

## 绪论

- 人类基因组：核基因组： $3 \times 10^9$ bp（每个染色体 55-250Mb）  
线粒体基因组：16569bp（每个细胞 800 个线粒体，每个线粒体 10 个基因组拷贝）
- 可变的 DNA 结构：A-DNA（脱水），B-DNA（水环境，活细胞），Z-DNA（连接 Pro，罕见）
- DNA 超螺旋：  
30 亿对碱基；拉直 2m 长；6 微米宽
- 基因组的序列组成【微观一级结构】
  - 重复序列 k-mers：
    - 🚩 高度重复序列（卫星 DNA/小卫星 DNA/微卫星 DNA/VNTR）；
    - 🚩 中度重复序列（逆转录转座子、DNA 转座子、长末端重复(LTRs)、非长末端重复(non-LTRs)、长间隔重复(LINEs)、短间隔重复(SINEs)）
  - 非重复序列：单一序列（80%基因）
  - 其他 DNA 元件：CpG 岛、G-quadruplex、启动子、TFBs、转录起始位点 TSS、终止子

PS:卫星 DNA:由大段重复的非编码组成的，卫星 DNA 是功能性中心粒主要组分，形成异染色质的主要结构；

小卫星 DNA:一类可变串联重复（variable number tandem repeat, VNTR），有一系列  $10 \sim 60$ bp 重复序列组成的 DNA 片段；

微卫星 DNA:简单序列重复(SSRs)或短串联重复序列，是重复序列的 2-6 碱基对 DNA。它是一种变数串联重复(VNTR)；

VNTR 是基因组中的一个位置，短核苷酸序列被组织成串联重复序列。这些可以在许多染色体上发现，并且经常显示个体之间的长度变化。

- non-codingRNA 基因：rRNA、tRNA、scRNA、snRNA、snoRNA、miRNA
- 生物基因中有哪些异常结构基因：重叠基因【基因内基因、反义基因】
- 假基因：指来源于功能基因但失去其活性的 DNA 序列，有沉默的假设基因，也有可转录的假基因

全基因组测序=》序列组装=》基因组注释 =》信息查询、基因搜寻、启动子分析、其他 DNA 元件 =》比较基因组学、RNA 序列分析、其他核酸分析软件使用