

分子遗传学

遗传的分子基础

1.DNA结构

1) A-DNA; B-DNA; Z-DNA

2) 三聚体DNA (dsDNA-ssRNA; dsDNA-ssDNA) 结构基础和功能

dsDNA-ssRNA三聚体

dsDNA-ssDNA三聚体

dsDNA-ssRNA三聚体

dsDNA-ssDNA三聚体

3) G4 (G-quadruplex) 结构基础、基因组分布、调控和功能

结构基础

基因组分布

调控

功能

2.DNA复制

复制的起始与延伸过程：

DNA 复制如何防止引入超螺旋？ 拓扑异构酶

如何解决复制导致的DNA末端磨损？

遗传的细胞基础

基因组学

基因表达调控

染色质活化与转录的关系：

染色质活化的“开关”作用：

表观遗传学

1. DNA甲基化

2. RNA甲基化

3. 组蛋白修饰

组蛋白修饰类型：

赖氨酸甲基化：

精氨酸甲基化：

组蛋白乙酰化

组蛋白变体

mRNA与蛋白质组学

非编码RNA

Short interfering RNA (siRNA)

Micro RNA (miRNA)

Double-stranded RNA (dsRNA)

Short heterochromatic RNA (shRNA)

Transcripts from repeated sequences (ALU, LTR)

Long non-coding RNA (lncRNA)

Circular RNA (circRNA)

遗传重组与基因编辑

1. 断裂基因：

2. 重叠基因：

3. 基因组

3.1. 概念

3.2. 遗传标记

3.3. 分子标记

3.4. 人类基因组计划

3.5. 人类基因组特征

人类基因组的组成比例、来源及其特点如下：

3.6. mRNA质量控制

3.6.1. 基因结构

3.6.2. mRNA预处理

3.6.3. RNA编辑

3.6.4. RNA降解

体细胞重组学说

抗体遗传

基因突变

DNA损伤修复

现阶段基因的概念：一段可以转录为功能性RNA的特定DNA（RNA）区域，它包含调节、转录和其它功能区域，如启动子、中止子；可以重迭、断裂的形式存在。

遗传的分子基础

1.DNA结构

1) A-DNA； B-DNA； Z-DNA

A-DNA、B-DNA和Z-DNA是DNA的 three forms（三种形态），它们在结构上有所不同，主要表现在螺旋的形状、碱基对的排列方式和螺旋的紧密程度上。

A-DNA

- 结构特点：
 - A-DNA是一种右手双螺旋结构。
 - 它的螺旋较宽，螺旋轴相对较直。
 - 每个螺旋包含11个碱基对。
 - 碱基对倾斜相对较大，导致螺旋的沟槽较宽。
- 环境：
 - A-DNA在相对干燥的条件下或在某些盐浓度较高的环境中更稳定。

B-DNA

- 结构特点：
 - B-DNA是最常见的DNA形态，也是一种右手双螺旋结构。
 - 它的螺旋较紧，螺旋轴呈轻微的扭曲。
 - 每个螺旋包含大约10.5个碱基对。
 - 碱基对相对垂直于螺旋轴，沟槽较窄。
- 环境：
 - B-DNA在细胞内的大多数环境中都非常稳定，是DNA在生理条件下的主要存在形式。

Z-DNA

- 结构特点：
 - Z-DNA是一种左手双螺旋结构。
 - 它的螺旋非常紧，呈明显的曲折形状。

- 每个螺旋包含大约12个碱基对。
- 碱基对呈现顺式倾斜，导致Z-DNA的沟槽非常狭窄。

- **环境：**

- Z-DNA在特定的序列上下文出现，通常是富含GC的序列，并且在这些序列中，当DNA分子处于扭曲张力较大时形成。

这三种DNA形态的转换与DNA分子的序列、环境条件（如离子浓度、pH值、温度）以及分子内部的应力状态有关。这些不同的结构形态可能影响基因的表达和调控，因为它们可能会改变DNA与蛋白质（如转录因子）的相互作用方式。此外，DNA形态的变化也可能在疾病发生和发展中起到作用。

	旋向	螺距 (nm)	碱基数 (每圈)	螺旋直径 (nm)	骨架 走行	存在条件
A型	右手	2.53	11	2.55	平滑	体外脱水
B型	右手	3.54	10.5	2.37	平滑	生理条件
Z型	左手	4.56	12	1.84	锯齿型	CG序列

2) 三聚体DNA (dsDNA-ssRNA; dsDNA-ssDNA) 结构基础和功能

RNA·DNA-DNA triplex来源与分布

lncRNA, 来源于与靶向双链 DNA不同的座位;

(反式) pRNA, 来源于调控基因的启动子区域;

(顺式) dsDNA/ssRNA triplexes 发生于RNA多聚嘧啶序列, 且 RNA序列中U相比C更容易形成三聚体复合物。

生物信息手段预测人类基因组中存在大量具有形成triplex 结构的基序, 并且大部分位于基因调控区域, 如: 启动子 区和增强子区;

功能: 基因表达调控; 同源重组。

lncRNA·DNA-DNA三聚体 功能: 通过lncRNA结合的染色质修饰蛋白调控基因表达 (PRC2: H3K27三甲基化修饰; P300 : 组蛋白乙酰转移酶)

dsDNA-ssRNA三聚体

1. 结构特点：

- 这种结构通常由双链DNA（dsDNA）和单链RNA（ssRNA）组成。
- 在这些复合物中，ssRNA可能与dsDNA的一个链形成碱基配对，而另一个DNA链则形成单链区域。
- 这种配对可以是完全的或部分的，取决于RNA和DNA序列的互补性。

2. 形成机制：

- 这些结构可能在病毒复制过程中形成，例如逆转录病毒，其中病毒RNA基因组被逆转录成DNA。
- 在某些情况下，RNA可以与DNA的模板链配对，形成RNA-DNA杂交区域。

dsDNA-ssDNA三聚体

1. 结构特点：

- 这种结构由双链DNA和单链DNA组成。
- 单链DNA可能与双链DNA的一个链形成互补配对，导致一个DNA双链和一个DNA-RNA杂交链的形成。

2. 形成机制：

- 这些结构可能在DNA复制、重组和修复过程中形成。
- 例如，在DNA复制叉处，新合成的DNA链可能与模板链形成短暂的杂交结构。

功能：

dsDNA-ssRNA三聚体

1. 病毒复制：

- 在逆转录病毒中，dsDNA-ssRNA三聚体是逆转录过程的一个重要中间体，其中病毒RNA被逆转录酶转化为DNA。

2. 基因表达调控：

- RNA-DNA杂交结构可以影响基因的转录和剪接过程。

3. RNA干扰：

- 在某些生物中，RNA-DNA杂交结构可能参与RNA干扰途径，调控基因表达。

dsDNA-ssDNA三聚体

1. DNA复制：

- 在DNA复制过程中，ssDNA可能与模板链形成短暂的杂交，以促进新链的合成。

2. DNA修复:

- 在同源重组修复过程中, ssDNA可能与同源DNA序列形成三聚体结构, 促进DNA损伤的修复。

3. 基因转换:

- 在某些生物中, dsDNA-ssDNA三聚体可以促进基因间的序列交换, 增加遗传多样性。

这些三聚体结构的生物学意义在于它们能够调节核酸的相互作用和功能, 影响基因表达、DNA复制和修复等关键生物学过程。对三聚体DNA结构的深入研究有助于我们更好地理解这些过程, 并可能为疾病治疗提供新的靶点。

3) G4 (G-quadruplex) 结构基础、基因组分布、调控和功能

G-quadruplex:富含鸟嘌呤的DNA (RNA) 序列通过氢键 (Hoogsteen) 和一价阳离子(K^+ , Na^+)辅助折叠形成的 稳定的局部四链右手螺旋, 在其螺旋的每层都有由鸟嘌呤组成的特征性共平面四边形 (G4) 。

G4 复合体的结构:

杂合型指的是 G4的形成既包含平行链, 又包含反向平行链; 因此: G4平面的4个糖苷键排列方式既包含顺式也包含反式

G4 复合体分布与结构调控 :

算法预测了人类基因组中超过370,000个G4序列基序, 尤其与基因组调控相关的区域(如端粒、启动子和非翻译区域)观察到普遍富集。

除此之外: 免疫球蛋白基因Switch区, 核糖体DNA, 线粒体 DNA, 复制起始区, LINE-1反转录转座子, 转录中的DNA:RNA杂交体 G4

rG4: RNAG4 (lncRNA)

G4 的形成依赖于G4 motif, 如: GGGNGGG(Myc promoter), TTAGGG (telomere) , 某些启动子序列可以在一个重叠区域或在单独区域形成多个G4复合体。G4结构的形成与分布依赖于细胞类型和生理状态, 是高度动态的。

Telomeric G4结构的调控: 通过蛋白相互作用以细胞周期 依赖性的方式进行调控 。

G4 复合体功能 :

Transcription

Replication

Translation

Genome instability

Inhibite homologous recombination

Epigenetic regulation, connecting to cancer.

如：小鼠端粒延长调节因子解旋酶1 (RTEL1)被证 明可以通过解除端粒G4s来维持端粒的完整性。

G4 复合体的潜在应用：

G4 ligand: 能够结合并稳定G4结构的蛋白质

由于G4结构能够抑制多种癌基因的表达，已经作为肿瘤治疗的靶点开展了研究。

端粒G4可以通过抑制端粒酶活性的小分子靶向治疗癌症，而原癌基因c-MYC启动子的G4能够控制基因转录，可以靶向影响转录。

G4 (G-quadruplex) 结构是一种由富含鸟嘌呤 (G) 的核酸序列形成的稳定的四链结构。

结构基础

1. 构成单元：

- G4结构由四个鸟嘌呤碱基通过Hoogsteen氢键相互配对形成平面，称为G-四分体 (G-tetrad)。
- 这些G-四分体堆叠在一起，形成四链结构，通常由一个或多个G-四分体层组成。

2. 茎和环：

- G4结构中，G-四分体之间由单链DNA或RNA的环 (loop) 连接，这些环可以是单核苷酸或多核苷酸。
- 四个G-四分体之间的连接可以形成茎 (stem)，由碱基配对形成。

3. 离子稳定：

- G4结构的稳定性通常由钾离子 (K⁺) 或其他单价阳离子通过离子桥与G-四分体的鸟嘌呤碱基相互作用来维持。

基因组分布

1. 富集区域：

- G4结构在基因组中的分布是不均匀的，通常富集在基因的调控区域，如启动子、转录起始位点、剪接位点、终止子和端粒。

2. 序列特征：

- 形成G4结构的DNA序列通常具有连续的鸟嘌呤重复序列，如G₃N₁-G₃N₁-G₃N₁-G₃，其中N代表任意核苷酸。

调控

1. 蛋白质相互作用：

- 许多蛋白质可以与G4结构相互作用，包括G4解旋酶、G4结合蛋白和其他调控因子，这些蛋白质可以影响G4结构的稳定性和功能。

2. 表观遗传调控：

- G4结构可以通过影响染色质结构和DNA的可及性来调控基因表达。

功能

1. 基因表达调控：

- G4结构可以通过阻碍转录因子和RNA聚合酶的访问来抑制基因转录。
- 在某些情况下，G4结构也可以促进基因表达。

2. DNA复制和修复：

- G4结构在DNA复制叉前形成，可能会阻碍复制过程，导致复制叉停滞或DNA损伤。
- 在DNA修复过程中，G4结构可能参与同源重组或非同源末端连接。

3. 端粒维护：

- 端粒区域的G4结构在维持端粒稳定性和功能方面起着重要作用。

4. 遗传多样性：

- G4结构可能导致DNA重组和重排，增加遗传多样性。

5. 疾病相关性：

- G4结构的异常可能与多种疾病相关，包括癌症、神经退行性疾病和某些遗传性疾病

2.DNA复制

什么是DNA复制：在每次细胞分裂过程中、整个基因组都 精确地复制一次。（无论细胞含有单条或多条染色体。）

复制子：DNA中发生一次复制的单位被称为复制子（replicon），复制子含有的基本单元：复制起始位点和终止 位点。原核生物基因组，如：细菌等，只有一个复制子。

质粒、细胞器或病毒的基因组也仅有一个复制子，但是常常 完成多拷贝复制。

真核生物染色体含有多个复制子，每条染色体有多个复制起点，并且，一般来说多个复制子是同时启动复制的。每个复制子大小不同，如：约40kb（酵母），约100kb（哺乳动物）。复制速度约2000bp/min，而细菌约为50000bp/min。

复制的起始与延伸过程：

1) ORC复合体识别复制起始位点：（yeast: ARS1; human: non-specific）；

- 2) 前复制起始复合体 (Pre-RC) 的形成: 在M期~G1期的早期, ORC将CDC6和CDT1募集到复制起点, 并募集MCM2-7形成双6聚体 (OCCM)
- 3) 复制起始复合体 (RC) 的形成: 在S期, CDC45和DNA复制复合物Go-Ichi-Ni-San (GINS) 与MCM2-7六聚体结合形成CDC45 / MCM2-7 / GINS (CMG) 复合物, 激活MCM的解旋酶活性。
- 4) 复制延伸: (elongation) MCM的解旋酶激活后促进DNA解链 (前导链后随链的复制 分别使用DNA聚合酶 ϵ 和 δ , 而RNA引物合成使用的是DNA聚合酶 α)
- 5) 复制完成 (maturation)

DNA 复制如何防止引入超螺旋? 拓扑异构酶

拓扑异构酶 (Topoisomerases) 改变DNA拓扑结构 (缠绕圈数), 而不改变DNA上核苷酸排列方式的酶. DNA拓扑结构: DNA双链序列不变的情况下, 缠绕数的差异所形成的结构。

拓扑异构酶与单链或双链 DNA 结合, 切断 DNA 的磷酸骨架。

这种断裂的中间产物使 DNA 得以解开或缠绕, 在这些过程结束后, DNA 骨架再次被重新封闭上。

I 型拓扑异构酶切断 DNA 双螺旋的一条链, 引发松弛, 允许另一条链穿过断点, 然后切断的链重新闭合。

II 型拓扑异构酶切断一条 DNA 双螺旋链的两条链, 将另一条未断开的 DNA 螺旋线穿过它, 然后将切断的链重新闭合。

The type and function of Topoisomerases

Topoisomerase	IA	IB	IIA	IIB
Metal Dependence	Yes	No	Yes	Yes
ATP Dependence	No	No	Yes	Yes
Single- or Double-Stranded cleavage?	SS	SS	DS	DS
Cleavage Polarity	5'	3'	5'	5'
Change in L	± 1	$\pm N$	± 2	± 2

真核生物拓扑异构酶 I 和 II 是越来越多抗癌药物的靶点，这些药物通过阻断 DNA 断裂后的连接反应来抑制TOPI和TopII酶 喜树碱，羟基喜树碱，拓扑替康，伊立替康

Topoisomerase II inhibitors, such as anthracyclines (蒽环类), are amongst the most widely used anti-cancer agents. 例如：依托泊苷能嵌入到DNA-TOP2cc复合物之间，阻止 DNA重新连接。

Merbarone (硫巴龙) 能够抑制TOP2的切割功能，阻止 DNA切割的发生。

Topoisomerase II inhibitors 能够导致广泛的染色体异常、进而导致细胞周期阻滞于G2期，或者导致细胞凋亡。

如何解决复制导致的DNA末端磨损？

- 1) Special telomere sequence: tandem repeats of TTAGGG (human) 2) T环加D环结构 链突出3'末端回折并侵入到双链之中，替换原有双链中的一条形成两个环，大的T环，小的D环。
- 3) G4 structure 4) Shelterin 复合体（端粒蛋白复合体：含TRF1的复合体、含TRF2的复合体、TPP1-POT1）端粒末端双环结构通过与TRF1和TRF2复合体结合构成 Shelterin 复合体，保护DNA末端、阻止DNA损伤反应（ATR/ATM信号）其中TRF1 复合物参与端粒长度控制，而 TRF2 复合物则起到端粒保护端盖的作用。
- 5) Telomerase, a specific enzyme with integrated RNA template. (TERT, TERC)

端粒的概念

端粒酶复合体

端粒的组成与结构：

T环和D环的形成

端粒结合蛋白：Shelterin 复合体

避免复制导致的DNA末端磨损

端粒酶延长染色体端粒机制

通过端粒酶的逆转录酶活性向染色体前导链端粒末端 3' 突出的DNA上添加TTAGGG重复序列。 经过端粒酶所催化的端粒末端“合成-转位-合成”的循环 不断添加端粒DNA重复序列单元到前导链端粒末端。 最后，通过新的冈崎片段合成延长后随链。 如果端粒酶活性下降会发生？

端粒缩短、端粒酶活性调控与细胞 衰老及疾病

端粒病： 调节端粒长度平衡的基因缺陷可能导致各种疾病，统称为 端粒病，患者表现出多种症状，如早衰和癌症风险增加， 表明端粒平衡在人类健康中的重要性。 在端粒病患者中发现了端粒酶和端粒维持基因的突变。

端粒长度干预

总结问题：

1. How do the cells maintain DNA double helix conformation stable?

细胞如何保持DNA双螺旋结构的稳定性？

- DNA双螺旋的稳定性依赖于氢键以及碱基堆积力。氢键使得碱基对（如A-T和C-G）之间有相对弱但足够的结合力，而碱基堆积是指碱基对之间的相互作用进一步增强DNA结构的稳定性。此外，拓扑酶、组蛋白等蛋白质的作用有助于维持DNA的整体稳定结构。

2. DNA-RNA, DNA-DNA Triplex

- DNA-RNA 和 DNA-DNA 三链结构是非典型的DNA构象。在一些调控机制中，特别是基因表达和修复过程中，DNA可以形成三链结构，如RNA与DNA的互补链形成三链复合物或DNA自身形成的三链结构。这种结构可能在基因调控中起作用，尤其是在特定的DNA序列区域。

3. G4 structure

G4结构（G-四链结构）

- G-四链结构是一种由富含鸟嘌呤的DNA序列形成的特殊四链结构。这种结构在端粒、启动子等区域有较多出现，可能参与调控基因表达、染色体稳定性和端粒维护。G4结构的形成与解构在细胞中是通过特定的酶（如解旋酶）进行调控的。

4. How do the cells solve the problem of overwinding of DNA during replication?

细胞如何解决DNA在复制过程中过度缠绕的问题？

- 细胞通过拓扑异构酶（topoisomerase）解决DNA在复制时的超螺旋问题。拓扑异构酶可以切割并重新连接DNA链，释放缠绕和张力，确保复制叉正常前进。

5. Telomeres and TERT complex (TERC)

端粒和端粒酶复合物（TERC）

- 端粒是染色体末端的重复DNA序列，端粒酶复合物是端粒酶（TERT）和端粒RNA组分（TERC）的组合。TERT是负责将端粒延长的酶，TERC则是一个RNA模板，指导TERT进行端粒延长。该复合体对于保持端粒长度和细胞的增殖能力至关重要。

6. Why are telomeres essential for chromosome stability? What is the role of telomerase during cell division?

端粒为何对染色体稳定性至关重要？端粒酶在细胞分裂中起什么作用？

- 端粒在染色体末端充当“保护帽”，防止染色体末端的损伤和不正常融合。如果没有端粒，染色体会逐渐缩短，导致基因丢失。端粒酶通过延长端粒，确保细胞可以经历更多次分裂而不会丢失重要的遗传信息。

7. Why do the telomeres shortening occur during differentiation of the somatic cell?

为什么在体细胞分化过程中会发生端粒缩短？

- 在体细胞分化过程中，端粒酶的活性显著降低，导致端粒不能被延长。每次细胞分裂时，端粒都会缩短，这种现象随着分裂次数的增加而累积，最终导致细胞衰老或凋亡。

遗传的细胞基础

1. 基因和染色体行为的平行性：

减数分裂过程中染色体的行为与等位基因的分离、自由组合相偶联。

减数分裂前期I中发生的DNA断裂和染色体配对与互换行为和非等位基因间的连锁互换相偶联。

减数分裂中染色体行为是遗传分离、自由组合、连锁与互换的基础！

2. 染色体的结构特征：

染色质：首先由Flemming(1882年)提出，指间期细胞核内能被碱性染料着色的物质。现在指由DNA、组蛋白、非组蛋白和少量RNA组成的线性复合结构。

染色体：指细胞在有丝分裂或减数分裂过程中，由染色质聚缩而成的棒状结构，是DNA螺旋化的最高形式。

3.染色体结构：形态结构

着丝粒的位置，端粒，次缢痕，随体，染色单体

染色体的功能实现：着丝粒，端粒，复制起点

染色体包装：染色体的超微结构：核小体(nucleosome)—基本单位，螺线管结构，也叫染色质纤维，支架环结构（如：果蝇染色体—多线染色体，染色单体。

4.常染色质：细胞间期核内纤维折叠盘曲程度小，分散度大，染色较浅且具有转录活性。在间期核中，染色质以一个直径在5–24nm之间的无序链的存在形式，以不同的浓度密度排列在一起。染色质体积浓度(CVCs, chromatin volume concentration): $\text{chromatin volume} / 120\text{-nm}^3 \text{ nuclear volume}$.描述间期核中染色质分布的不均一性

在间期核中，CVC范围为12% ~ 52%，平均值为 $30 \pm 10\%$ 。核膜上的异染色质结构区域：CVC(37 – 52%)，靠近核中心的亚细胞核区域：CVC(12 – 21%)。

异染色质：细胞间期核内纤维折叠盘曲紧密，呈凝集状态（Strongly condensed），染色较深且没有转录活性。

组成型异染色质：指各类细胞的全部发育过程中都处于凝缩状态的染色质。大多位于着丝粒区和端粒区，不具有转录活性或具有较低转录活性。

兼性异染色质：指在特定细胞的某一发育阶段所具有的凝缩状态的染色质。X染色质，其它调控基因表达的染色体区域。

ChromEMT：可视化和重建细胞核中染色质超微结构和3D组织。

局部常染色质–异染色质的结构调控：

启动子区域的高甲基化导致染色体凝缩和基因转录的沉默

组蛋白的高甲基化及低乙酰化导致染色体凝缩和基因转录的沉默,

5.异染色质定位与发育

核内膜 (INM) 可通过与 HDAC 或其他转抑制因子、异染色质结构蛋白相互作用来抑制基因表达。异染色质基因座通过与核周膜蛋白结合而定位。在核内膜, 从而产生不同的异染色质灶, 这些部位能够封存异染色质结构蛋白染色质的核内膜周边定位也受发育控制, 在分化细胞中对维持谱系特异性的表达模式更为关键。

6.哺乳动物基因组三维(3D)层级结构单元为: 染色体疆域: 间期细胞核中每条染色体占据独立分离的区域, 称之为染色体疆域(CTs)。基因排列密集区域位于核基质区域, 基因稀少区域位于核膜内侧; 这种非随机分布的染色体疆域, 具备保守性, 调控DNA复制、基因转录、DNA损伤修复、细胞分化等。一旦发生3D基因组结构异常会导致疾病发生。染色质区室是由基因组表观状态所决定的较大的结构单元, 与染色质活性密切相关,表达活跃区室 (A) 和表达沉默区室 (B)、拓扑关联结构域 (TAD) 是细胞核内稳定存在的独立空间结构单元,在局部范围内介导基因的表达调控; 1Mb。染色质环是直接调控基因表达的最精细的结构和功能单元,通常由启动子与远端增强子互作形成,在介导基因的转录激活中发挥着重要作用。(中位数: 185kb, CTCF、Cohesin、TF等)。

总结问题:

- 1) 真染色质和异染色质在染色体上是如何分布的?
- 2) 局部染色体结构的构建如何受到调控。
- 3) 基因变异是如何形成的?
- 4) 染色体区域及其在基因调控中的意义。
- 5) 里昂假说。

基因组学

基因表达调控

概念：

结构特征与调控特征：

调控的时空性：阶段特异性、组织细胞特异性

调控的方式与途径：

原核生物：乳糖操纵子和色氨酸操纵子

真核生物：转录水平的调控、转录后水平的调控、翻译水平上的调控

生物学意义：

染色质结构的调控：

1) 染色质重塑 与核小体相位改变

染色质重塑是指染色质结构在不改变DNA序列的情况下发生改变的过程。染色质重塑复合物通过以下方式改变染色质结构：

1. **核小体滑动**：染色质重塑复合物可以改变核小体在DNA上的位置，使得某些基因区域变得更容易或更难与转录因子和其他调控蛋白接触。
2. **核小体移除或重新组装**：染色质重塑复合物可以移除或重新组装核小体，从而暴露或隐藏DNA序列，影响基因表达。
3. **改变核小体结构**：通过改变组蛋白与DNA的结合方式，影响核小体的稳定性及其与DNA的结合紧密程度。

核小体相位改变是指在基因的转录区，核小体的位置和排列可以发生特定的变化，以适应转录过程的需要。核小体通常以特定的相位（例如，一个核小体覆盖约147个碱基对，DNA在组蛋白八聚体上的卷曲产生约10个碱基对的链接DNA）排列在DNA上。相位改变可能包括：

1. **相位移动**：核小体在DNA上的位置可能会发生移动，导致相位改变，这样转录起始位点附近的DNA可以被转录机器访问。
2. **相位展开**：在转录活跃的区域，核小体可能会部分或完全解开，使得转录因子和RNA聚合酶可以直接接触到DNA。
3. **相位重排**：核小体相位改变可能导致核小体阵列的整体重排，从而影响染色质的高级结构。

2) 染色质的区间性

染色质的区间性（Chromatin compartmentalization）是指染色质在细胞核内不是均匀分布的，而是被组织成不同的区域或区间，每个区域具有不同的结构和功能特性。这些区间性的特征主要体现在以下几个方面：

1. 常染色质与异染色质：

- **常染色质**：是一种较为松散的染色质结构，基因在这个区域是活跃的，可以被转录。
- **异染色质**：是一种较为紧密的染色质结构，基因在这个区域通常是沉默的，不易被转录。

2. 核小体定位和密度：

- 在某些区域，核小体紧密排列，形成致密的染色质结构，这些区域通常转录活性较低。
- 在其他区域，核小体排列较为疏松，形成较为开放的染色质结构，这些区域转录活性较高。

3. 染色质域：

- 染色质可以被划分为不同的域，这些域可以是大的结构单元，如染色体带（chromosome bands）或染色质区室（chromatin compartments）。
- **A区室**：通常与活跃的基因表达相关，染色质较为开放。
- **B区室**：通常与基因沉默相关，染色质较为紧密。

4. 拓扑关联结构域（TADs）：

- TADs是染色质在三维空间中形成的结构域，它们通过染色质互作将基因调控元件和它们的靶基因组织在一起，即使在长距离上也能保持相互作用。

5. 染色质环：

- 染色质可以形成环状结构，这些环可以将远距离的基因调控元件和基因启动子拉近，促进基因的转录调控。

染色质区间性的生物学意义包括：

- **基因调控**：通过将活跃和沉默的基因区域分隔开，细胞可以更有效地调控基因表达。
- **基因组稳定性**：紧密的染色质结构有助于保护DNA免受损伤，而开放的染色质结构则便于DNA修复。
- **细胞分化**：在细胞分化过程中，特定的基因需要在特定的时间和空间被激活或沉默，染色质区间性有助于这一过程的精确调控。
- **染色体组织**：染色质的区间性有助于染色体的组织和维护，避免染色质之间的交叉和紊乱。

染色质区间性的研究是当前表观遗传学和基因组学领域的前沿课题，对于深入理解基因表达调控和细胞核内的复杂生物学过程具有重要意义。

3) 染色质模板的转录 染色质活化是转录前的基因调整的开关

染色质模板的转录是指RNA聚合酶在染色质上特定的DNA序列上开始合成RNA的过程。在这个过程中，染色质活化（Chromatin Activation）是转录前的一个重要步骤，它涉及到基因表达调控的“开关”机制，

以下是一些关键的点：

染色质活化与转录的关系：

1. 染色质重塑：

- 在转录开始前，染色质重塑复合物会作用于染色质，使得核小体结构变得松散，从而暴露出DNA模板，使得RNA聚合酶和其他转录因子可以访问到基因的启动子区域。

2. 组蛋白修饰：

- 染色质活化通常伴随着组蛋白的修饰，如乙酰化、磷酸化、甲基化等。这些修饰可以改变染色质的结构，使其变得更加开放，有利于转录的进行。
- 例如，组蛋白H3的赖氨酸4甲基化（H3K4me3）通常与转录活跃的基因相关联。

3. 转录因子结合：

- 转录因子可以识别并结合到特定的DNA序列上，这些序列通常位于基因的启动子或增强子区域。转录因子的结合可以招募其他蛋白质复合物，包括染色质重塑复合物和RNA聚合酶，从而启动转录过程。

4. DNA甲基化：

- 在某些情况下，DNA甲基化的减少也与染色质活化相关。低甲基化状态通常与基因的活跃转录相关联。

染色质活化的“开关”作用：

- 开启转录：**染色质活化通过上述机制“开启”特定基因的转录，使得基因可以被RNA聚合酶读取并转录成RNA。
- 关闭转录：**相反地，染色质钝化（Chromatin Repression）则通过使染色质结构更加紧密，阻止转录因子和RNA聚合酶的访问，从而“关闭”基因的转录。

染色质活化是基因表达调控中的一个复杂过程，它涉及到多种分子机制的协同作用，包括染色质重塑、组蛋白修饰、DNA甲基化、转录因子结合等。这些机制共同决定了基因在特定时间和空间条件下是否会被转录。通过这种方式，细胞可以精确地控制基因表达，以适应不同的生理需求和发育阶段。

DNA调控：

DNA调控是指细胞内通过各种机制来控制基因表达的过程。

1) 基因重排（Gene Rearrangement）：

- 基因重排是指在DNA序列中，基因片段发生改变或重新排列的过程。这通常发生在免疫系统的B细胞和T细胞中，以产生多样性的抗体和T细胞受体。
- 在B细胞中，V(D)J重组是基因重排的一种形式，它涉及免疫球蛋白基因的可变区（V）、多样性区

(D) 和连接区 (J) 的重排。

- 基因重排也可以发生在癌症细胞中，可能导致癌基因的激活或肿瘤抑制基因的失活。

2) 基因的顺势调控元件 (Cis-Regulatory Elements) :

- 顺势调控元件是指位于基因附近的DNA序列，它们可以影响该基因的表达。这些元件包括启动子、增强子、沉默子和绝缘子等。
- 启动子是RNA聚合酶识别和结合的区域，启动转录过程。
- 增强子是可以增强基因转录的DNA序列，它们可以位于基因的上游或下游，甚至可以在基因内。
- 沉默子是抑制基因转录的序列。
- 绝缘子则可以阻止或限制增强子和沉默子之间的相互作用。

3) 反式因子 (Transcription Factors) :

- 反式因子是指那些可以结合到DNA上并调控基因表达的蛋白质。它们可以是激活因子，也可以是抑制因子。
- 反式因子通常含有DNA结合域，能够识别并结合到特定的DNA序列上。
- 它们可以响应细胞内的信号，调控基因的转录，从而影响细胞的生理过程。

4) 组蛋白易位 (Histone Translocation) :

- 组蛋白易位是指组蛋白在染色质中的位置发生变化。组蛋白是染色质的基本结构蛋白，它们的修饰和位置变化可以影响染色质的结构和基因的表达。
- 例如，组蛋白乙酰化和磷酸化可以促进染色质结构的开放，从而促进基因的转录。

5) DNA甲基化与去甲基化 (DNA Methylation and Demethylation) :

- DNA甲基化是指DNA分子上的胞嘧啶碱基被甲基团 ($-CH_3$) 修饰的过程，通常发生在胞嘧啶的第五位碳原子上，形成5-甲基胞嘧啶。
- DNA甲基化通常与基因表达的抑制有关，特别是在基因启动子区域的高甲基化可以阻止转录因子的结合，从而抑制基因的转录。
- 去甲基化是指移除DNA上的甲基团，这通常与基因的活化相关，因为去甲基化可以使染色质结构变得更加开放，便于转录因子的结合。

RNA调控:

RNA调控是基因表达调控的重要组成部分，涉及从初级转录本到成熟RNA分子的多个加工和修饰步骤。

1. hnRNA加工 (Heterogeneous Nuclear RNA Processing) :

- hnRNA (异核RNA) 是指从DNA模板上初步转录出来的RNA前体，也称为前mRNA (pre-mRNA) 。
- 加工过程包括以下几个步骤:

- **5'端加帽**：在hnRNA的5'端添加一个7-甲基鸟苷（m7G）帽子结构，这个过程称为加帽反应。帽子结构有助于RNA的稳定性和mRNA的核输出，以及翻译的起始。
- **3'端聚腺苷酸化**：在hnRNA的3'端添加一串腺苷酸，形成poly(A)尾。这个过程有助于mRNA的稳定性和翻译效率。
- **剪接**：去除内含子（非编码序列）并将外显子（编码序列）连接起来，形成成熟的mRNA。

2. 选择性剪接（Alternative Splicing）：

- 选择性剪接是指从一个前mRNA分子中通过不同的剪接方式产生多个不同的mRNA转录本，从而编码不同的蛋白质。
- 这种机制极大地增加了蛋白质的多样性，因为一个基因可以通过选择性剪接产生多个蛋白质异构体。

3. RNA编辑（RNA Editing）：

- RNA编辑是指在RNA分子上发生碱基的插入、删除或替换，这些改变可能会改变编码的氨基酸序列，从而影响蛋白质的功能。
- RNA编辑通常由特定的酶催化，如ADAR（腺苷酸脱氨酶活性RNA）家族的酶。

4. RNA修饰（RNA Modification）：

- RNA修饰是指RNA分子上的核苷酸发生化学修饰，这些修饰包括甲基化、假尿苷化、硫醇化等。
- 这些修饰可以影响RNA的稳定性、结构、功能和蛋白质翻译效率。

5. mRNA的非翻译区（Untranslated Regions, UTRs）：

- mRNA的非翻译区包括5'非翻译区（5' UTR）和3'非翻译区（3' UTR），这些区域不编码蛋白质，但对基因表达调控至关重要。
- 5' UTR：
 - 包含翻译起始元件，如Kozak序列，参与翻译的起始。
 - 包含调控元件，如内部核糖体进入位点（IRES），在某些病毒和细胞应激条件下允许翻译的起始。
- 3' UTR：
 - 包含调控元件，如miRNA结合位点，参与mRNA的降解和翻译抑制。
 - 包含多腺苷酸化信号序列，影响mRNA的稳定性和翻译效率。

翻译的调控：

1. 扫描模型（SD序列）：

- **扫描模型**是翻译起始的一个机制，特别是在原核生物中，而在真核生物中，这个过程略有不同。

- **SD序列**（Shine–Dalgarno序列）是一个短的核苷酸序列，通常位于原核生物mRNA的5'非翻译区（5' UTR），紧邻起始密码子（通常是AUG）的上游。
- **工作原理：**
 - 在翻译起始时，小核糖体亚单位结合到mRNA的5'端。
 - SD序列与核糖体小亚单位中的16S rRNA上的反SD序列互补配对，从而引导核糖体定位到正确的起始密码子位置。
 - 随后，大核糖体亚单位结合，形成完整的翻译起始复合物，开始蛋白质的合成。
- 在真核生物中，没有SD序列，起始密码子的识别是通过eIF4F复合物和eIF1A等翻译起始因子来实现的。

2. mRNA甲基化（m7G）：

- **mRNA甲基化**是指在mRNA的5'端加上一个7-甲基鸟苷（m7G）帽子结构，这个过程称为加帽反应。
- **功能：**
 - **翻译起始：** m7G帽子结构是翻译起始所必需的。它有助于核糖体小亚单位识别mRNA，并促进大亚单位与之结合，从而起始翻译。
 - **mRNA稳定性和核输出：** 帽子结构保护mRNA免受降解，并促进mRNA从细胞核运输到细胞质。
 - **剪接：** 在某些情况下，帽子结构也参与前mRNA的剪接过程。
- **调控作用：**
 - m7G帽子结构的存​​在与否以及其完整性与翻译效率密切相关。帽子结构的缺失或损伤可能会降低翻译效率或导致mRNA的降解。
 - 此外，一些翻译后调控机制，如某些RNA结合蛋白，可以识别并结合到帽子结构，从而影响翻译的起始。

翻译调控是一个复杂的网络，涉及多种因子和机制，SD序列和m7G帽子结构只是其中的两个重要方面。细胞通过这些机制精细地调节蛋白质合成，以适应不同的生理需求和外部环境。

表观遗传学

表观遗传学可以定义为对基因活动的任何可遗传的影响，但不涉及 DNA序列的改变。

基因选择性转录的调控：DNA甲基化，基因印记，组蛋白共价修饰，染色质重塑

基因转录后的调控：非编码RNA（miRNA, LncRNA, circRNA, 反义RNA）

内含子，核糖开关

1. DNA甲基化

DNA甲基化的主要特点包括：

1. **发生位置**：在哺乳动物中，DNA甲基化主要发生在胞嘧啶碱基的第五位碳原子上，形成5-甲基胞嘧啶。这种甲基化通常发生在特定的序列上下文中，如CpG二核苷酸（即一个胞嘧啶后面紧跟着一个鸟嘌呤）。
2. **功能**：
 - **基因表达调控**：DNA甲基化通常与基因表达的抑制相关。甲基化的CpG位点更可能被紧密包裹在染色质中，使得转录因子和其他调控蛋白难以访问，从而抑制基因表达。
 - **基因组稳定性**：DNA甲基化有助于维持基因组的稳定性，比如通过抑制转座元件的活动，减少染色体重排。
 - **细胞分化**：在细胞分化过程中，DNA甲基化模式的改变有助于维持细胞特异性基因表达模式。
3. **酶**：DNA甲基化由一组特定的酶催化，称为DNA甲基转移酶（DNMTs）。在哺乳动物中，主要的DNMTs包括DNMT1、DNMT3A和DNMT3B。
4. **去甲基化**：与甲基化过程相对，去甲基化是指从DNA分子上去除甲基基团的过程。去甲基化可以通过主动的酶促反应或被动地通过DNA复制过程中不保留甲基化模式来实现。
5. **与疾病的关系**：DNA甲基化的异常与多种疾病有关，包括癌症、神经退行性疾病和发育异常。例如，在癌症中，某些基因的过度甲基化可能导致肿瘤抑制基因的沉默，而全局性的低甲基化可能与基因组的不稳定性和肿瘤的形成有关。

DNA甲基化是表观遗传学的一个重要方面，它不改变DNA序列本身，但可以影响基因的表达和功能，从而在细胞分化和发育过程中起到关键作用。

2. RNA甲基化

RNA甲基化是指在RNA分子上添加甲基基团（-CH₃）的化学修饰过程。这种修饰属于RNA表观遗传调控的重要部分，对基因表达调节、RNA稳定性和功能发挥关键作用。

常见的RNA甲基化类型

1. **N6-甲基腺嘌呤（m6A）**：这是最广泛研究的RNA甲基化形式，主要存在于信使RNA（mRNA）中。m6A修饰影响RNA的剪接、核输出、稳定性和翻译效率。

2. **5-甲基胞嘧啶 (m5C)**：这种修饰发生在转运RNA (tRNA)、核糖体RNA (rRNA) 和 mRNA 中，影响RNA的结构和功能。
3. **N1-甲基腺嘌呤 (m1A)**：主要存在于tRNA和rRNA中，参与调节蛋白质合成的准确性。

功能与机制

- **基因表达调控**：RNA甲基化可以改变RNA的二级结构，影响RNA与蛋白质的相互作用，从而调节基因表达。
- **RNA稳定性**：甲基化修饰可以保护RNA不被降解，延长其寿命。
- **翻译调控**：通过影响核糖体的结合和移动，RNA甲基化调节蛋白质的翻译效率。

关键酶类

- **甲基转移酶 ("写入者")**：如METTL3、METTL14，负责将甲基基团添加到RNA上。
- **去甲基化酶 ("擦除者")**：如FTO、ALKBH5，能够移除RNA上的甲基化修饰。
- **识别蛋白 ("读取者")**：如YTH家族蛋白，专门识别并结合甲基化的RNA，介导其生物学效应。

3. 组蛋白修饰

组蛋白是一类富含碱性氨基酸（如赖氨酸和精氨酸）的蛋白质，它们在真核生物的细胞核中与DNA结合，形成核小体，这是染色质的基本结构单位。组蛋白在维持染色质结构、调控基因表达和参与DNA复制与修复等生物学过程中发挥着关键作用。

以下是关于组蛋白的几个主要特点：

1. **种类**：在大多数真核生物中，有五种主要的组蛋白类型，分别是：
 - H1：连接核小体，帮助形成更高阶的染色质结构。
 - H2A、H2B、H3和H4：这些组蛋白形成了一个八聚体的核心，即组蛋白八聚体，它由两个分子各自的H2A、H2B、H3和H4组成，DNA双螺旋则围绕这个核心进行缠绕。
2. **核小体结构**：核小体由约147个碱基对的DNA缠绕在组蛋白八聚体上，形成了一个“珠串”结构。H1组蛋白则位于核小体的DNA入口处，帮助稳定核小体结构。
3. **碱性特性**：组蛋白的碱性特性使它们能够与带有负电荷的DNA紧密结合。
4. **进化保守性**：组蛋白在进化上非常保守，这意味着从酵母到人类，组蛋白的核心结构变化不大。
5. **修饰**：组蛋白可以通过多种共价修饰进行调节，如乙酰化、甲基化、磷酸化、泛素化和SUMO化等。这些修饰可以改变染色质的结构，影响基因表达。
6. **功能**：

- **染色质包装**：组蛋白帮助将长链DNA包装成紧凑的染色质结构，以便在细胞分裂时能够正确分配遗传物质。
- **基因调控**：组蛋白的修饰可以调节基因的活性，决定哪些基因被表达，哪些基因被沉默。
- **DNA修复和复制**：组蛋白在DNA损伤修复和复制过程中也起着重要作用。

组蛋白修饰是指在组蛋白的氨基酸残基上发生的共价修饰，这些修饰**不改变蛋白质的氨基酸序列**，**直接或间接影响转录活性**，能调节染色质结构和基因表达。

组蛋白修饰类型：

1. **乙酰化 (Acetylation)**：在组蛋白的**赖氨酸残基**上添加乙酰基团，通常与**基因活化**相关。
2. **甲基化 (Methylation)**：在组蛋白的**赖氨酸**或**精氨酸**残基上添加甲基基团，可以与**基因活化或抑制**相关，具体作用取决于修饰的位置和程度。往往长期存在。HKMTs (组蛋白赖氨酸甲基转移酶) – PRMTs (组蛋白精氨酸甲基转移酶)。可逆的生化反应。分子效应—增加赖氨酸的疏水力。生物学功能—基因转录活化，基因转录沉默，X染色体失活，异染色质致密维持。
3. **磷酸化 (Phosphorylation)**：在组蛋白的**丝氨酸**、**苏氨酸或酪氨酸**残基上添加磷酸基团，通常与**基因活化**相关。
4. **泛素化 (Ubiquitination)**：在组蛋白上连接泛素蛋白，通常与**基因抑制**和**染色质重塑**相关。
5. **Sumo化 (Sumoylation)**：在组蛋白上连接小泛素样修饰蛋白 (SUMO)，与**基因表达的调控**相关。
6. **糖基化 (Glycosylation)**：在组蛋白上添加糖基团，这种修饰在组蛋白中的功能尚不十分清楚。
7. **丁酰化 (Butyrylation)**：在组蛋白的**赖氨酸**残基上添加丁酰基团，与**基因表达调控**有关。
8. **丙酰化 (Propionylation)**：在组蛋白的**赖氨酸**残基上添加丙酰基团，其功能目前了解较少。
9. **脱氨化 (Deamination)**：组蛋白**赖氨酸**残基上的脱氨作用，影响组蛋白电荷和基因表达

赖氨酸甲基化：

1. **类型**：赖氨酸甲基化可以是单甲基化 (monomethylation, 添加一个甲基基团)、二甲基化 (dimethylation, 添加两个甲基基团) 或三甲基化 (trimethylation, 添加三个甲基基团)。
2. **位点**：在组蛋白分子中，多个赖氨酸位点可以发生甲基化，包括组蛋白H3的K4、K9、K27、K36和K79，以及组蛋白H4的K20。
3. **功能**：
 - **基因活化**：例如，组蛋白H3的K4甲基化通常与活跃的基因表达相关联。
 - **基因抑制**：组蛋白H3的K9和K27的三甲基化通常与基因沉默和异染色质形成相关。
 - **染色质结构**：甲基化可以影响染色质的紧密程度，从而调节基因的可访问性。

4. 酶：

- **甲基转移酶**：负责将甲基基团转移到赖氨酸残基上，如SET1、SUV39H1、SUV39H2等。
- **去甲基化酶**：负责移除赖氨酸上的甲基基团，包括赖氨酸特异性去甲基化酶（LSD1）和JmjC域蛋白。

5. **调控**：赖氨酸甲基化的状态可以通过酶的活性调控来动态调整，以响应细胞内的信号或环境变化。

6. **疾病相关性**：赖氨酸甲基化的异常与多种疾病有关，包括癌症、神经退行性疾病和发育障碍等。

精氨酸甲基化：

1. 类型：

- **单甲基化**：在精氨酸的胍基上添加一个甲基基团。
- **对称二甲基化**（即Arginine Dimethylation）：在两个胍基氮原子上各添加一个甲基基团，通常标记为Rme2s。
- **非对称二甲基化**：在一个胍基氮原子上添加两个甲基基团，通常标记为Rme2a。
- **三甲基化**：在精氨酸的胍基上添加三个甲基基团。

2. 功能：

- **基因表达调控**：精氨酸甲基化可以影响染色质结构和转录因子与DNA的结合，从而调控基因表达。
- **信号传导**：甲基化的精氨酸可以作为蛋白质间相互作用的信号，影响信号传导途径。
- **蛋白质稳定性**：甲基化可能影响蛋白质的稳定性和寿命。

3. 酶：

- **蛋白质精氨酸甲基转移酶（PRMTs）**：催化精氨酸甲基化过程。根据催化的甲基化类型，PRMTs可以分为类型I、II和III。

4. 去甲基化酶：

- **蛋白质精氨酸去甲基化酶（PADs）**：负责去除精氨酸上的甲基基团，逆转甲基化修饰。

5. 生物学意义：

- **发育**：精氨酸甲基化在胚胎发育和细胞分化过程中起着重要作用。
- **疾病**：异常的精氨酸甲基化与多种疾病相关，包括癌症、神经退行性疾病和心血管疾病。

6. 组蛋白和非组蛋白：

- 虽然精氨酸甲基化在组蛋白上非常重要，但它也发生在许多非组蛋白上，包括转录因子、信号蛋白和核孔复合物蛋白等。

组蛋白乙酰化

通常发生在组蛋白的赖氨酸（K）上；一类可逆的生化反应：A. 组蛋白脱乙酰酶, HAT (>30) B. 组蛋白脱乙酰酶, HDAC (18)。分子效应—中和赖氨酸上的正电荷，增加组蛋白与DNA的排斥力；生物学功能A. 基因转录活化，B. DNA损伤修复

组蛋白变体

组蛋白变体（Histone Variants）是指组蛋白家族中的一些成员，它们与标准的组蛋白有所不同，通常在序列、结构和功能上有所变异。这些变体在染色质的结构和功能调控中扮演着特殊的角色，它们可以通过替换核小体中的标准组蛋白或改变染色质的环境来影响基因表达、DNA复制和修复等生物学过程。

1. H2A变体：

- **H2A.Z**：在许多生物中都有发现，参与基因表达调控和染色质重塑。H2A.Z的替换通常与活跃的转录区域相关。
- **macroH2A**：在哺乳动物中，这种变体与异染色质的形成和基因沉默有关。

2. H3变体：

- **H3.3**：与标准的H3相比，H3.3在基因体的活跃区域富集，参与基因表达的调控和DNA损伤响应。

3. **H4**变体：

- **H4.X**：在某些生物中存在，但其功能尚不十分清楚。

4. H1变体：

- **H1变体**：存在多种H1的变体，它们在染色质结构和基因表达调控中具有不同的作用。

组蛋白变体的功能包括：

- **基因表达调控**：通过替换核小体中的标准组蛋白，组蛋白变体可以改变染色质的紧密度，影响转录因子的访问和基因的转录状态。
- **染色质重塑**：组蛋白变体可以促进染色质的动态变化，有助于染色质结构的重塑。
- **DNA复制和修复**：某些组蛋白变体在DNA复制和修复过程中发挥作用，确保遗传信息的准确传递。
- **异染色质形成**：特定的组蛋白变体参与异染色质的形成，有助于维持基因沉默和基因组稳定性。

组蛋白乙酰化、甲基化、以及DNA甲基化的关系：

A. MBD蛋白结合甲基化的DNA，招募HDACs，组蛋白发生去乙酰化，招募HMT，后者使组蛋白甲基化，转录沉默。

B. 组蛋白无乙酰化修饰，MBD结合甲基化的DNA，再与SET结合，甲基化组蛋白。

C. 甲基化的组蛋白尾部招募DNMT，引发基因长期沉默。

mRNA与蛋白质组学

非编码RNA

Short interfering RNA (siRNA)

- **功能：** siRNA是一种小分子RNA，通常由21–23个核苷酸组成，用于沉默特定基因的表达。
- **机制：** siRNA通过RNA干扰（RNAi）途径与互补的mRNA结合，导致mRNA的降解或翻译抑制。

Micro RNA (miRNA)

- **功能：** miRNA也是一种小分子RNA，大约22个核苷酸长，它通过结合到mRNA的3'非翻译区（3' UTR）来调控基因表达。
- **机制：** miRNA可以抑制mRNA的翻译或促进其降解，从而调控蛋白质的产生。
- 高度保守性，其他种系中找到同源体
- 一个可以调控多个，多个可以共同调控一个

Double-stranded RNA (dsRNA)

- **定义：** dsRNA是指具有两条互补链的RNA分子，可以是天然的或人工合成的。
- **功能：** 在RNA干扰过程中，dsRNA可以被Dicer酶切割成siRNA，进而沉默基因表达。

Short heterochromatic RNA (shRNA)

- **功能：** shRNA是一种合成RNA分子，通常设计用于特定基因的长期沉默。
- **机制：** shRNA通过RNA干扰途径发挥作用，通常被克隆到表达载体中并转染到细胞中。

Transcripts from repeated sequences (ALU, LTR)

- **定义：** 这些RNA分子是由细胞基因组中的重复序列转录产生的，如ALU重复序列和LTR（长末端重复序列）。

- **功能：**这些转录本可能参与调控基因表达或影响基因组稳定性。

Long non-coding RNA (lncRNA)

- **定义：**Long non-coding RNAs (lncRNAs) 是一类长度超过200个核苷酸的非编码RNA分子，它们不编码蛋白质。
- **功能多样性：**
 - **基因表达调控：**lncRNAs可以在多个层面上调控基因表达，包括染色质重塑、转录和剪接过程。
 - **染色质组织：**一些lncRNAs参与染色质结构的组织和修饰，例如通过结合到特定的蛋白质或DNA区域，影响基因的活性。
 - **细胞核与细胞质间的运输：**某些lncRNAs能够将RNA或蛋白质分子从细胞核运输到细胞质。
 - **miRNA海绵：**一些lncRNAs可以作为miRNA海绵，吸附特定的miRNA分子，从而解除miRNA对其靶标的抑制作用。
 - **蛋白质复合体组装：**lncRNAs可以作为支架，促进蛋白质复合体的形成。
- **特点：**
 - **组织特异性：**许多lncRNAs在特定的组织或发育阶段表达。
 - **细胞类型特异性：**它们在不同的细胞类型中可能具有不同的功能。
 - **稳定性：**与mRNA相比，lncRNAs通常更稳定，可能在细胞内持续较长时间。

Circular RNA (circRNA)

- **定义：**circRNA是一种闭环形式的RNA分子，由前mRNA的剪接过程产生。
- **功能：**circRNA具有多种功能，包括调控基因表达、作为miRNA海绵吸附miRNA、影响剪接过程等。
- **特点：**circRNA在细胞中比线性RNA更稳定，不易被降解。

遗传重组与基因编辑

一段可以转录为功能性RNA的特定DNA（RNA）区域，它包含调节、转录和其它功能区域；可以重迭、断裂的形式存在。

mRNA、rRNA、tRNA、miRNA、snRNA、scRNA、lncRNA、cirRNA、eRNA

大多数原核基因与蛋白质共线性，因此是不间断的基因

在大多数真核细胞中，主要基因含有内含子，因此是中断基因

对于哺乳动物来说，90%的编码基因含有内含子，每个基因平均有8个内含子，而且内含子长度通常比外显子长很多。

1. 断裂基因：

基因中断（也称为断裂基因）是指包含称为外显子的 DNA 片段的基因，这些外显子被表达为 RNA 和蛋白质，并被称为内含子的 DNA 片段所中断，这些内含子不被表达。

外显子是保守的，但内含子既在大小上也在序列上不是都保守。除了内含子区域位于外显子两侧的情况。

内含子区域两侧的外显子区域也显示出相对较高的保守性。

例：个体 DHFR 基因内含子的长度差异很大。然而，外显子的序列、定位位置和大小相对保守。

断裂基因的结构是保守的

基因的保守性：

- 1) 基因的结构高度保守；
- 2) 基因的外显子序列高度保守；
- 3) 外显子序列侧翼的内含子序列高度保守；
- 4) 分支点100%保守

真核生物通常含有分裂基因，且分裂基因的比率不同，物种等级越高，分裂基因的比率越高。

人类基因外显子长度相对短而集中（50–200bp）

人类基因内含子相对长且分布广泛（>80bp）

2. 重叠基因：

某些 DNA 序列编码多于一种蛋白质或 RNA 产物

1. 使用相同的 ORF 但不同的转录或翻译起始位点；
2. 使用不同的 ORF；
3. 使用不同的链；Ras 和 Rassf1
4. 一个基因位于其他基因内部

3. 基因组

3.1. 概念

基因组 (Genome) 指生物的整套染色体所含有的全部DNA序列, 更进一步说: 生物所具有的携带遗传信息的遗传物质总和称为基因组。配子中所含有的全部遗传信息。

Karyotype (核型): 核型是指染色体组在有丝分裂中期的表型, 包括: 染色体数目、大小、形态特征的总和。Karyotyping (核型分析): 染色体核型分析是以分裂中期染色体为研究对象, 根据染色体的长度、着丝点位置、随体的有无等特征, 借助显带技术对染色体进行分析、比较、排序和编号。

遗传图也叫连锁图, 它是指描述基因组中基因或遗传标记之间相对距离的图谱(遗传距离描述方式: $1\text{ cM}=1\%\text{重组率}$)。

3.2. 遗传标记

遗传标记种类: 能够快速识别的遗传性状、基因、一段DNA序列或基因产物。

遗传标记用途: 用于鉴定与某一遗传性状或疾病密切相关的基因。

遗传多态 (Genetic polymorphism):

群体中基因的不同等位形式, 其中最少的等位基因频率大于1%时称之为多态。

3.3. 分子标记

分子标记 (Molecular marker) 是位于染色体上已知位置的 DNA 或基因序列片段, 用作识别工具。

染色体上可识别的特定DNA序列, 包括: RFLP; VNTR; STR; SNP

RFLP (限制性片段长度多态性)

通过限制性内切酶通过将DNA切割成片段来检测基因组DNA序列变异的方法。

用途: 基因组制图、疾病基因的遗传定位、确定疾病风险、基因指纹、法医鉴定和亲子鉴定等方面的重要工具。第一代分子标记。

VNTR (可变数目串联重复序列)

基因组中短核苷酸串联重复的位置, 这些可以在许多染色体上找到, 并且通常显示出长度多态性, 在遗传学、生物学研究、法医学和 DNA 指纹识别中 useful, 优点: 分布广、数量多、不需要杂交检测。

STR — DNA中的短串联重复 (微卫星)

重复模式的长度范围从2至10 bp，通常在非编码内含子区域，目前在人类中发表了超过10,000个STR序列

基因组，确定遗传的流行方法 (方法便捷)，计算特定STR在给定基因座的重复次数，可以创建独特的遗传特征，亲子鉴定

假设STR基因座位A含有4个等位基因，A1、A2、A3、

A4基因频率分别为10%、10%、40%、40%。则基因型

A1/A3个体的频率是多少？

$FA1/A3 = 10\% \times 40\% = 0.04 = 4\%$

假设目的基因座位的基因型频率在群体中相同，如果同时

检测20个基因座位的STR基因型，则任意2个个体间基因

型完全相同的概率有多少？ $= (4\%)^{20}$

Haplotype (单体型)：

来源于单个亲本的位于同一条染色体上的一组基因,位于一条染色体上的一组(几个)单核苷酸多态性 (SNP)

位于一条染色体上的一组遗传标记

数量众多，并且可以通过不涉及凝胶电泳的方法进行打印。

SNP 检测更快，因为它基于寡核苷酸杂交分析或 DNA 测序

3.4. 人类基因组计划

遗传多态：群体中基因的不同等位形式，其最少的等位基因频率大于1%时称之为多态。

绘制高精度基因组图谱

遗传标记和分子标记：RFLP、VNTR、STR、SNP

人类基因组计划的准备，第一代 第二代 第三代 分子标记，在上述分子标记的基础上绘制4张图，包括：

细胞图，遗传图，物理图和序列图

Strategy1: Genetic map-based sequencing, 自上而下，先后绘制细胞图谱、遗传图谱、物理图谱

Strategy 2: Whole genomic shotgun, 自下而上, 无需要绘制繁琐的各种图谱, 但依赖于大规模测序技术和高性能计算机拼接。

3.5. 人类基因组特征

- **C-Value:** C值是指一种生物体单个倍体基因组的DNA含量。虽然C值与物种的复杂性无直接关系, 但它用于比较不同生物之间的基因组大小。
- **Composition of human genome (人类基因组的组成):**
- **Unique sequence (独特序列):** 在人类基因组中, 独特序列是指那些只存在一次或极少重复的DNA片段, 通常包含大部分编码蛋白的基因序列。
- **Repetitive sequence (重复序列):** 基因组中有大量重复的DNA片段, 这些片段可能没有特定功能, 或者调控基因表达。
- **Highly repetitive sequence (高度重复序列):** 这些序列多次重复, 往往包括卫星DNA, 它们在基因组中的功能不完全明确, 但可能与染色体结构的稳定性相关。
- **Intermediate repetitive sequence (中度重复序列):** 这些序列的重复次数介于高度重复序列和独特序列之间, 常常包括基因家族或转座子等重复元件。
- **Multigene family (多基因家族):** 多基因家族是指一组结构相似的基因, 这些基因来源于共同的祖先基因, 并通过基因重复事件产生。它们可能具有类似或相关的功能。序列相似, 来源相同, 聚集或分散分布在基因组。
- **Pseudogene (假基因):** 假基因是指因突变失去功能的基因, 它们可能曾经是功能性基因, 但由于突变累积, 无法编码蛋白质。
- **CpG island (CGI, CpG岛):** CpG岛是指富含CpG二核苷酸的DNA序列, 通常位于基因的启动子区域, 并与基因表达的调控有关。CpG岛通常未被甲基化, 这有助于基因的活性转录。

CpG岛的分布与基因的关系及其甲基化的意义

1. CpG岛的分布与基因的关系:

- **非随机分布:** CpG岛在基因组中的分布具有显著的非随机性, 常见于基因的启动子区域, 但不限于此。CpG岛在基因启动子区域的存在有助于基因的启动和表达调控。
- **编码区相关性:** 除了启动子区域外, CpG岛还可以分布在基因的编码区附近 (如首个外显子), 与基因的表达和转录密切相关。

2. CpG岛的保护作用:

- **基因保护:** CpG岛在进化过程中受到保护, 不容易发生突变。这使得它们在基因表达的调控中起到重要作用, 特别是在启动子区域, CpG岛的存在能保证关键基因的正常转录。

3. 基因组中CpG频率低于预期:

- 在人类基因组中, CpG二核苷酸的实际频率远低于预期值, 实际频率约为1%, 而理论上的预期

值（根据碱基组成比例）应为4%左右。这种低频现象被认为是由于CpG二核苷酸易于发生甲基化及后续的脱氨反应，导致C转化为T，从而使CpG序列逐渐消失。

4. CpG岛甲基化的意义：

- **正常组织中的作用：**在正常组织中，CpG岛通常缺乏甲基化，尤其是位于基因启动子区域的CpG岛，这有助于保持基因的活性。然而，某些生理过程中，如X染色体失活和基因组印记，CpG岛的甲基化对基因的沉默起着至关重要的作用。
- **细胞分化中的作用：**不同组织之间的甲基化状态具有差异性，这表明CpG岛甲基化是细胞分化和发育过程中重要的转录调控机制。通过甲基化调控基因的表达，可以保证不同细胞类型在不同发育阶段有各自特定的功能。
- **恶性细胞中的异常甲基化：**在癌症等恶性细胞中，CpG岛的异常甲基化经常导致基因的不当沉默，特别是抑癌基因的沉默，这会使细胞失去对生长和增殖的正常控制，最终导致肿瘤发生。

5. **总结：** CpG岛在基因组中分布具有非随机性，特别是在基因启动子区域，它们参与基因的表达调控。CpG岛的甲基化通常与基因沉默相关，这在正常生理过程（如基因组印记和X染色体失活）中至关重要。然而，在恶性细胞中，CpG岛的异常甲基化会导致基因功能的丧失，尤其是抑癌基因的沉默，从而促进癌症的发生。

人类基因组的组成比例、来源及其特点如下：

1. 蛋白编码基因 (Protein-coding genes) – 约1.5%

来源：这些序列编码用于构建细胞功能和结构的蛋白质。

特点：

在基因组中占比非常小（约1.5%）。

蛋白编码基因通过转录产生mRNA，并最终翻译成功能性蛋白质。

人类基因组大约包含20,000到25,000个蛋白编码基因。

不同基因的表达通过转录因子、增强子等调控元件严格调控。

2. 重复序列 (Repetitive Sequences) – 约50%

来源：包括转座子、卫星DNA等。

特点：

转座子 (Transposable elements)：占人类基因组约45%，是可以在基因组内自我复制并插入新位置的DNA片段。

LINEs (长散布核元件)：约占基因组的20%。这些序列较长，并且包含能够自我复制的逆转录酶。

SINEs (短散布核元件)：约占基因组的13%。这些较短的序列不能独立复制，通常依赖于LINE的酶功能。

LTR (长末端重复序列) 和 DNA转座子：分别占基因组的8%和3%。

卫星DNA (Satellite DNA)：由大量串联重复的简单序列组成，通常位于染色体的着丝粒和端粒区

域，有助于维持染色体的结构稳定性。

这些重复序列通常不编码蛋白质，但有时可能参与基因表达调控或对基因组进化产生影响。

3. 基因间非编码DNA (Intergenic Non-coding DNA) – 约24%

来源：存在于基因之间的非编码区。

特点：

不直接编码蛋白质，但可能包含调控元件，如增强子、沉默子和微小RNA (miRNA) 位点。

在调控基因表达、染色质结构和基因组完整性方面起重要作用。

4. 基因内非编码序列 (Intronic and Regulatory Non-coding Sequences) – 约25%

来源：存在于基因内部的内含子 (introns) 和调控区域。

特点：

内含子 (Introns)：占基因组约20%。内含子是编码序列（外显子）之间的非编码区，它们在转录后从mRNA中剪切出来，不翻译成蛋白质。

调控区域：包括基因启动子、增强子和沉默子，调控基因何时、何地以及以何种程度表达。

这些序列对基因的正确表达至关重要，错误的调控可以导致疾病（如癌症）。

5. 其他非编码RNA基因 (Non-coding RNA Genes) – 约1.5%

来源：非编码RNA基因包括rRNA、tRNA、miRNA和长链非编码RNA (lncRNA) 等。

特点：

这些RNA不编码蛋白质，但在转录后调控、蛋白质合成和基因表达调控中扮演关键角色。

miRNA和lncRNA对基因表达的精细调控起到重要作用，可能与疾病（如癌症和代谢疾病）相关。

6. 伪基因 (Pseudogenes)

来源：由于基因复制或突变丧失功能的基因残余。

特点：

通常不再具有功能，无法编码有效的蛋白质。

有时可能保留部分调控功能或被转录为非功能性RNA。

题目：

1. How the Genetic map-based sequencing was performed?

• Genetic map (linkage map), physical map, transcription map, sequence map:

○ 基因图谱测序是根据遗传图谱和物理图谱相结合的方法来完成的。

- **Genetic map (遗传图谱)**：遗传图谱通过观察遗传标记在家系中的遗传共分离（连锁）情况来构建。通过确定不同遗传标记之间的重组频率，计算出这些标记在染色体上的相对位置。
- **Physical map (物理图谱)**：物理图谱通过实际的DNA物理长度（以碱基对数目表示）来显示遗传标记在基因组中的位置。这些标记通常由限制性酶切位点、YAC/BAC克隆等方法确定。

- **Transcription map (转录图谱):** 转录图谱标示了基因组中转录活跃区域的位置, 指明了哪些区域包含编码蛋白质的基因。
- **Sequence map (序列图谱):** 序列图谱是基因组的碱基对顺序, 展示了DNA的确切序列。

这种基于图谱的测序方法通过将遗传、物理和转录图谱与DNA序列相结合, 逐步缩小目标区域, 最终获得准确的基因组序列。

2. Genetic marker: RFLP, VNTR, STR, SNP, Haplotype

- **RFLP (限制性片段长度多态性):** 利用不同个体中DNA序列的差异对限制性内切酶的切割位点造成的不同, 产生不同长度的DNA片段, 用于基因分型和作图。
- **VNTR (可变数目串联重复):** 指一段重复序列的拷贝数目在不同个体间变化, 常用于个体识别和亲子鉴定。
- **STR (短串联重复):** 类似于VNTR, 但重复的序列通常更短, 常用于法医分析和基因连锁作图。
- **SNP (单核苷酸多态性):** 基因组中的单个碱基的变异, 是研究遗传变异、个体化医疗和群体遗传学的重要标记。
- **Haplotype (单倍型):** 一组位于同一染色体上的等位基因的组合, 它们通常一起遗传下来, 常用于遗传连锁分析和疾病研究。

3. Repetitive sequence, LINE, SINE

- **Repetitive sequence (重复序列):** 指基因组中多次重复出现的DNA序列。重复序列包括非编码重复序列和功能性重复序列。
- **LINE (长散布核元件):** 是一类长的自我复制的转座子, 能通过RNA中介逆转录复制自己并插入基因组的不同位置, 占人类基因组的相当一部分。
- **SINE (短散布核元件):** 是短的自我复制的转座子, 不能编码功能性蛋白质, 常依赖LINE逆转录复制。

4. Multiple gene family, pseudogene

- **Multiple gene family (多基因家族):** 由结构相似的基因组成, 通常由共同的祖先基因通过重复事件产生, 这些基因可能具有相似或相关的功能, 例如血红蛋白基因家族。
- **Pseudogene (假基因):** 曾经是功能性基因, 但由于突变或其他原因失去了编码功能, 无法产生功能性蛋白质。

5. CGI and the meaning of CGI methylation

- **CGI (CpG island, CpG岛):** 是富含CpG二核苷酸的DNA片段, 通常出现在基因启动子区域。未甲基化的CpG岛与基因活跃转录相关。
- **CGI methylation (CpG岛甲基化):** CpG岛甲基化通常抑制基因表达, 甲基化的CpG岛可以导致基因沉默, 这种现象在细胞分化和疾病(如癌症)中常见。

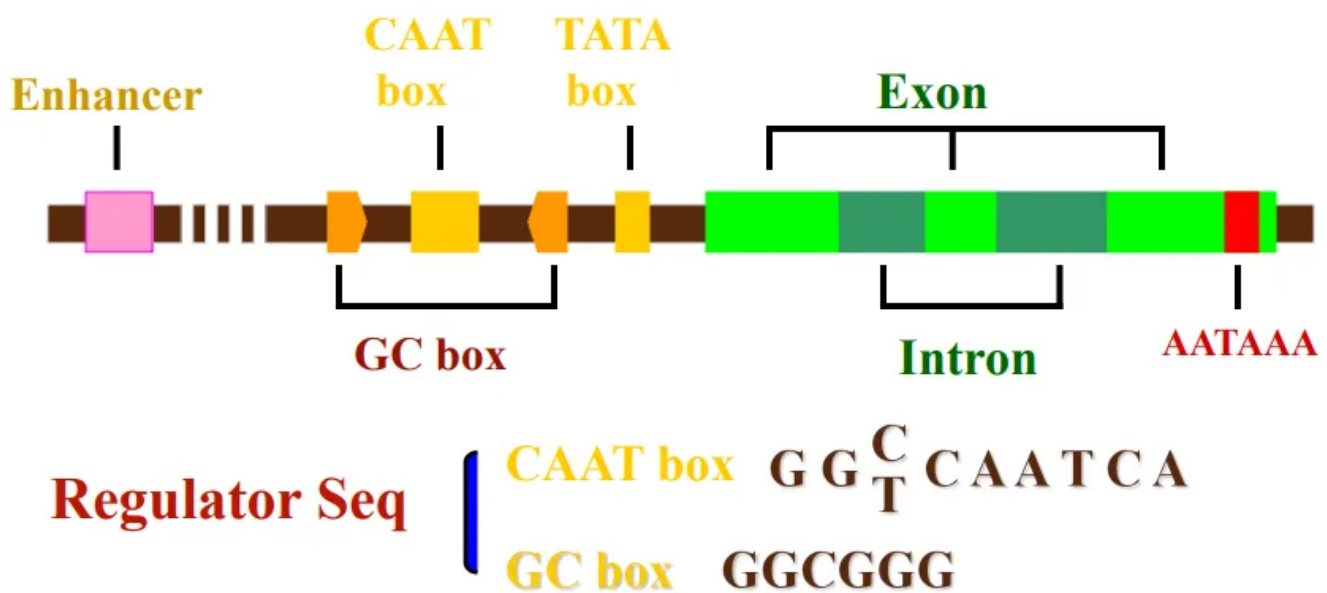
6. What can you learn from the contents of human genome?

- 从人类基因组内容中可以学到：
 - a. 基因的结构和功能，包括编码蛋白质的基因和调控元件。
 - b. 重复序列的分布及其对基因组稳定性的影响。
 - c. 转座子的存在及其对基因组进化的贡献。
 - d. 单核苷酸多态性(SNP)和其他遗传标记的分布，它们有助于揭示个体遗传变异和疾病易感性。
 - e. 基因表达调控机制，如CpG岛甲基化对基因沉默的影响。
 - f. 人类与其他物种的进化关系，以及基因家族和假基因在进化中的作用。

3.6. mRNA质量控制

3.6.1. 基因结构

Split gene



3.6.2. mRNA预处理

超级增强子：常常在细胞命运决定中**增强细胞分化表型**

2013年，由美国白头生物研究所Richard A Young 提出，首次报道231个SE 特点：

- (1) 跨度大8–20kb，（TE，数百bp到2kb）
- (2) 更高密度激活组蛋白修饰（H3K27ac, H3K4me1）
- (3) 结合更多mediator/P300等蛋白

(4) 高密度结合谱系特异性转录因子及转录激活 相关染色体标记

(5) 促进高水平基因表达

作用：分化中的命运决定、癌症、分化等

小核核糖核蛋白（small nuclear ribonucleoproteins, **snRNPs**）是由小核RNA（small nuclear RNA, **snRNA**）和特定的蛋白质组成的核糖核蛋白复合体，主要功能是在细胞核内参与**RNA剪接**（splicing）过程。snRNPs 是剪接体（spliceosome）的重要组成部分，剪接体负责将前体 mRNA（pre-mRNA）中的内含子（intron）切除，并连接外显子（exon）形成成熟的mRNA。

snRNP的结构

snRNP复合体包括：

1. **snRNA**：小核RNA是snRNP的核心成分，通常长约100–300个核苷酸。人类最常见的snRNA有U1、U2、U4、U5和U6。每种snRNA在RNA剪接中都有不同的作用。
2. **蛋白质**：snRNP还包含一组特定的蛋白质，这些蛋白质有助于snRNA在剪接体中发挥作用。核心蛋白质家族称为Sm蛋白，这些蛋白与snRNA形成环状复合物，稳定snRNP的结构。

snRNP的功能

1. **参与pre-mRNA的剪接**：snRNP是剪接体的组成部分，通过识别pre-mRNA中的关键位点（例如5'剪接位点、3'剪接位点和支点序列），帮助将内含子去除，确保外显子正确连接。
 - **U1 snRNP**：识别pre-mRNA中的5'剪接位点。
 - **U2 snRNP**：结合在支点序列（branch point sequence）处，帮助形成剪接体的活性中心。
 - **U4/U6和U5 snRNPs**：形成剪接体的催化核心，协助内含子的切除和外显子的连接。
2. **RNA分子加工中的其他角色**：除了在mRNA剪接中发挥作用，某些snRNPs还参与核RNA的加工、成熟和调控，例如核仁中的小核仁RNA（snoRNP）与核糖体RNA（rRNA）的修饰有关。

snRNP相关疾病

snRNP的异常功能与一些疾病相关，最常见的是**系统性红斑狼疮（SLE）**，这是一种自身免疫性疾病，患者体内会产生针对snRNP成分的自身抗体，尤其是针对U1 snRNP的抗体。这些抗体与snRNP相互作用，破坏正常的RNA加工过程，导致多种免疫系统紊乱。

总结：

1. 5'端加帽（5'capping）

- **酶和意义：**5'加帽是mRNA加工的第一个步骤，由一系列酶（包括RNA 5'三磷酸酶、鸟苷转移酶和甲基转移酶）完成。加帽的意义在于：
 - 保护mRNA不被核酸外切酶降解。
 - 有助于核糖体识别mRNA，促进翻译的开始。
 - 参与mRNA的核输出过程。

2. 剪接（Splicing）

- **U1和U2 snRNAs如何促进正确剪接：**在mRNA剪接过程中，U1 snRNA识别并结合pre-mRNA中的5'剪接位点，而U2 snRNA识别支点序列（branch point sequence），通过这些结合，U1和U2 snRNAs帮助形成剪接体（spliceosome），确保正确的内含子去除和外显子连接。
- **跨外显子识别复合物如何参与剪接：**跨外显子识别复合物（exon junction complexes, EJCs）是在外显子连接处形成的蛋白质复合物。这些复合物不仅帮助定义正确的剪接位点，还与mRNA的转运、翻译调控和非正常终止的核糖体回收机制（nonsense-mediated decay, NMD）有关。

3. 加尾（Polyadenylation）

- **信号是什么，帮助切割因子识别并在pre-mRNA末端添加poly(A)尾：**在pre-mRNA的3'非翻译区（3'UTR），存在一个AAUAAA的保守序列，称为**多聚腺苷酸加尾信号**。这个信号由一组剪切和加尾因子识别，包括CPSF（cleavage and polyadenylation specificity factor），它们

在特定位点切割mRNA，并在其3'端添加一段多聚腺苷酸（poly-A）尾巴，增强mRNA的稳定性并调控其翻译效率。

这些过程共同保证了mRNA的正确加工和功能，为基因表达的调控提供了重要的控制手段。

3.6.3. RNA编辑

RNA修饰是指在RNA分子上发生的化学修饰，这些修饰对于RNA的稳定性、功能、运输以及与蛋白质的相互作用等方面都至关重要。RNA修饰种类繁多，以下是一些主要的RNA修饰类型：

1. 甲基化（Methylation）：

- **m6A**（N6-甲基腺苷）：这是真核生物mRNA中最常见的修饰，主要发生在转录后。
- **m5C**（5-甲基胞嘧啶）：这种修饰在tRNA、rRNA和一些mRNA中都很常见。
- **m1G**（1-甲基鸟苷）：通常在tRNA和rRNA中发现。

2. 假尿苷化（Pseudouridylation）：

- 将尿嘧啶（U）转变为假尿嘧啶（Ψ），在rRNA和tRNA中较为常见。

3. 尿苷化（Uridylation）：

- 在某些RNA分子上添加尿苷残基，常见于某些mRNA的3'非翻译区。

4. 核糖甲基化：

- **2'-O-核糖甲基化**：在mRNA、tRNA和rRNA的某些部位发生。
- **核糖的其它位点甲基化**：如1'-O-甲基化、3'-O-甲基化等。

5. 脱氨作用（Deamination）：

- **腺苷到肌苷（A-to-I）编辑**：在mRNA中，腺苷脱氨成为肌苷，改变编码的氨基酸。
- **胞苷到尿苷（C-to-U）编辑**：胞苷脱氨成为尿苷。

6. 磷酸化：

- 在RNA分子的某些部位添加磷酸基团。

7. 核苷酸插入或删除：

- 在某些情况下，RNA分子中可以发生核苷酸的插入或删除。

总结：

（一）断裂基因结构（基因和侧翼序列）

- **断裂基因**指的是真核生物中包含**内含子**和**外显子**的基因结构。内含子是转录后需要被剪除的非编码区，而外显子则是编码蛋白质的部分。
- **侧翼序列**是位于基因的编码区前后的非编码区域，通常包含启动子、增强子等调控基因表达的序列。

(二) Pre-mRNA加工 (capping、splicing、polyadenylation基本原理) , alternative splicing

- **Capping (加帽)**：在pre-mRNA的5'端添加一个7-甲基鸟嘌呤帽子结构，这有助于保护mRNA并促进翻译的开始。
- **Splicing (剪接)**：pre-mRNA通过剪接体 (spliceosome) 移除内含子，并连接外显子，形成成熟的mRNA。
- **Polyadenylation (加poly(A)尾)**：在mRNA的3'端添加多聚腺苷酸尾巴，增加mRNA的稳定性，并调控翻译。
- **Alternative splicing (可变剪接)**：通过选择不同的剪接位点，从同一个基因中生成多种不同的mRNA产物，增加了基因的多样性和调控能力。

(三) RNA编辑概念、原理 (酶)、类型及作用

- **RNA编辑**：RNA编辑是指转录后的RNA序列被特定酶修饰或改变，通常会改变核苷酸的排列，进而影响最终的蛋白质产物。
- **原理与酶**：RNA编辑的常见机制包括腺苷 (A) 被编辑为肌苷 (I) ，或胞嘧啶 (C) 被编辑为尿嘧啶 (U) 。参与编辑的关键酶有腺苷脱氨酶 (ADARs) 和胞嘧啶脱氨酶 (C-to-U editing enzymes) 。
- **类型**：
- **A-to-I编辑**：主要发生在哺乳动物中，参与调节基因表达和调控神经递质受体等功能。
- **C-to-U编辑**：常见于线粒体mRNA或特定的核基因中。
- **作用**：RNA编辑可以改变mRNA的编码信息，影响蛋白质结构和功能，从而调节细胞的生理过程。

3.6.4. RNA降解

1) mRNA稳定性的维持

这张幻灯片的主题是 **mRNA稳定性维持的机制**。以下是具体内容的简要说明：

1. 一般途径

- mRNA的修饰
- mRNA状态（突变或翻译状态）

2. 顺式作用元件

- 快速降解信号：
 - AU富集元件（ARE），例如 AUUUA 序列，位于 3' 非翻译区（UTR）。例如：c-Fos、c-Myc、白细胞介素4和6。
- 极端降解信号：
 - 酵母Matla，其编码区包含一个 MIE（mat1不稳定元件），长度为65个核苷酸。

具体构成包括：

1. **5' 端帽结构（5' cap）**：mRNA的5'端有一个甲基化的7-甲基鸟嘌呤帽（m7G），这个帽子结构有助于保护mRNA免受核酸外切酶的降解，同时也是翻译起始的重要信号。

2. **3' 端多聚腺苷酸尾（poly-A tail）**：mRNA的3'端通常会有一段由腺苷酸（A）组成的多聚腺苷酸尾，它在mRNA的稳定性中起重要作用，随着翻译的进行，poly-A尾会逐渐缩短，最终导致mRNA降解。

3. **eIF4E 和 eIF4G 蛋白的相互作用**：eIF4E是结合在5'帽结构上的蛋白质，eIF4G则是一个桥接蛋白，它与eIF4E结合后，还可以与结合在3' poly-A尾上的poly-A结合蛋白（PABP）相连，形成mRNA环的结构。这种环状结构能够促进核糖体的循环利用，从而提高翻译效率。

2) 真核细胞RNA降解途径

这张幻灯片的内容是关于 **真核细胞RNA降解途径**，其关键点如下：

1. 一般mRNA降解途径

去腺苷化 去腺苷化酶 PAN复合体,CCR4-NOT复合体,PARN复合体

去帽（Decapping）

mRNA的5'端有一个7-甲基鸟嘌呤帽（m7G cap），这个帽子结构在mRNA翻译和稳定性中起着重要作用。在降解之前，mRNA的5'帽结构会被去除，这一过程由去帽酶（decapping enzyme，如DCP1和DCP2复合体）完成。去掉帽结构后，mRNA变得不稳定，更容易被外切核酸酶降解

一般mRNA降解途径通过先去帽、去腺苷化，再由外切核酸酶进行5'到3'或3'到5'的降解

mRNA脱帽反应的几个重要因素，具体包括以下几点：

1. Poly(A) 尾状态

- Poly(A)结合蛋白通过与5'端帽复合物的物理相互作用（通过eIF4G），抑制脱帽，从而稳定帽结构并增强翻译。

2. 翻译状态

- mRNA需要从翻译活跃状态转变为翻译不发生的状态（伴随mRNP结构的变化）才能进行脱帽。

3. 帽结构状态

- 脱帽之前，CAP蛋白复合物必须从帽结构上解离。

4. 脱帽复合物的组装

- 需要特定因子参与脱帽复合物的定位、组装和活化。

2. 异常RNA降解途径

- 识别并降解异常的mRNA或翻译效率低下的mRNA。
- 这一过程提高了RNA合成的质量控制。

异常RNA降解途径，也称为RNA监视途径（RNA surveillance pathways）。

A. 无义介导的mRNA降解（Nonsense-mediated mRNA decay, NMD）

- **作用：**降解带有过早终止密码子的mRNA（即mRNA中存在提前终止翻译的无义密码子）。
- **功能：**防止产生过短或功能异常的蛋白质，这些蛋白质可能对细胞有害。

过程：

1. 前体mRNA的核输出

- 剪接后的mRNA包含多个外显子（如Ex1、Ex2、Ex3），在这些外显子连接处形成外显子连接复合物（Exon Junction Complex, EJC）。这些复合物标记了剪接位置，mRNA通过核孔复合体（Nuclear pore complex）从细胞核转运到细胞质中。

2. 初次翻译

- 在细胞质中，mRNA进入翻译过程的第一轮，称为Pioneer round of translation。在这一阶段，核糖体沿着mRNA移动，进行翻译。

- 如果在翻译过程中，核糖体遇到了正常终止密码子（Normal Ter），并且EJC位于终止密码子后方较远的区域，翻译会正常完成。

3. 无义终止（PTC, Premature Termination Codon）有残留的EJC就被识别出来了

- 如果mRNA上存在提前终止密码子（PTC），翻译过程中核糖体会在PTC位置停下，这一信号会触发NMD途径。

- EJC标记了mRNA的正常剪接位点，当核糖体在PTC处终止时，距离太近的EJC无法被移除，系统识别出这一异常并开始降解该mRNA。

4. 降解过程

- NMD机制通过各种降解酶，如去帽酶（decapping enzyme）和外切核酸酶（exonucleases），迅速降解带有PTC的mRNA，防止产生有害的截短蛋白质。

5. 稳定状态下的翻译

- 对于不含PTC的mRNA，核糖体正常翻译到末端，EJC复合物被移除，mRNA进入稳定状态下的翻译循环

在mRNA的翻译过程中，核糖体沿着mRNA链从5'端向3'端移动，按照mRNA上的密码子顺序合成蛋白质。这个过程是高度保守的，适用于所有的生物体。

1. 核糖体首先结合到mRNA的5'帽结构（m7G cap）。

2. 开始扫描mRNA，直到找到**起始密码子（AUG）**。

3. 从起始密码子开始，核糖体沿着mRNA的5'到3'方向移动，每次读取一个三核苷酸的密码子，合成相应的氨基酸链。

4. 当核糖体遇到****终止密码子（UAA、UAG、UGA）****时，翻译结束。

B. 无终止mRNA降解（Nonstop mRNA decay, NSD）

- **作用：**降解缺少终止密码子的mRNA。

- **功能：**确保所有mRNA序列能够被正常翻译终止，防止产生异常延长的蛋白质。

C. 卡顿mRNA降解（No-go mRNA decay, NGD）

- **作用：**降解核糖体翻译过程中停滞的mRNA。

- **功能：**清除由于mRNA二级结构或翻译错误导致核糖体无法继续前进的mRNA，维持翻译过程的顺畅。

D. 核糖体延伸介导的降解（Ribosome extension-mediated decay, REMD）

- **作用：**降解核糖体翻译超过终止密码子并进入3'非翻译区（UTR）的mRNA。

- **功能：**避免核糖体误读终止密码子，确保翻译按计划正确的在终止位点结束。

这些降解途径在细胞中起到**质量控制**的作用，确保异常mRNA不会被错误翻译为有害蛋白质，从而维持基因表达的准确性和细胞功能的稳定性。

3. 特异性mRNA降解途径

- 依赖于RNA顺式作用元件（cis elements），对特定信号做出应答。
- 特定信号包括：激素、细胞周期、病毒感染、分化、营养供应、应激等。

4. 细胞质内特定的mRNA降解位点

- 处理小体（P bodies）。

处理小体（Processing bodies, P-bodies）是细胞质内特定的mRNA降解和储存位点，参与mRNA的代谢和调控。这些小体是富含mRNA降解、翻译抑制和RNA干扰相关蛋白的细胞器，通常通过荧光显微镜可以观察到它们在细胞质中的颗粒状结构。

处理小体的组成和功能

1. 组成成分：

- 去帽酶（Decapping enzymes, 如DCP1、DCP2）：这些酶负责去除mRNA 5'端的帽结构，使其更容易被降解。
- 5'到3'外切核酸酶（Xrn1）：一旦mRNA被去帽，它负责从5'端开始降解mRNA。
- 翻译抑制因子（如GW182、Argonaute proteins）：这些蛋白质与mRNA结合，抑制其翻译。
- RNA结合蛋白和调控蛋白：如Lsm蛋白复合物和Edc3，这些蛋白调控mRNA的稳定性和降解。

2. 主要功能：

- mRNA降解：P小体是mRNA去帽和降解的主要场所。当mRNA的5'帽结构被去除，随之被外切核酸酶降解。
- 翻译抑制：P小体中富含翻译抑制蛋白，未被翻译的mRNA在P小体中聚集，使得这些mRNA暂时不进行翻译，但仍可能处于稳定状态并在合适条件下重新进入翻译途径。
- RNA干扰（RNA interference, RNAi）：P小体是miRNA和siRNA等小分子RNA调控mRNA翻译和降解的中心场所。在P小体中，miRNA通过与Argonaute蛋白结合，抑制靶mRNA的翻译或促进其降解。

处理小体的调控

P小体的形成和功能受到细胞内外环境信号的调控。比如，在应激条件下，P小体数量会增加，以处理更多需要被降解或抑制的mRNA。此外，P小体与**应激颗粒（stress granules）**具有一定的关联，后者是与翻译抑制有关的细胞质内另一类颗粒结构，二者可以互相传递mRNA。

处理小体是细胞内mRNA降解和翻译抑制的重要场所，能够调控mRNA的稳定性、翻译状态及降解途径。通过聚集mRNA和相关蛋白质，P小体在基因表达的调控

体细胞重组学说

抗体遗传

1. 抗体分子的基本结构：

- 抗体（免疫球蛋白）由两条重链（Heavy Chain）和两条轻链（Light Chain）构成，形成Y字形结构。

- 每条重链大约由450个氨基酸组成，分子量大约为50 kD。
- 抗体的轻链和重链之间通过二硫键（S-S键）连接，每条轻链约由200多个氨基酸组成。
- 抗体的可变区决定了它对抗原的特异性，而恒定区决定了它的效应功能。

高变区：体细胞高突变形成

抗体（免疫球蛋白，Ig）由两条重链（Heavy Chain）和两条轻链（Light Chain）通过**二硫键（S-S键）**连接形成Y字形结构。

- VH Domain 和 VL Domain：抗体的可变区（Variable Region），位于重链和轻链的末端，负责识别和结合特定抗原。
- 功能区 是Ig分子中由链内二硫键连接形成的球状区域，负责执行特定功能。抗体的轻链和重链各包含多个功能区：
 - 轻链有两个功能区：可变区（VL）和 恒定区（CL）。
 - 重链包含四个功能区：可变区（VH）和三个恒定区（CH1、CH2、CH3）在不同类型的抗体（如IgG、IgA、IgM等）中结构和功能有所不同。
 - 抗体的可变区负责结合抗原，特异性识别外来物质（如病毒、细菌等）。
 - 恒定区决定了抗体的效应功能，如激活补体系统、与Fc受体结合等。

2. 抗体的类型：

- 根据重链的不同，抗体可分为五类：IgA、IgD、IgE、IgG、IgM，每种抗体在免疫反应中起不同的作用。例如：

- IgM：通常是早期免疫反应中首先生成的抗体。
- IgG：是最主要的血清抗体，参与长期免疫保护。

3.基因重排 (Gene rearrangement) 是抗体初次多样性产生的关键机制。通过基因重排，免疫系统能够产生大量不同的抗体来识别并对抗各种病原体。基因重排主要发生在B淋巴细胞的发育过程中，特别是在重链和轻链基因的V（可变区）、D（多样性区，仅在重链中）和J（连接区）片段的重新排列中。

抗体基因重排的基本原理：

1. 免疫球蛋白 (Ig) 基因结构：

- 抗体的重链和轻链基因位于不同的染色体上，每条链的基因分为多个片段：
- 重链 (Heavy chain) 基因包含 **V、D、J 和 C（恒定区）** 片段。
- 轻链 (Light chain) 基因仅包含 **V、J 和 C（恒定区）** 片段，没有D片段。

2. 重链基因重排：

- **V-D-J重排**：在B细胞的早期发育阶段，**首先**发生的是重链基因的重排。来自不同V、D、J片段的组合被选择并重排在一起，形成可变区。随后，可变区与恒定区（C区）连接，完成重链的编码。

- 通过V、D、J片段的不同组合，B细胞可以产生大量不同的重链序列。

3. 轻链基因重排：

- **V-J重排**：与重链类似，轻链基因的V和J片段通过重排形成可变区，然后与恒定区连接，形成完整的轻链。

- 轻链也有多种片段组合，这进一步增加了抗体的多样性。

4. 片段组合和接合多样性：

- **片段组合多样性**：通过重链的V、D、J片段和轻链的V、J片段的组合，产生数百万种不同的抗体可变区序列。

- **接合多样性**：在片段重排过程中，加入或删除核苷酸（如通过端连接多样性（N-terminus addition））可以进一步增加抗体的多样性。

5. 体细胞超突变 (Somatic hypermutation)：

- 在抗原刺激后，B细胞的抗体基因可在**体细胞超突变**过程中积累突变，从而增加抗体与抗原的亲和力。这种突变发生在成熟B细胞的抗体基因可变区，进一步优化抗体与抗原的结合。

等位排斥（Allelic Exclusion）是免疫系统中一种重要机制，它确保每个B淋巴细胞（B细胞）只表达来自一个等位基因的抗体重链和轻链。这一机制的核心作用是避免一个B细胞同时产生多种不同特异性的抗体，从而保证抗体对抗原的特异性。

等位排斥的基本原理：

1. B细胞抗体基因的两套等位基因：

- 每个B细胞都有两套免疫球蛋白（抗体）基因，分别来自父母双方。B细胞的重链和轻链基因各有两份拷贝。

- 如果B细胞同时从两个等位基因（父本和母本）重排出不同的抗体基因，那么这个B细胞就可能表达两种不同特异性的抗体。这会导致对抗原识别的混乱，从而影响免疫反应的特异性。

2. 重链的等位排斥：

- 在B细胞发育过程中，首先发生的是重链基因的重排。通常，B细胞会随机选择其中一个等位基因（父本或母本）进行V-D-J重排，如果重排成功并生成了功能性抗体重链，该重链会立即表达，并抑制另一个等位基因的重排。这种抑制确保了B细胞只使用一个等位基因生成重链。

- 如果第一次重排失败，B细胞会尝试重排另一个等位基因上的重链。如果两次重排都失败，该B细胞通常会凋亡。

3. 轻链的等位排斥：

- 重链重排成功后，轻链的V-J片段开始重排。轻链也有两套等位基因（来自父母的两条染色体）。类似重链，轻链的重排通常只在一个等位基因上发生。如果重排成功并生成了功能性轻链，另一个等位基因的重排会被抑制。

- 如果轻链的重排不成功，B细胞会尝试在另一个等位基因上进行重排。

4. 机制控制：

- 信号反馈：当B细胞成功生成功能性抗体分子（即重链和轻链成功重排并表达）时，通过抗体表达与细胞膜上的受体信号通路，抑制另一个等位基因的重排。

- 效应器蛋白：参与等位排斥的具体分子机制还包括一些信号转导蛋白和转录因子，它们帮助维持对未使用等位基因的抑制。

等位排斥的意义：

1. 抗体特异性：等位排斥确保每个B细胞只能表达一种抗体，针对单一特异性抗原进行应答。这是抗体高度特异性、精确识别抗原的重要保障。

2. **避免混乱信号：**如果没有等位排斥，B细胞可能同时表达两种或多种抗体，导致识别多个抗原，这可能引发免疫反应失调，甚至产生自身免疫问题。

3. **免疫系统的效能：**通过等位排斥，免疫系统能够集中精力生成针对特定病原体的抗体，而不是在一个细胞内产生多个抗体。这提高了免疫反应的效能。

基因重排的意义：

1. **抗体多样性：**基因重排为每个B细胞产生独特的抗体提供了基础。通过V、D、J片段的重排，可以产生数百万种不同的抗体分子，确保免疫系统能够识别和中和广泛的病原体。

2. **特异性：**每个B细胞只进行一次基因重排，确保每个B细胞只能产生一种特异性的抗体，从而在抗原识别中高度精准。

3. **适应性免疫：**基因重排是适应性免疫反应的核心机制之一，赋予免疫系统高度灵活性，使其能够应对未曾遇到的病原体。

基因重排是抗体多样性产生的核心机制，通过V、D、J片段的重排和接合多样性，B细胞可以生成大量不同的抗体。这个过程确保了免疫系统能够识别广泛的病原体，并在体细胞超突变的帮助下产生高亲和力的抗体，从而有效保护机体。

基因突变

DNA损伤修复