**核酸----遗传信息的载体**

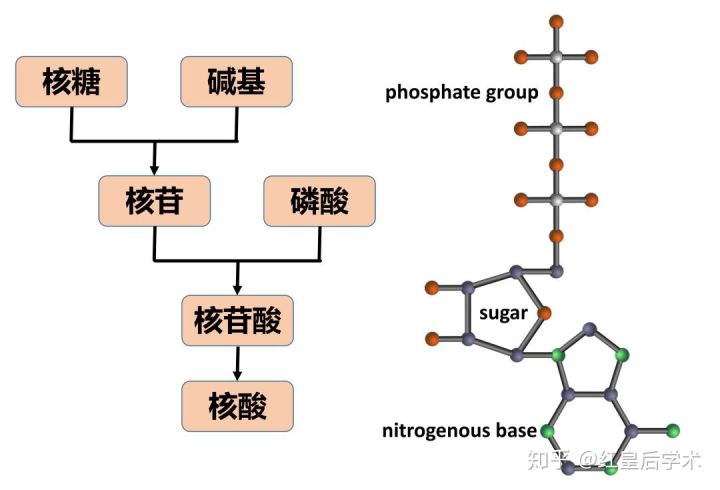
## 核 酸

核酸 (Nucleic acids) 是一类生物大分子化合物，是构成生命的最基本物质之一。**核酸存在于所有生命中**，包括动植物细胞、细菌、古菌、真核生物线粒体和叶绿体、病毒以及类病毒。

**核酸是生物的最基本组成物质和生物学研究的基础物质**，针对核酸的实验研究是现代生物学和医学的重要部分，同时其也是基因组学、生物工程技术以及制药工业的基础。

## ​核酸的基本结构

**核苷酸是组成核酸的基本单位**，即组成核酸分子的单体，核酸由许多核苷酸分子聚合而成。

核酸的组成结构

**一个核苷酸分子由一分子含氮的碱基、一分子戊糖和一分子磷酸组成**。构成核苷酸的**碱基分为嘌呤和嘧啶两类**，前者主要指腺嘌呤 (A) 和鸟嘌呤 (Ｇ)，后者主要指胞嘧啶 (C)、胸腺嘧啶 (Ｔ) 和尿嘧啶 (U)。

嘌呤环上的N-9或嘧啶环上的N-1是构成核苷酸时与戊糖形成糖苷键的位置，其与戊糖环C-1位置所连的羟基形成糖苷键，**一分子碱基和一分子戊糖组成的物质称为核苷**。核苷通过戊糖环的C-3或C-5位置与磷酸基团相连，故构成核酸的核苷酸可分别视为3'-核苷酸或5'-核苷酸。

核酸是由众多核苷酸聚合而成的多聚核苷酸，**相邻两个核苷酸之间的连接键即3',5'-磷酸二酯键**，这种链接可理解为核苷酸糖基上的C-3位羟基与相邻5'-核苷酸的磷酸残基之间，或者核苷酸糖基上C-5位的羟基与相邻3'核苷酸的磷酸残基之间形成的两个酯键，多个核苷酸残基以这种方式连接而成的分子就是核酸。

## 核酸的分类

根据化学组成不同，**核酸可分为两种，脱氧核糖核酸 (DNA) 和核糖核酸 (RNA)**。所有活细胞中均同时含有DNA和RNA，而病毒大多数情况下只含有DNA或RNA的一种。

**DNA和RNA的区别：**

* DNA中的戊糖为脱氧核糖，RNA为核糖，核糖在戊糖环的C-2位点含有一个羟基，脱氧核糖没有，这个羟基使得RNA更不稳定，更容易被水解;
* 通常情况下DNA是双链而RNA是单链，也有一些特殊情况，比如某些病毒的基因组是双链RNA，也有些病毒的基因组为单链DNA；
* 在DNA中与A互补的碱基是T，而RNA中为U;
* RNA具有多种生物学功能，根据其功能不同可以分为tRNA、rRNA、mRNA、sRNA、miRNA等；
* DNA的双链结构包括线性双链DNA和环状双链DNA；
* RNA虽然是单链，但是其可以形成链内的部分双链结构，例如tRNA。

## DNA

## DNA是生物的遗传物质

### DNA的发现

在1868年，瑞士生物学家**Fridrich Miescher**首次从外科绷带上的脓细胞中分离得到一种富含磷元素的酸性化合物，因**存在于细胞核中**而将它称为“核质 (nuclein)”。

在核质发现20年后的1889年，**Richard Altmann**正式将这种物质命名为“核酸 (nucleic acids)”。

### 格里菲斯实验

1928-1936年，Frederick Griffith的实验证明了**遗传物质可以从一个菌株转移到另一个细菌菌株中**。

此实验是利用*两个不同的肺炎双球菌品系，一种是III-S型 (平滑型，有毒性)，另一种是II-R型 (粗糙型，无毒性)*。无论是III-S型还是II-R型，被高温灭活后均无毒性，不会引起宿主死亡。

**实验主要分为四种不同的处理方式：**



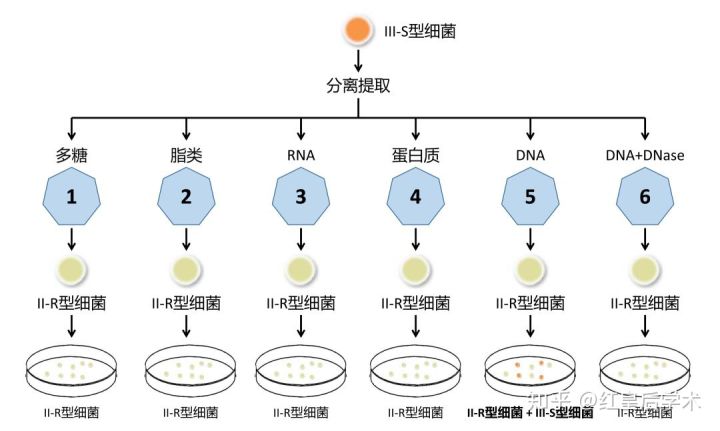
**处理方式实验结果**II-R型菌株小鼠存活III-S型菌株小鼠死亡加热杀死的III-S型菌株小鼠存活II-R型菌株+加热杀死的III-S型菌株小鼠死亡

在其中一种处理方式中，格里菲斯将来自III-R品系的细菌以高温杀死，再将其残骸与活的II-R品系混合。实验结果显示此组合可将宿主老鼠杀死，而且从这些死亡的老鼠体内，可分离出活的III-S品系和II-R品系。因此，格里菲斯提出一项结论，认为**II-R品系被死亡的III-S品系所含的一种转型因子（transforming principle）所“转型”成为具有致命性的III-S品系**。

### 肺炎双球菌转化实验

在1944年，Oswald Avery、Colin MacLeod和Maclyn McCarty的在Griffith实验的基础上进一步通过**肺炎双球菌的体外实验，证明了DNA是生物的遗传物质**。

他们从III-S型活菌体内分别提取DNA、RNA、蛋白质和荚膜多糖，将它们分别和II-R型活菌混合均匀后，进行体外转化实验。

肺炎双球菌转化实验

实验结果表明，**从III-S型菌体中分离得到的DNA能够被II-R型细菌摄取，使其转化为III-S型细菌，这一过程后来被称作DNA的转化 (Transformation)**。II-R型向III-S型转化的转化率与所加的III-S型细菌DNA的浓度和纯度呈正相关，若将DNA预先用DNA酶降解，转化现象就不会发生。

*但是由于当时提取的DNA依然含有一定量的蛋白质，因此并未完全排除蛋白质作为遗传物质的可能性。*

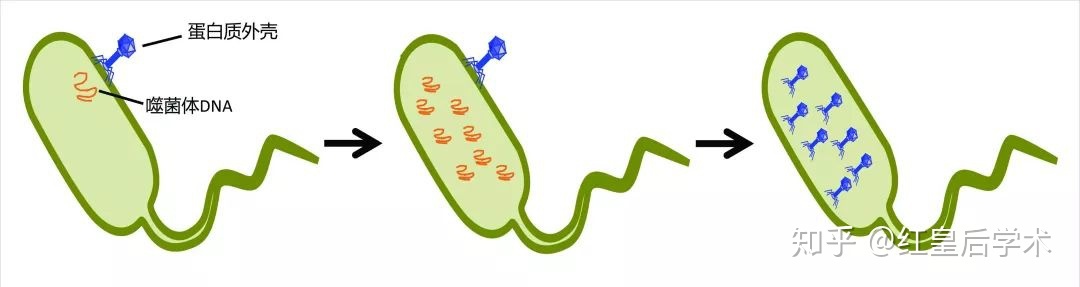
### T2噬菌体侵染实验

1952年，Alfred D. Hershy和Martha Chase的**T2噬菌体侵染实验完全证实了DNA是遗传物质**。

噬菌体T2的结构十分简单，其有一个蛋白质的外壳，DNA包裹其中，当噬菌体T2感染大肠杆菌时，它的尾部会吸附在菌体上，该过程会导致被侵染的细菌体内形成大量的T2噬菌体，菌体裂解后，释放出几十个乃至几百个与感染细菌一样的噬菌体T2。

构成蛋白质的氨基酸中，甲硫氨酸和半胱氨酸含有硫元素，而DNA中不含硫，所以硫元素只存在于T2噬菌体的蛋白质外壳中，相反磷元素主要存在于DNA中，至少占T2噬菌体含磷量的99%。

Alfred Hershy和Martha Chase分别**用35S和32P标记T2噬菌体的蛋白质和DNA**，之后用这些噬菌体侵染没有放射性同位素标记的宿主菌，被35S标记的噬菌体所感染的宿主菌细胞内检测不到35S的存在，而大多数35S出现在宿主菌细胞的外面，也就是说，**35S标记的噬菌体蛋白质外壳在感染宿主菌细胞后，并未进入宿主菌细胞内部而是留在细胞外面**，被32P标记的噬菌体感染宿主细胞后，测定宿主菌的同位素，发现**32P主要集中在宿主菌细胞内，所有噬菌体感染宿主细胞时进入细胞内的主要是DNA**。

T2噬菌体侵染实验

## DNA的结构

### DNA的一级结构

DNA的基本结构为脱氧核糖核苷酸，其**通过核糖的3'端与磷酸基团的5'端通过3',5'-磷酸二酯键相连形成DNA的核苷酸链，即为DNA的一级结构**。

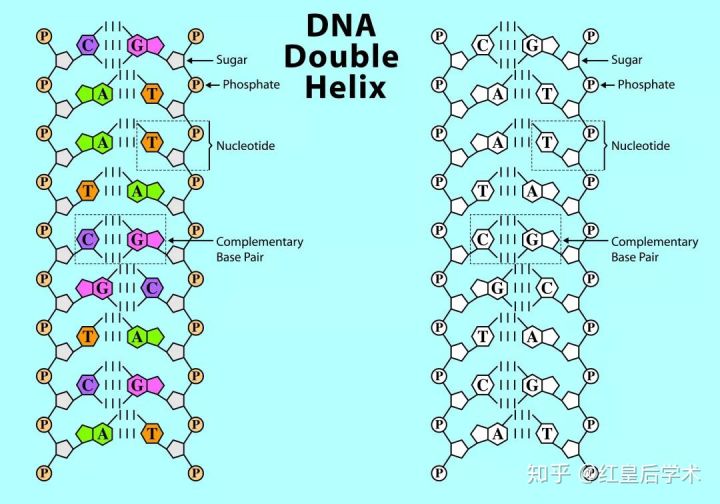
核酸链具有方向性，其两个末端分别是5'末端和3'末端，5'末端含磷酸基团，3'末端含羟基。在DNA一级结构中**4种碱基排列顺序有其特定含义，即遗传信息**，DNA的一级结构决定了遗传信息的种类和数量。

### DNA的二级结构

1949年，Erwin Chargaff发现了DNA中含有4种碱基，A、T、G、C，并且提出了**Chargaff rule**。在任何DNA中，其碱基的数目均符合如下规则：

* A = T
* G = C
* A + C = G + T
* A + G = T + C

1953年，**James Watson和Francis Crick发现了DNA的双螺旋结构**，DNA双螺旋结构模型的确立为遗传学进入分子水平奠定了基础，是现代分子生物学的里程碑。

DNA的二级结构

DNA分子具有规则的双螺旋结构，是由两条相互平行且反向右旋的脱氧核苷酸长链所构成，两条核苷酸链的碱基通过氢键两两配对，配对规则为A对T、G对C。

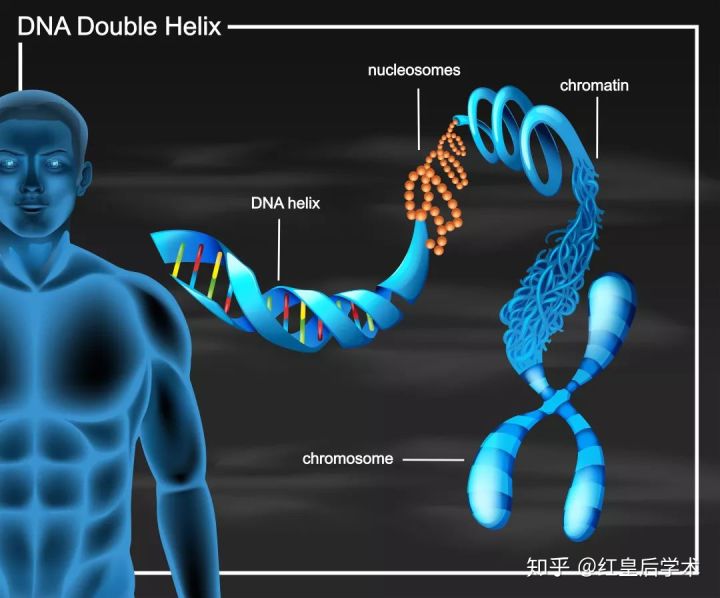
**DNA二级结构的特征：**

* 反向平行的两条链缠绕一个中心轴，两条链均为右手螺旋；
* 碱基在内侧，核糖和磷酸在外侧，碱基平面与中心轴垂直，糖环与中心轴平行，形成大沟和小沟；
* 平均直径2 nm，碱基堆积距离0.34 nm，螺距3.4 nm，螺旋旋转一周正好为10个碱基；
* 两条DNA链依靠彼此碱基之间形成的氢键结合在一起；
* A与T相配对，形成2个氢键，G与C相配对，形成3个氢键；
* 碱基对之间的氢键和碱基的堆积力维持DNA双螺旋结构的稳定。

**DNA二级结构还存在三股螺旋DNA**，三股螺旋DNA中通常是一条同型寡核苷酸与寡嘧啶核苷酸-寡嘌呤核苷酸双螺旋的大沟结合，**三股螺旋DNA存在于基因调控区等重要区域**。

### DNA的三级结构

**DNA的三级结构是指DNA链进一步扭曲盘旋形成超螺旋结构，例如环状双链DNA**。某些病毒和噬菌体的DNA、细菌质粒DNA、线粒体和叶绿体DNA等都是环状双链DNA。

DNA的结构

### DNA的四级结构

由于真核生物的基因组DNA非常大，因此通常与蛋白质结合，经过多层次反复折叠和压缩，**以染色体形式存在于细胞核中，即DNA的四级结构**。线性双螺旋DNA折叠的第一层次是形成核小体，在此基础上核小体又进一步盘绕折叠，最后形成染色体。

## 基因和基因组

遗传信息以核苷酸顺序的形式贮存于DNA分子中，它们以功能单位在染色体上占据一定的位置构成基因，因此，搞清DNA顺序无疑是非常重要的。

**基因的现代分子生物学概念是指能编码有功能的蛋白质多肽链或合成RNA所必须的全部核酸序列，是核酸分子的功能单位**。

一个基因通常包括编码蛋白质多肽链或RNA的编码序列，保证转录和加工所必须的调控序列以及5'端、3'端非编码序列，另外在真核生物基因中还含有内含子等核酸序列。

**基因组（genome）是指一个细胞或病毒的所有基因及其间隔序列**，储存了一个物种所有的遗传信息，在病毒中通常是一个核酸分子的碱基序列，单细胞原核生物中是它仅有的一条染色体的碱基序列，而多细胞真核生物中是一个单倍体细胞内所有染色体的碱基序列。

**基因组的概念是指碱基序列，是核酸的一级结构，不要与DNA的三级和四级结构相混淆**。

基因通过指导蛋白质的合成进而直接控制生物体的性状。**生物中基因序列的差异称为基因型 (genotype)，而生物体性状的差别称为表型 (phenotype)**，例如在格里菲斯实验中，II-R型细菌通过转化摄入III-S型细菌的DNA，该过程改变了宿主细胞的基因型，这些III-S型细菌的DNA在宿主转化细胞中的表达，使得原本不具有毒性的细胞获得毒性，这是改变了这些细胞的表型。分子生物学、遗传学等学科的首要目的是研究生物基因型与表型之间的关联，以及通过基因型确定生物表型的内在机制。

## RNA

## RNA的功能

RNA是以DNA的一条链为模版，以碱基互补配对原则，转录而形成的一条单链，**RNA的主要功能是实现遗传信息在蛋白质上的表达，是遗传信息向表型转化过程中的桥梁**。

**RNA的具体功能包括：**

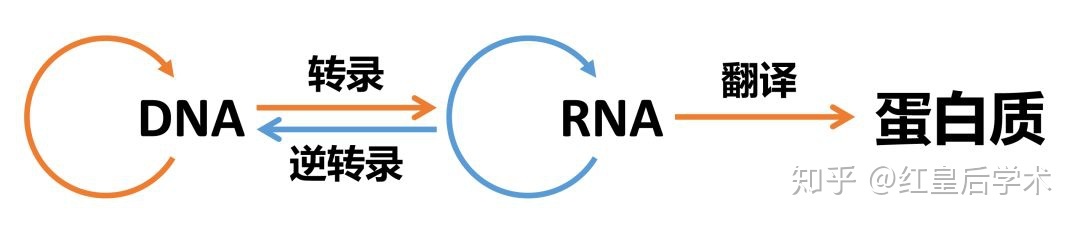
* 控制蛋白质的合成；
* 作用于RNA转录后加工和修饰；
* 基因表达与细胞功能的调节；
* 生物催化与其它细胞持家功能；
* 遗传信息的加工与进化。

*在生物体内，大部分RNA都不负责编码蛋白质，在真核生物中超过99%的RNA为非蛋白质编码RNA，其中最主要的非编码RNA是tRNA和rRNA，它们负责蛋白质合成的翻译过程，其它非编码RNA参与基因调控、RNA加工等其它过程*。

## 中心法则

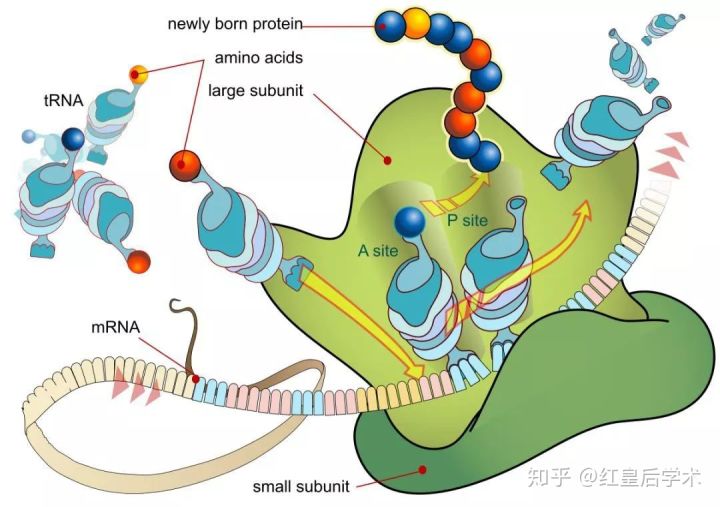
中心法则是现代生物学中最重要最基本的规律之一，其在探索生命的普遍规律中发挥了巨大的作用，极大的推动了现代生物学的发展。

**中心法则（genetic central dogma）：**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，此外，遗传信息也可以从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均遵循的法则；在某些病毒中RNA的自我复制和以RNA为模版逆转录成DNA的过程是对中心法则的补充。

中心法则

中心法则合理地说明了在细胞的生命活动中两类大分子的联系和分工：**核酸的功能是储存和转移遗传信息，指导和控制蛋白质的合成；而蛋白质的主要功能是进行新陈代谢活动以及作为细胞结构的组成成分**。

### 转录和翻译过程

蛋白质翻译过程

**转录 (Transcription) 是遗传信息由DNA转换到RNA的过程，是mRNA及其它非编码RNA (tRNA、rRNA等) 的合成步骤**。在转录过程中，一个基因会被读取、复制为mRNA，这个过程由RNA聚合酶和转录因子所共同完成。

**剪切：在真核细胞中，原始转录产物 (mRNA前体) 还要被加工，使之成为成熟的mRNA**，该过程中一个或多个内含子序列被除去，选择性剪切的机制使之可产生出不同的成熟mRNA分子，并非所有mRNA都要经过这种剪切，**剪切在原核细胞中是不存在的**。

**翻译：成熟的mRNA与核糖体结合，进而开始蛋白质的合成**。在此过程中，核糖体会以三个密码子来读取mRNA上的信息，一般是从AUG开始。**原核细胞没有细胞核，其转录和翻译可同时进行，而在真核细胞中，转录位于细胞核中，翻译位于细胞质中**。

## RNA的结构

绝大部分RNA分子都是线状单链，但是RNA分子的某些区域可自身回折进行碱基互补配对，形成局部双螺旋，而非互补区则膨胀形成环 (loop)，这种短的双螺旋和环区域称为发卡结构 (hairpin)，**发卡结构是RNA中最普通的二级结构形式，二级结构进一步折叠形成三级结构，RNA只有在三级结构时才能称为有活性的分子**。RNA也能与蛋白质形成核蛋白复合物，RNA的四级结构是RNA与蛋白质的相互作用。

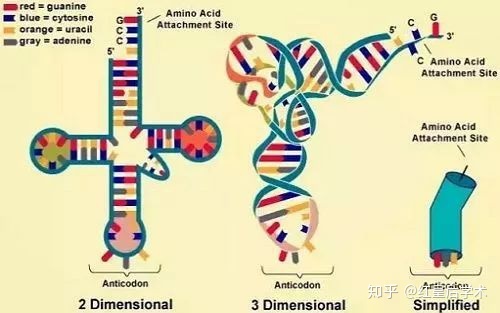
## RNA的分类

RNA具有多种功能，根据其所发挥的功能不同，不同的RNA被赋予了相应的名称。例如，RNA在蛋白质合成过程中发挥重要作用，其中，**tRNA携带和转移活化的氨基酸，mRNA是合成蛋白质的模版，rRNA是细胞合成蛋白质的主要场所**。其它还有诸如参与基因表达调控的miRNA、lncRNA，参与mRNA加工和转运的snoRNA等。

此外，根据RNA链的长度，可以将RNA分为小RNA (samll RNA) 和长RNA (long RNA)，**小RNA的长度小于200个碱基，超过200个碱基的RNA称为长RNA**，长RNA主要包括lncRNA和mRNA，小RNA包括5.8S rRNA、5S rRNA、tRNA、miRNA、siRNA、snoRNA、piRNA、tsRNA和srRNA。

### tRNA

tRNA是一类长度大约为80个碱基的小RNA，其在蛋白质翻译过程中，将特定的氨基酸转运至核糖体，以保证蛋白质多肽链的延伸。tRNA约占总RNA的15%，细胞内每种氨基酸都有其相应的一种或几种tRNA，因此**tRNA种类很多，在细菌中约有30～40种，在动物和植物中约有50～100种tRNA**。

tRNA的结构

**tRNA的一级结构：**tRNA是单链分子，其3'端为CCA-OH，5'端多为pG，其有30%左右的碱基是基本不变的，也就是说这些碱基是保守的，可变区域的碱基通常与氨基酸和密码子的识别有关，用于识别和转运特定的氨基酸以及与mRNA结合进行蛋白质合成。

**tRNA的二级结构：**tRNA的二级结构为三叶草形，通过碱基互补配型形成局部双螺旋构成四臂四环结构。

**tRNA的三级结构：**tRNA的三级结构为倒L形，tRNA的3'端位于倒L形一横的末端，负责结合氨基酸，倒L形一竖的末端为反密码子环，能与mRNA上对应的密码子互补配对。

### rRNA

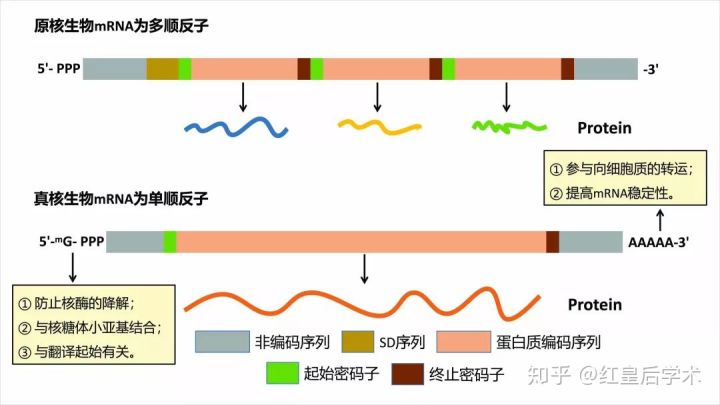
rRNA分子作为骨架与多种核糖体蛋白结合成核糖体作为蛋白质合成的场所，**rRNA是细胞中含量最多的RNA，占细胞总RNA的80%左右**。

**原核生物主要的rRNA有三种，即5S、16S和23S rRNA**，5S含有120个核苷酸，16S含有1500个核苷酸，而23S含有2900个核苷酸。**真核生物则有4种rRNA，即5S、5.8S、18S和28S rRNA**，其分别具有120、160、1900和4700个核苷酸。所有生物体的核糖体都由大小不同的两个亚基所组成，原核生物为50S和30S亚基，30S含有16S rRNA和21种蛋白质，50S含有23S和5S rRNA以及34种蛋白质。真核生物为60S和40S亚基，40S含有18S rRNA和33种蛋白质，60S含有28S、5.8S和5S rRNA以及49种蛋白质。

*S为沉降系数，当用超速离心测定一个粒子的沉降速度时，此速度与粒子的大小直径成比例*。

### mRNA

mRNA的功能是在蛋白质分子合成过程中，将基因组DNA的遗传信息 (即碱基排列顺序) 传递至核糖体，使核糖体能够以碱基排列顺序掺入互补配对的tRNA分子，进而合成正确的肽链，实现遗传信息向蛋白质分子的转化。

原核与真核生物mRNA的结构

**原核生物mRNA结构简单，往往含有几个功能上相关的蛋白质的编码序列，可翻译出几种蛋白质，为多顺反子**。在原核生物mRNA中编码序列之间有间隔序列，可能与核糖体的识别和结合有关；在5'端和3'端有与翻译起始和终止有关的非编码序列，原核生物mRNA中没有修饰碱基，5'端没有帽子结构，3'端没有多聚腺苷酸的尾巴 (polyA尾巴)。**原核生物mRNA的半衰期比真核生物要短得多，转录后1min，mRNA降解就开始**。

**真核生物mRNA为单顺反子结构，即一个mRNA分子只包含一条多肽链的信息**。在真核生物成熟的mRNA中5'端有mGpppN的帽子结构，帽子结构可保护mRNA不被核酸外切酶水解，并且能与帽结合蛋白结合识别核糖体，与翻译起始有关；3'端有polyA尾巴，其长度为20～250个腺苷酸，其功能可能与mRNA的稳定性有关。

### 其它RNA分子

**核内小分子RNA (snRNA)** 是细胞核内核蛋白颗粒的组成成分，参与mRNA前体的剪切以及成熟的mRNA由核内向细胞质中转运的过程。

**核仁小分子RNA (snoRNA)** 是一类指导其它RNA化学修饰的小RNA分子，主要作用于rRNA、tRNA和snRNA。snoRNA分为两类，C/D box snoRNA主要介导甲基化，H/ACA box snoRNA主要介导假尿苷酸化。

**microRNA (miRNA)** 是一种具有茎环结构的非编码RNA，长度一般为20～24个核苷酸，其功能为基因表达沉默和转录后的基因表达调控。其通过碱基互补配对与mRNA的特定区域结合，使发生下列变化从而抑制其表达：

1. 将mRNA裂解为两段；
2. 缩短mRNA的polyA尾，使其变得不稳定，更易被降解；
3. 降低mRNA与核糖体的结合，从而降低其翻译效率。

**siRNA (Small interfering RNA)** 是一类双链RNA分子，长度通常在20-25个碱基左右，其功能类似于miRNA，能够抑制特定基因的表达，其通过与特定基因的mRNA某一区域的互补配对，激活mRNA的降解，使其无法进行蛋白质的翻译过程，目前多用于RNA干扰技术沉默特定基因以分析其功能。  
  
**lncRNA (Long non-coding RNA)** 为长度超过200个碱基，但是不具有蛋白质编码能力的转录本。lncRNA主要的功能是在细胞核和细胞质中与转录因子相互作用调节mRNA的转录，另外lncRNA还参与转录后调控机制，其可以使miRNA远离它作用的靶基因。

**环状RNA (circRNA)** 是生物体内存在的一类不具有5'末端帽子和3'末端polyA尾巴，并能通过共价键形成闭合环形结构的RNA分子。circRNA含有大量的miRNA结合位点，可以竞争性结合miRNA发挥miRNA的海绵作用，从而间接调控miRNA下游靶基因的表达。另外，circRNA还可以通过结合RNA结合蛋白，形成RNA-蛋白复合物，调控基因的转录。

**细菌sRNA**是细菌中长度在50～500个碱基的非编码RNA，它们主要位于基因间区或蛋白质编码基因的5'和3'-UTR区域。据估算，细菌基因组中sRNA的数量可能占5%左右。这些sRNA参与很多生物学过程，如细菌的转录调控、RNA的加工和修饰、mRNA的稳定、mRNA的翻译、蛋白质降解、质粒复制和细菌感染等。

参考

<https://zhuanlan.zhihu.com/p/77606144>