

# Alu 家族在人类基因组中的作用

王 伟 王亚平 \*

(南京大学医学院, 南京 210093)

**摘要** Alu 家族是灵长类动物特有的且是最重要的短散在元件(short interspersed elements, SINEs), 经过 6 千 5 百万年的进化, Alu 序列在基因组中约有 120 万份拷贝, 占基因组的 10% 以上。Alu 家族在基因组中有很多功能, 如介导重组、基因插入和删除、甲基化和 A-to-I 的编辑作用、调控转录和翻译、选择性剪接等等。Alu 家族的变异与疾病和进化存在密切关系。

**关键词** Alu 家族; 人类基因组; 基因重组; 反转录转座; 选择性剪接

Alu 家族是哺乳动物的短散在元件(short interspersed elements, SINEs)中最大的家族。因一半以上的 Alu 序列双链 DNA 可被限制性核酸内切酶 *AluI* 切割而得名, 切割位置在 170 bp 附近。Alu 序列由 300 bp 左右的短序列所组成, 经过 6 千 5 百万年的进化在基因组中有约 120 万份拷贝。Alu 家族曾经被称为自私的垃圾序列, 但目前对 Alu 家族有了新的较深入的认识<sup>[1]</sup>。本文针对人类基因组中 Alu 家族的研究进展做一综述。

## 1 Alu 家族

Alu 序列的结构由左右两个单体组成, 右侧单体比左侧单体长 31 bp, 每个单体含有一个 RNA 聚合酶 III 启动子, 然而只有左单体中的启动子才有活性。Alu 序列的两翼存在长为 7~20 bp 的正向重复序列, 在 Alu 序列的 3' 端还存在一个长几十 bp 的 poly(A), 在两单体之间有一个富含 A 的区域, 同时 Alu 序列中还存在较多的 CpG 岛<sup>[2]</sup>。Alu 家族可能来自病毒源性的 7SL RNA, 在灵长类动物分支出来时基因组中由 7SL RNA 基因的 5' 和 3' 端融合产生原始的 Alu 序列单体, 再经由复制、缺失等变异, 演化形成现在的 Alu 家族。人类基因组中超过 10% 的序列为 Alu 序列, Alu 序列拷贝的平均间隔约 4 kb<sup>[3]</sup>。人类 Alu 家族具有多态性, 主要分为 Alu-J、Alu-S 和 Alu-Y 三大亚家族。

## 2 Alu 家族的功能

目前, 关于 Alu 家族在基因组中的功能人们已经有了一些认识。它对基因组的结构具有很大的影响, 在多层次上参与基因表达及其调控。

### 2.1 对 DNA 结构及复制的影响

**2.1.1 反转录转座** Alu 家族通过反转录转座机制扩增。哺乳动物反转录转座子有 3 种类型: 病毒超家族, 长散在元件(long interspersed elements, LINEs) 和不编码功能蛋白质的非病毒超家族。Alu 家族属于非病毒超家族的 SINEs, 需要利用 LINE1 家族编码的逆转录酶和核酸内切酶。

Alu 序列的插入可改变 DNA 结构, 主要表现为插入多态性。这种多态性不仅表现为插入位置不同, 也表现为插入/缺失的多态性。Alu 序列插入虽与疾病发生相关, 但并非每一个 Alu 序列插入都是致病风险因素。Alu 序列在血管紧张素转换酶(ACE)基因中的插入即对防止心脏疾病的发生具有保护作用, 而 Hamdi 等<sup>[4]</sup>也曾报道 Alu 序列插入对年龄相关性视网膜黄斑变性(age-related macular degeneration, AMD) 具有防护作用。

Alu 反转录转座介导 DNA 序列的删除(ARD)。近期发现, 除通过同源重组介导基因序列的插入和缺失外, Alu 序列还可通过反转录转座导致基因组不稳定。Salem 等<sup>[5]</sup>首先报道, Alu-Yg6 插入 3 号染色体的同时, 伴随插入位点约 300 bp 的缺失。Carter 等<sup>[6]</sup>也发现, 年轻的 Yb7 亚家族成员在基因组中的插入伴随约 300 bp 的 DNA 缺失。Pauline 等<sup>[7]</sup>在比较人类和黑猩猩基因组的基础上, 评估了反转录转座对基因组的影响, 发现了 33 个 Alu 序列反转录转座介导的 DNA 序列的缺失, 共删除基因组 DNA 约 9 000 bp。有资料显示, 在整个灵长类进化过程中约有 3 000 个 DNA 序列缺失事件与 Alu 序列的反转录转座有关, 共删除

收稿日期: 2007-03-27 接受日期: 2007-06-12

\* 通讯作者。Tel: 025-83686495, E-mail: wangyap@nju.edu.cn

约 900 kb 的 DNA。Alu 反转录转座介导 DNA 序列缺失的可能机制为: L1 内切酶依赖的反转录转座; L1 内切酶非依赖的反转录转座; 在 DNA 断裂处的内部启动反转录转座; 随机位点启动的反转录转座。

**2.1.2 基因重组** Alu 序列往往充当不等同源交换的靶点, 其两侧的基因片段常由此出现缺失或扩增。由于 Alu 序列相互间的近邻分布, 在高密度分布 Alu 序列的区域很容易发生重组, 从而引发染色体易位。

另外, 在重组位点或上下游 20~50 bp 处发现 Alu 核心序列中有 26 bp 的保守序列。Alu 核心序列与 *E. coli* 中促进 RecBCD 介导重组的 X 序列的同源性, 提示核心序列的 26 bp 在某些情况下可作为催化同源重组蛋白质的结合位点。有研究表明在 Alu 序列中含有转位蛋白结合位点的同源序列<sup>[8]</sup>。转位蛋白与特定序列相结合并参与局部 DNA 的解螺旋。且转位蛋白的结合使得染色体局部区域很容易受到核酸酶的攻击并有发生重排的可能。这样可以推测 Alu 序列不仅是因为有大量拷贝而被动地参与重组, 而且还主动促进了重组的发生。另一方面, 非同源染色体上 Alu 序列间的相互吸引, 导致了两个基因位点的接近, 可能使基因重组机会增加。

如果 Alu 序列的不等交换发生在同一条染色体内部, 就可能引起序列的缺失或扩增; 如果发生在两条染色体之间, 将会引起较复杂的染色体重排事件。例如, 两个 Alu-Sx 序列介导的不等同源重组删除了长臂猿基因组中约 100 kb 的 DNA 序列, 这为 Alu 序列不等同源重组在哺乳动物基因组进化中的重要作用提供了新的证据<sup>[9]</sup>。Alu 序列重组导致基因删除与很多人基因疾病相关。Sen 等<sup>[10]</sup>比较了人类和黑猩猩基因组, 确定了从大约 6 百万年前人类 - 黑猩猩分离至今人类基因中这种重组作用的数量, 鉴定出 492 个位点人类基因序列的特异性删除(计约 400 kb 大小) 归因于 Alu 序列重组, 这在人类基因组的插入 / 删除谱中占有重要的分量。

最近还发现一种 Alu 序列介导的非重组性 DNA 大片断缺失, 两个反向同源性强的 Alu 序列形成茎-环结构导致 DNA 大片断的删除, 上游和下游断裂位点都位于 GC 二核苷酸处<sup>[11]</sup>。

**2.1.3 启动 DNA 复制** 在人类基因组中, 由于 Alu 家族含有大量 CpG 岛, 而 GC 富含区较其他区域更易启动 DNA 复制, 故 Alu 序列富含区容易启动 DNA 复制<sup>[12]</sup>。

**2.1.4 甲基化修饰** Alu 序列中 CpG 二核苷酸含

量丰富, 这为人类基因组提供了甲基化的位点。Alu 序列可以启动甲基化的扩展, Alu 序列对甲基化的影响较复杂, 有人研究 Alu 序列的密度和与 CpG 岛的距离对甲基化扩展的影响<sup>[13]</sup>。有人在甲基化区寻找对甲基化扩展有促进或抑制作用的特殊模式<sup>[14]</sup>, 也有其他假说模式但具体机制仍不清楚<sup>[15]</sup>。与 Alu 序列结合的组蛋白 H3 的第 9 位赖氨酸(H3-Lys9)上常发生甲基化。H3-Lys9 的甲基化可能是 Alu 序列的甲基化引起的, Alu 序列通过其甲基化来抑制 Alu 序列的重组<sup>[16]</sup>。

## 2.2 对 RNA 结构及转录的影响

**2.2.1 A-to-I 编辑** A-to-I 编辑是一种 RNA 编辑, 变双链区的腺嘌呤为次黄嘌呤。3 个独立的小组最近对 A-to-I 编辑位点统计发现人类转录组的次黄嘌呤局限于 Alu 序列中。这 3 个小组对照 mRNA(EST 和 cDNA 数据库)和人类基因组序列检测 A-G 替代来确定编辑位点<sup>[17~19]</sup>。超过 90% 的 A-I 替代发生在 RNA 的 Alu 序列中。A-I 替代可发生在转录有义或反义 Alu 序列。编辑的碱基主要与保留的内含子、延长的 UTR 以及那些不能与已知基因对应的 RNA 相关<sup>[20]</sup>。

A-to-I 编辑反应在体内由作用于 RNA 的腺苷脱氨酶(ADAR)家族成员催化, ADAR 在编辑位点上没有严格的序列要求, 但双链 RNA 的形成是编辑的前提条件<sup>[21]</sup>。对经过编辑的 mRNA 分析比较表明, Alu 序列的编辑热点为其一致序列的第 27、28、136 和 162 碱基位点的 A<sup>[18]</sup>。相邻碱基可能影响这些位点的编辑频率, 编辑位点紧邻的上游第一个多为 T, 而 G 十分少见<sup>[17]</sup>。

首先, mRNA 中的 Alu 序列优先被编辑可能是由于 Alu RNA 的二级结构可形成较长的双链区域<sup>[22]</sup>。Athanasiadis 等<sup>[19]</sup>确定了一个 Alu 序列 A-to-I 编辑和附近 Alu 序列的插入存在联系。他们证明当两个反向 Alu 序列间隔小于 2 kb 时编辑很容易发生。这形成了一种模式, 即两个邻近的 Alu 序列产生双链 RNA, 形成 ADAR 的理想底物。这种 A-to-I 编辑模式如图 1 所示<sup>[23]</sup>。对周期蛋白基因 M3 内含子 2 和 NFκB1 基因内含子 16 的研究发现两个反向 Alu 序列间碱基配对发生在分子内却不发生在分子间<sup>[22]</sup>, 这正与上述模式吻合。

虽然 RNA 编辑的确切作用仍不能确定, 但它可能在几个水平上影响了基因的表达。因为次黄嘌呤 I 不能与尿嘧啶 U 配对, 而是与胞嘧啶 C 配对, 编辑作

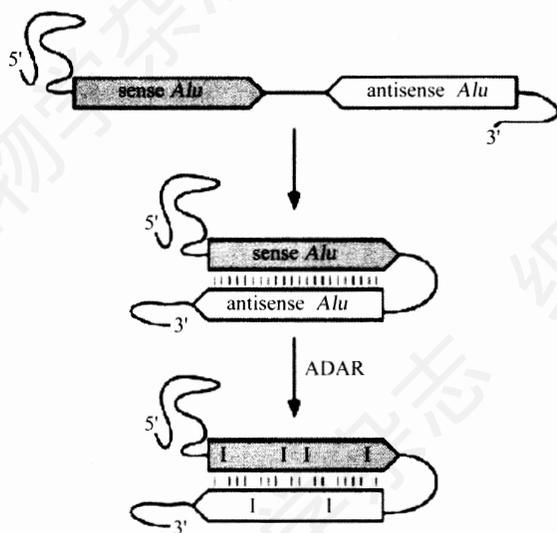


图1 Alu 序列内的 A-to-I 编辑<sup>[23]</sup>

一段含两个反向 Alu 序列的 mRNA 先发生自身碱基配对, 然后 ADAR 使这段长双链 RNA 产生 A-to-I 替代(用 I 表示)。

用可能因此破坏了 RNA 的二级结构, 而影响 RNA 的稳定性。在另一方面, 翻译和剪接时 I 被识别为 G, A-to-I 编辑可以产生编码序列的氨基酸替代或对内含子剪接位点修改而导致蛋白质翻译的提前终止或移框突变。在几个神经或精神疾病相关基因分析中, 已发现有异常编辑现象存在, 可能产生重要的生理意义<sup>[24,25]</sup>。最近发现, 在 15 种人类脑组织 RNA 的内含子序列或 3'-UTR 的重复序列中存在很多 A-to-I 编辑的靶位点。

### 2.2.2 Alu 序列作为顺式作用元件对转录的影响

Alu 序列可作为顺式作用元件与转录因子或激素结合。基因转录启动位点上游区域富含 Alu 序列, Alu 一致序列含有许多转录因子(TF)的假定结合位点。因此一些 TF 有 Alu 序列相关 TF 结合位点存在于很多基因转录启动位点上游。有几个 TF 的假定结合位点大多数都存在于 Alu 序列。进化过程中 TF 结合位点可能通过转座序列插入启动子区域。有意思的是, 很多相应的 TF 是发育早期基因的, 而发育相关基因上游是含 Alu 较少的, 相比较来说翻译和蛋白质生物合成相关基因是富含 Alu 序列的。这些 TF 可能起到抑制作用, 抑制分化过程中的增殖<sup>[26]</sup>。BC200 RNA 是灵长类神经元中一种小的非信使 RNA, 它在进化过程中吸收两个独立的 Alu 序列插入并且提供了转录的上游启动子<sup>[27]</sup>。锌指转录因子 SP1、雌激素受体  $\alpha$  和转录因子 YY1 的结合区与启动子中 Alu 序列的 CpG 岛重叠。对两个数据库中大于 5 000 个启动子的分析发现大量(大于 1 000)的人类基因被 Alu

序列的 CpG 岛控制<sup>[28]</sup>。

**2.2.3 细胞核 RNA 的二级结构** 在细胞核 RNA 上存在的 Alu 序列的多份拷贝, 能使其产生特定的二级结构。实际上, 在哺乳动物的细胞核 RNA 中, 以反向重复序列形式存在的 Alu 序列是大部分 RNA 二级结构的基础。

**2.2.4 转录后加工 - 选择性剪接** 比较人类基因组 DNA 与 cDNA 和 EST 数据库, Sorek 等<sup>[29]</sup>及其同事们确定 5% 的选择性剪接产生的新外显子是由 Alu 序列引入。对组成性外显子的同样分析显示没有任何 Alu 序列, 他们认为所有含 Alu 的外显子都是选择性剪接的结果。Alu 一致序列含有 9 个潜在 5' 剪接位点和 14 个潜在 3' 剪接位点。一致序列的 32 个潜在剪接位点中有 19 个位于反义链, 这意味着 Alu 序列插入在有义链不转录的双链有更多的机会成为外显子。这项研究证实 84% 的含 Alu 外显子来自反义 Alu 序列。对含 Alu 的外显子序列比较研究显示出潜在剪接位点被选用的几率是不同的。最常用到的位点是 275 和 279 位点作为 3' 剪接位点, 第 158 位点作为 5' 剪接位点。对选择性剪接位点周围序列突变的进一步比较研究可以确定一致序列突变对剪接位点的控制作用<sup>[30,31]</sup>。

虽然这些研究提供了 Alu 序列参与的选择性剪接的分子基础, 但是其调节可能更加复杂。内含子的辅助剪接信号在剪接点的选择中有重要作用, 但人们对这些序列了解很少。最近血管紧张素转换酶基因第 16 内含子的研究中发现外显子中的反义 Alu 序列含有几个辅助剪接增强子序列<sup>[32]</sup>。有义链起源的 Alu 序列插入含有强抑制外显化作用的剪接沉默子序列<sup>[33]</sup>。因为剪接受体样序列在外显子的剪接沉默子中大量存在, 这些信号可以抑制灵长类基因组中大量的假位点。

这种外显化需要在 Alu 序列产生功能性 5' 剪接位点。5' 剪接位点序列与小的核 RNA U1 碱基配对控制选择性剪接。在起源于 Alu 序列的 GC 内含子区, U1 snRNA 与 5' 剪接位点间碱基配对的强度控制着选择性剪接的剪接选择<sup>[31]</sup>。Galit Lev-Maor 等<sup>[34]</sup>编辑和分析了人类外显化的 Alu 序列数据, 揭示了一个机制在选择性剪接时控制 3' 剪接位点的选择。还鉴定了一些产生 Alu 外显化的突变。因为含 Alu 序列的外显子都是选择性剪接的, 这增加了转录组的剪接多样性同时也保持了原始蛋白的完整性。图 2 显示这种选择性剪接具有多样性同时仍能产生原始蛋白。外显化的 Alu 序列可以这样获得功能和与原始

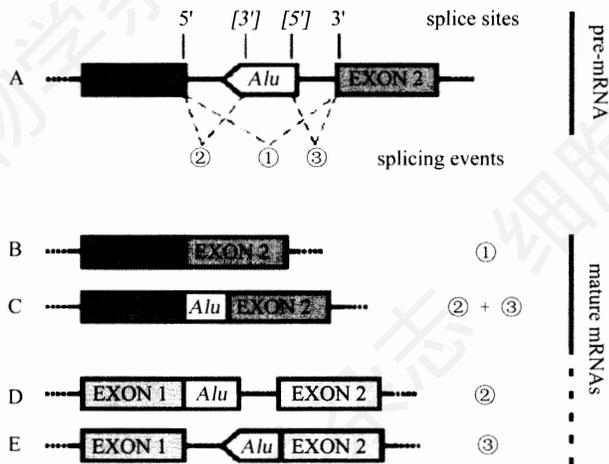


图2 Alu 序列选择性剪接<sup>[23]</sup>

(A)一段内含子中含 Alu 序列的 pre-mRNA, 其上标明了 4 个剪接位点(5'、[3']、[5']、3')及 3 种可能的剪接①、②、③。(B)①方式剪接形成的成熟 mRNA。(C)②和③方式剪接形成的成熟 mRNA。(D)②方式剪接形成的成熟 mRNA。(E)③方式剪接形成的成熟 mRNA。

蛋白不同的功能分化。当 Alu 外显子以组成性剪接出现时, 编码原始蛋白的转录本就被彻底破坏了, 这提供了遗传疾病的基础。

**2.2.5 mRNA 的稳定性** ARE(AU 富含序列)模式由重复的AUUU和末尾的A组成[(AUUU)<sub>n</sub>A], 存在于 3' 端非翻译区(UTR)。ARE 的重要生物学意义是它们使某些 mRNA 稳定性下降。有人发现在人类 mRNA 的 3' UTR 中, 有半数的长 ARE 是起源于 Alu 序列互补的多聚 T 区域, 并提出起源于 Alu 序列的 ARE 产生并保持的可能机制<sup>[35]</sup>。

### 2.3 对翻译水平的影响

Alu 序列内部含有来自 7SL RNA 的 RNA 聚合酶 III 启动子 A 和 B 框。在正常生长条件下 Alu RNA 在细胞质中处于很低水平, 但在环境压力条件下如病毒感染、暴露于环己酰亚胺或热休克时表达可暂时增加, 之后又可能恢复原表达水平。这种可控制的调节加强了 Alu RNA 于环境压力条件下在细胞代谢中扮演的特殊作用。过度表达的 Alu RNA 刺激哺乳动物细胞中共转化的报道基因翻译<sup>[36]</sup>。研究者最初认为 Alu RNA 可以作为依赖蛋白激酶的双链 RNA(PKR)的抑制子。他们认为 Alu RNA 结合 PKR 并阻止其磷酸化, 这样也会阻碍真核生物起始因子 2 $\alpha$ (eIF2 $\alpha$ )磷酸化。由于磷酸化的 PKR 具有翻译起始抑制作用, 因此 Alu RNA 可能是翻译的激活子。然而近期发现 Alu RNA 低浓度可以激活 PKR, 高浓度则可以起到抑制作用<sup>[37]</sup>。

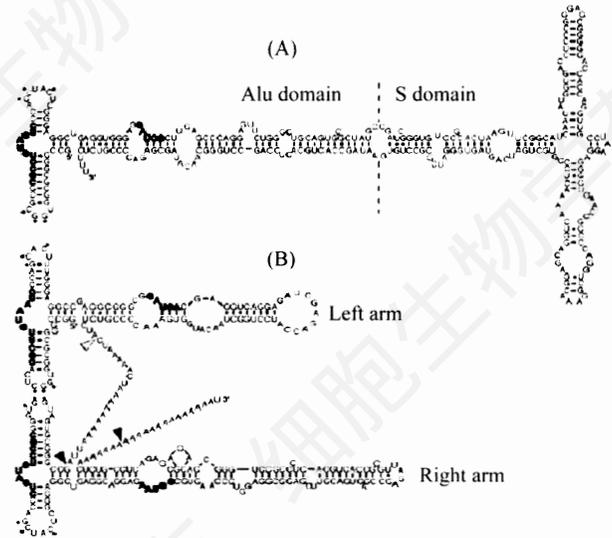


图3 Alu RNA 和 SRP RNA 的二级结构<sup>[38]</sup>

(A) SRP RNA 的二级结构分为两个功能结构区域(S 和 Alu)。S 区含有信号序列, 可以和 SRP 结合; Alu 区使转录在延伸阶段暂时延迟。粗体表示与 SRP9/14 结合的位点。(B) Alu RNA 的二级结构, 分为左右两臂, 白三角表示左臂的 3' 端, 黑三角表示右臂的 5' 和 3' 端。

典型的 Alu RNA 是由相似但不同的两臂组成, 中间由 A 富含区连接, 末端存在一段短 poly(A)。每一条臂都与 SRP RNA 的 Alu 区在序列和二级结构上相似, 并且在胞内和胞外都可以结合 SRP 相关蛋白 SRP9/14。Alu RNA 和 SRP RNA 的二级结构见图 3。SRP9/14 和 Alu RNA 结合形成 Alu RNP 复合物, 后者与裸露的 Alu RNA 比较在蛋白质翻译过程中作用不同。在一个无细胞体外翻译体系中, Alu RNA 能刺激所有报道 mRNA 翻译, 而 Alu RNP 则一般作为抑制蛋白翻译的物质。Alu RNP 和 Alu RNA 对蛋白质翻译的相反作用似乎难以理解, 但从 Alu RNA 与 SRP9/14 结合后的构象变化很容易解释这些结果。虽然 Alu RNP 和 SRP 有很高的结构相似性, 但它们通过不同的方式影响翻译: SRP 通过阻碍翻译延长起到抑制作用, Alu RNP 则主要依靠减少翻译的始动来达到抑制目的<sup>[23]</sup>。有人证明 Alu RNP 和 Alu RNA 在翻译过程中与核糖体的联系不稳定, 基于多核糖体形式和同步翻译的实验分析显示 Alu RNP 和 Alu RNA 调节翻译的起始<sup>[38]</sup>。

转录物的 UTR 中也常发现 Alu 序列, 它们被 RNA 聚合酶 II 转录作为 RNA 的一部分。在 5' 和 3' 端的 UTR 还存在调节翻译起始的位点。有研究证实了这一机制: 人类生长激素受体(hGHR) mRNA 的 5'-UTR 中的 Alu 序列可以调节这个 mRNA 的翻译<sup>[39]</sup>; ZNF177 mRNA 是一种未知功能的锌指蛋白, 其 5'-

UTR 的反向部分 Alu 序列有很强的抑制翻译作用<sup>[40]</sup>。

上述资料显示, Alu 序列在调节蛋白质翻译中至少有两种不同方式: 当被聚合酶 III 转录并装配成 Alu RNP 后它们可以作为反式调节因子, 当 5'- 和 3'-UTR 被聚合酶 II 转录后可以作为顺式调节序列<sup>[23]</sup>。

### 3 展望

截至目前, 人们提出了很多关于 Alu 家族功能的假说, 其中部分有待新的实验证实。越来越多的证据表明, Alu 家族在人类进化和疾病发生、发展中的起到重要作用, 对 Alu 家族的研究将有助于发现人类进化的奥秘, 并可能为寻找 Alu 家族变异相关性疾病的治疗方法提供方向。

#### 参考文献(References)

- [1] Makalowski W. *Science*, 2003, **300**: 1246
- [2] Khitrinskaya I Iu et al. *Mol Biol (Mosk)*, 2003, **37**: 382
- [3] Smit AF. *Curr Opin Genet Dev*, 1999, **9**: 657
- [4] Hamdi HK et al. *Biochem Biophys Res Commun*, 2002, **295**: 668
- [5] Salem AH et al. *Mol Biol Evol*, 2003, **20**: 1349
- [6] Carter AB et al. *Hum Genomics*, 2004, **1**: 167
- [7] Pauline A et al. *J Mol Biol*, 2005, **348**: 791
- [8] Jeffs AR et al. *Hum Mol Genet*, 1998, **7**: 767
- [9] Nakayama K et al. *Genome Res*, 2006, **16**: 485
- [10] Sen SK et al. *Am J Hum Genet*, 2006, **79**: 41
- [11] Xie F et al. *Blood Cells Mol Dis*, 2006, **36**: 385
- [12] Jurka J et al. *Proc Natl Acad Sci USA*, 2004, **101**: 1268
- [13] Kang MI et al. *Genomics*, 2006, **87**: 580
- [14] Feltus FA et al. *Genomics*, 2006, **87**: 572
- [15] Caiafa P et al. *J Cell Biochem*, 2005, **94**: 257
- [16] Kondo Y et al. *J Biol Chem*, 2003, **278**: 27658
- [17] Kim DD et al. *Genome Res*, 2004, **14**: 1719
- [18] Levanon EY et al. *Nat Biotechnol*, 2004, **22**: 1001
- [19] Athanasiadis A et al. *PloS Biol*, 2004, **2**: 391
- [20] Levanon K et al. *EMBO Rep*, 2005, **6**: 831
- [21] Valente L et al. *Prog Nucleic Acid Res*, 2005, **79**: 299
- [22] Kawahara Y et al. *FEBS Letters*, 2006, **580**: 2301
- [23] Häsler J et al. *Nucleic Acid Res*, 2006, **34**: 5491
- [24] Palladino MJ et al. *Cell*, 2000, **102**: 437
- [25] Gurevich I et al. *Neuron*, 2002, **34**: 349
- [26] Polak P et al. *BMC Genomics*, 2006, **7**: 133
- [27] Ludwig A et al. *J Mol Biol*, 2005, **350**: 200
- [28] Oei SL et al. *Genomics*, 2004, **83**: 873
- [29] Sorek R et al. *Genome Res*, 2002, **12**: 1060
- [30] Lev-Maor G et al. *Science*, 2003, **300**: 1288
- [31] Sorek R et al. *Mol Cell*, 2004, **14**: 221
- [32] Lei H et al. *Nucleic Acids Res*, 2005, **33**: 3897
- [33] Lei H et al. *Mol Cell Biol*, 2005, **25**: 6912
- [34] Lev-Maor G et al. *Science*, 2003, **300**: 1288
- [35] An HJ et al. *BMC Genomics*, 2004, **5**: 97
- [36] Rubin CM et al. *Nucleic Acids Res*, 2002, **30**: 3253
- [37] Williams BR. *Oncogene*, 1999, **18**: 6112
- [38] Häsler J et al. *Nucleic Acids Res*, 2006, **34**: 2374
- [39] Landry JR et al. *Genomics*, 2001, **76**: 110
- [40] Goodyer CG et al. *J Mol Endocrinol*, 2001, **27**: 357

## Roles of Alu Family in Human Genome

Wei Wang, Ya-Ping Wang\*

(Medical Department of Nanjing University, Nanjing 210093, China)

**Abstract** Alu family is the most important primate-specific short interspersed elements (SINEs). Over the last 65 million years, Alu repeats have propagated into a big family, which has about 1.2 million copies and comprises over 10% of the human genome. Alu repeats have many functions such as recombination, retrotransposition-mediated insertion and deletion, DNA methylation, A-to-I editing, regulation of transcription and translation, alternative splicing and so on. The mutation of Alu repeats is connected to the diseases and evolution.

**Key words** Alu family; human genome; gene recombination; retrotransposition; alternative splicing

Received: March 27, 2007 Accepted: June 12, 2007

\*Corresponding author. Tel: 86-25-83686495, E-mail: wangyap@nju.edu.cn