陆剑教授讲座笔记

会议回顾

2023年12月21日，中山大学中山医学院王金凯教授邀请了北京大学的陆剑教授进行了线上学术报告“蛋白质翻译调控的进化基因组学研究”。陆剑教授长期从事比较基因组学和演化生物学研究，将组学和演化生物相结合，系统解析真核生物基因表达调控因子的进化特征与驱动机理，并探索其在细胞演化及物种适应性进化中的作用。近年来从事新冠病毒演化工作，获得“全国科技系统抗击新冠肺炎疫情先进个人” 称号，相关工作获得北京市科技进步奖二等奖（第一完成人）。

会议内容

本次会议陆剑教授的研究报告集中于RNA表达调控的元件对于生物的发育的影响。在果蝇的早期发育过程中，并没有新的转录事件发生，而是受调控于两种基因的调控因子（Bicoid和Nanos）的此消彼长的一个过程。这两个基因翻译的蛋白质都可以分别调控很多mRNA的翻译。这个例子启示我们生物发育过程中所需要的大量基因表达往往只需要很少的基因便可以进行调控。

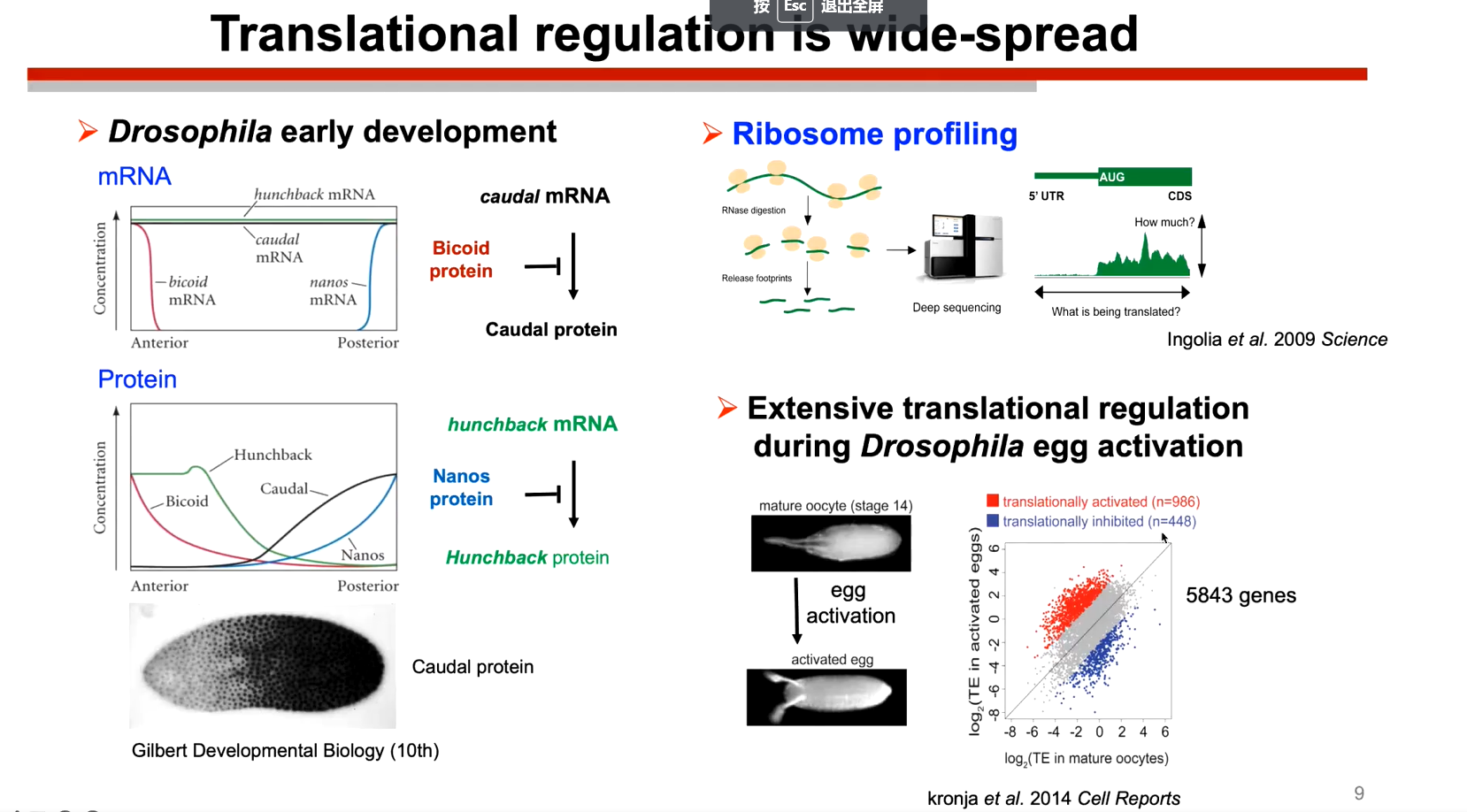


图1 翻译的广泛调控

随后陆教授提到在真核生物中，mRNA的一个重要特征就是具有一个很长的5`UTR区，其中蕴含着大量的调控元件。而陆剑教授的课题组主要聚焦于以下问题：

1. 哪些元件和因子参与了调控过程
2. 这些元件如何起源和演化
3. 演化模式和功能之间存在何种关联

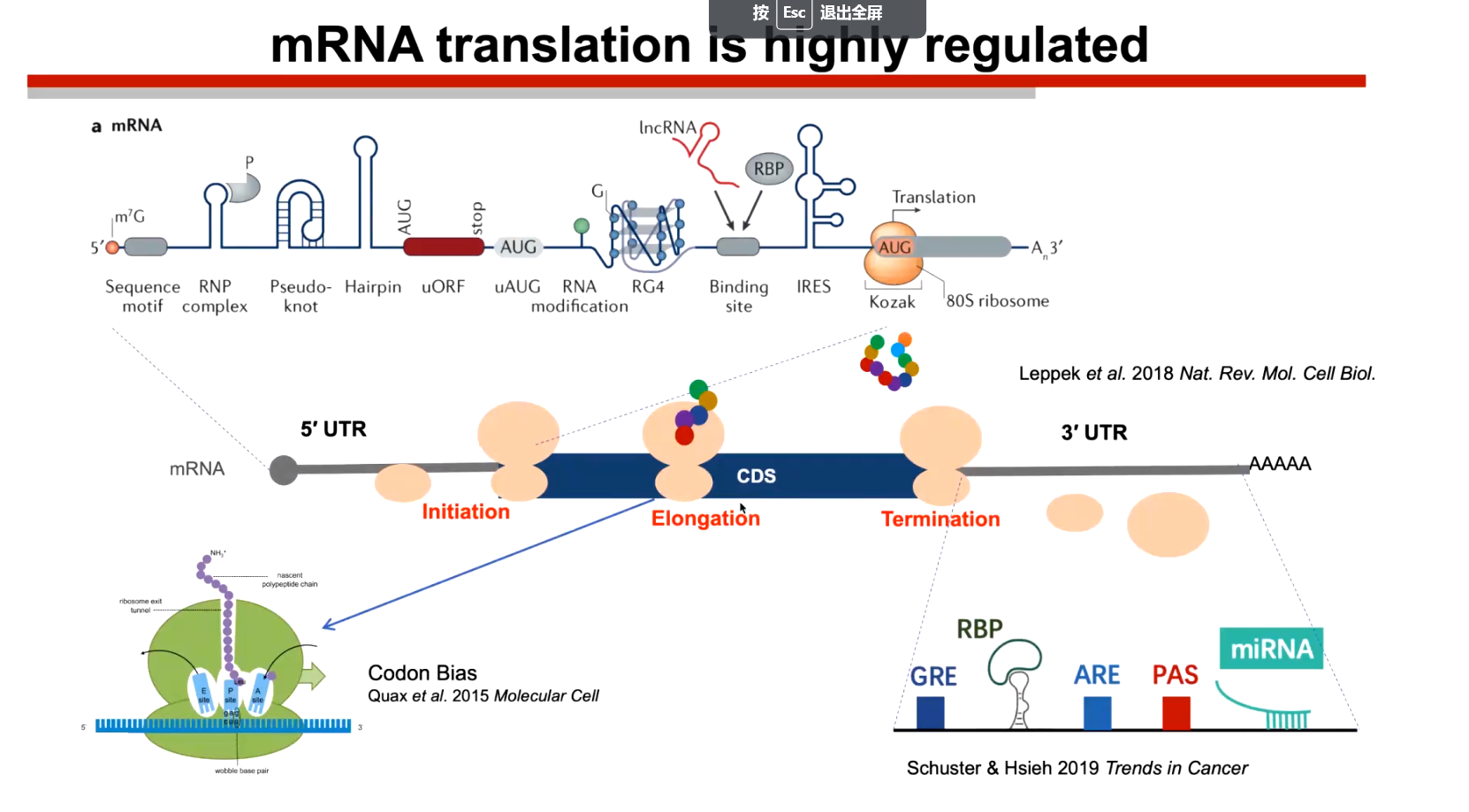


图2 顺式调控元件

本次会议陆教授主要汇报了他们课题组在uORF上面的一些研究成果。真核生物的翻译过程并非是线性的，其mRNA会形成一个假环状结构来提升翻译效率。当翻译起始复合体形成之后，43S PIC便会结合到帽子附近开始扫描，若遇到翻译起始密码子则还是进行翻译延伸。然而陆教授指出AUG密码子并非仅存在于CDS区域，其在5`UTR区域的一些很小的读码框之中也有可能大量存在。而问题就在于翻译复合体如何识别这些真假AUG从而进行正常的翻译。

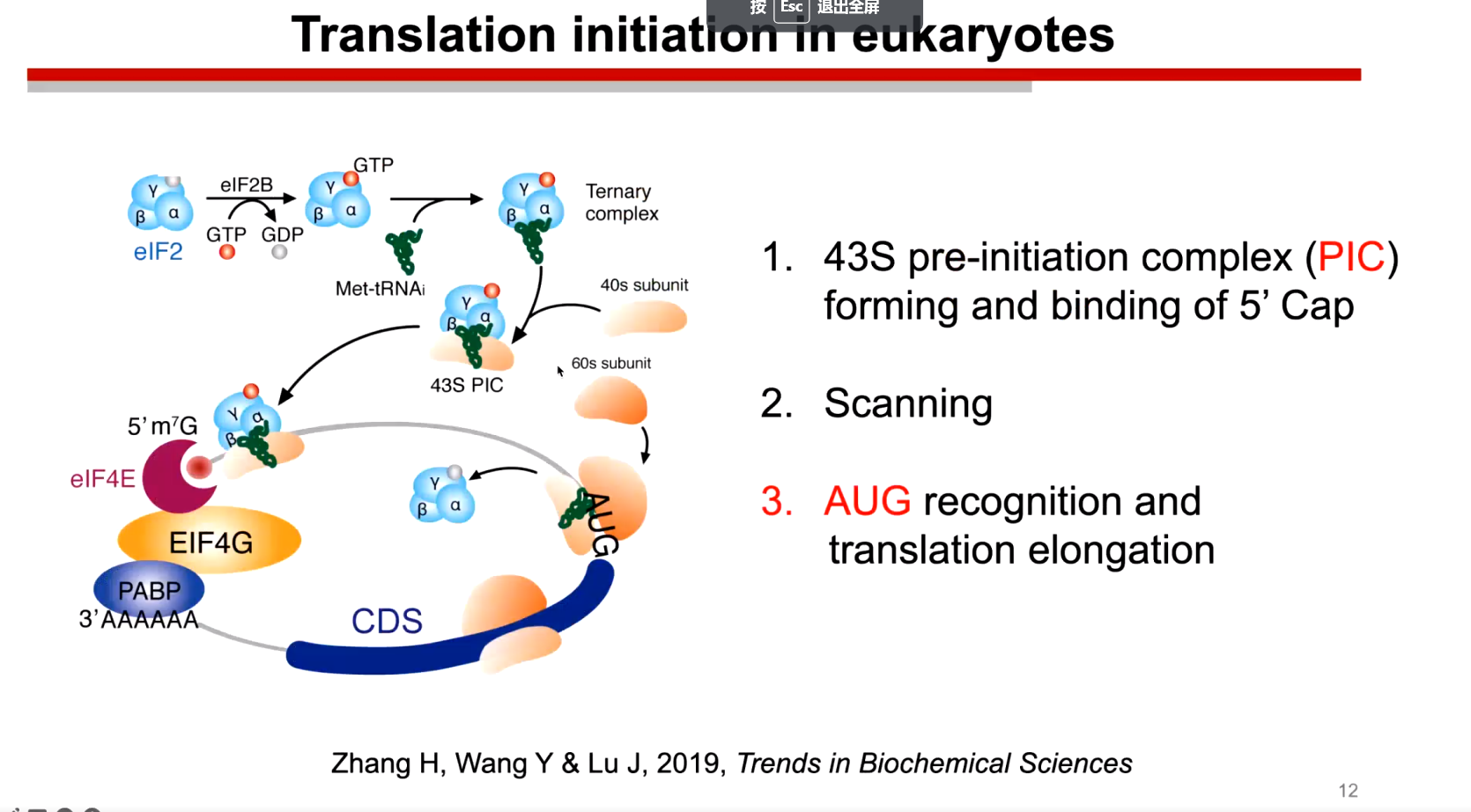


图3 真核生物翻译的假环状结构及其翻译起始机制

陆教授提出了三种假设，分别是：

1. uORF不会导致翻译起始，所以对下游CDS区的翻译也没有影响
2. uORF会导致翻译起始，同时也会抑制下游CDS区的翻译
3. uORF完全拦截了核糖体，导致下游CDS区的翻译被完全抑制

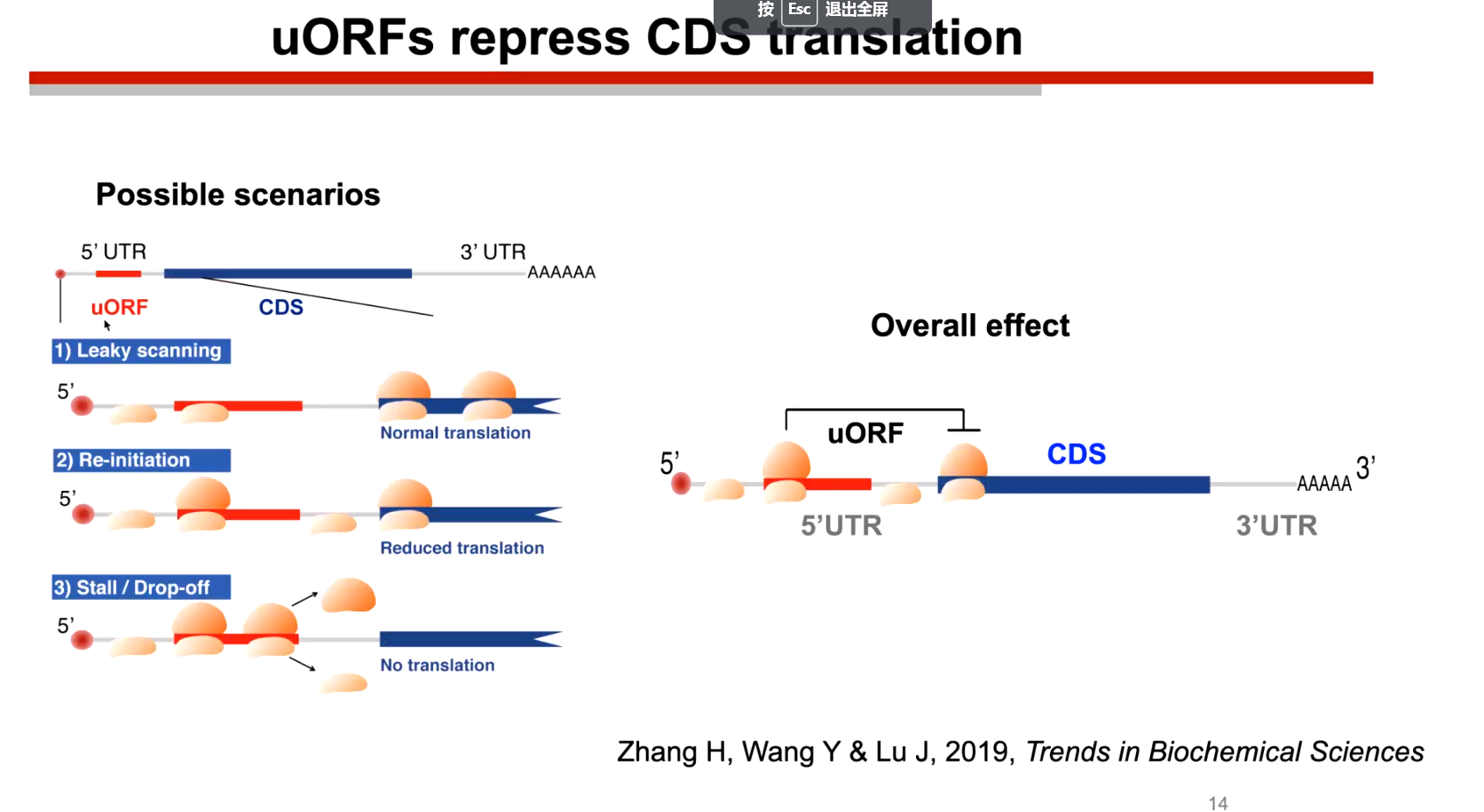


图4 uORF的读码判定假设

在大多数情况下，uORF对于下游CDS区的蛋白翻译都是起到抑制作用，并且uORF在基因组中的分布也极为广泛，占到了55%左右。并且根据Johnstone等人的研究，uORF在生物进化过程中是高度保守的，这说明uORF对于翻译的抑制并非百害而无一利。那么uORF到底在什么情况下是有利的，什么情况下是有害的呢。

针对这样的问题，陆教授课题组对果蝇发育周期的不同阶段都进行了核糖体图谱技术和mRNA-seq技术的测量，从而获得mRNA核糖体的覆盖度。他们共计测量了6000多个管家基因和7000多个阶段特异表达的基因，找到了2万8千多个uORF。根据翻译信号的由强到弱将这些uORF分为四类。陆教授提到根据细胞能量经济学的观点，细胞需要尽可能的节约出能量来对那些急需的蛋白进行大量的翻译，而这些基因就含有非常少量的uORF，而另外那些并不怎么需要的蛋白的基因则含有大量的uORF以抑制其翻译。

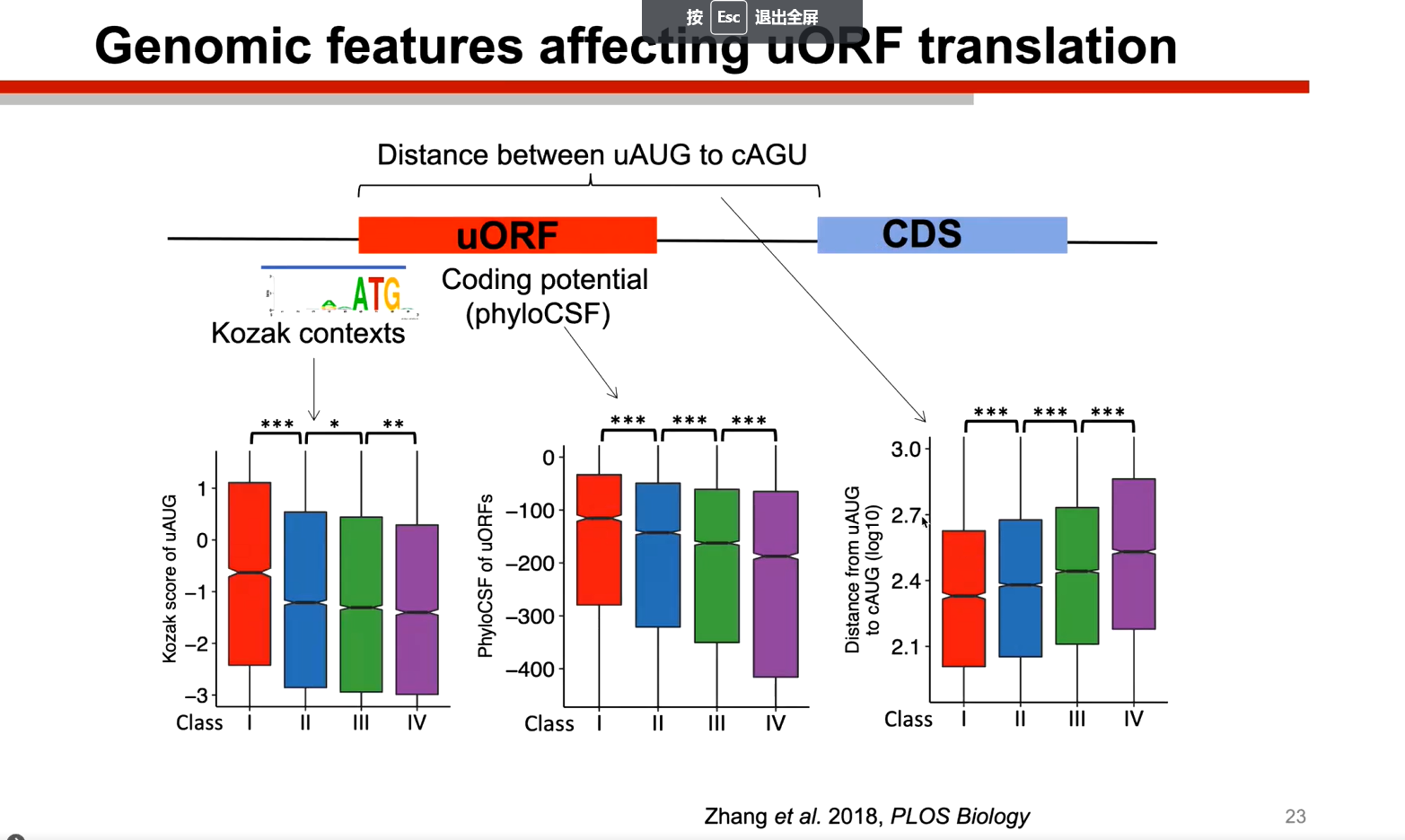


图5 影响uORF功能发挥的因素

紧接着陆教授介绍了影响uORF活性的一些因素。首先就是uORF起始密码子附近的Kozak序列的优化程度，这个机制类似于CDS区域的起始密码子，uORF起始密码子活性越高，则CDS受到的抑制越强烈。此外，uORF离CDS区域越近，也越容易抑制CDS区域的翻译。

那么uORF在进化过程中是如何演变的呢？陆教授团队采用了一种叫McDonald-Kreitman test的算法，用以计算uORF在进化过程中所面临的正向选择压力。可以看到uORF在产生之初是大量有害的，于是生物演化过程中对其进行了大量突变使其失活，而在筛选之后剩余的则是有利的，于是我们可以看到其在后期演化过程中是高度保守的。此外越是不需要的基因则演化越是保守。

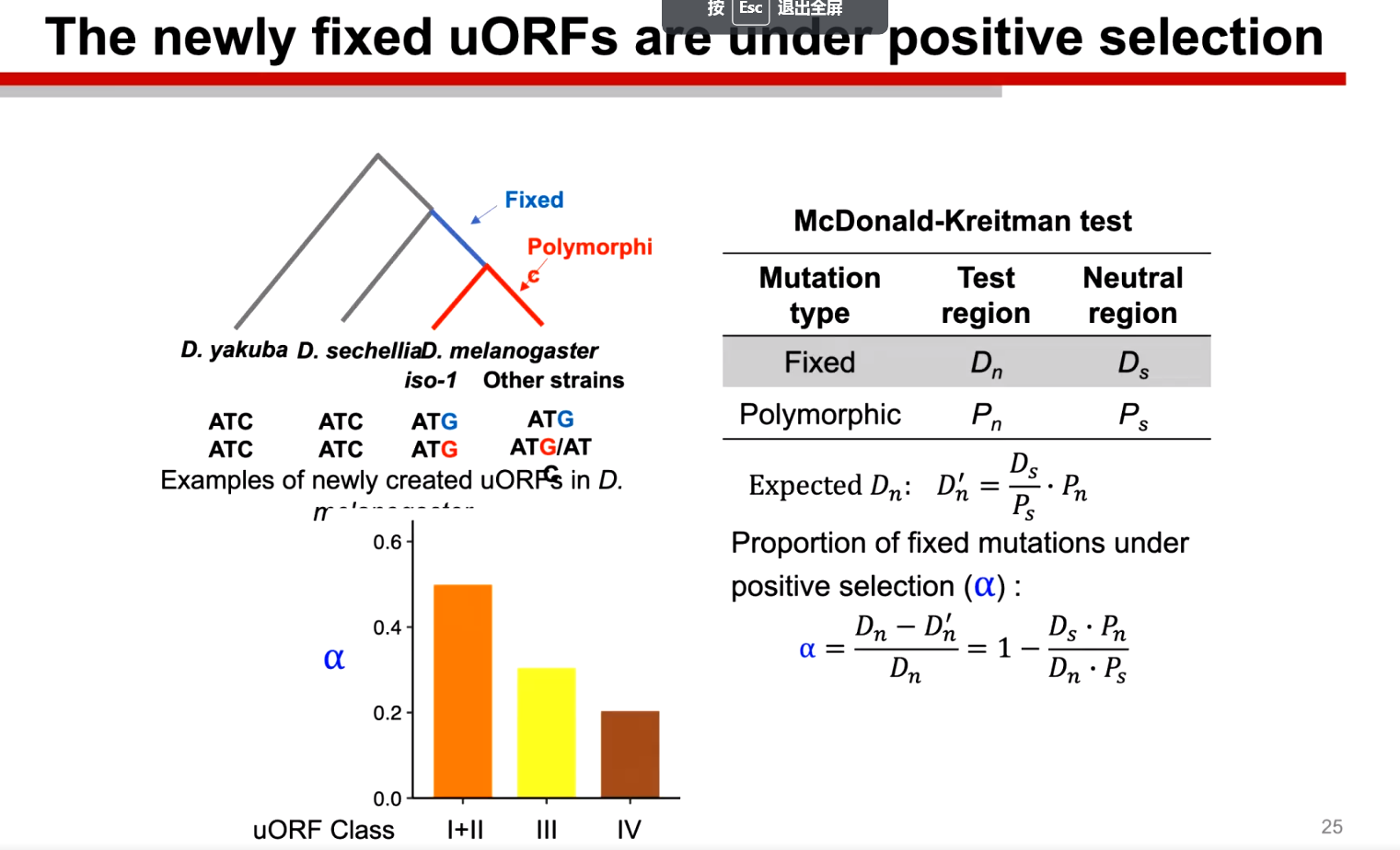


图6 uORF在生物演化中的正向选择

陆教授接下来为我们讲解uORF发挥作用的具体机制。首先uORF的数量叠加会大幅的增强其抑制效用，其次uORF自身的翻译效率也会在果蝇的不同发育阶段受到不同程度的影响，而这会反过来影响CDS区域的翻译。然后，可变剪切在5`UTR的剪切很可能影响了最终mRNA上uORF的数量，从而影响其对于CDS区域翻译的抑制程度。

陆教授对不同物种的不同基因组中的uORF分别进行了预测和观察统计，发现所有物种中的OE（观察预测比）值均小于1，且在不同物种中的差异较大。OE值越大则代表该物种中残存下来的uORF是经历过多次选择的有利变化，例如人的基因组中的uORF的OE值就比果蝇中的大，表明人类经历过的自然选择更加有效。此外，陆教授发现uORF的序列本身并非高度保守，但是在其起始信号和终止信号附近高度保守，这表明uORF所表达出来的多肽并不重要，uORF主要通过翻译起始这个事件本身来抑制后续CDS区域的翻译。

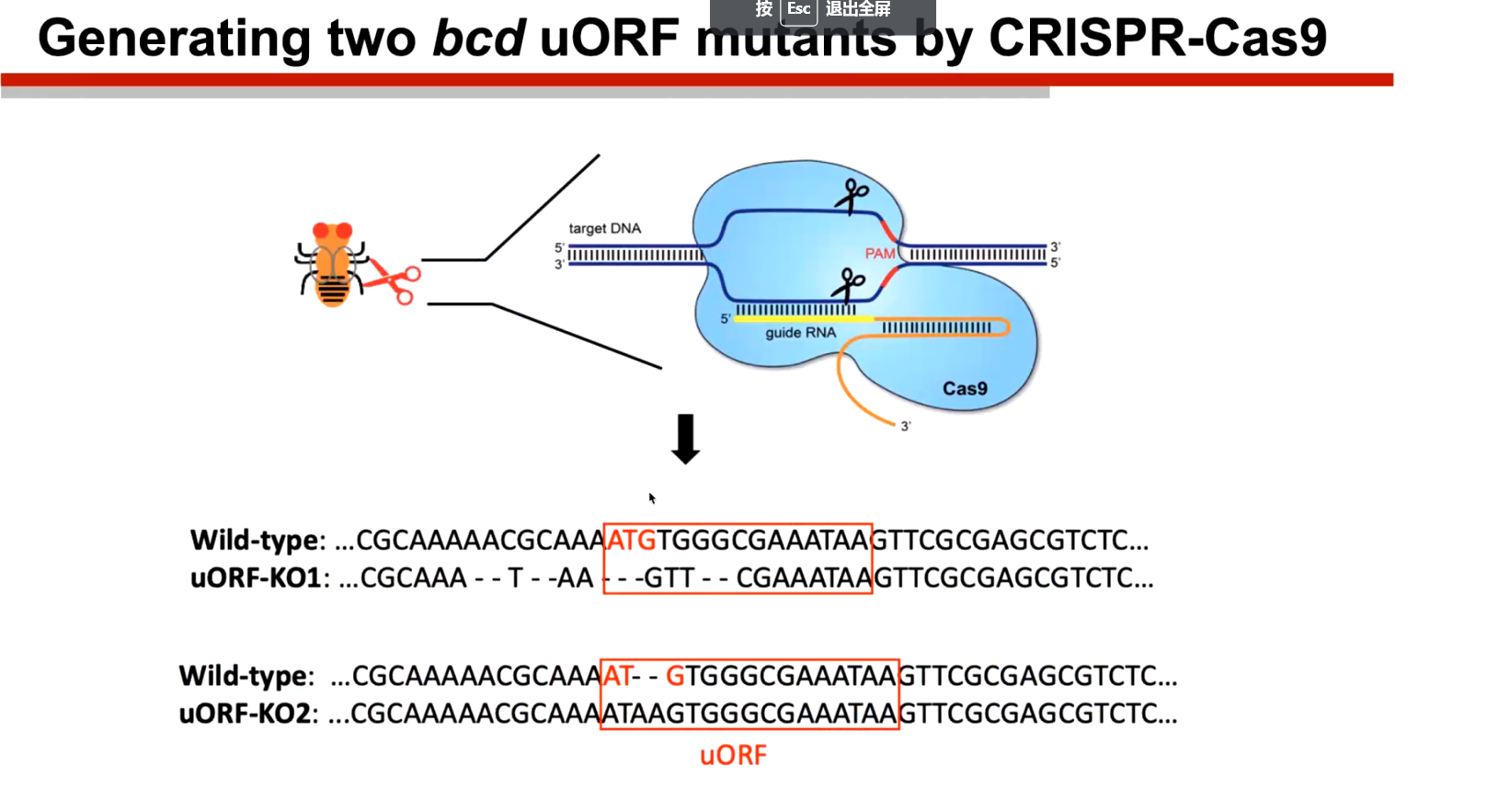


图7 uORF起始信号的突变设计

陆教授课题组的孙同学将uORF中的起始密码子进行了突变，发现CDS区域的翻译明显上升，但是这对于果蝇的发育和繁殖反而不利。不仅如此，他们还发现这种突变还会缩短果蝇的昼夜节律周期。

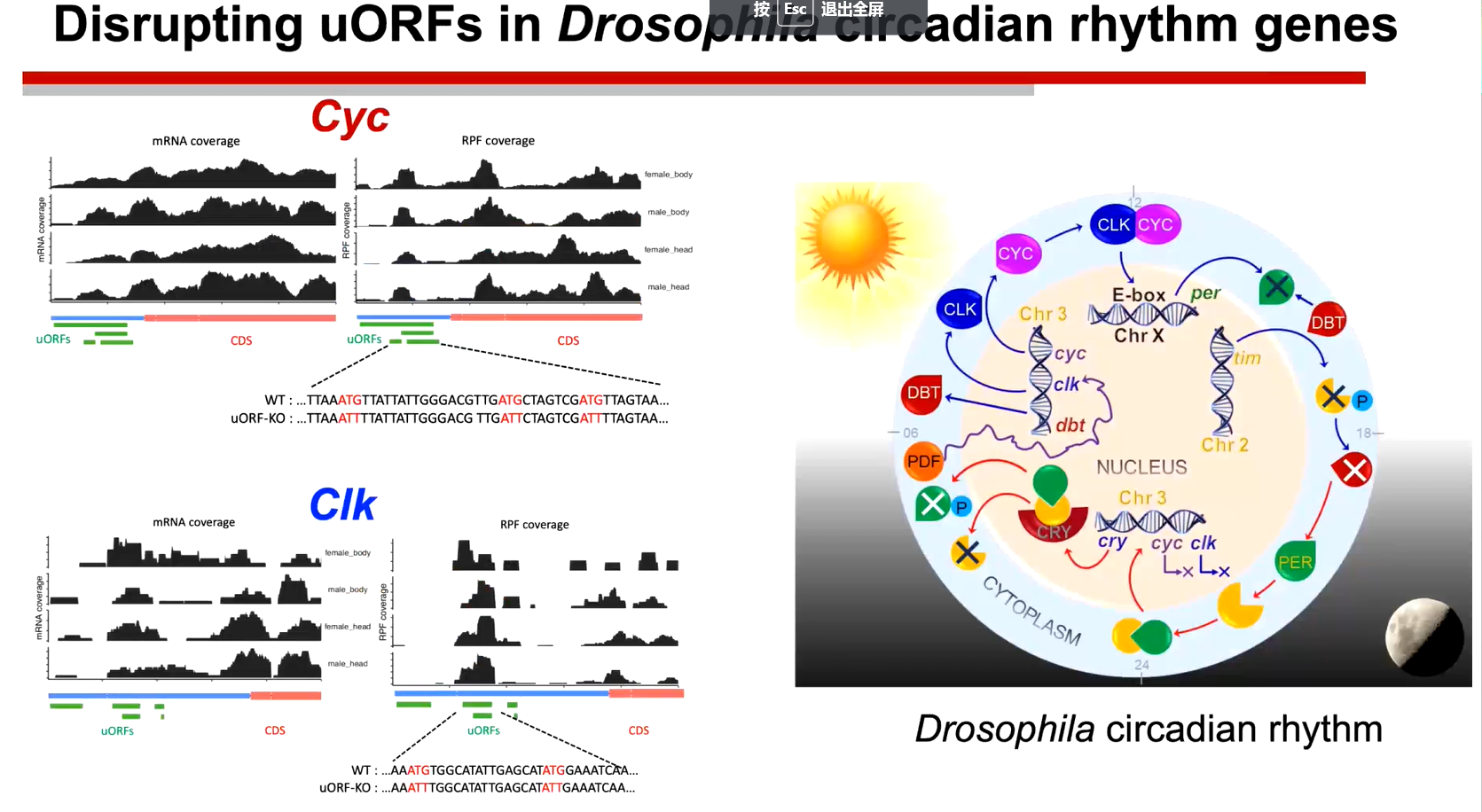


图8 uORF起始密码子突变对于生物钟的影响

陆教授课题组在全国各地收集黑腹果蝇体系，现在实验室已经保有1万5千多个体系，并基于这些数据开发了工具研究对环境的适应能力和基因变基因组变异间的关联。

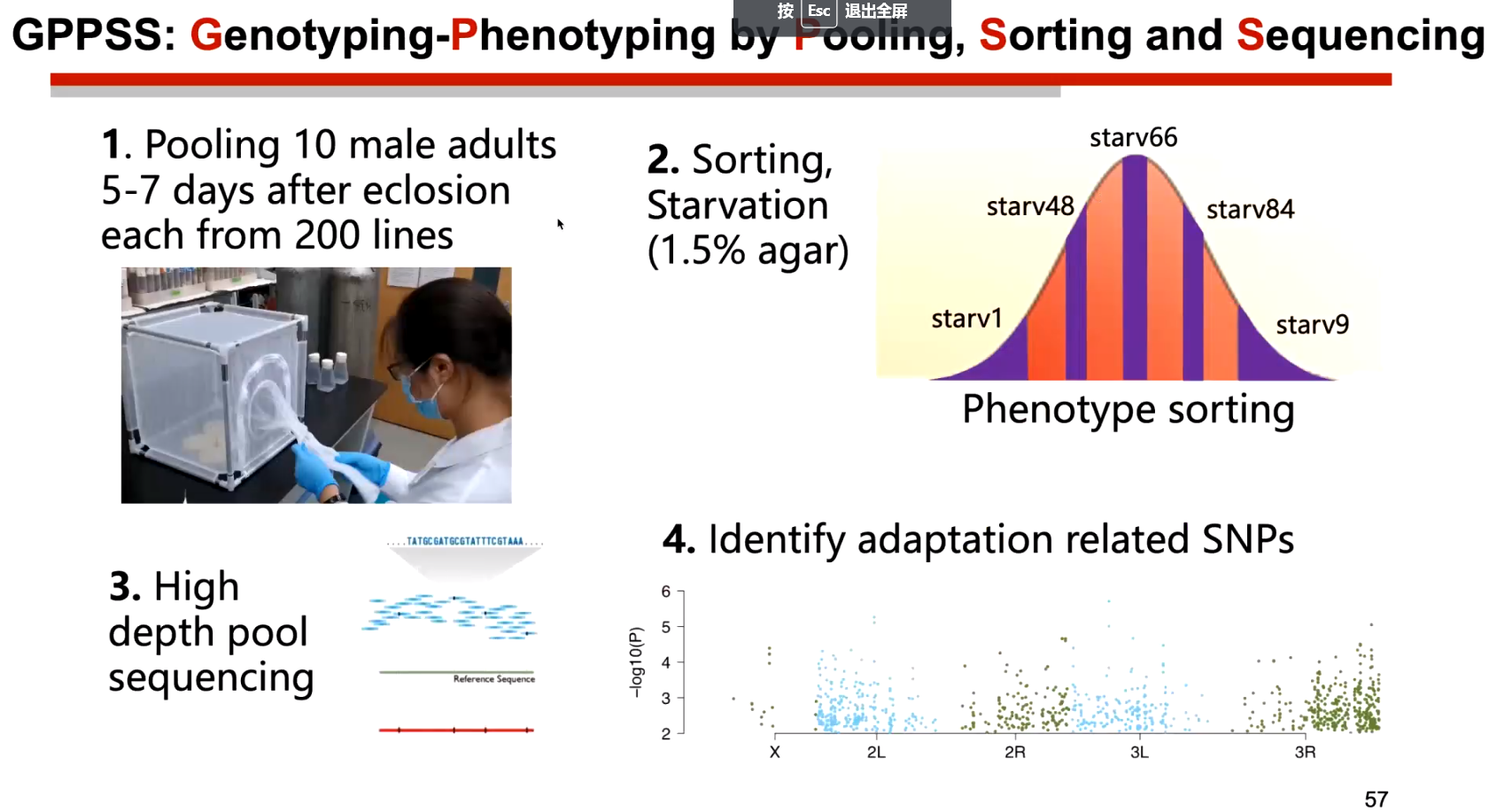


图9 对环境的适应能力和基因变基因组变异间的关联

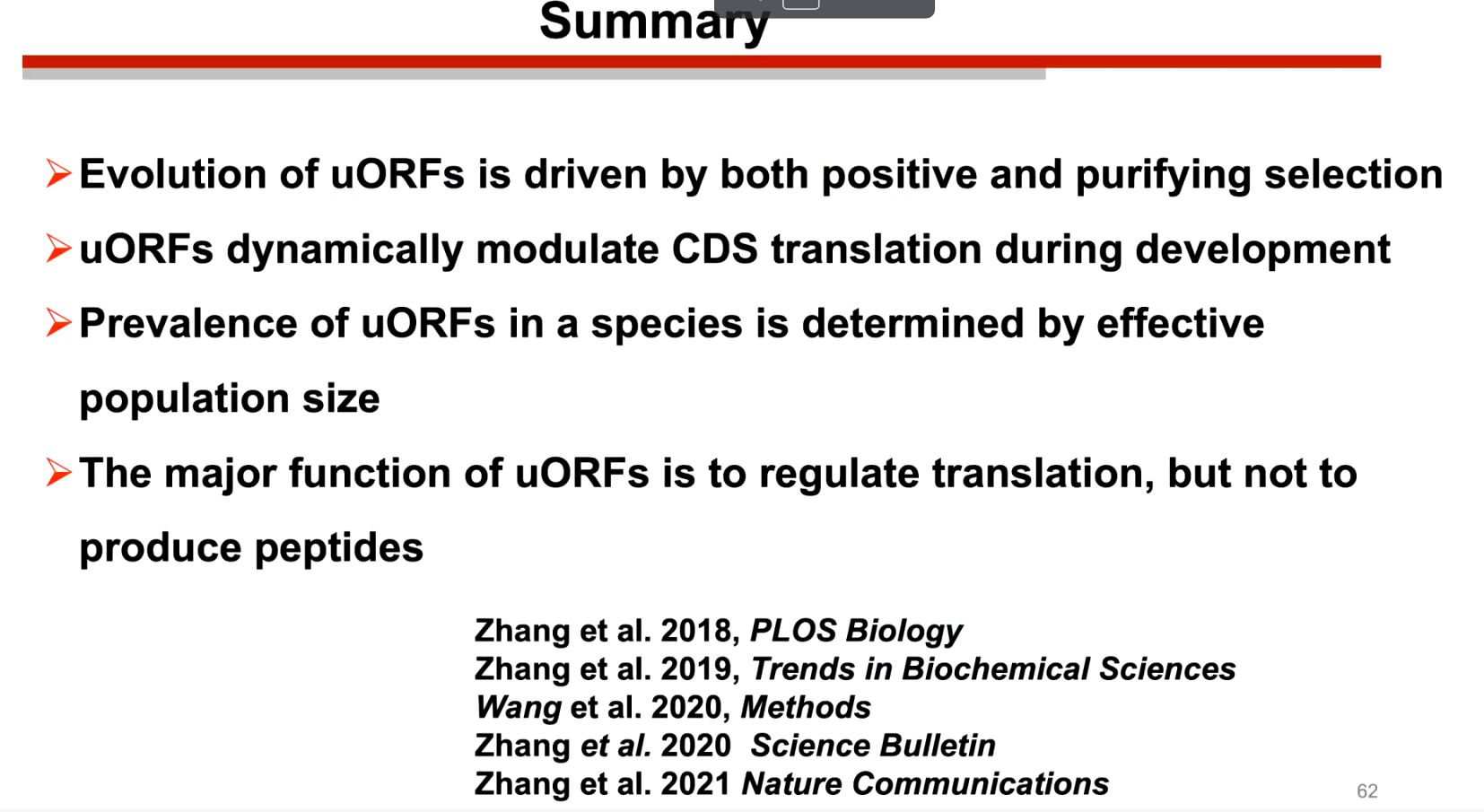


图10 uORF研究内容总结

uORF研究内容总结：

1. uORF的进化是由正选择和纯化选择共同驱动的XUORF在开发过程中动态调节CDS翻译
2. uORF在一个物种中的流行是由有效种群大小决定的
3. uORF的主要功能是调节翻译，而不是产生多肽。

接下来陆教授向我们介绍他们利用核糖体图谱技术来推测新冠病毒品系演化过程的研究。我们知道在不同生物体内或者不同的组织细胞内对于同一种氨基酸的不同同义密码子是有着不同的偏好的，而陆教授团队则通过捕捉在某一mRNA区域核糖体数量占比来推测核糖体在该区域的翻译速率。他们发现在病毒演化的过程中会逐渐抛弃掉那些翻译速率较慢的密码子，改为较快的密码子以提高整体的蛋白翻译效率。与此同时，由于病毒是寄生生物，必须使用宿主的原材料，决定了新冠病毒更加倾向于使用人类体内非偏好使用的密码子，以求得更多的tRNA资源。

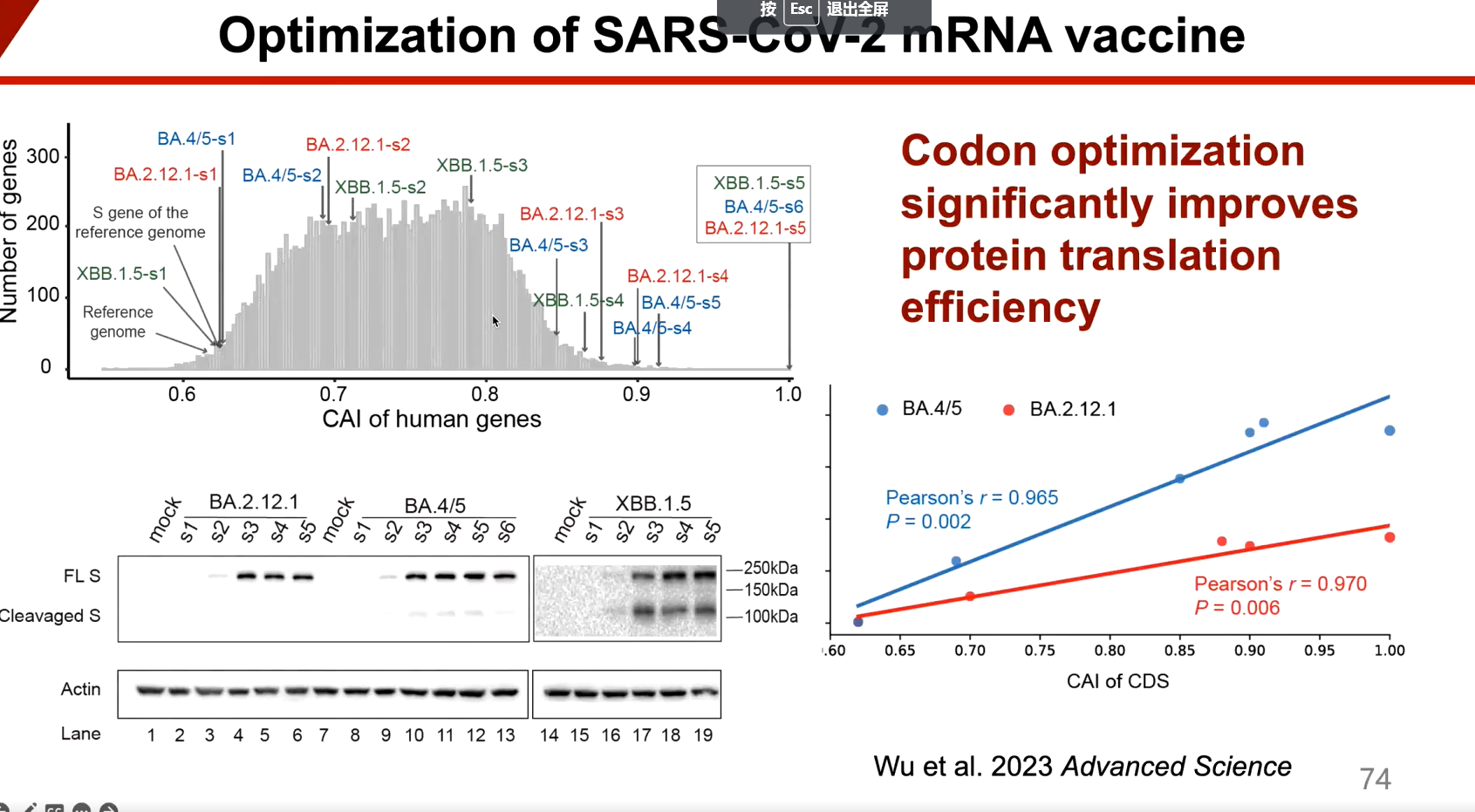


图11 新冠病毒疫苗的优化

提问

会议中许多老师和学生与教授进行互动，就目前领域上的问题进行探讨，例如：

问题：如何判断uORF是直接对下游发挥调控作用还是先翻译出一个多肽再发挥调控作用

陆剑教授回答道这个问题的回答可能就是他们课题组的一家之言了。在基因组中事实上具有很多个mRNA的拷贝，其中可能真的具有某些能够翻译出一段小多肽的uORF。但是现阶段他们课题组还是更加倾向于uORF直接对下游发挥调控作用，即它翻译出什么东西不重要，但是它翻译这个事件本身很重要。