

实用生物信息技术课程交流

——*Introduction to Bioinformatics*

专 业：生物化学与分子生物学

研究所：中国农业科学院研究生院

指导老师：罗静初 闫艳春

E-mail:704240667@qq.com

CONTENTS



课程简介



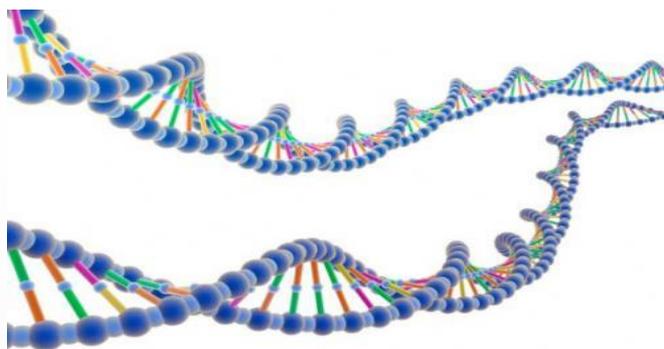
生物信息技术在科研中的应用



收获与体会



Part one



生物信息技术

采用信息科学技术，借助数学、生物学的理论、方法，对各种生物信息（包括核酸、蛋白质等）的收集、加工、储存、分析、解释的一门学科。

收集、加工、储存：计算机科学家
分析、解释：生物学家

“实用生物信息技术”

常用的生物信息分析方法与实际科学研究紧密结合



**his nation under god shall have an
ew birth of freedom and
that government of the people by
the people and for the people
shall not perish from
the earth**

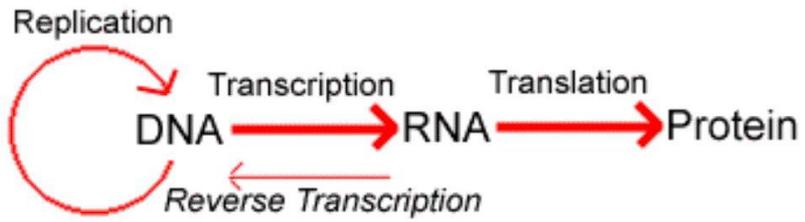




his nation under god shall have a new birth of fr
eedom and that government of the p
eople by the people and for the peo
ple shall not perish from the ear

that this Nation, under GOD, shall have a
new birth of freedom; and that government
of the People, by the People and for the
People shall not perish from the earth





Bioinformatics





Part two



01

Primer design (引物设计)

02

RNA-seq (转录组测序)

03

CRISPR/Cas9 (基因编辑系统)

04

Others (其他应用)



课题背景

“抗生素对斑马鱼生长发育的影响”

抗生素：抗肿瘤类米托蒽醌（Mitoxantrone, MA）

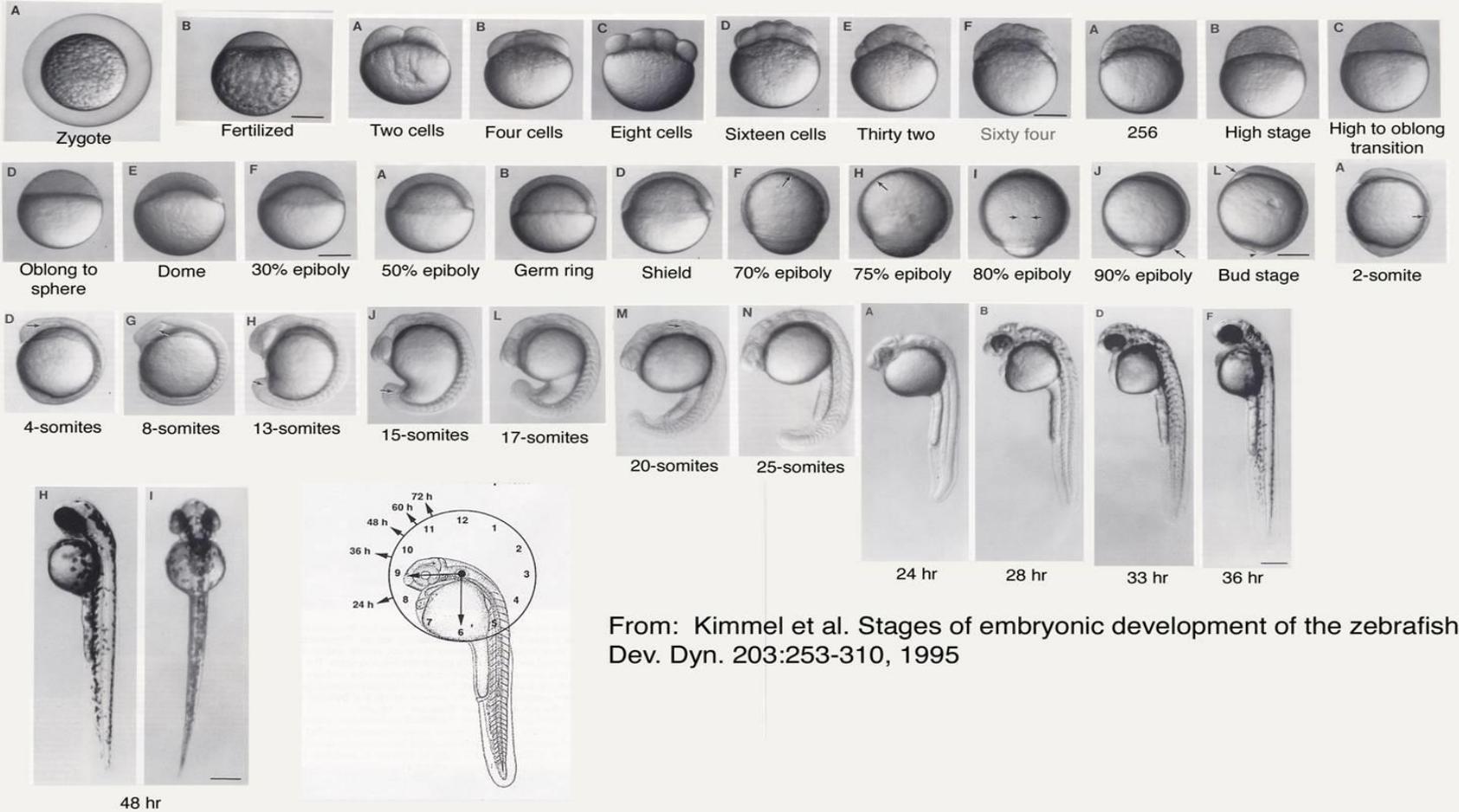
抗肿瘤作用是嵌入DNA与其交叉联结，导致单链和双链断裂，抑制RNA聚合酶，阻止DNA，RNA的合成，作用于增殖周期和非增殖周期细胞，对S期最敏感。有一定的心脏毒性、骨髓抑制以及对肝、肾功能有所损害等副作用。

斑马鱼(Danio rerio, zebrafish)

斑马鱼体型小，胚胎具有光学透明性，易于观察；作为模式生物，其具有清晰的遗传背景；25对染色体，全基因组测序工作已于2005年完成；与人类基因相似度高达70%，可用于人类疾病的研究。斑马鱼早期胚胎对多数药物具有相当强的敏感性，具有20多种不同表现的反应指标，适合药物的毒性评价及筛选。



课题背景



From: Kimmel et al. Stages of embryonic development of the zebrafish
Dev. Dyn. 203:253-310, 1995



课题背景

1.构建血管荧光标记斑马鱼，观察药物对其血管发育的影响；

2.抗氧化酶活性检测，酶活性及功能分析；

3.转录组测序（RNA-seq），分析差异表达的基因；

4.RT-PCR和dd-PCR验证相关基因表达的差异；

5.数据库中筛选骨骼发育相关的差异基因；

6.通过CRISPR/Cas系统验证基因的功能；





Primer design

斑马鱼基因serpine1 : 抑制血管生成和新陈代谢

NCBI Reference Sequence: NM_001114559.1

FP: AGTGAAGATGGAGTGGAGGTAG

RP: TCTGGCTGGCTGAAGTCTATC

NCBI Primer-Blast进行引物验证

Primer pair 1

	Sequence (5'->3')	Length	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	AGTGAAGATGGAGTGGAGGTAG	22	58.63	50.00	2.00	0.00
Reverse primer	TCTGGCTGGCTGAAGTCTATC	21	59.24	52.38	3.00	3.00

Products on target templates

>[NM_001114559.1](#) Danio rerio serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1 (serpine1), mRNA

product length = 131

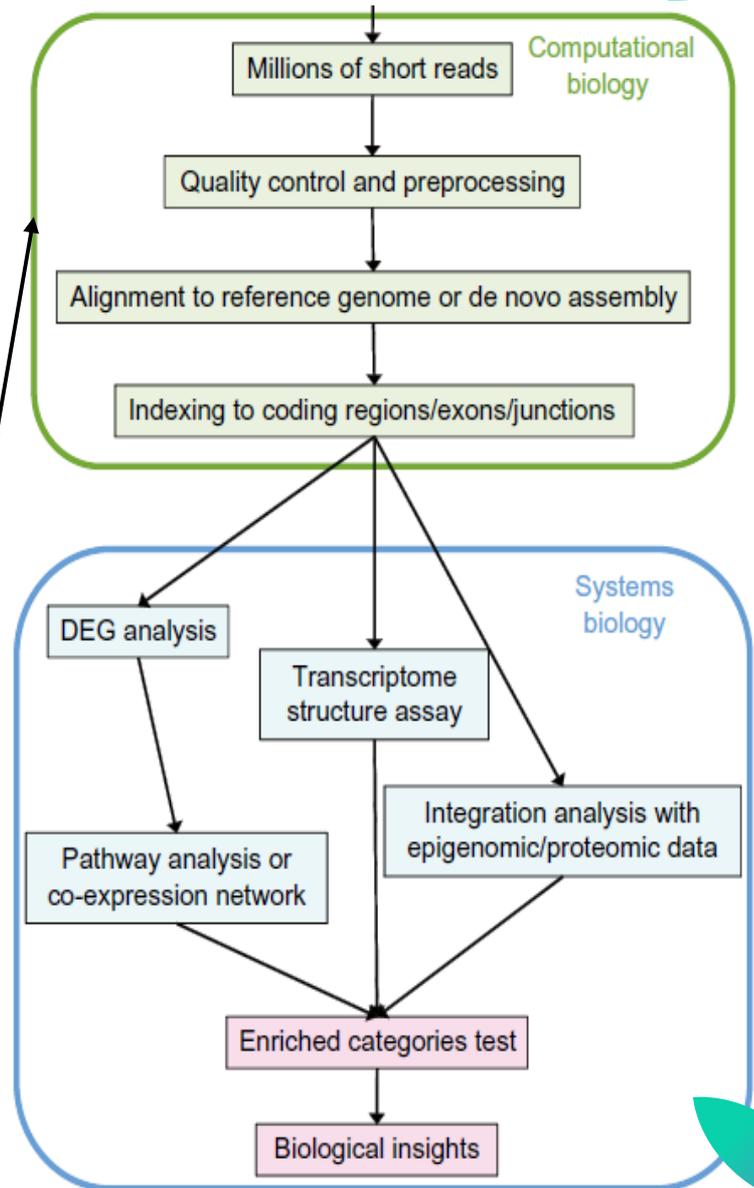
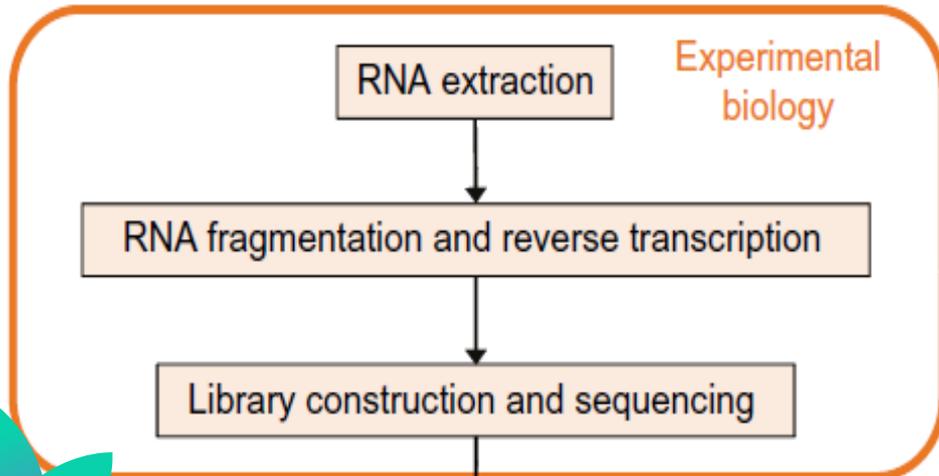
```
Forward primer 1 AGTGAAGATGGAGTGGAGGTAG 22
Template       402 ..... 423
```

```
Reverse primer 1 TCTGGCTGGCTGAAGTCTATC 21
Template       532 ..... 512
```

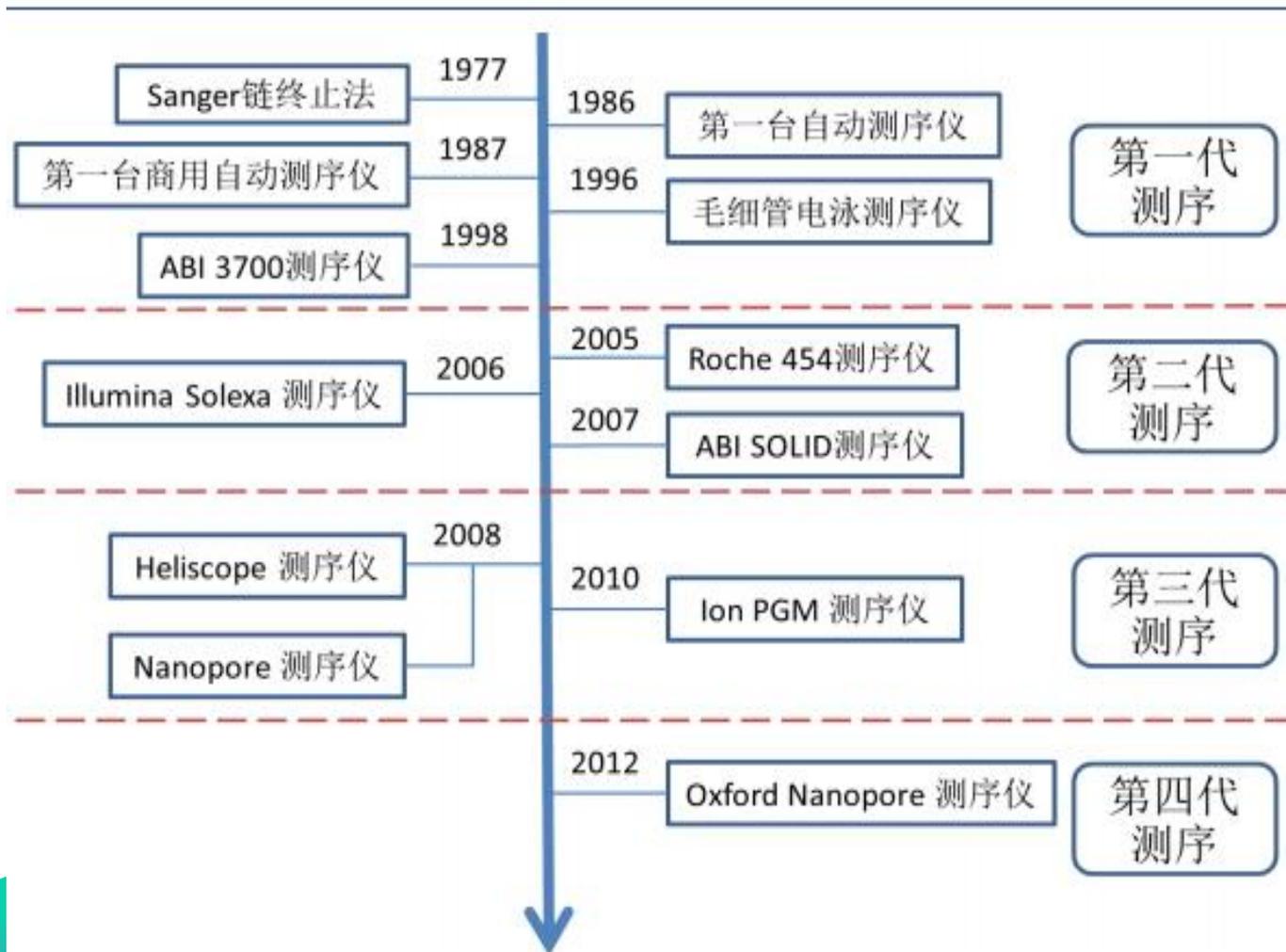


RNA-seq

转录组是某个物种或者特定细胞类型产生的所有转录本的集合。转录组研究能够从整体水平研究基因功能以及基因结构，揭示特定生物学过程以及疾病发生过程中的分子机理。



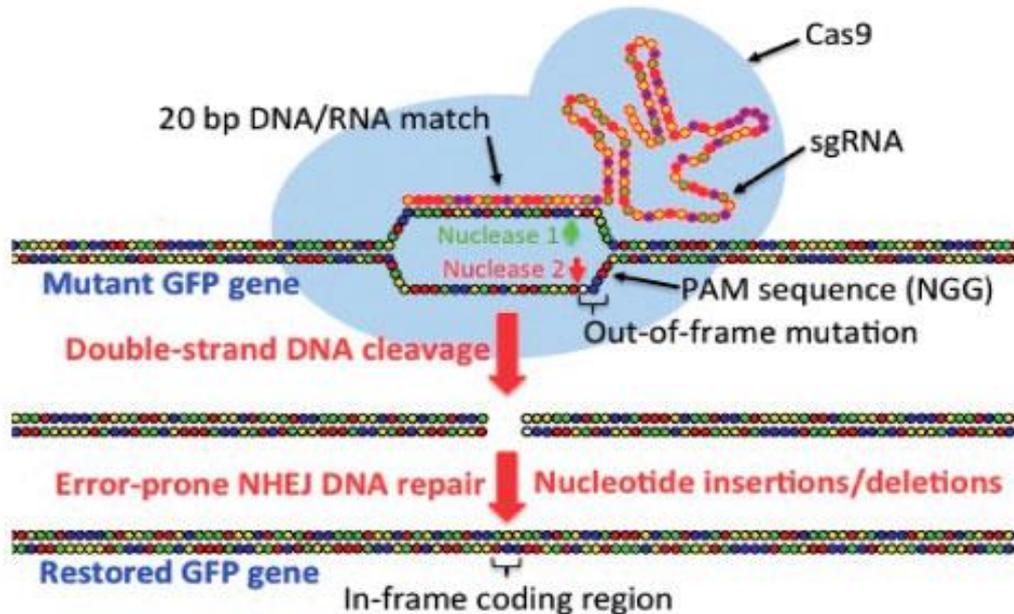
测序技术发展历程



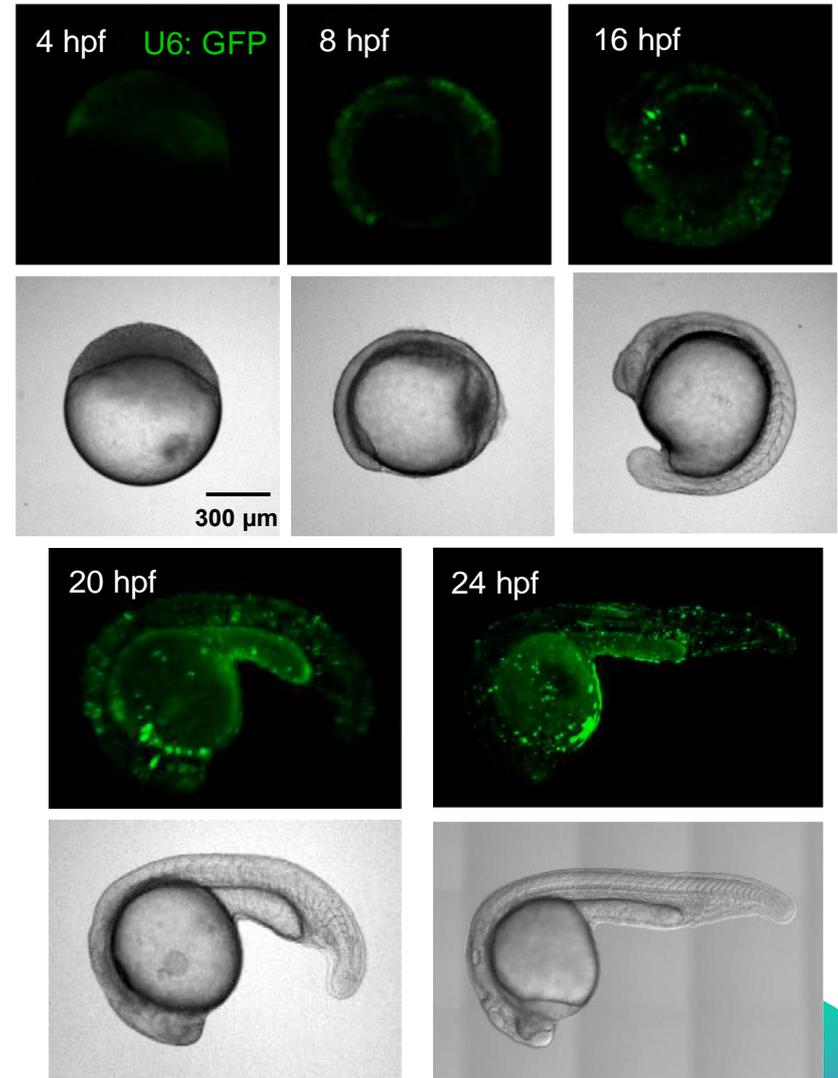
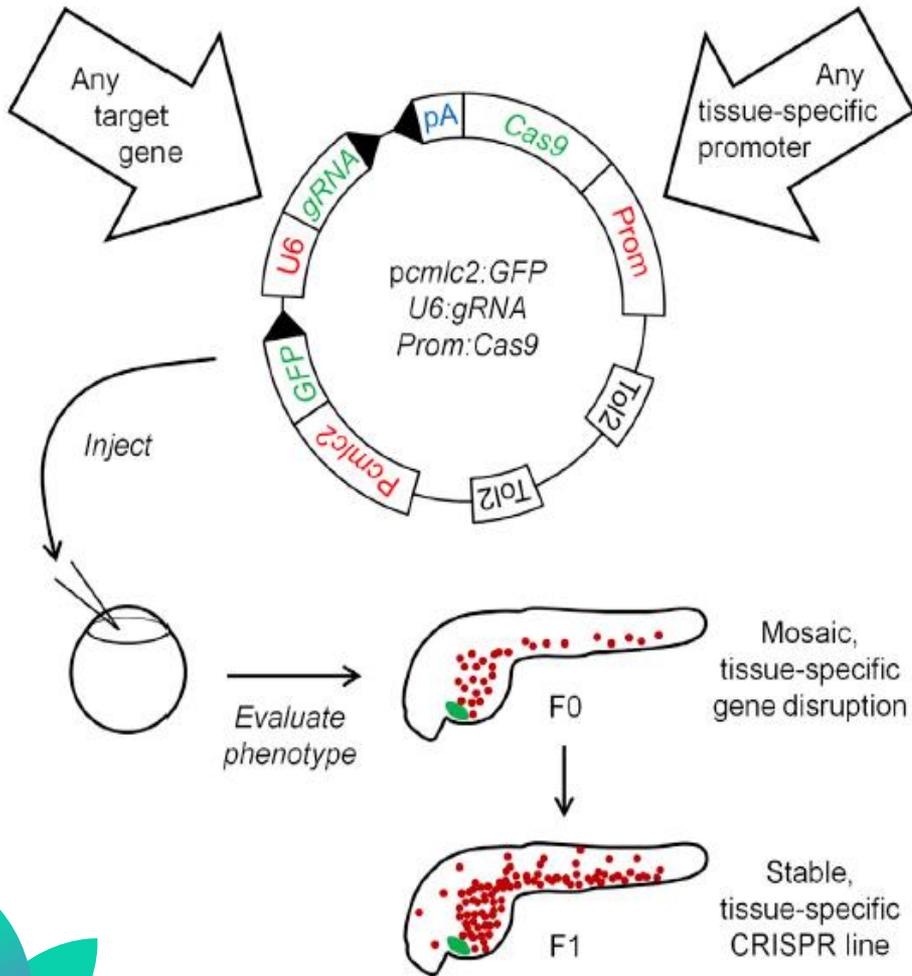
CRISPR/Cas9

——Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/ CRISPR-associated (Cas)9

这是一种人工核酸内切酶基因改造系统，它由基于细菌的一种获得性免疫系统改造而成。



CRISPR/Cas9 Vector System





CRISPR/Cas9 Vector System

在**NCBI**或**ENSEMBLE**中找到目标基因CDS区，分析相应的基因组结构，明确CDS的外显子部分。可选择将待敲除位点放在起始密码子ATG后的外显子上。

获取对应的蛋白序列，查找序列中的M（起始位点）。然后在**Protein BLAST**中找到该蛋白家族的保守结构域，从保守序列前端选择敲除靶点。

将选择的靶位点对应到编码区外显子上，然后在**crispr.mit.edu/**的页面下设计识别靶位点的一对DNA Oligo。

gRNA-*entpd5*:

CTATAAGAAGTCCGAGCATGguuuuagagcuagaaauagcaaguuaaaauaaggcuaguc
cguuaucaacuugaaaaaguggcaccgagucggugcuuuuuu

下划线部分20nt为gRNA，特异性识别目标基因组序列（特异序列）

小写字母部分为sgRNA骨架，用于稳定sgRNA，Cas9和基因组DNA复合物（保守序列）

<http://crispr.mit.edu/>





Others

1.蛋白质结构分析

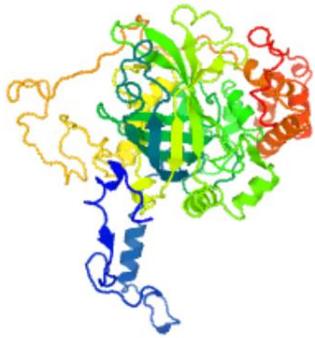
Protein name	Catalase (CATA)
Organism	Zebrafish (<i>Danio rerio</i> ,7955)
Protein information	http://www.uniprot.org/uniprot/Q9PT92
Protein length	526aa
Entry ID	Q9PT92
Feature key	Active sites 75-75aa Active sites 148-148aa Metal binding 358-358aa
Sequence	>CATA_DANRE MADDREKSTDQMKLWKEGRGSQRPDVLTTGAGVPIGDKLNAMTAGPRGPLLVDVFTDE MAHFDRE RIPERVV H AKGAGAFGYFEVTHDITRYSKAKVFEHVGGKTPIVVRFSTVAGEAGSPD TVRDPRGFAVKFYTDEGNWDLTG N NTPTFFIRDLLSPSFIHSQKRNPQTHLKDPMVWDFW SLRPESLHQVSFLFSDRGIPDGYRHMNGYGSHTFKLVNAQGGQPVYCKFH YKTNQGIKNIPVEEA DRLAATDPDYSIRDLYNAIANGNFPSWTFYIQVMTFEQAENWKWNPFDLTKVWSHKEFPLIPV GRFVLNRNPVNYFAEVEQLAFDPSNMPPGIEPSDKMLQGRLFS Y PDTHRHRLGANYLQLPV NCPYRTRVANYQRDGP MCMHDNQQGAPNYYPNSFSAPDVQPRFLESKCKVSPDVARYNSAD DDNVTQVRTFFTTQVLNEAERERLCQNMAGHLKGAQLFIQKRMVQNLMAVHSDYGNRVQALL DKHNAEGKKN TVHVYSRGGASAVAAASKM

Others

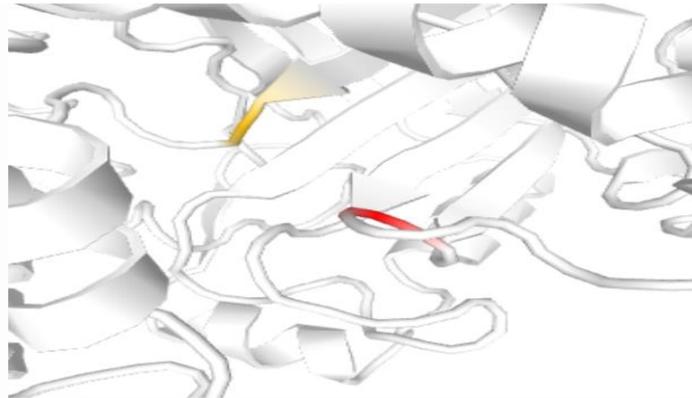
1.蛋白质结构分析

对于结构已知的蛋白可以在数据库中得到蛋白质信息，如结构，活性中心以及底物等。可以直接用分析软件进行解析，如**Pymol**，**Swiss-PDB viewer**，**Chimera**。

对于结构未知的蛋白质，则需要进行**同源模建**，通过相似性分析，查找同源蛋白等方式进行结构构建，然后再进行预测分析。



CATA_DANRE ; Catalase; EC=1.11.1.6;
Danio rerio (Zebrafish) (Brachydanio rerio)
Resolution: 1.65 (X-RAY)



.ASN148.CA

.HIS75.CA



.TYR358.CA

Swiss-PDB viewer



Others



2. 分子进化树同源性分析

NCBI Uniprot

	Entry ID	Organism	Length	Identical
1	Q9PT92	Danio rerio	526	100%
2	P24270	Mus musculus (小鼠)	527	80.4%
3	P04762	Rattus norvegicus (家鼠)	527	79.6%
4	O97492	Canis lupus (狼)	527	79.5%
5	P00432	Bos taurus (牛)	527	80.1%
6	O62839	Sus scrofa (野猪)	527	79.3%
7	Q2I6W4	Callithrix jacchus (绒毛猴)	527	78.5%
8	Q64405	Cavia porcellus (豚鼠)	527	78.9%
9	Q5RF10	Pongo (黑猩猩)	527	78.1%
10	P04040	Homo sapiens	527	77.7%



Part three

收获与体会：

- 1.几个实用网站：<http://abc.cbi.pku.edu.cn/>（课程主要使用网站）
<http://www.ncbi.nlm.nih.gov/>（NCBI）
<http://weblab.cbi.pku.edu.cn/>（实用在线分析工具）

- 2.一个优秀的学习小组

- 3.理解方式的改变

- 4.探索学习能力的提升

- 5.经历和知识同等重要





建议

- 1.带着兴趣去学习
- 2.重视基础
- 3.结合着自己的研究方向
- 4.取长补短
- 5.课后练习





THANKS