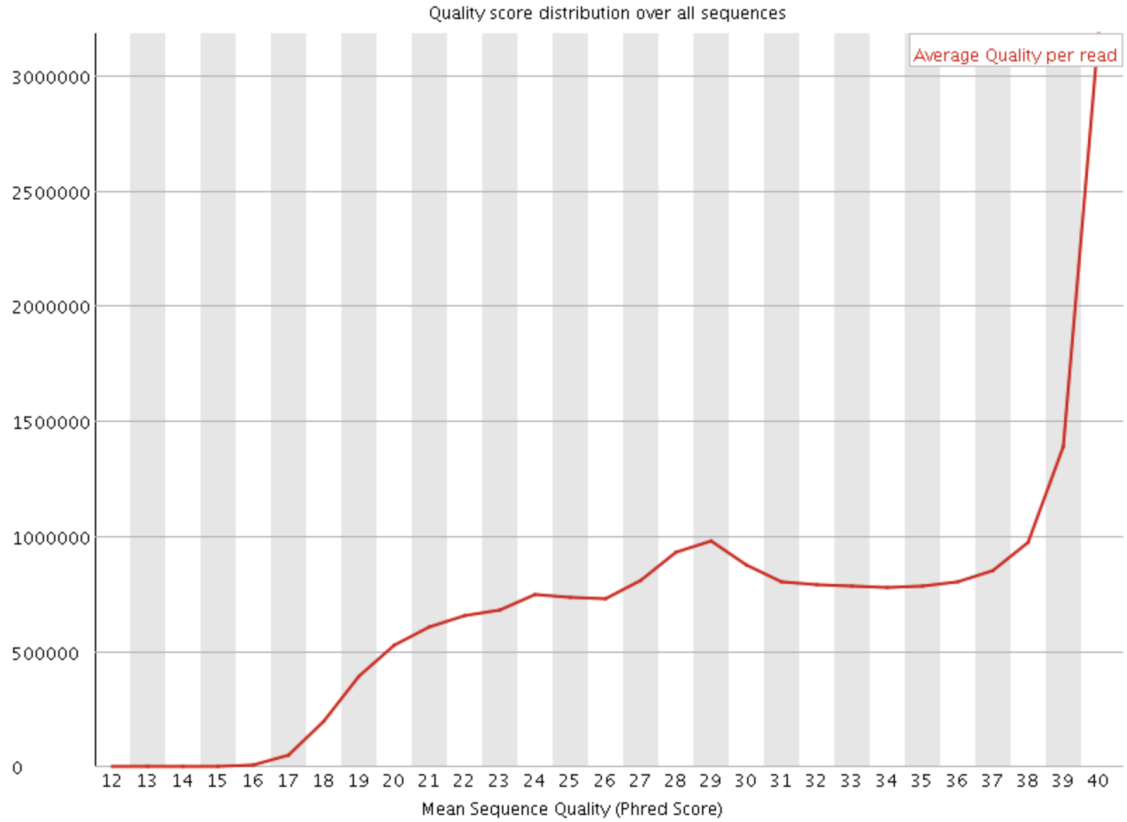
北京大学 PEKING UNIVERSITY

Pol2测序数据的质量分析

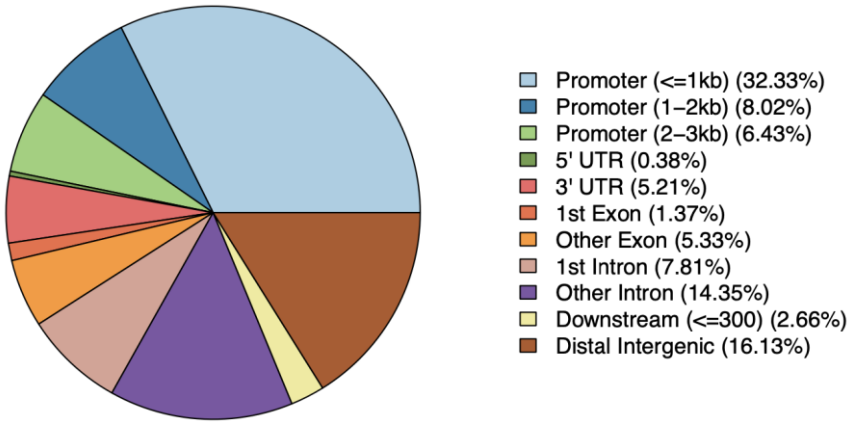
Basic Statistics

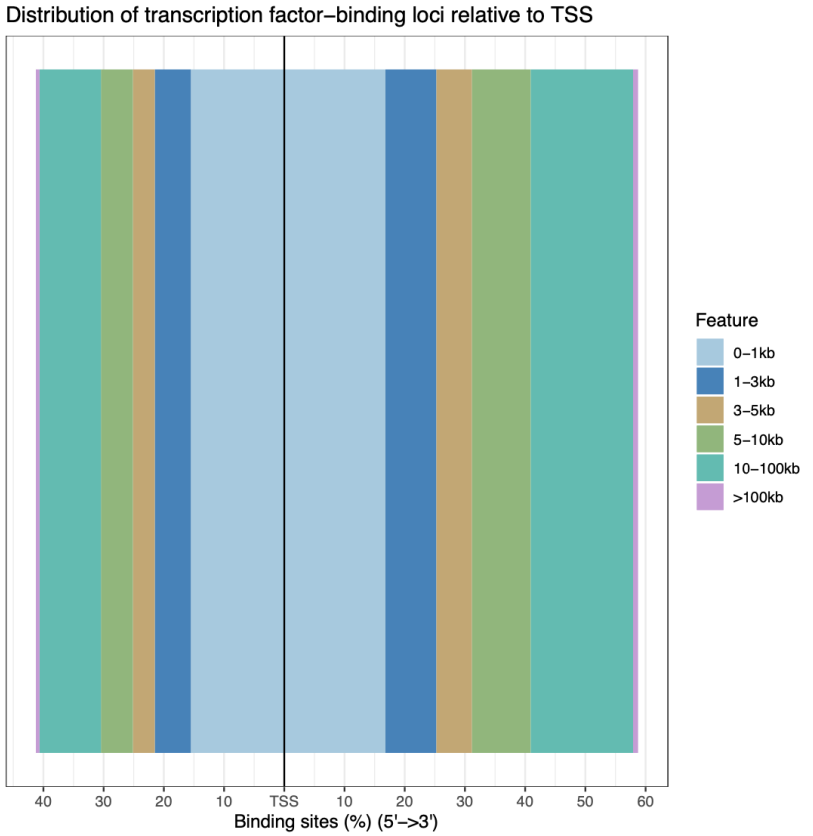
Measure	Value
Filename	20190715_ChIP-Pol2_untreated_input_FKDL190751184-1a-9_1.fq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	20100186
Sequences flagged as poor quality	0
Sequence length	150
%GC	55

Per sequence quality scores



Pol2在基因组上的分布信息





北京大学 PEKING UNIVERSITY

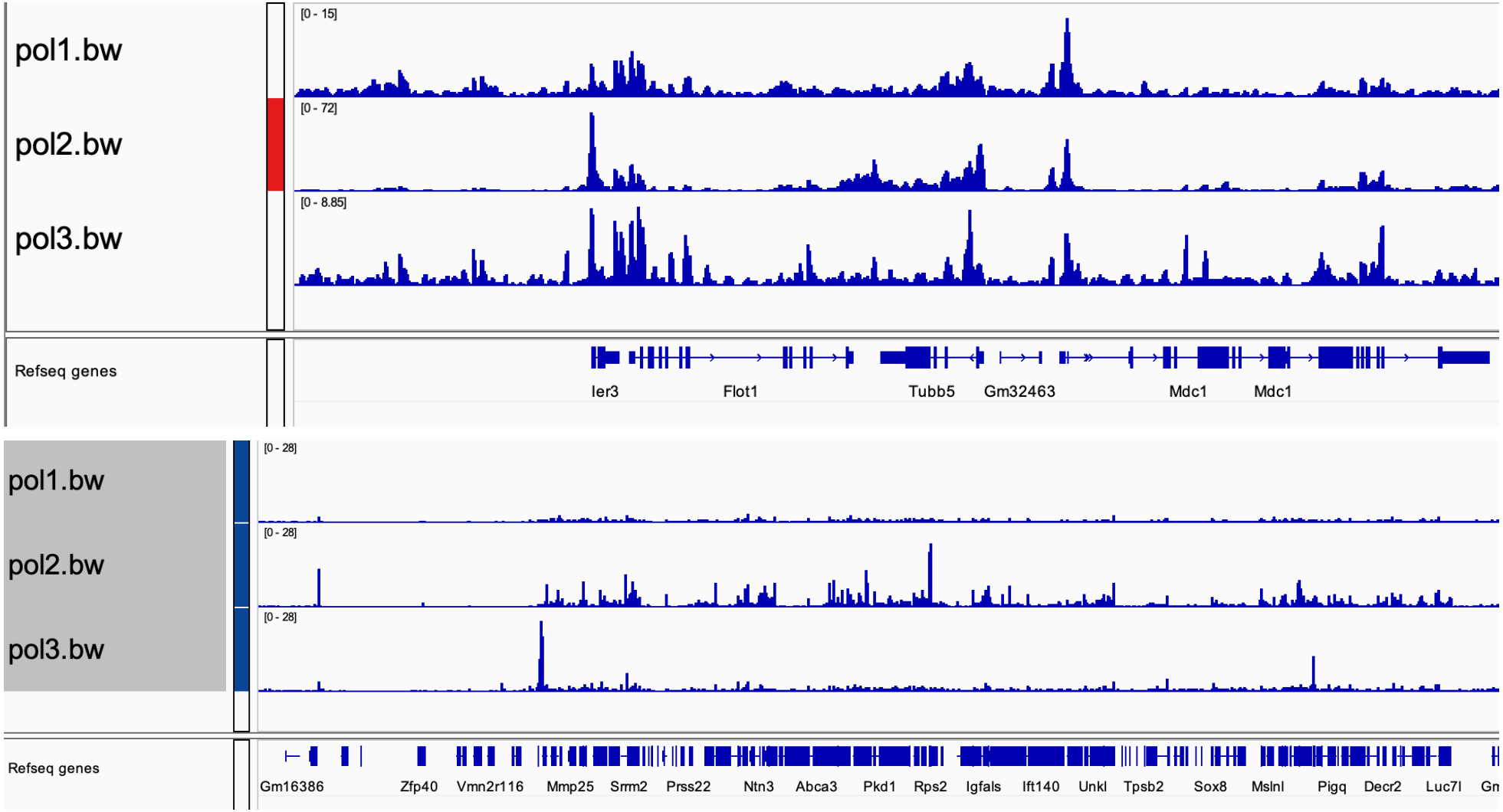
Pol1;Pol2;Pol3在基因组上的结合位点富集的motif

RNA-POL2

Rank	Motif	P-value	log P-pvalue	% of Targets	% of Background	STD(Bg STD)	Best Match/Details
1		1e-184	-4.241e+02	24.67%	20.29%	54.0bp (58.3bp)	Maz(Zf)/HepG2-Maz-ChIP-Seq(GSE31477)/Homer(0.898) More Information | Similar Motifs Found
2		1e-67	-1.557e+02	29.47%	26.61%	55.6bp (57.8bp)	Elk1(ETS)/Hela-Elk1-ChIP-Seq(GSE31477)/Homer(0.695) More Information | Similar Motifs Found
3		1e-55	-1.288e+02	23.98%	21.56%	55.9bp (57.7bp)	MZF1/MA0056.2/Jaspar(0.735) More Information | Similar Motifs Found
4		1e-40	-9.214e+01	17.33%	15.54%	56.2bp (56.8bp)	THRb(NR)/Liver-NR1A2-ChIP-Seq(GSE52613)/Homer(0.731) More Information | Similar Motifs Found
5		1e-37	-8.585e+01	12.38%	10.89%	56.1bp (61.4bp)	FOXA1(Forkhead)/MCF7-FOXA1-ChIP-Seq(GSE26831)/Homer(0.694) More Information | Similar Motifs Found
6		1e-35	-8.224e+01	22.59%	20.71%	56.2bp (58.7bp)	Zfx/MA0146.2/Jaspar(0.685) More Information | Similar Motifs Found
7		1e-35	-8.174e+01	15.85%	14.23%	55.3bp (60.0bp)	POL012.1_TATA-Box/Jaspar(0.812) More Information | Similar Motifs Found
8		1e-34	-8.041e+01	9.84%	8.54%	56.7bp (58.2bp)	MYF5/MA1641.1/Jaspar(0.731) More Information | Similar Motifs Found
9		1e-34	-7.865e+01	16.17%	14.56%	55.6bp (56.7bp)	ISL2/MA0914.1/Jaspar(0.644) More Information | Similar Motifs Found
10		1e-32	-7.426e+01	8.34%	7.19%	56.2bp (57.3bp)	YY2/MA0748.2/Jaspar(0.886) More Information | Similar Motifs Found

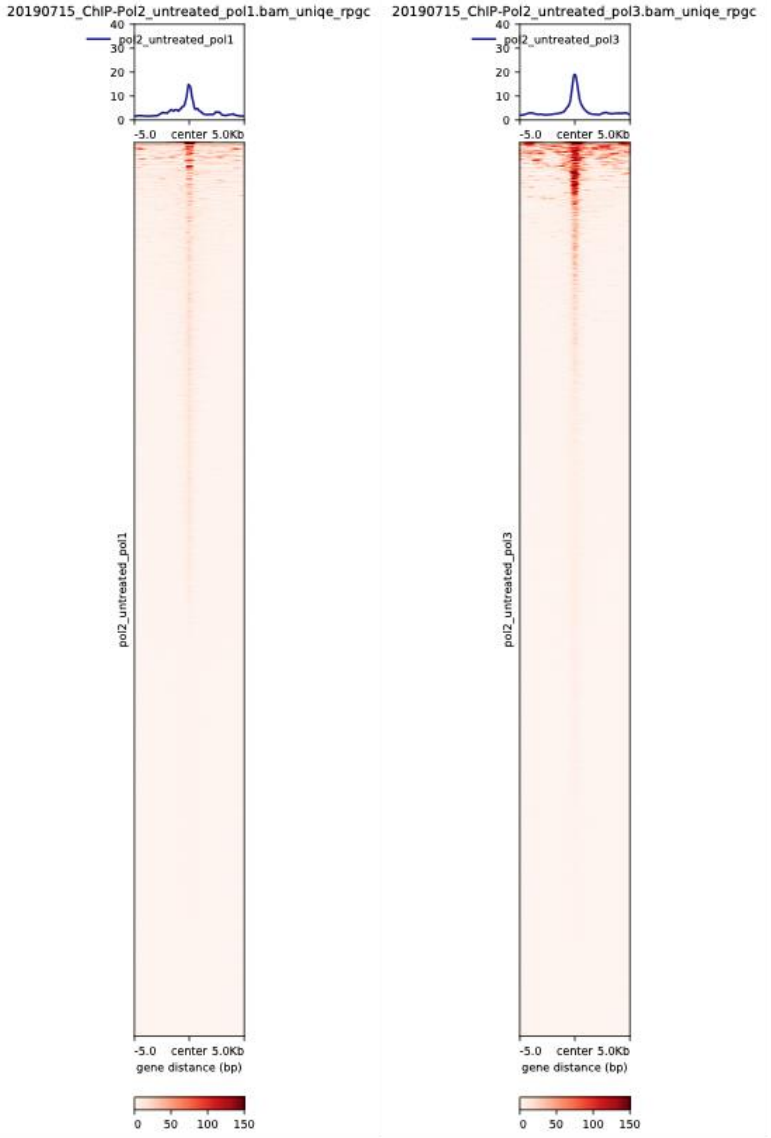
北京大学 PEKING UNIVERSITY

Pol1;Pol2;Pol3在基因组上的共定位



高通量测序

Pol1;Pol2;Pol3在基因组上的共定位



PART 03

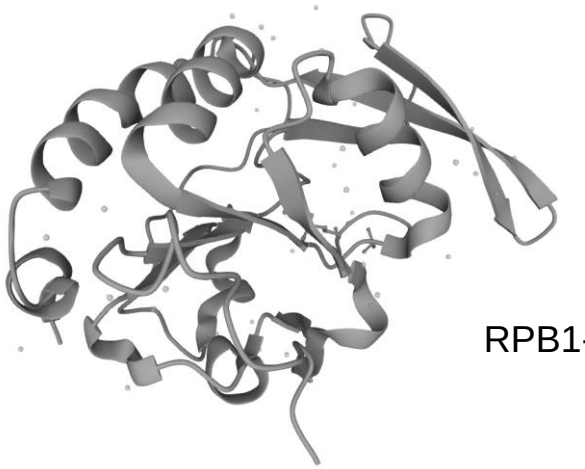
同源性比较

Homology Comparison

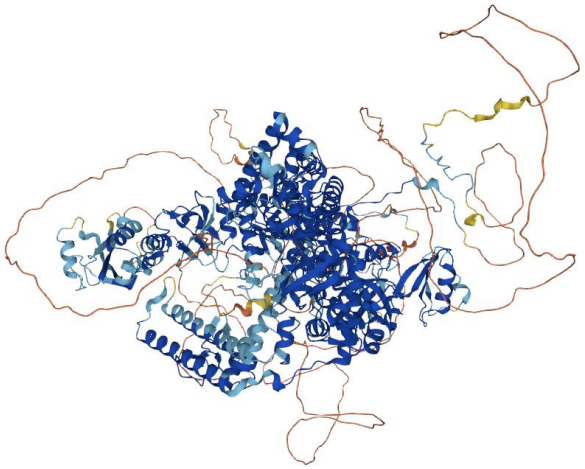
RNA polymerase II

- RNAP II是由12个亚基组成的550 kDa复合物，是研究最多的RNA聚合酶类型。
- 真核核心RNA聚合酶II首次通过转录分析纯化纯化后的酶通常有10-12个亚基(12个在人类和酵母中)。
- RPB1是RNA聚合酶II的最大亚基。它包含一个羧基末端结构域(CTD)，由52个七肽重复序列(YSPTSPS)组成。
- RPB3参与RNA聚合酶II组装亚基合成后，RPB2和RPB3出现一个亚复合物，该复合物随后与RPB1相互作用。

同源性比较



RPB1-HUMAN



RPB1-MOUSE

Model Confidence:

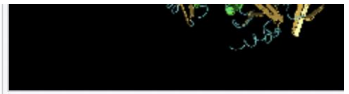
- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.

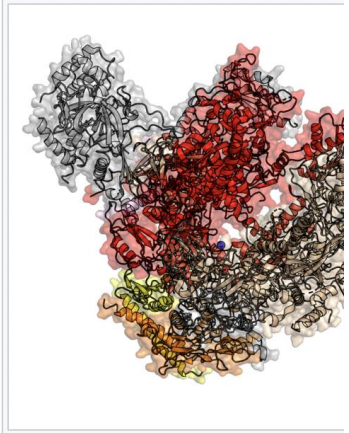
Subunits [edit]

The eukaryotic core RNA polymerase II was first purified using transcription assays.[9] The purified enzyme has typically 10–12 subunits (12 in humans and yeast) and is incapable of specific promoter recognition.[10] Many subunit–subunit interactions are known.[11]

- **DNA-directed RNA polymerase II subunit RPB1** — an enzyme that in humans is encoded by the POLR2A gene and in yeast is encoded by RPO21. RPB1 is the largest subunit of RNA polymerase II. It contains a carboxy terminal domain (CTD) composed of up to 52 heptapeptide repeats (YSPTSPS) that are essential for polymerase activity.[12] The CTD was first discovered in the laboratory of C.J. Ingles at the University of Toronto and by JL Corden at Johns Hopkins University. In combination with several other polymerase subunits, the RPB1 subunit forms the DNA binding domain of the polymerase, a groove in which the DNA template is transcribed into RNA.[13] It strongly interacts with RPB8.[11]
- **RPB2** (POLR2B) — the second-largest subunit that in combination with at least two other polymerase subunits forms a structure within the polymerase that maintains contact in the active site of the enzyme between the DNA template and the newly synthesized RNA.[14]
- **RPB3** (POLR2C) — the third-largest subunit. Exists as a heterodimer with another polymerase subunit, POLR2J forming a core subassembly. RPB3 strongly interacts with RPB1–5, 7, 10–12.[11]
- **RNA polymerase II subunit B4 (RPB4)** — encoded by the POLR2D gene[15] is the fourth-largest subunit and may have a stress protective role.
- **RPB5** — In humans is encoded by the POLR2E gene. Two molecules of this subunit are present in each RNA polymerase II.[16] RPB5 strongly interacts with RPB1, RPB3, and RPB6.[11]
- **RPB6 (POLR2F)** — forms a structure with at least two other subunits that stabilizes the transcribing polymerase on the DNA template.[17]
- **RPB7** — encoded by POLR2G and may play a role in regulating polymerase function.[18] RPB7 interacts strongly with RPB1 and RPB5.[11]
- **RPB8 (POLR2H)** — interacts with subunits RPB1–3, 5, and 7.[11]
- **RPB9** — The groove in which the DNA template is transcribed into RNA is composed of RPB9 (POLR2I) and RPB1.
- **RPB10** — the product of gene POLR2L. It interacts with RPB1–3 and 5, and strongly with RPB3.[11]
- **RPB11** — the RPB11 subunit is itself composed of three subunits in humans: POLR2J (RPB11–a), POLR2J2 (RPB11–b), and POLR2J3[19] (RPB11–c).
- **RPB12** — Also interacts with RPB3 is RPB12 (POLR2K).[11]



RNA polymerase II of Saccharomyces cer
consisting of all 12 subunits.[4]



Eukaryotic RNA–polymerase II from *Saccharomyces cerevisiae*, PDB ID.[8] Su
colored: RPB3 — orange , RPB11 — yellow
— wheat, RPB1 — red, RPB6 — pink, the
subunits are colored gray.

北京大学 PEKING UNIVERSITY

人源-鼠源RNA polymerase II亚基的核酸序列比对

```
# Program: needle
# Rundate: Sat 15 Jan 2022 03:42:26
# Commandline: needle
#    -auto
#    -stdout
#    -asequence emboss_needle-I20220115-034223-0465-95536483-p1m.asequence
#    -bsequence emboss_needle-I20220115-034223-0465-95536483-p1m.bupfile
#    -datafile EDNAFULL
#    -gapopen 10.0
#    -gapextend 0.5
#    -endopen 10.0
#    -endextend 0.5
#    -aformat3 pair
#    -snucleotide1
#    -snucleotide2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EAW90181.1
# 2: EAW90181.1
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 5913
# Identity:    5913/5913 (100.0%)
# Similarity:  5913/5913 (100.0%)
# Gaps:           0/5913 ( 0.0%)
# Score: 29565.0
#
#
#=======================================

EAW90181.1         1 ATGCACGGGGGTGGCCCCCCCTCGGGGGACAGCGCATGCCCGCTGCGCAC     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EAW90181.1         1 ATGCACGGGGGTGGCCCCCCCTCGGGGGACAGCGCATGCCCGCTGCGCAC     50

EAW90181.1        51 CATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGATGAACTGAAGCGAA    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EAW90181.1        51 CATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGATGAACTGAAGCGAA    100

EAW90181.1       101 TGTCTGTGACGGAGGGTGGCATCAAATACCCAGAGACGACTGAGGGAGGC    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
```

北京大学 PEKING UNIVERSITY

人源-鼠源RNA polymerase II亚基的蛋白序列比对

```
# Program: needle
# Rundate: Sat 15 Jan 2022 03:27:25
# Commandline: needle
#    -auto
#    -stdout
#    -asequence emboss_needle-I20220115-033830-0357-93585740-p1m.asequence
#    -bsequence emboss_needle-I20220115-033830-0357-93585740-p1m.bsequence
#    -datafile EBLOSUM62
#    -gapopen 10.0
#    -gapextend 0.5
#    -endopen 10.0
#    -endextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: RPB1_HUMAN
# 2: RPB1_MOUSE
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1970
# Identity:    1968/1970 (99.9%)
# Similarity:  1968/1970 (99.9%)
# Gaps:           0/1970 ( 0.0%)
# Score: 10404.0
#
#
#=======================================

RPB1_HUMAN         1 MHGGGPPSGDSACPLRTIKRVQFGVLSPDELKRMSVTEGGIKYPETTEGG     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
RPB1_MOUSE         1 MHGGGPPSGDSACPLRTIKRVQFGVLSPDELKRMSVTEGGIKYPETTEGG     50

RPB1_HUMAN        51 RPKLGGLMDPRQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFL    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
RPB1_MOUSE        51 RPKLGGLMDPRQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFL    100

RPB1_HUMAN       101 VKTMKVLRCVCFFCSKLLVDSNNPKIKDILAKSKGQPKKRLTHVYDLCKG    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
```

北京大学 PEKING UNIVERSITY

人源-鼠源RNA polymerase II亚基的系统树

Results for job tcoffee-I20220115-034411-0577-83989174-p1m

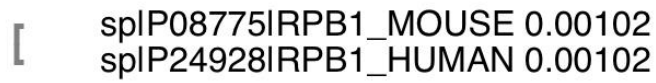
Alignments | Result Summary | Phylogenetic Tree | Results Viewers | Submission Details

Download Phylogenetic Tree Data

Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Branch length: ○ Cladogram ◉ Real



Tree Data

```
(sp|P08775|RPB1_MOUSE:0.00102,sp|P24928|RPB1_HUMAN:0.00102);
```

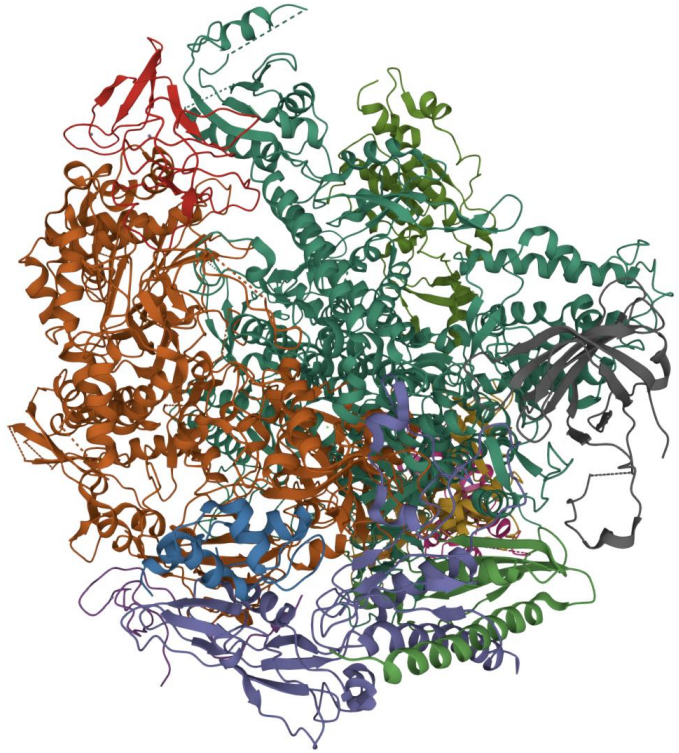
PART 04

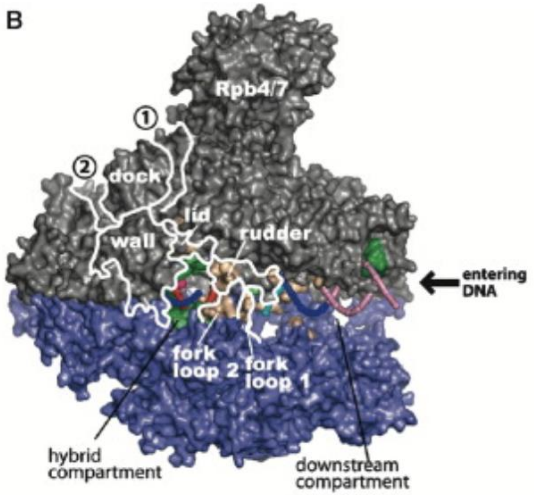
蛋白质构象

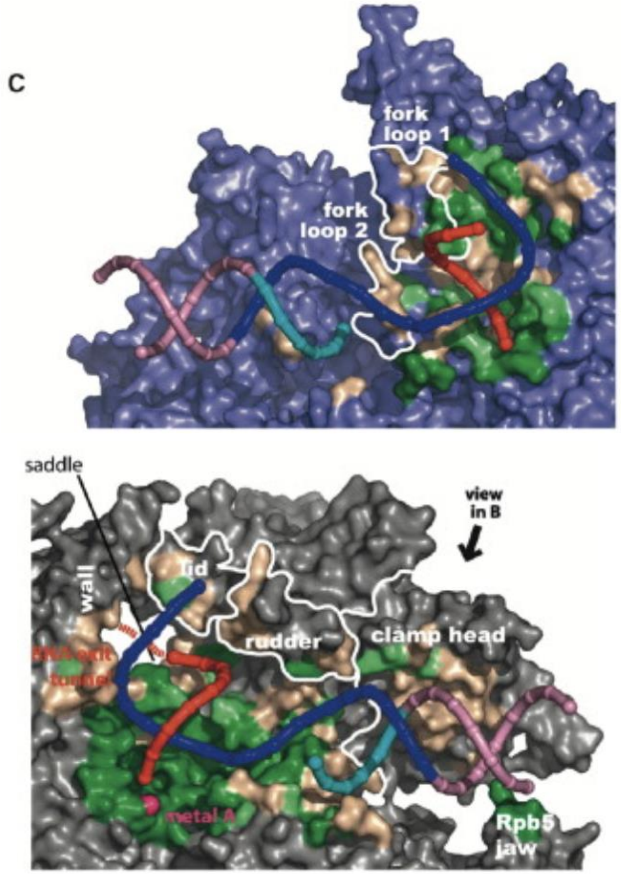
Protein Structure

RNA polymerase II

RNA聚合酶II (Pol II)复合物的组成部分，由12个亚基组成，其中包括一个10多肽催化核心和异二聚Rpb4/7复合物。





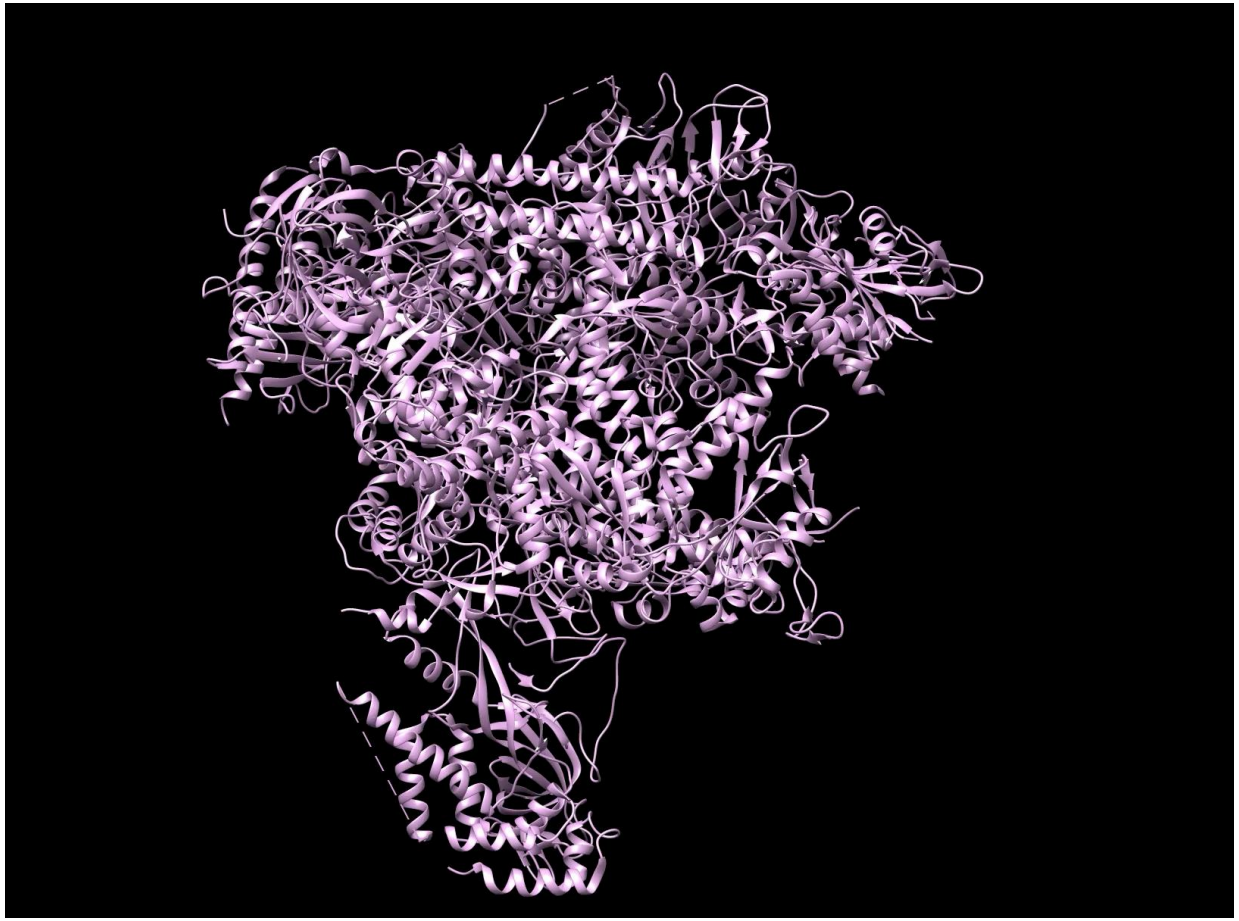


Kettenberger, H. , Cramer, P. etc. (2004). Complete RNA polymerase II elongation complex structure and its interactions withNTP and TFIIs. Molecular Cell, 16(6).

蛋白酶构象

北京大学 PEKING UNIVERSITY

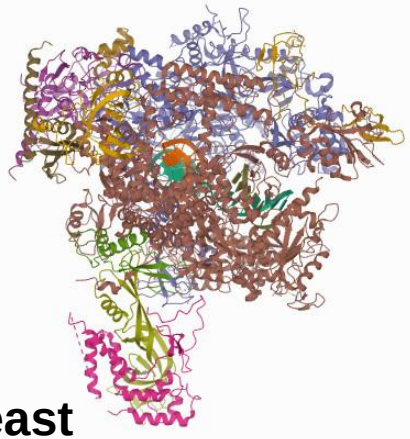
Conformational changes of RNA polymerase II between transcription initiation and elongation



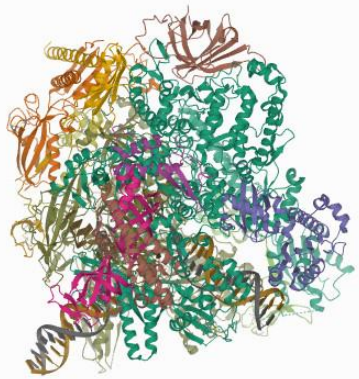
蛋白质构象

北京大学 PEKING UNIVERSITY

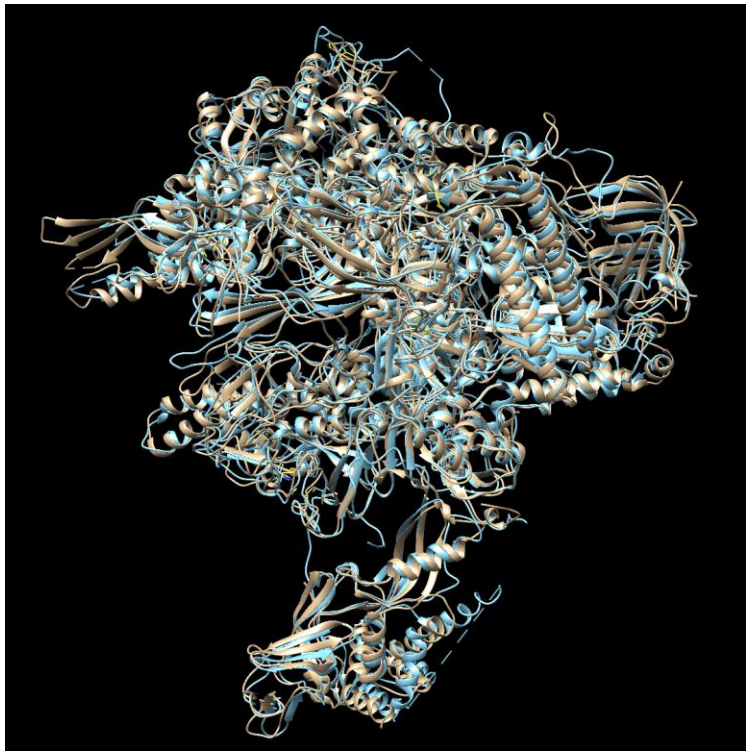
Conformational changes of RNA polymerase II between yeast and mammal during transcription



yeast



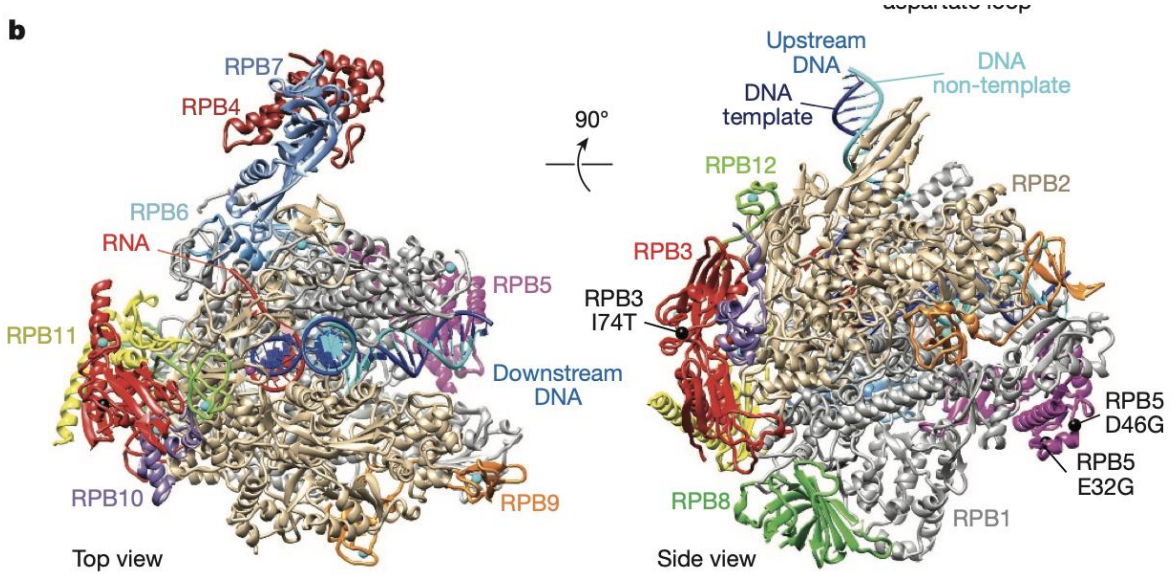
mammal

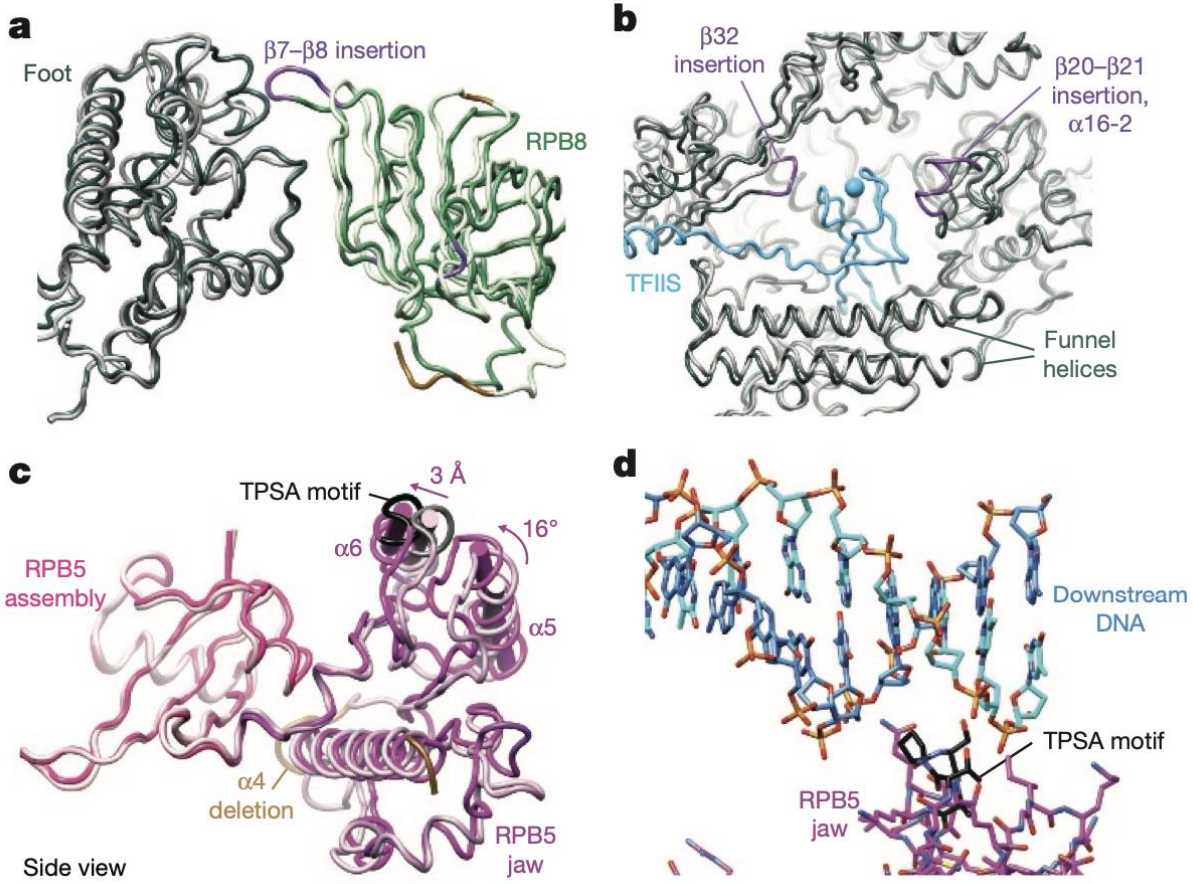


Matchmaker 1Y1W, chain A (#1) with 5FLM, chain A (#0), sequence alignment score = 4811.3
with these parameters:
chain pairing: bb
Needleman-Wunsch using BLOSUM-62
ss fraction: 0.3
gap open (HH/SS/other) 18/18/6, extend 1
ss matrix: (O, S): -6 (H, O): -6 (H, H): 6 (S, S): 6 (H, S): -9 (O, O): 4
iteration cutoff: 2
RMSD between 1156 pruned atom pairs is 0.977 angstroms; (across all 1389 pairs: 1.846)

蛋白质构象

Conformational changes of RNA polymerase II between yeast and mammal during transcription





思想自由 兼容并包

Bernecky, C. ,Cramer, P. etc. (2016). Structure of transcribing mammalian RNA polymerase II. Nature.

请老师同学批评指正

小组分工：

张　毅 RNA聚合酶背景、结构、功能整合

牛　迪 RNA聚合酶高通量测序分析

李姝含 聚合酶同源性比较与系统发生分析

刘雅迪 聚合酶蛋白构象预测与分析