

附录： 与核苷酸和蛋白质序列相关的特征关键词表

表 1 与核苷酸序列相关的特征关键词表

关键词	说明
allele	相关的个体或菌株含有相同基因的稳定的其它形式, 该形式区别于这一位置的现有的序列(和或许其它序列)
attenuator	存在调节转录的终止的 DNA 区域, 它控制了一些细菌操纵子的表达; (2) 位于启动子和第一个结构基因之间, 引起转录的部分终止的序列区段
C_region	免疫球蛋白轻和重链的恒定区, 和 T-细胞受体 α , β , 和 γ 链; 根据特定的链可包括一个或多个外显子
CAAT_signal	CAAT 盒; 位于可能参与 RNA 聚合酶结合的真核生物转录单位的起始点的 75bp 上游的保守序列的一部分; 共有序列=GG(C 或 T)CAATCT
CDS	编码序列; 对应于蛋白质中的氨基酸序列的核苷酸的序列(位置包括终止密码子); 特征包括氨基酸概念上的翻译
Conflict	在这一位点或区域, 单独确定的“相同”序列有所不同
D-loop	置换环; 线粒体 DNA 内的一个区域, 其中 RNA 的短的序列与 DNA 的一条链配对, 代替了这一区域的原始配对 DNA 链; 也用于说明在 RecA 蛋白质催化的反应中, 侵入的单链替代双链 DNA 的一条链的区域
D-segment	免疫球蛋白重链的多变区, 和 T-细胞受体的 β 链
Enhancer	顺式-作用序列, 它增强了(一些)真核生物启动子的作用, 并能在任一方向和与启动子相关的任何位置处(上游或下游)起作用
Exon	编码剪接 mRNA 部分的基因组区域; 可以含有 5' UTR, 所有 CDS, 和 3' UTR
GC_signal	GC 盒; 位于真核生物转录单位起始点上游的保守的富含 GC 区域, 可以以多重拷贝或任一方向存在; 共有序列=GGCGG
gene	鉴定为基因的生物学意义的区域, 并已经指定名称
iDNA	间插 DNA; 通过几种重组中的任何一种能被消除的 DNA
intron	被转录的 DNA 区段, 但通过同时剪接位于其两侧的序列(外显子)即可从转录本内部将其除去

J_segment	免疫球蛋白轻链和重链的连接区段, 和 T-细胞受体 α , β 和 γ 链
LTR	长的末端重复, 在确定序列的两端直接重复的序列, 类型典型地见于逆转录病毒中
mat_peptide	成熟的肽或蛋白质的编码序列; 翻译后修饰之后成熟的或最终的肽或蛋白质产物的编码序列; 位置不包括终止密码子(与相应的 CDS 不同)
misc_binding	不能用任何其它 Binding 关键词(primer_bind 或 protein_bind)表述的与另一个组成成分共价或非-共价结合的核酸中的位点
misc_difference	特征序列与记载中存在的有所不同, 并且不能用任何其它不同关键词(conflict, unsure, old_sequence, mutation, variation, allele 或 modified_base)表述
misc_feature	不能用任何其它的特征关键词表述的具有生物学意义的区域; 新的或少见的特征
misc_recomb	任何一般性的, 位点特异性的或复制的重组事件的位点, 该位点中有不能用其它重组关键词(iDNA 和 virion)或来源关键词的修饰词(/transposon, /proviral)表述的双螺旋 DNA 的断裂和愈合
misc_RNA	不能用其他 RNA 关键词 (prim_transcript, precursor_RNA, mRNA, 5' clip, 3' clip, 5' UTR, 3' UTR, exon, CDS, sig_peptide, transit__peptide, mat_peptide, intron, polyA_site, rRNA, tRNA, scRNA 和 snRNA) 限定的任何转录本或 RNA 产物
misc_signal	含有控制或改变基因功能或表达之信号的任何区域, 所述信号不能用其他 Signal 关键词 (promoter, CAAT_signal, TATA_signal, -35_signal, 10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator 和 rep_origin)表述
misc_structure	不能用其他 Structure 关键词(stem_loop 和 D-loop)表述的任何二级或三级结构或构象
modified_base	被指示的核苷酸是经修饰的核苷酸, 应由被指示的分子(在 mod_base 修饰词意义中给出)所取代
mRNA	信使 RNA; 包括 5' 非翻译区(5' UTR), 编码序列(CDS, 外显子)和 3' 非翻译区(3' UTR)
mutation	在此位置处, 相关品系的序列中具有突然的, 可遗传的变化
N_region	在重排的免疫球蛋白区段之间插入的额外的核苷酸
Old_sequence	在此位置处, 所表述的序列修改了此序列以前的版本
PolyA_signal	聚腺苷酸化之后内切核酸酶裂解 RNA 转录本所必需的识别区域; 共有序列 = AATAAA
PolyA_site	RNA 转录本上的位点, 通过转录后聚腺苷酸化该位点将被加上腺嘌呤残基
Precursor_RNA	仍不是成熟的 RNA 产物的任何 RNA 种类; 可包括 5' 剪切区(5' clip), 5' 非翻译区(5' UTR), 编码序列(CDS, 外显子), 间插序列(内含子), 3' 非翻译区(3' UTR), 和 3' 剪切区(3' clip)

prim_transcript	初级（最初的，未加工的）转录本；包括 5' 剪切区（5' clip），5' 非翻译区（5' UTR），编码序列（CDS, 外显子），间插序列（内含子），3' 非翻译区（3' UTR）和 3' 剪切区（3' clip）
prim_bind	起始复制, 转录或逆转录的非—共价的引物结合位点; 包括合成的例如 PCR 引物元件的位点
Promoter	参与 RNA 聚合酶的结合以启动转录的 DNA 分子区域
protein_bind	核酸上非—共价的蛋白质结合位点
RBS	核糖体结合位点
repeat_region	含有重复单位的基因组区域
repeat_unit	单个重复元件
rep_origin	复制起点; 复制核酸以得到两个相同拷贝的起始位点
RRNA	成熟的核糖体 RNA; 将氨基酸装配成蛋白质的核糖核蛋白颗粒（核糖体）中的 RNA 成份
S_region	免疫球蛋白重链的开关区; 它参与重链 DNA 的重排, 导致来自相同 B—细胞的不同免疫球蛋白类的表达
Satellite	短的基本重复单位的很多串联重复（相同或相关的）; 大多数具有的碱基组成或其它性质与基因组的一般水平不同, 这使得它们与大部分（主带）的基因组 DNA 分离开来
ScRNA	小的细胞质 RNA; 几个小的细胞质 RNA 分子中的任何一个存在于真核生物的细胞质和（有时）核中
sig_peptide	信号肽编码序列; 被分泌的蛋白质的 N—末端结构域的编码序列; 此结构域涉及新生多肽与膜的结合; 前导序列
SnRNA	小的核 RNA; 很多小的 RNA 种类中的任何一个都被局限于核中; 几个 snRNA 参与剪接或其它 RNA 加工反应
source	鉴定序列中特定范围的生物来源; 此关键词是强制性的; 每一项至少要有一个跨越整个序列的单一来源关键词; 每个序列可允许有一个以上的来源关键词
stem_loop	发卡结构; 由 RNA 或 DNA 单链的相邻（反向）互补序列之间的碱基—配对形成的双螺旋区域
STS	序列标记位点: 表述基因组上作图界标并能通过 PCR 检测的短的, 单拷贝 DNA 序列; 通过测定 STS 系列的次序即可作出图谱的基因组区域
TATA_signal	TATA 盒; Goldberg-Hogness 盒; 在每个真核生物 RNA 聚合酶 II 转录单位起点前约 25bp 处发现的保守的富含 AT 的七聚体, 它可能涉及使酶定位以正确地起始; 共有序列=TATA (A 或 T) A (A 或 T)
terminator	或者位于转录本的末端或者与启动子区域相邻的 DNA 序列, 该序列可导致 RNA 聚合酶终止转录; 也可以是阻抑蛋白的结合位点
transit_peptide	转运肽编码序列; 核编码的细胞器蛋白质 N—末端结构域的编码序列; 此结构域参与将蛋白质翻译后运送到细胞器中

tRNA	成熟的转移 RNA, , 小的 RNA 分子 (75—85 个碱基长), 介导核酸序列翻译成氨基酸序列
unsure	作者不能确定此区域的准确序列
V_region	免疫球蛋白轻链和重链的可变区, 和 T 细胞受体 α , β 和 γ 链; 编码可变的氨基末端部分; 可由 V__segment, D_segment, N_region 和 J_segment 组成
V_segment	免疫球蛋白轻链和重链的可变区段, 和 T 细胞受体 α , β 和 γ 链; 编码大多数可变区 (v_region) 和前导肽的最后几个氨基酸
variation	含有来自相同基因的稳定突变的相关系列 (例如 RFLP, 多态性等), 在此 (和可能其它) 位置处所述相同基因与被表述的不同
3' clip	在加工过程中被切下的前体转录本 3' 端大部分区域
3' UTP	不被翻译成蛋白质的成熟转录本的 3' 末端区域 (终止密码子之后)
5' clip	在加工过程中被切下的前体转录本 5' 端大部分区域
5' UTP	不被翻译成蛋白质的成熟转录本的 5' 末端区域 (起始密码子之前)
_ 10 _signal	Pribnow 盒; 细菌转录单位起点上游约 10bp 处的保守区域, 它可能参与结合 RNA 聚合酶; 共有序列=TatAaT
_ 35 _signal	细菌转录单位起点上游约 35bp 处的保守六聚体; 共有序列=TTGACa[] 或 TGTTGACA[]

表 2 与蛋白质序列相关的特征关键词表

关键词	说明
CONFLICT	不同的论文报道了不同的序列
VARIANT	作者报道存在序列变体
VARSLIC	由可选择的剪接产生的序列变体的表述
MUTAGEN	经实验操作已改变的位点
MOD_RES	残基的翻译后修饰
ACETYLATION	N-末端或其它
AMIDATION	通常位于成熟的活性肽的 C-末端
BLOCKED	不能被测定的 N-或 C-末端封闭基团
FORMYLATION	N-末端甲硫氨酸的
GAMMA-CARBOXY-GLUTAMIC	天冬酰胺, 天冬氨酸, 脯氨酸或赖氨酸的

ACID HYDROXYLATION	
METHYLATION	通常为赖氨酸或精氨酸的
PHOSPHORYLATION	丝氨酸，苏氨酸，酪氨酸，天冬氨酸或组氨酸的
PYRROLIDONE CARBOXYLICACID	已形成内部环内酰胺的 N-末端谷氨酸
SULFATATION	通常为酪氨酸的
LIPID	脂质组成成分的共价结合
MYRISTATE	通过酰胺键与蛋白质成熟形式的 N-末端甘氨酸残基或内部的赖氨酸残基结合的豆蔻酸基团
PALMITATE	通过硫酯键与半胱氨酸残基或通过酯键与丝氨酸或苏氨酸残基结合的棕榈酸基团
FARNESYL	通过硫酯键与半胱氨酸残基结合的法尼基
GERANYL-GERANYL	通过硫酯键与半胱氨酸残基结合的香叶基-香叶基基团
GPI__ANCHOR	与蛋白质成熟形式 C-末端残基的 α -羧基相连的糖基-磷脂酰肌醇 (GPI) 基团
N__ACYL DIGLYCERIDE	原核生物脂蛋白成熟形式的 N-末端半胱氨酸，所述脂蛋白具有酰胺-键联的脂肪酸和通过酯键连接了两个脂肪酸的甘油基
DISULFID	二硫键；“FROM”和“TO”终点表示通过一个链-内二硫键连接的两个残基；如果“FROM”和“TO”终点是完全相同的，则二硫键是链-间键，而说明书领域示出交联的性质
THIOLEST	硫醇酯键；“FROM”和“TO”终点表示通过硫醇酯键连接的两个残基
THIOETH	硫醚键；“FROM”和“TO”终点表示通过硫醚键连接的两个残基
CARBOHYD	糖基化位点；碳水化合物（如果已知）的性质在说明书领域给出
METAL	金属离子的结合位点；说明书领域示出金属的性质
BINDING	任何化学基团（辅酶，辅基，等等）的结合位点；基团的化学性质在说明书领域给出
SIGNAL	信号序列的范围（前肽）
TRANSIT	运转肽的范围（线粒体，叶绿体或微体）
PROPEP	前肽的范围
CHAIN	成熟蛋白质中多肽链的范围
PEPTIDE	被释放的活性肽的范围

DOMAIN	序列中感兴趣的区域的范围；所述区域的特征在说明书领域给出
CA_BIND	钙—结合区域的范围
DNA_BIND	DNA—结合区域的范围
NP_BIND	核苷酸磷酸酯结合区域；核苷酸磷酸酯的特征示于说明书领域
TRANSMEM	转膜区域的范围
ZN_FING	锌指区域的范围
SIMILAR	与另一个蛋白质序列具有相似性的区域；与那个序列有关的精确的资料在说明书领域给出
REPEAT	内部序列重复的范围
HELIX	二级结构；螺旋，例如 α —螺旋，3（10）螺旋，或Pi-螺旋
STRAND	二级结构； β —链，例如氢键连接的 β —链，或分离的 β —桥中的残基
TURN	二级结构转角，例如H—键连的转角（3—转角，4—转角或5—转角）
ACT_SITE	涉及酶活性的氨基酸
SITE	序列中任何其它感兴趣的位点
INIT_MET	已知序列以起始密码子甲硫氨酸开始
NON_TER	序列末端的残基不是末端残基；如果应用于位置1，这表示第一个位置不是完整分子的N—末端；如果应用于最后一个位置，这表示此位置不是完整分子的C—末端；对此关键词没有说明书领域
NON_CONS	非连串残基；表示序列中的两个残基不是连串的，在它们之间有很多未测序的残基
UNSURE	序列的不确定性；用于表述不能确定序列排列的序列区域