



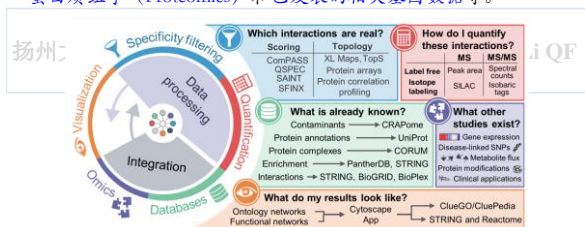
基因功能的研究策略主要包括哪些呢？

第八章 基因功能研究技术

- 1、应用生物信息学分析软件或数据库进行结构和功能的预测；
- 2、基因的体内表达规律分析；
- 3、亚细胞定位分析；
- 4、转基因研究；
- 5、基因编码产物相互作用蛋白的研究。

1 应用生物信息学分析软件或数据库进行结构和功能的预测

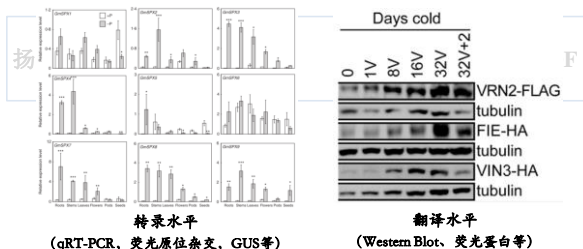
生物信息学 (Bioinformatics) 是以计算机为工具对生物信息进行储存、检索和分析的科学，包含基因组学 (Genomics)、蛋白质组学 (Proteomics) 和已发表的相关基因数据等。



根据基因和蛋白质序列，获得基因结构、同源性等生物信息，初步预测基因功能。

2 基因的体内表达规律分析

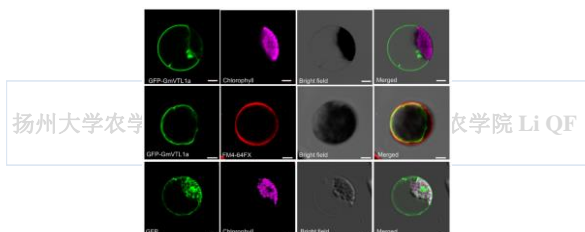
在mRNA和蛋白质两个水平上对基因的时空表达谱进行分析，揭示基因在植物体内的表达规律。



转录水平
(qRT-PCR, 荧光原位杂交, GUS等)

翻译水平
(Western Blot, 荧光蛋白等)

3 亚细胞定位分析



- 可以验证前期预测是否正确；
- 可以帮助进一步分析基因功能、蛋白互作及作用机理。

4 转基因研究

对基因进行生物信息学预测、表达分析和亚细胞定位分析后，对其功能有一个大致推断，但需要通过遗传材料进一步验证。基因功能研究至少包括以下几种类型的遗传材料：

- 恢复基因表达：功能互补
- 增强基因表达：过表达
- 减弱或去除基因表达：抑制/敲除

基因功能或表达调控

1. 恢复基因表达：功能互补

扬州大学农学院 Li QF

扬州大学农学院 Li QF

2. 增强基因表达：过表达

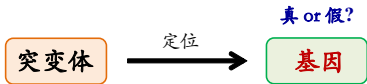
3. 减弱或去除基因表达：抑制/敲除

1. 功能互补 (functional complementation)

通过基因定位，发现某个基因突变可能是导致突变体表型改变的原因。

由自然突变、EMS诱变、辐射诱变、T-DNA插入、转座子插入等方式获得的突变体往往存在较多的背景突变。

如何验证定位到的基因就是造成表型改变的基因，而不是定位错了呢？

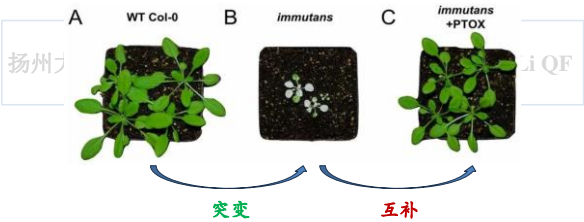


如克隆到的基因发生了隐性突变，则其表型可能是突变导致该基因功能丧失或减弱造成的。将野生型基因（候选基因）编码区连同表达调控区域（包括启动子、UTR等）通过转基因导入突变体，如果互补株表型恢复为野生型，则说明定位到的基因是造成突变体表型改变的目标基因。

原始载体名称	载体类别	骨架	原核抗性	真核抗性
pBWA(V)H-Tnos	promoter:CDS	pCAMBIA1300	KAN	潮霉素
pBWA(V)B-Tnos	promoter:CDS	pCAMBIA3300	KAN	草丁腈
pBWA(V)K-Tnos	promoter:CDS	pCAMBIA2300	KAN	G418
pBWA(V)M-Tnos	promoter:CDS	pUBI-C4300	KAN	甘露糖

- 全长：基因启动子+基因（外显子+内含子+UTR）
- 核心：基因启动子+编码区（CDS）

在突变体中转入野生型PTOX基因使其表型恢复



但显性突变或功能获得性突变的突变体往往不能通过转入正常版基因使之表型恢复到野生型水平。

可尝试把突变基因型转入野生型，确认转入突变基因的野生型植株会不会出现类似于突变体的表型，进而验证表型改变是不是由该基因突变造成的。

基因功能或表达调控

1. 恢复基因表达：功能互补

扬州大学农学院 Li QF

扬州大学农学院 Li QF

2. 增强基因表达：过表达

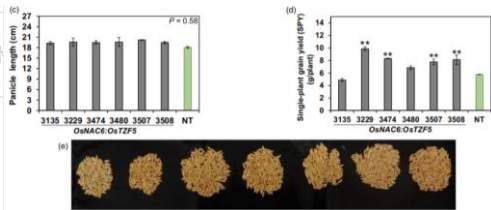
3. 减弱或去除基因表达：抑制/敲除

2. 过表达 (overexpression)

有些基因在基因组中有多个结构同源、功能类似的拷贝 (基因家族), 冗余性比较高, 敲除其中一两个基因后, 其它家族成员可代替其功能, 所以敲除个体没有任何表型或缺陷。
如何研究这些基因的功能呢?

过表达 (overexpression)

诱导型启动子: 指在诱导条件刺激下能大幅度快速调控转录活性的启动子。当植物发育到特定阶段或在特定的生长环境与组织器官中接收诱导信号时, 此类型启动子可以快速调控目的基因表达, 同样也可以在胁迫信号消失时停止目的基因的表达。

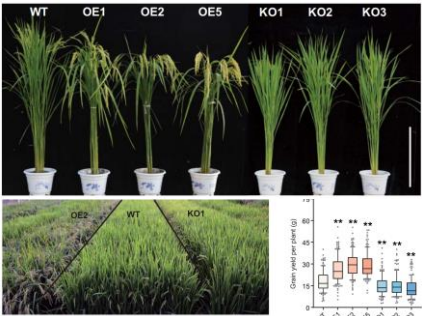


利用逆境诱导型启动子OsNAC6调控水稻OsTZF5在干旱条件下的表达, 避开OsTZF5对植株生长发育的不利影响, 实现干旱下产量提升。

过表达 (overexpression): 融合目的基因的全长序列与强启动子 (如: CaMV35S、Ubiquitin等), 通过遗传转化, 获得基因过量转录和翻译、表达产物大量积累的个体。

元件名称	名称	用途
启动子	CMV, EF1a, hSyn, TRE3g等。	广谱性、组织特异性和诱导型启动子启动转录
常用标签	GFP, EGFP, YFP, mCherry, DsRed2, tdTomato, BFP-tag, CFP等	荧光标签
	His, flag, GST, HA, Myc等	亲和标签
增强子	WPRE, SV40等	加强基因转录
抗性基因	KanR, AmpR等	原核细胞抗性筛选
	PuroR, Neo, Diphtheria toxin, Actidione等	真核细胞抗性筛选
连接元件	2A (T2A, P2A, E2A等), IRES	用来连接2个表达基因

- 广谱/组成性启动子使基因异位表达;
- 诱导型启动子



过表达高产基因DREB1C, 同时提高作物氮素利用和光合效率, 增产>30%
(OsDREB1C表达受光和低氮状态诱导)

基因功能或表达调控

3.1 基因沉默 (gene silencing) 技术

1. 恢复基因表达: 功能互补

扬州大学农学院 Li QF

2. 增强基因表达: 过表达

3. 减弱或去除基因表达: 抑制/敲除

① 反义 (antisense) 技术

扬州大学农学院 Li QF

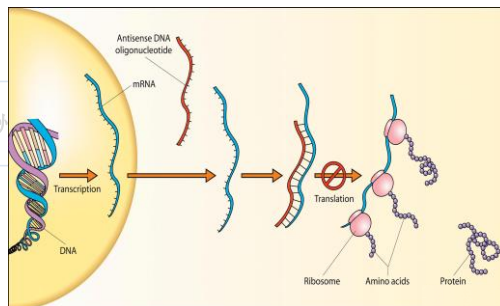
② RNA干扰/RNA干涉 (RNA interference, RNAi)

① 反义 (antisense) 技术

通过**碱基互补配对原理**，利用人工或生物合成的特异互补性DNA或RNA片段（或其修饰产物），干扰基因的转录、mRNA剪接加工与输出、翻译等各个环节，抑制或封闭靶基因表达，从而调节细胞的生长分化等。

主要包括：反义寡核苷酸技术（Antisense oligonucleotides, ASON）、核酶技术（Ribozyme）、小干扰RNA等。

反义技术的原理



反义技术的优劣

优点：

- 针对靶基因mRNA/DNA序列设计反义核酸，**特异性极高**；
- 靶基因序列已知，反义核酸仅有15~30个碱基，**结构简单**，**容易设计和体外大量合成**。

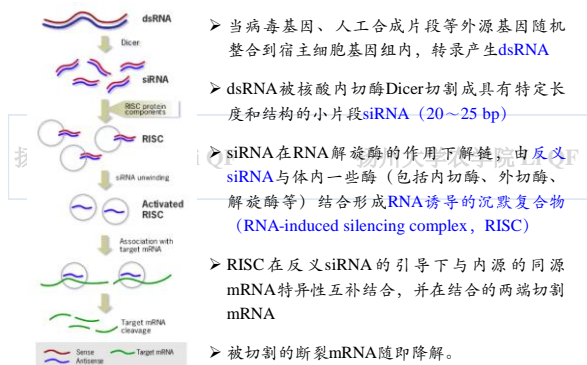
缺点：

- 反义核酸的稳定性较差；
- 产生细胞毒性；
- 包含外源DNA插入—转基因。**

② RNA干扰技术 (RNA interference, RNAi)

RNA干扰 (RNAi) 是指在进化过程中高度保守的、由**双链RNA (double-stranded RNA, dsRNA)**诱发的、**同源mRNA高效特异性降解**的现象。利用该现象开发的技术称为RNA干扰技术。

RNAi作用机制示意图



RNAi技术的优劣

优点：

- 不改变遗传物质**，可用于研究致死基因的功能；

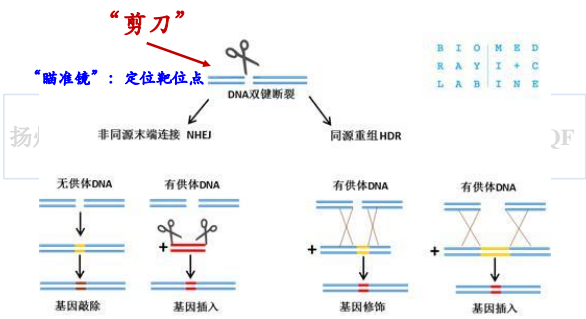
缺点：

- 易脱靶**，特别是基因组内有多个高度同源基因的情况；
- 不稳定性**，依赖RNA干扰来降低基因表达，而RNA不稳定。

3.2 基因组/基因编辑 (genome/gene editing) 技术

依赖于经过基因工程改造的核酸酶，也称“分子剪刀”，在基因组中特定位置产生位点特异性DNA双链断裂 (double-strand breaks, DSBs)，诱导生物体通过非同源末端连接 (NHEJ) 或同源重组 (HR) 来修复DSB，因为修复过程容易出错，从而导致靶向突变。这种靶向突变就是基因编辑。

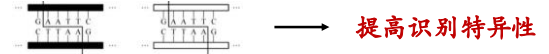
基因组编辑的原理



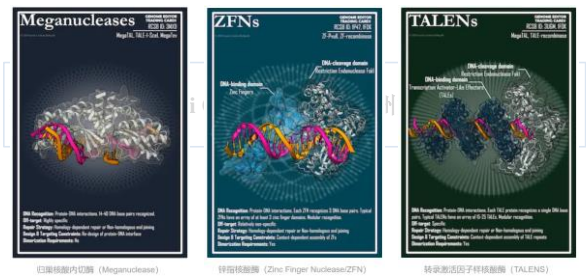
“剪刀”：核酸酶

基因编辑的关键是在基因组内特定位置创建DNA双链断裂 (DSBs)。常用的限制酶在切割DNA方面有效，但其通常在多个位点进行识别和切割，特异性较差。

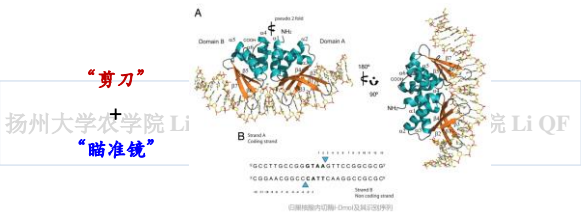
为克服这一问题并创建特定定位点的DSB，人们对四种不同类型的核酸酶 (Nucleases) 进行了生物工程改造：巨型核酸酶 (Meganuclease)、锌指核酸酶 (ZFNs)、转录激活因子样效应因子核酸酶 (TALEN)、成簇规律间隔的短回文重复序列及其相关蛋白系统 (CRISPR/Cas)。



基因编辑工具的发展—诸侯

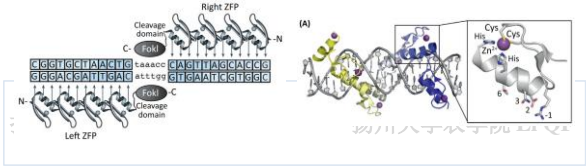


最早的基因编辑工具—巨型核酸酶 (Meganuclease)



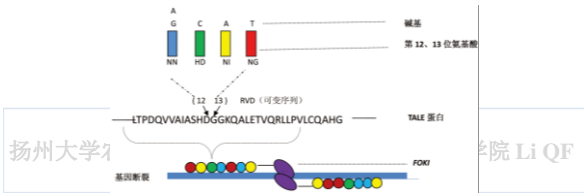
巨型核酸酶 (Meganuclease)：是一种脱氧核糖核酸内切酶，其特征在于它的识别位点较大 (12至40个碱基对的双链DNA序列)。该位点通常在任何给定的基因组中仅发生一次。是最特异的天然核酸酶，但不具有普适性。巨型核酸酶效率低，每1000 nt才识别一个潜在靶标，且精确度不可预测。

锌指核酸酶技术 (ZFN)



- 1984年，在非洲爪哇的转录因子中发现锌指蛋白，经过人工改造并连接上核酸内切酶后，发展为锌指核酸酶技术 (ZFN)
- “瞄准镜”：锌指蛋白 (zinc finger protein, ZFP) 组，由3-4个Cys2-His2锌指蛋白串联组成，每个ZFP识别并结合一个特异的三联体碱基
- “剪刀”：FokI核酸内切酶，可以通过二聚体化特异性地切割目的基因产生DSB

转录激活因子样效应物核酸酶技术 (TALEN)

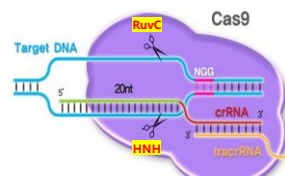


- 2007年，德国科学家在植物黄单胞菌 (*Xanthomonas*) 中发现一种特殊的分泌蛋白质——**转录激活因子样效应物 (Transcription activator-Like effector, TALE)** 能够结合植物宿主基因组并激活转录。2009年，TALE结合DNA的机制被阐明；2012年，TALEN技术出现
- “瞄准镜”**：DNA的特异性识别和结合区域——**TALE重复序列**
- “剪刀”**：与ZFN相同的**FokI核酸酶**

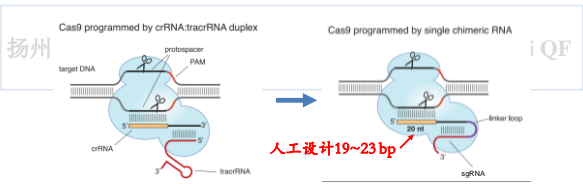
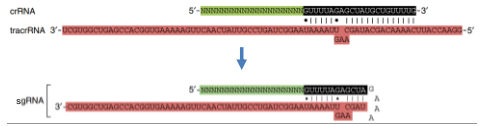
基因编辑工具的发展与演化——称王



CRISPR/Cas技术



- 成簇规律间隔的短回文重复序列 (CRISPR) 是原核生物基因组内的一段重复序列，与 CRISPR 相关 (CRISPR associated, Cas) 蛋白组成 CRISPR/Cas系统。
- 常用的Cas核酸酶是Cas9，含2个核酸酶结构域 (RuvC和HNH)。**HNH结构域**负责断裂通过互补靶向的DNA单链，切割位点位于PAM序列上游3nt处。**RuvC结构域**负责断裂另一条DNA链，切割位点位于PAM序列上游3~8nt处。



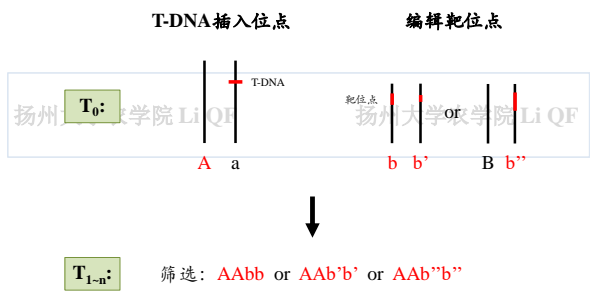
- 2012年Jinek等人，将crRNA-tracrRNA融合形成一条嵌合的单分子指导RNA (single-guide RNA, sgRNA)，简化形成CRISPR/Cas系统。

ZFN、TALEN、CRISPR 的不同之处对比

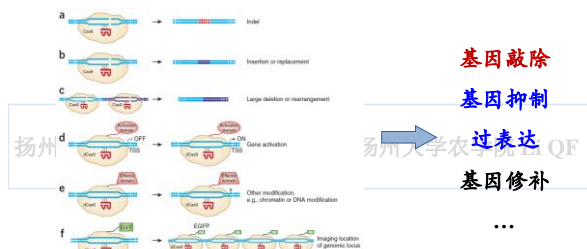
	ZFN	TALEN	CRISPR系统
识别模式	蛋白质—DNA	蛋白质—DNA	RNA—DNA
靶向元件	ZF array蛋白	TALE array蛋白	sgRNA
切割元件	Fok I 蛋白	Fok I 蛋白	Cas9蛋白
识别长度	(3-6)x 3 x 2 bp	(12-20) x 2 bp	20 bp
识别序列特点	以3 bp为单位	5'前T为 T	3'序列为 NGG
优点	平台成熟、效率高于被动同源重组	设计较ZFN简单、特异性高	靶向精确、脱靶率低、细胞毒性低、廉价 多基因
缺点	设计依赖上下游序列、脱靶率高、具有细胞毒性	细胞毒性、模块组装过程繁琐、需要大型测序工作、一般大型公司才有能力开展、成本高	靶区前无 PAM 则不能切割、特异性不高、NHEJ 依然会产生随机毒性
能否编辑RNA	不可以	不可以	可以

更快、更准、更简单、更高效

去除基因改造的痕迹



基于CRISPR/Cas9的各种应用



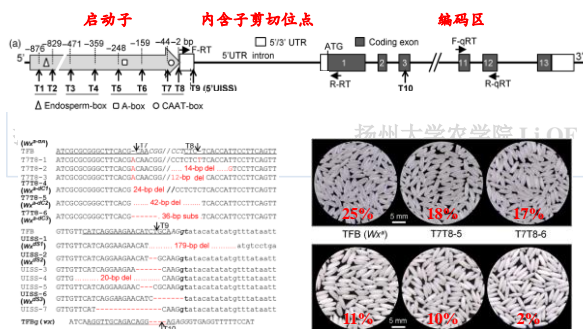
基因敲除

基因抑制

过表达

基因修补

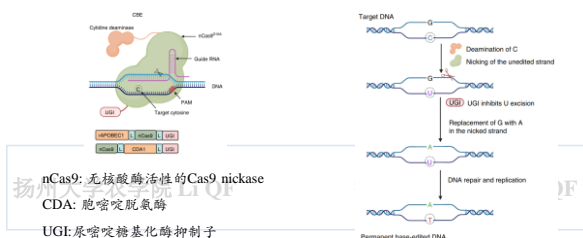
CRISPR/Cas9技术的应用—Wx基因



碱基编辑技术 (Base Editing/Base Editor, BE)

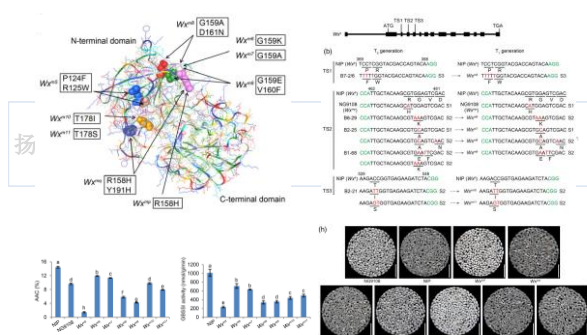
基于CRISPR/Cas系统自2016年发展起来的单碱基精准编辑技术。可用于研究某个(些)氨基酸残基对蛋白质的结构、催化活性以及结合配体能力的影响等。依据碱基编辑酶的不同可分为胞嘧啶碱基编辑器 (cytosine base editor, CBE)、腺嘌呤碱基编辑器 (adenine base editor, ABE)、C-to-G碱基编辑器 (CGBE)、C-to-A碱基编辑器 (CABE) 等。

特点: 不产生DSB, 实现单碱基的精准编辑, 脱靶效应低

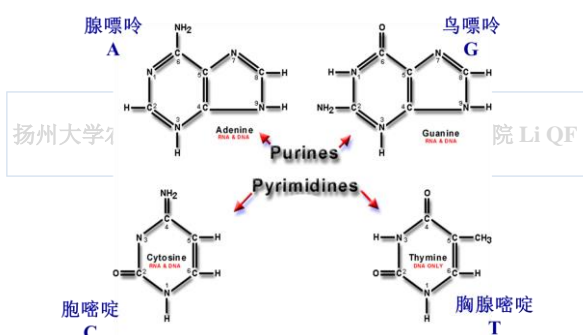


- gRNA引导nCas9识别并结合靶位点, 但不产生DSB;
- CDA使靶向位点的胞嘧啶 (C) 脱氨形成尿嘧啶 (U);
- UGI抑制U的清除, 使U在DNA复制过程中被T替代, 同时互补链上G被A替换。

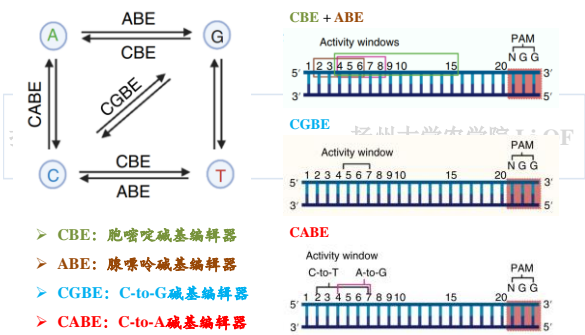
碱基编辑的应用—Wx基因



碱基编辑技术的类型和对应的活性窗口



碱基编辑技术的类型和对应的活性窗口



引导/先导编辑技术 (Prime Editing, PE)

PE是能够在基因组的靶位点处实现任意碱基替换和小片段精准删除、插入的新型基因组编辑工具，是基因组编辑领域的重大变革。核心包括两部分：

效应蛋白：逆转录酶与具有单链切割活性的nCas9(H840A)的融合蛋白；

pegRNA：在sgRNA的基础上，对其3'端进行延长，增加逆转录模板(reverse transcriptase templates, RTT)和引物结合位点(primer binding site, PBS)。

不产生DSB，“任意”编辑，但目前效率较低

基因编码产物相互作用蛋白的研究

蛋白质是生命功能最直接的执行者，虽然一些蛋白质可以独立的完成使命，但是大部分蛋白都需要一些伴侣分子的协助一起完成任务或者形成复合物之后才能充分发挥它的功能。所以，了解蛋白质与蛋白质之间、蛋白质与核酸分子之间的相互作用，能够帮助我们更好的了解细胞的生命活性，揭示隐藏在表象下的调控机理。

蛋白-蛋白相互作用的筛选和点对点验证方法



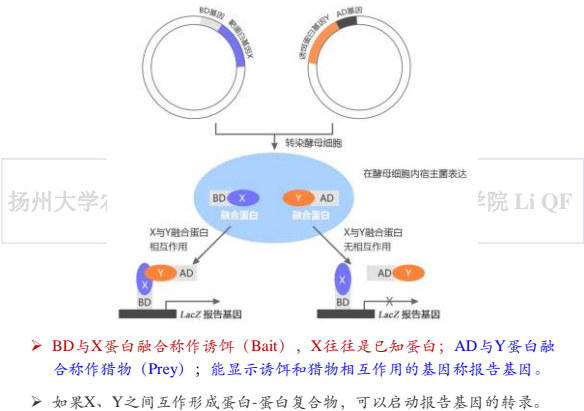
01 酵母双杂交技术

酵母双杂交 (Yeast two-hybrid, Y2H) 是一种直接在细胞内外检测蛋白质-蛋白质相互作用的方法。

其产生是基于对真核细胞转录因子特别是酵母转录因子 GAL4 的研究。GAL4 包括两个彼此分离但功能必需的结构域：DNA 结合域 (DNA binding domain, DNA-BD) 和转录激活域 (Activation domain, AD)。

DNA-BD 和 AD 单独作用不能激活转录反应，但当二者在空间上充分接近时，则呈现完整的 GAL4 转录因子活性，并激活下游基因转录。

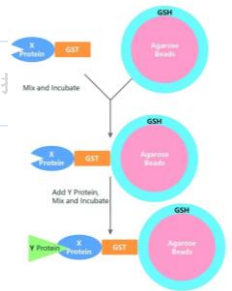
适合高通量蛋白质互作的筛选，但假阳性较多。



02 Pull-down 实验

Pull-down 实验是一种验证酵母双杂交结果的**体外**试验技术。基本原理是将靶蛋白-GST融合蛋白亲和和固定在谷胱甘肽亲和树脂上，充当“诱饵蛋白”；目的蛋白溶液过柱，可从中捕获与之相互作用的“捕获蛋白”（目的蛋白），洗脱结合物后通过SDS-PAGE电泳分析和western blot检测，从而证实两种蛋白间的交互或筛选相应的目的蛋白。

方法简单易行，操作方便。



03 免疫共沉淀

免疫共沉淀 (Co-Immunoprecipitation, Co-IP) 是一种用于确定两种蛋白质在**体内**相互作用的方法。

原理是使用靶蛋白的特异性抗体，可以在捕获靶蛋白的同时，将与靶蛋白互作的蛋白一起捕获，进而可鉴定蛋白复合体中的其他成员。

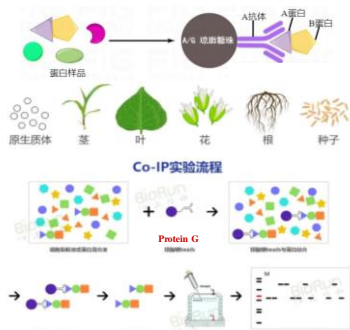
- 特异性抗体捕获样品中的靶蛋白，形成抗体-靶蛋白复合物；
- 利用能够与抗体结合的**支持物** (Protein A或G等) 固定和沉淀复合物，同时与靶蛋白互作的蛋白也沉淀下来；
- 洗去与靶蛋白结合的非特异性蛋白，通过SDS-PAGE、Western blot 等进行分析鉴定。

04 双分子荧光互补验证蛋白互作

- 双分子荧光互补 (Bimolecular fluorescence complementation, BiFC) 技术是一种基于蛋白质互补技术发展起来的，能够用于**体内或体外**检测蛋白质相互作用的技术。

- 其本质是将荧光蛋白拆分成可独立折叠的N端和C端两个亚基，并分别与两个待研究的目标蛋白 (A, B) 融合表达。如果这两个荧光蛋白具有相互作用的关系，就会把荧光蛋白两个亚基拉近，从而促使两个亚基组装成完整功能的荧光蛋白。

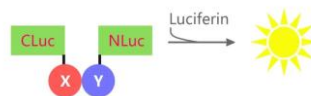
- 常用于测定两种蛋白质的体内互作；也可用于确定与特定蛋白质相互作用的未知蛋白 (与质谱连用，也称IP-MS)。



05 荧光素酶互补验证蛋白互作

- 荧光素酶互补实验 (Luciferase Complementation Assay, LCA) 以荧光素为底物来检测荧光素酶的活性。荧光素酶催化荧光素发生氧化，在560nm波长发出生物荧光。

- 将荧光素酶蛋白分成N端和C端2个功能片段 (NLuc和CLuc)，分别与目标蛋白X和Y融合，如果2个目标蛋白互作，则NLuc和CLuc在空间上会足够靠近并正确组装成荧光素酶，分解底物产生荧光。



双分子荧光互补验证蛋白互作原理图

主要蛋白-蛋白相互作用方法的比较

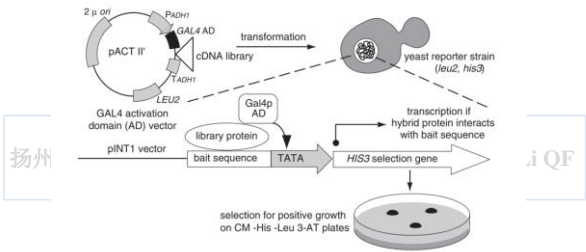
名称	互作情况	优点	缺点
酵母双杂	体外	适合高通量蛋白质互作筛选	假阳性较多，且不利于核外蛋白研究
Pull down	体外	鉴定蛋白的直接作用蛋白，非同接作用复合体；GST等标记操作简单	可能因电荷或强度和力等原因造成假阳性
免疫共沉淀	体内	可鉴定特定蛋白复合物中未知的蛋白组分	蛋白质的结合可能不是直接结合，且操作繁琐，技术要求高
BiFC	体内 体外	直观、快速；仪器要求不高，数据简单；只检测荧光有无，背景干净，检测灵敏	对温度敏感；不能实时观察蛋白互作或蛋白复合物形成；同时会有假阳性
荧光素酶互补	体外	高灵敏度、可量化、操作简单	存在假阳性

蛋白-核酸相互作用的筛选和点对点验证方法



酵母单杂

酵母单杂 (Yeast one hybrid, Y1H) 是在酵母双杂的基础上发展而来的，用于分析蛋白质和DNA之间的相互作用。基本原理：诱饵 (Bait) - 目标DNA报告基因融合；猎物 (Prey) - 目标转录因子和酵母转录激活域 (AD) 融合。将二者转入到合适的酵母菌株中，如果TF在酵母核的环境中与DNA结合，那么AD就会诱导报告蛋白的表达。



➤ 需要注意的是，无论TF是激活因子还是抑制因子，酵母AD都会激活报告因子。

凝胶迁移阻滞实验

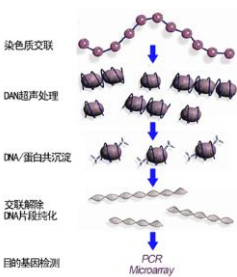
凝胶迁移阻滞实验 (electrophoretic mobility shift assay, EMSA) 是一种用于研究转录因子和其相关DNA结合序列相互作用的技术，可用于定性和半定量分析。基本原理是蛋白质可以与末端标记的核酸探针结合，电泳时DNA-蛋白质复合物比裸露的DNA电泳迁移慢，即表现为相对滞后。这一技术最初用于研究DNA结合蛋白，目前也可用于研究RNA结合蛋白和特定的RNA序列的相互作用。



染色质免疫沉淀

染色质免疫沉淀（Chromatin immunoprecipitation, ChIP）是一种研究体内DNA与蛋白质相互作用的重要工具。

在活细胞状态下将蛋白质-DNA复合物稳定，使用酶解法或者超声法随机打断DNA形成具有一定长度的染色质片段，利用免疫学方法分离出蛋白并纯化DNA，通过建库、测序以及生信分析获得更多和蛋白互作的基因结合信息（ChIP-seq），也可以采用qPCR检测目的蛋白和某些基因是否有结合(ChIP-PCR)。

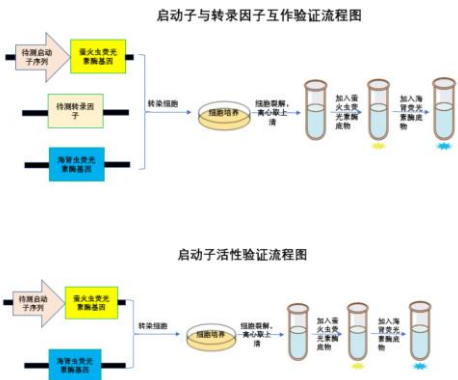


评估ChIP实验结果可靠性的关键在于设置对照：实验内对照（Input DNA, 断裂后的基因组DNA）；阳性抗体对照；阴性抗体对照。

双荧光素酶报告实验

双荧光素酶报告（Dual-Luciferase Reporter）是一种检测转录因子与目的基因启动子区DNA相互作用的方法。荧光素底物会在萤火虫荧光素酶作用下会发光（波长540-600nm），且光强能反应荧光素酶的表

达。将待测启动子等序列与萤火虫荧光素酶基因融合表达，荧光素酶的蛋白表达量与启动子活性相关。为减少实验误差，同时转染一个能稳定表达的海肾荧光素酶的质粒作为内参，催化腔肠素氧化产生蓝光（波长460-540nm），所以叫双荧光素酶报告基因系统。



蛋白-核酸相互作用的研究方法比较

技术	DNA/RNA EMSA	酵母单杂	染色质免疫沉淀 (ChIP)
相互作用情况	体外	体外	体内
核酸标记方法	生物素	无需	无需
填料基质	无需	无需	Protein A/G磁珠 Protein A/G琼脂糖
检测方法	Western blot	酵母生长情况 及测序	qPCR或 二代测序 (NGS)
目标	鉴定与DNA/RNA 序列结合的蛋白质	鉴定启动子和 转录因子相互作用	监测转录调控 (qPCR) 或鉴定靶序列 (NGS)