

Alu 서열과 분자생물학적 특징

- 총 설 -

박은실 · 홍경원 · 김희수*

부산대학교 생명과학부

Received November 24, 2004 / Accepted December 4, 2004

Alu sequences and molecular features. Eun-Sil Park, Kyung-Won Hong and Heui-Soo Kim. *Division of Biological Sciences, College of Natural Sciences, Pusan National University, Busan 609-735, Korea* – During the past 65 million years, *Alu* sequences have been amplified through RNA-polymerase III derived transcripts, and have reached the copy number of about 1.4 million in primate genomes. They are the largest family among mobile genetic elements in human genome and consist of ten percent of the human genome. *Alu* sequences are thought to be functionless genetically, but many researchers have proved new function and disease implication. *Alu* elements make the genome insertional mutation, *Alu*-mediated recombination events, and unexpected splicing site and change gene structures, protein sequences, splicing motifs and expression patterns. In this review, the structure and origin of *Alu*, consensus sequences of *Alu* subfamilies, evolution and distribution of *Alu*, and their related diseases were described. We also indicated new research direction of *Alu* elements in relation to evolution and disease.

Key words – *Alu*, RNA-polymerase III, genetic mobile element, primate evolution

인간 게놈은 단백질을 암호화하는 다양한 유전자들과 다양한가동성 인자들로 구성된다. 2002년 인간 게놈 프로젝트가 완성되면서 이들 가동성 인자들이 인간 게놈의 상당부분을 차지하는 것으로 밝혀졌으며, 다양한 유전자 발현에 영향을 미쳐 여러 인간 질병과 높은 관련이 있음이 밝혀졌다. 가동성 인자는 그들의 길이와 구조에 따라 HERVs (Human Endogenous RetroViruses), LINEs (Long Interspersed Nuclear Elements), SINEs (Short Interspersed Nuclear Elements) 등의 RNA서열을 기본으로 하는 역전사 인자와 위성 유전자와 같은 DNA 인자로 나뉘어진다. 이들 중 SINEs은 *Alu*와 MIR을 포함하며, LINEs의 상대적인 이름으로 진핵생물의 게놈에 짧은 길이로 퍼져 있는 인자임을 의미하고 그 전체 길이가 500 bp 이하인 것으로 가동성 인자 중 가장 큰 그룹을 형성한다[2,52,53]. SINEs들 중 인간과 영장류에서만 존재하는 *Alu*는 영장류의 진화과정 동안 게놈 내에서 증폭되어 인간 게놈의 약 10%를 차지하며, 평균 3 kb당 하나가 존재하는 것으로 알려져 있다[20,36].

SINE에서 그들의 기원이 SRP RNA-관련 그룹인 *Alu*는 이들의 서열 내에 *Alu* I 제한효소 절단 부위를 가지기 때문인데, 특히 인간을 포함한 영장류에만 존재하는 SINEs의 서열을 다른 SINEs과 달리 *Alu*라고 재명명 하게 되었다[18]. *Alu* 인자들은 일반적으로 유전자의 3' 비번역 영역과 유전자사이 (intergenic) 영역에서 더욱 많이 발견되었다[35,50]. 그래서 과거에 *Alu*의 인간 게놈에서의 삽입을 무작위적이라고 보았

던 것과 달리, 최근에는 그들의 삽입에 관한 어떠한 연관성을 찾으려는 연구가 진행되고 있다. 이들은 RNA-중합효소III에 의해 생성된 이들 전사체가 다시 그들의 게놈에 삽입되는 기작을 통해 100개 이상의 복사수로 증폭되었다. 각 서열들은 각각의 패밀리에 의해 공유되는 인지 서열을 바탕으로 크게 3개의 패밀리로 나뉘지고, 이들은 영장류의 진화과정 동안 서로 다른 시기에 삽입되었음을 보여준다. 그들은 영장류가 분기되기도 전인 6천만년 전에 삽입된 것으로 보이는 *Alu*J, 가장 최근에 삽입된 것으로 보이는 *Alu*Y, 그리고 그들의 중간 시기에 삽입된 *Alu*S로 나뉘진다 [5,22].

*Alu*는 원시 원숭이를 포함하는 모든 영장류에서 발견되는 것으로 영장류를 제외한 다른 종에는 발견되지 않는다. 또한 *Alu*의 패밀리 중 인간 게놈에만 존재하고 다른 영장류 게놈에는 결실된 인간 특이적인 *Alu*가 보고되고 있다[1,2]. 그래서 *Alu*는 영장류의 진화에 중요한 역할을 할 것이라고 제안되었고 수많은 관련성에 대한 연구가 진행되었지만 구체적인 이들 역할은 논란에 쌓여 있다. 이들이 게놈 내에 축적되기 위해서는 역전사 효소에 의한 역전사 과정을 거치게 되지만, *Alu* 서열은 역전사 효소를 암호화 하지 않는다. 그러므로 *Alu*는 어디선가 그들이 역전사에 필요한 효소를 획득해야 하는데 *Alu*와 같이 인간 게놈에서 주요 가동성 인자중의 하나인 LINEs가 *Alu*의 역전사에 필요한 효소를 제공할 것이라 추측한다[13]. 역전사에 의해 새롭게 삽입된 *Alu* 서열은 그들이 숙주 DNA에 삽입되면서 양쪽 말단에 짧은 반복 서열을 가지게 되는데, 이들 말단 반복 서열(target site duplication; TSD)은 그들이 역전사에 의해 삽입된 것인지 유전자 전환(gene conversion)으로 생겨난 것인지 확인할 수 있는 지표가 된

*Corresponding author

Tel : +82-51-510-2259, Fax : +82-51-510-2962

E-mail : khs307@pusan.ac.kr

다[19,23].

Alu 서열은 유전자가 풍부한 영역에 우세하게 축적되지만 인간 게놈에 규칙적으로 삽입되는 것은 아니다[8,28]. 대부분이 유전자사이 영역이나 유전자를 암호화 하지 않는 인트론 영역에 존재하지만 최근의 많은 연구에서 *Alu* 서열의 단편들이 성숙한 mRNA에서 나타나고, 또한 일부는 단백질을 코딩하는 영역에 존재한다는 것이 밝혀졌다. 이로써 그들의 코딩 영역과 엑손/인트론에서의 위치에 관한 분석이 이루어졌고, 이들이 인간 게놈에서 '쓸모없는 DNA'가 아닌 중요한 역할을 한다는 접근이 계속 되어 오고 있다. 단순한 반복서열로만 생각되었던 이들 *Alu* 서열에 대한 기능적 연구를 통하여 이들이 유전자 구조와 단백질 서열의 변이를 초래하고 스플라이싱의 모티프를 제공함으로써 엑손화에 의해 질병을 유발할 수도 있음이 밝혀졌다[43]. 본 논문은 *Alu* 인자가 인간 게놈에 어떠한 영향을 미치고 있으며, *Alu*의 진화과정 및 질병과의 관련성 등에 대한 이해를 고찰하고 앞으로의 연구 방향을 제시하고자 한다.

본 론

Alu 인자의 구조 형성

최근 연구에서 *Alu*는 7SL RNA 유전자로부터 생겨났다는 것이 밝혀졌다. 7SL RNA 유전자는 소포체로 분비되는 단백질의 전좌를 매개하는 SRP (Signal Recognition Particle) RNA의 일종으로 리보솜 복합체의 일부를 형성한다[41,55]. 이러한 7SL RNA의 서열은 *Alus*의 서열 내에 상당부분이 보존되어 있다. 최초의 *Alu* 단량체가 게놈 상에 화석처럼 남아 있는데, 그것을 화석 *Alu* 단량체(FAM; fossil *Alu* monomer)라고 하며, 이들은 7SL RNA로부터 중앙의 'S' 도메인이 결실되어 있다. 이러한 FAM은 인간 게놈에 아주 적은 수로 존재한다. 그 후, FAM이 유전자 중복(gene duplication)을 거쳐서 두개의 서로 다른 단량체로 만들어진다. 그렇게 형성된 단량체는 *Alu* 서열상에서 위치에 따라 FLAM (free left arm monomer)와 FRAM (free right arm monomer)로 되는데, 이들 두 단량체가 합쳐져 최초의 이량체 구조의 *Alu*가 생성된다. 이들 FAM, FRAM, 그리고 FLAM은 인간 게놈에 소수 존재하며 점차적으로 진화하여 새로운 *Alu* 인자들을 생성하였다[35,41]. *Alu* 서열은 약 280bp 정도 되며, 두 개의 단량체로 구성되어 있고, 두 단량체 사이에 약 18% 정도의 차이를 보인다[54]. 그들의 진화과정을 통한 구조적 형성에 대한 모델은 그림 1a에 보여진다. 최초의 이량체 구조의 *Alu*에서 보이는 바와 같이 단량체 양쪽 말단에 poly A tail을 가지고, 그들이 이량화 되면서 한쪽 말단에 존재했던 poly A tail이 이량체의 중간에 위치하게 되었다. 또한 이량체의 말단에도 역시 poly A tail이 20~100bp 가량 존재함을 알 수 있다[50,52]. 두 이량체는 거의 유사한 구조와 서열을 가지지만 동일하지

는 않다. 왼쪽 단량체에는 RNA 중합효소III 프로모터가 존재하지만, 오른쪽 단량체에는 많은 돌연변이가 발생하여 결실되어 있다. 현재의 이러한 이량체 구조의 *Alu*의 구조는 생성된 후 다양한 진화과정을 통하여 형성되었다. 또한, *Alu*는 역전사되어 게놈 내에 삽입될 때 그들의 삽입 부위가 복제되어 양쪽 말단에 짧은 반복서열이 형성된다[35,41].

Alu 패밀리들의 명명

*Alu*가 일단 영장류 게놈 내에 삽입된 후 계속적인 증폭을 하면서 다양한 패밀리를 가지게 되는데, 그들을 명명하는 방법과 그 근거를 살펴보자. 새롭게 발견된 *Alu*에 대한 명명법은 Batzer 등에 의해 1996년에 보고되었다. 먼저 *Alu*는 그들의 증폭 시기에 따라 크게 3개의 패밀리로 나뉘었는데, 각각 *Alu*의 이름 내에 접미어로 J, S 그리고 Y가 첨가된다. 가장 오래된 것은 'J', 중간단계 'S' 그리고 최근의 'Y'로 명명한다[21]. 이들 패밀리를 나누는 기준은 인간을 포함한 영장류에서 PCR 증폭을 하였을 때, 어떤 계통까지 증폭이 되었는지에 기초한다. 여기에 추가로 각 패밀리들의 DNA 서열 사이에 공유되는 돌연변이인 진단서열 돌연변이(diagnostic mutation)에 의해 나누어 진다. *AluS*는 *AluJ*에 비해 그들끼리 공유되는 진단서열 돌연변이 11개를 가지고(Fig. 2a), *AluY*는 *AluS*에 비해 6개의 진단 서열 돌연변이를 가진다(Fig. 2b). 또한 그들은 그 패밀리 내에 또 다른 세부적인 패밀리를 가지는데, *AluY*의 경우 동정된 순서를 바탕으로 'AluY' 뒤에 접미어를 알파벳 소문자 순서(a,b,c...)로 정의한다. 거기에 추가로 *AluY*의 컨센서스 서열에서 추가적인 진단 서열 돌연변이의 개수를 추가적으로 표기한다. 예를 들어 Ya5는 처음으로 *AluY*의 패밀리로 동정되어 *AluY*만의 진단 사이트 돌연변이를 가질 뿐만 아니라 5개의 추가적인 진단서열 돌연변이를 가지는 것이다. Ya8은 Ya와 같은 패밀리로써 5개 외에 추가적인 3개의 돌연변이를 더 가지는 것이다. 이러한 방법에 의해 *Alu* 패밀리를 기본적으로 명명하고, 인간 이외의 영장류들에서 변이체가 동정되었을 경우 영장류의 종명과 속명을 이탤릭체로 추가하여 나타낸다. 예를 들어, 침팬지에서 Ya5의 변이체는 *Pan troglodytes*로부터 Ya5Ptr이 되는 것이다[1]. 이렇게 *Alu* 서열들은 동정되어 발표된 순서로 지금까지 Yi6 패밀리까지 밝혀졌다. *Alu* Ya5, *Alu* Ya8 그리고 *Alu* Yb8이 1996년 Batzer 등에 의해 처음으로 명명되었다. 그 후, 여러 다른 패밀리들이 동정되었는데 *Alu* Yc1은 그들의 진단서열 돌연변이가 하나여서 패밀리 분류과정에서 오류도 많았고 처음엔 *AluS* 패밀리로 보고되기도 하였다[23]. 또한 *Alu* Yd2와 *Alu* Ye가 1997년, 1998년에 각각 인터넷상의 반복 서열 데이터베이스인 리베이스(Rebase)에 업데이트 되었다[24]. 추가적으로 *Alu* Yd의 *Alu* Yd3a1과 Yd8은 Smit에 의해 2001년 업데이트 되었고, 그 후 3개의 *Alu* Yb9, Yc2[48]와 *Alu* Ya5a2[47]가 보고되었다. 최근에는 *Alu* Ya1, Ya4, Yb3a1,

Yd3, Yf1, Yg6, Yh9 그리고 Yi6 가 보고되었다[25].

Alu 패밀리 분류

Alu 인자가 이렇게 다양한 패밀리로 분류되기까지 그들의 진화에 관한 다양한 이론들이 존재했다. 이러한 진화이론들 중 하나로, 최근에 대표 유전자 모델(master gene model)이 적합하다고 받아들여진다. 그림 1b는 '대표 유전자 모델'에서 Alu 패밀리의 형성과 증폭의 과정을 보여주고 이를 통해 그들이 현재의 패밀리로 분류됨을 나타낸다. 이 과정을 보면, 아주 제한된 수의 '대표 유전자'가 여러 단계로 역전사 되는 과정에서, 그 '대표 유전자'에서의 진단서열 돌연변이는 새로운 Alu 패밀리의 '대표유전자'(B, C)를 발생시켰으며, 계속적으로 이들의 자손(B', C')은 새로운 주요 진단서열에서의 돌연변이에 의해 다른 패밀리로 분류되었다는 것이다. 이 자손 Alu 서열은 계속적으로 독립적인 돌연변이를 겪었고 또한 그들의 '대표유전자' 서열로부터 나뉘어지게 되었다(Fig. 1b) [11]. 영장류의 동일좌위에 대한 계통 분류학적 연구는 Alu 서열이 단계적으로 게놈 내에 삽입되었다는 것을 나타내었고, 이것은 여러 개의 겹쳐진 Alu 패밀리 서열들의 진화적 삽입시기를 측정하도록 도와주었다[5,27]. 이들 Alu는 LINE이 매개된 역전사에 의해 영장류 게놈 내에 널리 분포하게 된 것이라 생각 된다[2]. 이런 연구 결과는 Alu 서열의 단계적인 작은 돌연변이들이 성공적으로 고정됨으로써 진화되었다는 제안을 이끌었다. 가장 오래된 Alu관련 인자는 앞에서 언급한 그들의 구조적인 모태가 되는 FAM, FRAM 그리고 FLAM 서열이고, 가장 오래된 이량체 구조의 Alu 패밀리는 AluJ 로 이들에게는 'Jo' 와 'Jb'가 있으며 이들은 진화적 삽입시기는 약 6000만년에 이른다. 중간단계의 AluS는 'Sx', 'Sp', 'Sq', 'Sg' 그리고 'Sc'로 나뉘어 지고 진화적 삽입시기가 3000~5000만년 정도로 추정 된다. 가장 최근에 삽입된 AluY의 진화적 삽입시기는 1500만년 이하이다[41]. 지금까지 동정된 여러 Alu 서열들에 대한 그들의 상대적인 삽입시기를 영장류의 진화과정과 연관지어 정리하였다(Fig. 3). AluY는 가장 최근에 삽입되었기 때문에 일부는 고정되었지만 그렇지 않은 것도 있어서 인간 게놈 내에서 형질 다형성을 보인다[9]. 이것은 Alu 서열의 생물학적 특징뿐만 아니라 인간 게놈 다양성의 연구에 대한 유전적 지표가 될 수 있다. 그러므로 최근에 삽입된 Alu 서열은 인간 게놈에 널리 퍼져있는 Alu 서열 중 가장 유용한 유전적 지표가 될 수 있다[1]. 그러나 아직도 Alu의 삽입이 왜 다른 속도를 보이는지는 알려지지 않았고, 이것은 얼마나 역전사에 대해 연구되지 않았는지를 나타낸다고도 할 수 있다[52]. Alu인자에 대해 현재까지 립베이스에 업데이트된 그들 서열을 분석하여 유전적 거리를 바탕으로 계통도를 작성하였는데, 서로 다르게 그룹화되는 현상을 알 수 있다(Fig. 4). 따라서 Alu 인자들은 진화속도를 측정하는데 마커로도 사용될 수 있다.

인간 유전체에서 Alu 인자의 분포

Alu 서열은 인간 게놈서열의 10%를 차지하고, 평균 약 3kb 마다 하나씩 발생한다. 그러나 그들은 균일하게 분포하는 것은 아니다. Alu가 특별히 많이 분포하는 영역에 대한 연구가 보고되어 있지만, 아직 그 Alu 서열이 많이 분포하는 영역에서 특이적으로 어떻게 기능을 하는지는 알려지지 않았으며 왜 유전자가 풍부한 영역에 선택적으로 삽입되는지에 대한 기작 또한 논쟁 중에 있다[3,6,31,41]. Alu 서열은 유전자가 풍부한 영역에 우선적으로 축적된다고 보고되고 있지만 그로 인해 Alu 서열이 인간 게놈에 균일하게 또는 규칙적으로 삽입되는 것은 아니다. Alu 서열은 일반적으로 유전자의 3' 비번역 영역과 유전자사이 영역에서 발견되고 성숙한 mRNA를 코딩하는 영역에는 드물게 탐지된다[8,30,36].

Alu 서열은 유전자 번역 부위를 가지지 않고 Alu의 증폭에 필요한 역전사 효소가 없으므로 LINEs로부터 그들의 증폭에 필요한 요소를 빌리는 것이다[39]. LINEs는 DNA 내부 절단효소의 기능을 가지고, 역전사 효소를 코딩한다. Alu 서열의 증폭에 결정적이라 생각되는 역전사 효소를 Alu의 증폭에 제공하는 것이다[4]. 이에 따라, 최근에 삽입된 Alu와 L1의 분포는 그들의 증폭에 관련이 있음을 바탕으로 초기에 서로 비슷하였다고 사료되었다. 그러나 L1과 달리, Alu의 분포는 이후 GC 비율이 높은 영역으로 변환되어, 시간이 지남에 따라 Alu는 초기에 게놈에 불규칙하게 삽입되었던 것이 GC 비율이 높은 영역에만 축적되게 되었다[31,40]. 이런 축적은 50만년이하의 진화적 삽입시기를 보여주는, 최근에 삽입된 Alu에 특히 잘 적용된다[3,6,31]. Alu 서열은 이렇게 우선적으로 GC 비율과 유전자 밀도, 유전자 영역 중에서도 인트론의 밀도에 영향을 많이 받지만 이러한 상관관계를 통하여 그들의 삽입과 축적에 관한 규칙성을 찾을 수 있는 것은 아니다.

Alu 서열은 인간 게놈 내에 균등하게 분포하지 않을 뿐만 아니라 이들 3개의 패밀리끼리도 그들의 분포에서 차이를 보이는데, 중간단계의 AluS (전체 게놈의 6.4%)가 가장 많고 다음으로 AluJ (~2.5%)가 차지하며 최근에 삽입되어 가장 활동적인 AluY는 가장 적게 분포하고 있다. Alu 서열은 각 염색체마다 그들의 비율이 차이가 있는데 19번 염색체에 상대적으로 가장 많이 존재하며 Y 염색체에서 가장 적게 분포한다. 그러나 특이적으로 AluY 서열은 다른 염색체와 비교했을 때 Y 염색체에 상대적으로 많이 분포하였다[16]. 또한 Alu 서열은 각 패밀리들이 균일하게 분포하지 않고 서로 상호 응집하는 경향이 있어서 여러 서열들이 같이 뭉쳐져서 존재하는 경우가 흔한데 재미있게도 예외적으로 최근에 삽입된 AluY는 더 오래된 Alu 패밀리가 존재하는 곳과 별개로 존재함이 더욱 자주 발견된다[25,26,45].

Alu 패밀리의 진화

Alu 패밀리는 인간게놈에서 그들이 제거되었다는 기작을

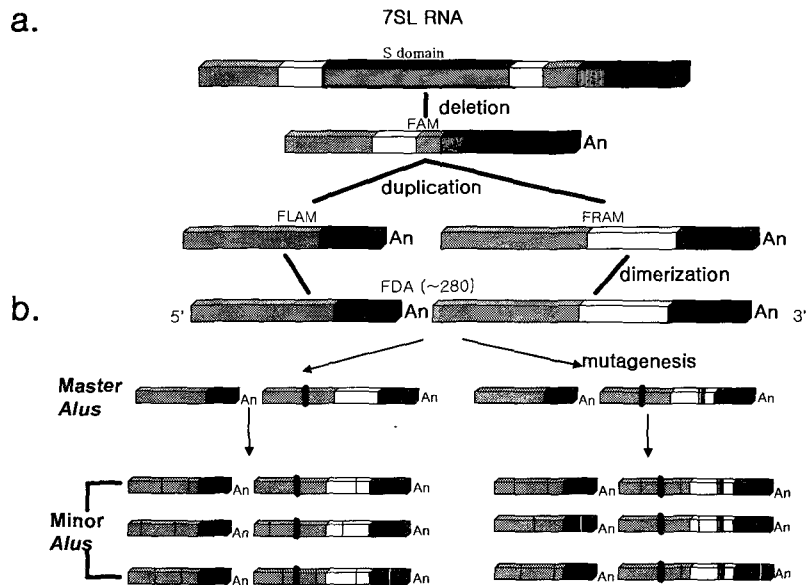


Fig. 1. Origin of Alu sequences and evolution of Alu subfamilies. a. Birth of dimeric Alu. b. Differentiation of Alu subfamilies. FAM; Free Alu Monomer, FLAM; Free Light Arm MonZomer, FRAM; Free Right Arm Monomer, FDA; First Dimeric Alu, An; poly A.

a.

AluJo	GGCGGGGCGC	GGTGGCTCAC	GCCTGTAATC	CCAGCACTTT	GGGAGGCCGA	GGCGGGAGGA	TTGCTTGAGC	70
AluJb	CA.....	70
AluSx	C...CA.C...	G	70
AluSp	C...CA.C...	G	70
AluSq	T...CA.C...	G	70
AluSc	C...CA...	G	70

Fig. 2. Continued

AluJo	CCAGGAGTTC	GAGACCAGCC	TGGGCAACAT	AGCGAGACCC	CGTCTCTACA	AAAAA-TACA	AAAA-TTAG	140
AluJb	G.T.A.	140
AluSx	T.....	C.....	G.T.A.	T.....	140
AluSp	T.G.....	AC.....	G.A.A.	T.....	140
AluSq	T.....	C.....	G.T.A.	T.....	140
AluSc	T.A...A.	T.....	C.....	G.T.A.	T.....	140

AluJo	CCGGGCGTGG	TGGCGCGCGC	CTGTAGTCCC	AGCTACTCGG	-GAGGCTGAG	GCAGGAGGAT	CGCTTGAGCC	210
AluJb	210
AluSx	A.....	A.....	A.....	210
AluSp	AT.....	A.....	A.....	A.....	210
AluSq	G.....	A.....	A.....	A.....	210
AluSc	T.....	A.....	A.....	210

AluJo	C-AGGAGTTC	GAGGCTGCAG	TGAGCTATGA	TCGGCCACT	GCACTCCA-	---GCCTG	GGCGACA-GA	280
AluJb	-G...G.	CG.....	280
AluSx	-G...GCG	T.....	CGA.....	280
AluSp	-G...GCG	T...G	CGA.....	T.....	A...A...	280
AluSq	-G...GCG	T.....	CGA.....	A...A...	280
AluSc	-G...GCG	T.....	CGA.....	CGAC-A...	280

AluJo	GCGAGACCCT	GTCTCA	296
AluJb	296
AluSx	T.C	296
AluSp	A.T.C	296
AluSq	A.T.C	296
AluSc	T.C	296

Fig. 2. Continued

b.

AluJo	GGCCGGGCGC	GGTGGCTCAC	GCCTGTAATC	CCAGCAGCTTT	GGGAGGCCGA	GGCGGGAGGA	TTGCTTGAGC	70
AluJb	CA.....	70
AluSg	CA---G	70
AluSc	CA---G	70
AluY	CA---G	70
AluYa5	CA---G	70
AluYa8	CA---G	70
AluYc2	CA---G	70
AluYd6	CA---G	70
AluYe5	CA---G	70
AluYf2	CA---G	70
AluYg6	A.....	CA---G	70
AluYi6	T.....	CA---G	70

AluJo	CCAGGAGTTC	GAGACCAGCC	TGGGCAACAT	AGCGAGACCC	CGTCTCTACA	AAAAA-TACA	AAAA-TTAG	140
AluJb	G.T.A.	140
AluSg	T.....	C.....	G.T.A.	140
AluSc	T.A.A.	T.....	C.....	G.T.A.	T.....	140
AluY	T.....	A.....	CT.....	C.G.T.A.	T.....	A.....	140
AluYa5	T.....	A.....	T.C.....	CT.A.C.G.T.A.	T.....	A.....	140
AluYa8	T.....	A.....	T.C.....	CT.A.C.G.T.A.	T.....	C.....	AAA.....	140
AluYc2	T.....	A.....	T.....	CT.A.G.T.A.	T.....	A.....	140
AluYd6	T.....	A.....	C.G.T.A.	T.....	AA.....	140
AluYe5	T.....	A.....	T.....	CT.C.G.T.A.	T.....	AA.....	140
AluYf2	T.....	A.....	T.....	CT.C.G.T.A.	T.....	AC.....	AA.....	140
AluYg6	T.....	A.....	T.....	CT.C.G.T.A.	T.....	AA.....	140
AluYi6	T.....	A.....	T.....	CT.C.G.T.A.	T.....	AA.....	140

AluJo	CCGGCGGTGG	TGGCGCGCGC	CTGTAGTCCC	AGCTACTCGG	-GAGGCTGAG	GCAGGAGGAT	GGCTTGAGCC	210
AluJb	210
AluSg	A.....	A.....	A.....	210
AluSc	T.....	A.....	A.....	210
AluY	G.....	A.....	G.G.A.	210
AluYa5	A.....	G.....	T.....	A.....	G.G.A.	210
AluYa8	A.....	G.....	T.....	T.....	A.....	G.G.A.	210
AluYc2	A.....	G.....	A.....	G.G.A.	210
AluYd6	CA.....	G.....	A.....	G.G.A.	210
AluYe5	A.....	G.....	A.....	G.G.A.	210
AluYf2	G.....	G.....	A.....	G.G.A.	210
AluYg6	A.....	A.....	A.....	G.G.A.	210
AluYi6	A.....	G.....	A.....	G.G.A.	210

AluJo	C-AGGAGTTC	GAGGCTGCAG	TGAGCTATGA	TCGCGCCACT	GCACTCCA	-----GCCTG	GGCGACA-GA	280
AluJb	-G.G.	CG.	-----	280
AluSg	-G.GCG	T.....	CGA.	-----	280
AluSc	-G.GCG	T.....	CGA.	-----	CGAC-A.	280
AluY	-G.GCG	CT.....	CGA.	-----	280
AluYa5	-G.GCG	CT.....	CGA.	C.....	-----	280
AluYa8	-G.GCG	CT.....	CGA.	C.....	-----	280
AluYc2	-G.GCG	CT.....	CGA.	-----	280
AluYd6	-G.A.GCG	CT.....	GGA.	A.....	C-----	280
AluYe5	CG.G.GCG	C.....	CGA.	-----	280
AluYf2	-G.GCG	CT.....	CGA.	-----	280
AluYg6	-G.GCG	CT.....	TCGA.	-----	280
AluYi6	-G.GCG	CT.....	CGA.	T.....	C-----	C.....	280

AluJo	GCGAGACCCT	GTCTCA	296
AluJb	296
AluSg	T.C	296
AluSc	T.C	296
AluY	T.C	296
AluYa5	T.C	296
AluYa8	T.C	296
AluYc2	T.C	296
AluYd6	A.....	T.C	296
AluYe5	T.C	296
AluYf2	T.C	296
AluYg6	A.T.C	296
AluYi6	T.C	296

Fig. 2. Diagnostic mutation of *Alu* subfamilies. a. *AluS* subfamilies, b. *AluY* subfamilies. Symbols on the left indicate subfamily names. Diagnostic sites were shaded. The dots represent same nucleotides as the *Alu Jo* consensus sequences and hyphens show alignment gaps. Mutations are shown as the correct base.

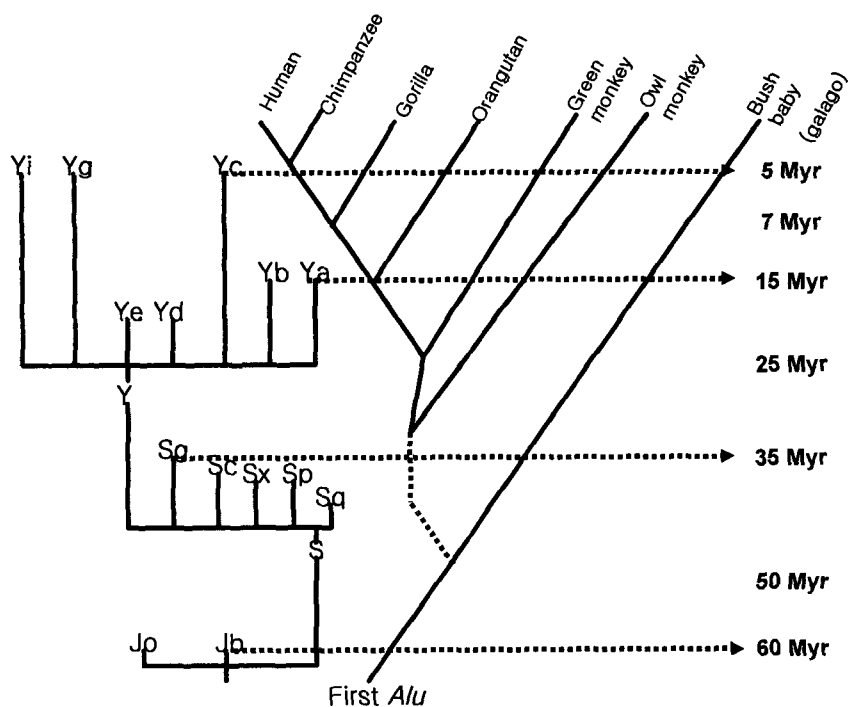


Fig. 3. Schematic description of *Alu* expansion during primate evolution, dotted arrows indicate insertion time of each *Alu* subfamilies.

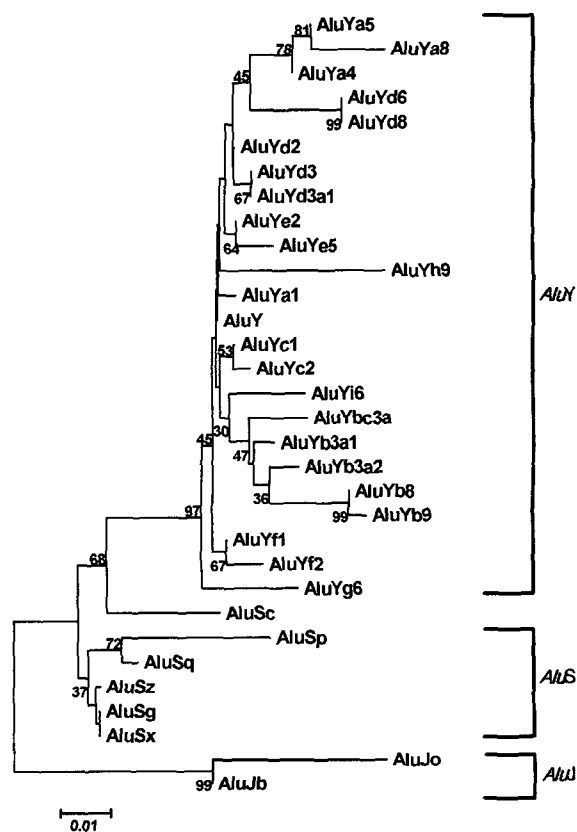


Fig. 4. Neighbor-joining (NJ) tree of all *Alu* consensus sequence. The branch lengths are proportional to the distances between the taxa. Above branch numbers indicate the bootstrap value.

발견할 수 없다는 점과 한번에 두 개의 *Alu*가 동시에 삽입될 수 없다는 점에서 분자 진화 연구에 있어 가장 어려운 요소인 호모플래시(homoplasy)에서 자유로운 성격을 지닌다. 이것은 계통분류학과 개체군 유전학에 대한 많은 의문점을 해결하는데 중요한 역할을 할 것이다[2].

Alu 패밀리의 증폭 속도는 규칙적이지 않으며, 이들은 모두 영장류의 진화과정 중 같은 시기에 삽입된 것이 아니라 단계적으로 삽입되었다. 그래서 영장류의 계놈에만 특별히 존재하는 *Alu* 서열의 진화과정은 영장류의 진화와 연관지어 많이 연구되고 있으며, 이로써 영장류의 계놈 내에 존재 유무에 따라 쉽게 그들의 삽입시기를 추측할 수 있다. 최근에 삽입된 일부의 *Alu* 서열은 침팬지와 인간이 분기한 이후인 5백만년 정도에 삽입되어 인간에만 그들의 분포가 제한되어 있고, 인간이 아닌 영장류에는 결실되어 있다[2]. 그러나 구체적인 진화적 나이를 측정하기 위해서는 그들의 컨센서스 서열에 대하여 CpG영역에서 발생하는 돌연변이수와 non-CpG영역에서 발생하는 돌연변이수로 구분하여 측정될 수 있다. 이것은 *Alu* 서열에서 CpG영역에서의 돌연변이율(0.15%/Myr)이 nonCpG영역에서의 돌연변이(0.015%/Myr) 보다 10배 가량 돌연변이율이 높고 *Alu* 인자는 그 서열 내부에 CG 서열이 많기 때문이다. 이런 방식으로 측정된 *Alu*Yi6와 *Alu*Yg6 서열은 진화적 나이가 330만년에서 460만년으로 측정되어 인간 계놈에만 존재하는 것으로 측정되었고, 이것은 PCR 실험의 결과에서 증명되었다[51].

인간 계놈에만 독특하게 나타나는 *Alu* 인자는 초기에 특정한 염기서열에서 추가적인 염기서열의 돌연변이를 공유한다는 것을 바탕으로 동정되었으며 이들은 인간계놈에서의 형질 다형성을 가진다. 이것은 침팬지에서 인간이 분기되고도 다시 인간 내에서의 개체군에 따른 분기가 되고 난 후인 약 10만년 전에 삽입된 것이라 추측될 수 있다[10,38]. 그리하여 이러한 *Alu* 패밀리들은 인간 계놈의 다양성에 대한 연구에 중요한 지표가 될 수 있다. 그림 3에서 각 *Alu* 패밀리에 대하여 영장류의 진화 과정에 따른 그들의 상대적인 삽입시기를 나타내었다. 그림에서 보이는 것처럼 *Alu* Yi, *Alu* Yg 그리고 *Alu* Yc는 인간 계놈에서만 존재하는 특이적인 염기서열이라 할 수 있다[2,51]. 이러한 인간 특이적인 *Alu*는 영장류와 인간사이의 비교연구를 위한 가치 있는 분야라 할 수 있다. 인간에만 특이적인 *Alu* 서열로 그들의 삽입시기가 5백만년 정도로 측정되었을 지라도 종종 그들의 패밀리 중 일부가 다른 원숭이의 계놈에 존재하는 경우가 있다. 이는 *Alu*의 진화에 대한 비전형적인 진화양상(nontraditional evolution)으로써 보통 2가지의 진화적인 기작으로 설명 된다. 첫번째로, 다른 *Alu* 패밀리에 속하는 인자에 의해 더 오래 전에 존재했던 *Alu* 인자 사이의 유전자 전환이 있을 수 있으며, 두번째로, 그 서열이 동일하지는 않지만 매우 가까운 *Alu* 서열간의 독립적인 삽입이 있었을 수 있다. 실제로 *Alu* Yi6와 *Alu* Yg6의

경우에서 대부분이 인간에만 특이적인 것으로 동정되었을지라도 일부의 경우 고등영장류에 존재하거나 구세계원숭이 그룹에 특이적으로 존재하는 경우가 있을 수 있다. 이는 영장류의 진화과정에서 고등영장류에 존재하는 다른 패밀리 서열과의 유전자 전환에 의해서 유발되거나, 구세계원숭이가 다른 영장류와 분기된 이후 그들의 계놈 내에 독립적으로 삽입되어 이후 구세계원숭이 모든 종에 존재하는 것이라 추정된다[49,51].

Alu 인자와 연루된 질병

최근의 연구에서 가동성 인자가 인간의 유전자의 ~4%로 단백질 코딩 영역에서 발견되고 *Alu* 인자가 이들 가동성 인자의 삽입의 약 1/3을 차지한다는 것이 보고되었다[43]. 이것은 인간의 계놈에 30,000개의 유전자가 있다고 가정한다면 이중 ~400개 유전자가 그들의 단백질 코딩 영역에 *Alu*의 단편을 포함한다고 할 수 있다. 초기에 *Alu* 인자는 그들이 숙주에 해롭지도 유용하지도 않는 기능이 없는 DNA로 생각되었으나 오랜 기간 동안 인간의 계놈 내에 적은 수의 대표 유전자로써 존재하였음은 그들이 어떤 기능을 수행할 수 있음을 제시한다[41]. 또한 *Alu* 서열은 mRNA의 단백질을 암호화하는 영역에서 드물게 나타나기 때문에 *Alu* 서열을 포함하는 특정한 엑손에 대해서는 더욱 관심이 커지게 되었다. *Alu* 서열이 인간의 계놈에 영향을 미치는 여러 기작 중 대표적인 기작으로는 삽입, *Alu*-*Alu*간 상동 재조합에 의한 결실, 복제, 전좌 등이 있다. *Alu* 인자가 유전자 발현 양상에 영향을 미치고, 단백질의 양상에도 영향력을 과시하여 *Alu* 서열의 중요성이 더욱 흥미롭다. 다양한 기작을 통하여 *Alu*와 연관된 질병이 발생하고 있으며, *Alu*에 의해 발생하는 여러 유전적 질병들에 대해 Table 1에 정리하였다.

Alu 인자에 의한 전사조절

Alu 인자의 삽입은 프로모터의 메틸화 단계를 변화시키거나, 그들 프로모터를 파괴시키거나, 스테로이드 호르몬 수용체에 대한 결합 사이트와 같이 추가적인 조절자 서열을 유도하는 등으로 일부의 *Alu* 패밀리를 포함한 유전자 전사체에 이상을 초래 할 수 있다[44]. 또, *Alu* 인자는 그들의 양방향에 많은 종결코돈을 가지고 있어 유전자의 전사를 방해한다. 따라서 *Alu* 서열의 삽입은 유전적 질병을 초래하고, 단백질의 변이를 유발한다[37]. 대표적인 예로서는 단백질의 코딩영역에 *Alu* Ya5가 삽입됨으로써 헤모필리아 B (haemophilia B)를 유발하였다[58].

Alu 인자는 정상적인 유전자의 발현에 영향을 미칠 수 있는데, 최근에 삽입된 *Alu*Y 패밀리는 기능적 retinoic acid에 대해 반응하여 잠재적으로 케라틴 K18과 같은 여러 유전자의 발현에 영향을 미친다[56]. 또한 *Alu* 인자는 인간의 콜레스테롤 에스테르트랜스퍼 단백질(CETP; colesteryl ester transfer protein)을 암호화하는 유전자의 상단부 3.4kb의 영

Table 1. *Alu* associated disease

Mechanism	Locus	Distribution	Subfamily	Disease	extra
Alu insertion	CaR	Familial	Ya4	Hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism	
	Mlvi-2	De novo (somatic?)	Ya5	Associated with leukemia	
	NF1	De novo	Ya5	Neurofibromatosis	
	PROGINS	About 50%	Ya5	Linked with ovarian carcinoma	
	IL2RG	Familial	Ya5	XSCID	
	ACE	About 50%	Ya5	Linked with protection from heart disease	
	Factor IX	A grandparent	Ya5	Hemophilia	
	EYA1	De novo	Ya5	Branchio-oto-renal syndrome	
	2 * FGFR2	De novo	Ya5 & Yb8	Apert's syndrome	
	Cholinesterase	One Japanese family	Yb8	Cholinesterase deficiency	
	APC	Familial	Yb8	Hereditary desmoid disease	
	Btk	Familial	Y	X-linked agammaglobulinaemia	
	C1 inhibitor	De novo	Y	Complement deficiency	
	BRCA2	De novo	Y	Breast cancer	
	GK	?	Y	Glycerol kinase deficiency	
Deletion	LDLR	Kindreds		Hypercholesterolemia	Germ-Line Disease
	a-globin gene cluster	Kindreds		a-thalassaemia	Germ-Line Disease
	Adenosine deaminase	One patient		ADA deficiency-SCID	Germ-Line Disease
	Apolipoprotein B gene	One patient		Hypo-betalipoproteinemia	Germ-Line Disease
	a-gal A	One patient		Fabry disease	Germ-Line Disease
	HPRT	One patient		Lesch-Nyhan syndrome	Germ-Line Disease
	Plat. Fibrinogen Receptor	Kindred		Glanzmann thrombasthenia	Germ-Line Disease
	a-Glucosidase	One patient		Glycogen storage disease type2	Germ-Line Disease
	GALNS	One patient		Mucopolysaccharidosis type IVA	Germ-Line Disease
	b-HEXA	Classic form of disease		Tay Sachs	Germ-Line Disease
	HEXB	27% of patients		Sandhoff's disease	Germ-Line Disease
	BRCA1	Somatic and kindreds		Breast and ovarian cancer	Cancer
	MLH1	Two kindreds		HNPCC	Cancer
	RB1	Common		Glioma brain tumors	Cancer
	Alu/Alu Recombination	MSH2 human DNA mismatch repair gene			Hereditary nonpolyposis colorectal cancer
FANCA				fanconi anemia	
Duplication	MLL dup (11) (q23)	Somatic		Acute myelogenous leukemia	Cancer
	Lys Hydrox.	Kindreds		Ehlers-Danlos syndrome	Germ-Line Disease
	Dystrophin	Kindred		Duchenne's muscular dystrophy	Germ-Line Disease
	C1 inhibitor gene	Kindred		Angioneurotic adema	Germ-Line Disease
Translocation	t (11;22)			Ewing's sarcoma	
	Ph translocation (9;22)			CML, Ph-positive acute leukemias	
	t (4;11)			Treatment-related acute lymphoblastic leukemia	
	t (14;18)	Somatic		Follicular lymphomas (FLs)	Cancer
	t (9;11)			De novo acute myeloid leukemia	
	t (2;8)			Burkitt lymphoma cell line	
	t (14;19)			Chronic lymphocytic leukemia (CLL)	
t (11;22)c			der (22) syndrome		

역에서 전사적 활성과 조직 특이적인 유전자 발현을 조절한다[32].

Alu에 의한 양자택일

Alu 인자가 게놈 내에서 영향력을 가지는 하나의 대표적인 예가 성숙한 mRNA내로 삽입된 Alu의 선택적 스플라이싱에 의한 엑손화이다(Fig. 5). 엑손화 과정은 스플라이싱 과정에 Alu 서열 내에 스플라이스 사이트를 인식하여 발생하는 것으로, 인간 게놈내의 양자택일된 서열의 약 5% 이상이 이러한 Alu에 의해 유도된다[33,43,54]. 이러한 경우로 인해 다양한 질병이 유발되는데, 특히 역방향에서 Alu 서열은 여러 영역에 스플라이싱의 서열을 가진다.

대표적인 Alu 서열은 9개의 잠재적인 5' 스플라이스 사이트(donor site)와 14개의 3' 스플라이스 사이트(acceptor site)를 가진다. 그러나 이들 스플라이스 사이트는 Alu 서열에 고르게 퍼져있는 것은 아니다. 오직 잠재적인 스플라이스 사이트 중 4개만이 Alu 서열의 플러스나선에 존재하고 반면에 마이너나선이 14개를 포함한다. 그래서 인트론에 있는 Alu 인자가 숙주 유전자의 전사체 방향과 반대 방향일 때 더욱 엑손으로 전환 될 수 있을 것이다[54]. 정상적인 mRNA의 스플라이싱은 이런 스플라이싱 사이트의 활성으로 인트론에 있는 Alu 서열의 점 돌연변이에 의해 비정상적인 단백질이 계속적으로 형성되어 Alpert syndrome 과 같은 임상적 질병을 일으키거나, 맥락막과 망막의 쇠약을 초래하기도 한다[28,42].

또한 β-글루쿠로니다아제(glucuronidase) 유전자에서 Alu로부터 유도된 새로운 엑손이 생기면서 원래 존재했던 엑손이 변화되는 Sly syndrome을 발생시키기도 한다[57]. Alu 인자를 매개로 하는 또 다른 예로, Factor VIII 유전자의 인트론 18내에 삽입된 Alu 인자는 19번 엑손의 불인식을 유도하여 Hemophilia A가 해로운 형태로의 전환되도록 한다[15].

Alu 인자의 상동적 재조합

멀리 떨어진 Alu 서열들간의 상동적 재조합이 발생하기도

한다. 이런 Alu-Alu재조합으로 가족성의 과콜레스테롤 혈증과 혈관부종과 같은 질병이 유발된다[7]. Alu 인자간의 전환/결실과 같은 복합적인 과정이 보고되고, 혈소판 피브로넥틴 수용체의 GPIIIa 유전자에서의 유전자내 Alu 서열에 의한 상동적 재조합으로 유전적 출혈이상인 Glamann's thrombasthenia가 유발되기도 한다[34]. 또한 Alu 인자의 재조합은 α-글로빈의 발현 조절물질을 컨트롤하여 α-thalassaemia를 유발하는 등의 인트론이나 엑손보다도 멀리 떨어진 그들의 조절 모티프에 영향을 미치는 경우도 발생한다[14]. Alu 인자의 매개로 인한 상동적 재조합은 엑손 유전자의 결실을 야기시켜 iduronate-2-sulfatase가 결핍되는 Hunter syndrome을 유발하기도 한다[46].

Alu의 삽입으로 인한 새로운 기능

Alu 인자는 다양한 질병과 연관되어 있을 뿐만 아니라, 때로는 Alu의 삽입으로 새로운 기능을 하거나 이전에 있던 기능을 변경하기도 한다. 이러한 예로 Casein Kinase2 (CK2)는 2α+2β의 단백질 사차구조의 효소로써 세린과 트레오닌을 인산화시켜 진핵세포의 생존에 필수적이다. 3개의 CK2 패밀리가 동정되었는데 74~90%의 염기서열 유사성을 보여 주고, 오직 카르복실 말단에서만 차이점을 보인다. 이들 카르복실 말단에 대한 결실로 인하여 세포에 특이적인 발현을 하는데 관련성을 찾지 못하였으나 CK2α-2의 성숙한 mRNA내에서 Alu를 포함하는 엑손이 존재하였다. 이들 C-말단 35 아미노산 서열이 CK2α-2가 핵에 국한되도록 조절하는데 필수적임이 확인되었다[17].

결론 및 전망

반복서열들에 의해 유도되는 트랜스포존은 인간게놈의 45%를 차지하고 있다. 이들 트랜스포존 중에서도 인간 게놈의 10%이상을 차지하는 서열이 Alu 패밀리이다. 많은 연구자들

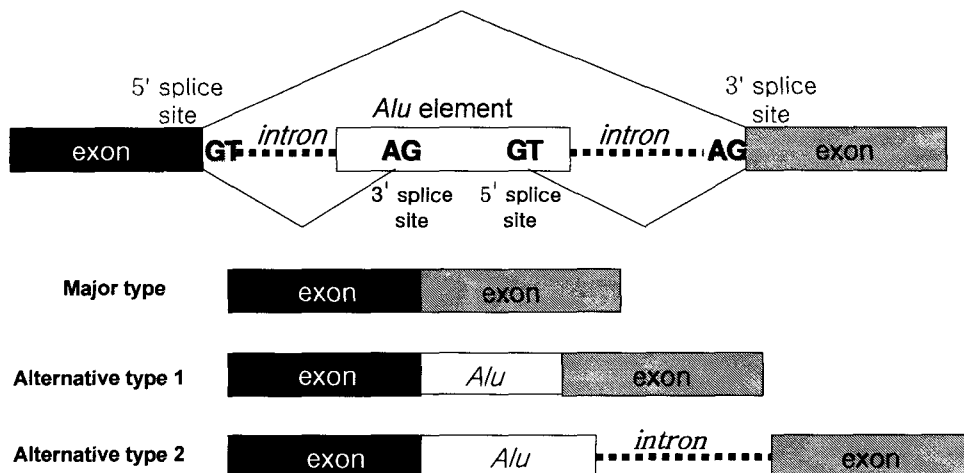


Fig. 5. Various alternative splicing patterns result from Alu insertion.

요 약

이 단백질, DNA를 포함하는 수많은 유전적 계놈 서열의 연구과정에서 인간 계놈의 널리 퍼져 있는 *Alu* 인자들을 접하게 된다. 이들은 인간을 포함한 영장류에만 특이적으로 존재하는 서열로써 인간의 진화과정을 연구하는 수많은 과학자들에게 흥미를 야기시킨다. 7SL RNA로부터 유도되어 단량체 *Alu*가 형성되고 이들의 이합화를 통하여 최근의 *Alu*의 형태를 이루게 되었다. 이후 이들 서열내의 RNA-중합효소III에 의해 생성된 전사체를 통하여 인간의 계놈 내에 100만이 넘는 복사수를 가지게 되었다. 특히 이들은 초기에 기능이 없는 것으로 여겨졌지만, 현재에는 이들 *Alu* 인자가 다양한 조절 기능과 관련이 있고 그들의 진화에 영향을 미치는 모든 계놈과 상호작용 하는 것으로 알려져 있다. 그리하여 *Alu* 인자는 유전자 발현 조절, 단백질 변이 그리고 새로운 기능을 창출해 내거나 변경해 내기도 한다. 그러나 아직도 많은 곳에서 *Alu*에 대한 논란은 그침이 없으며 그만큼 그들의 연구가치는 높아지고 있다. 아마도 그들은 더욱 많은 영역에서 유전적으로 인간을 포함하는 여러 영장류에서 그들의 계놈 서열에 영향을 미쳐 다양한 기능을 가지고 있을 것이며 차차 그들의 인간 계놈에서의 장단점이 밝혀질 것이다.

Alu 인자는 인간의 모든 염색체에 퍼져 있어 그들의 분포와 축적에 관여하는 인자를 찾는 것 또한 흥미있을 것이며, 그들의 삽입에 관해 무작위적 이면서도 유전자 빈도, GC 함량 그리고 인트론 빈도와 높은 상관관계($p < 0.0001$)를 가지는 비무작위적인 선택에 대한 해석 또한 주목할 만하다. 이렇게 *Alu* 인자는 계놈에서 전체 유전자의 3/4가 *Alu* 와 연관되어 있으며 이로써 여러 방향으로 유전적 다양성을 발생시켜 갈 것이다. 그들의 증폭은 인간 계놈에서 더욱 많은 가동성 인자의 패밀리들을 발생시킬 것이고, 이렇게 많은 *Alu* 인자들이 인간 계놈 내로 삽입되어 해로운 돌연변이를 유발하기도 한다.

Alu 인자는 다양한 기능을 가지고 인간 계놈에 영향을 미치며, *Alu*로부터 유도된 단백질은 여러 방법으로 진화에 이롭고 새로운 기능을 수행함으로써, 영장류의 종 분화에 중요한 역할을 할 것이다. 그럼에도 불구하고, 인간 계놈에서 쓸모없던 서열들의 의미를 찾은 중대한 발견은, 유전자 발현에 대한 영향이나 유전적 다양성에서의 역할이 아직도 과소평가되고 있는 실정이다. 우리는 지금도 계속하여 증폭해 가고 있을 *Alu* 인자에 대한 끊임없는 연구로 그들의 새로운 패밀리인 최근에 삽입된 *Alu* 서열을 찾아내고 그들이 다른 영장류, 특히 인간과 가장 유사한 침팬지에서조차 존재하지 않는 인간 특이적인 패밀리를 동정하여 영장류의 진화과정 중 그들의 역할에 대한 연구를 수행하고 있다. 또한 인간과 침팬지가 분기된 이후에 인간의 계놈 내에 삽입하여 유전적 다양성을 초래하는 새로운 *Alu* 패밀리를 동정하여 그들의 삽입 시기를 규명하고, 그들의 인간집단 내에서 계놈 서열에의 영향을 통한 질병과의 관련성 등의 여러 기능적 연구를 수행하고자 한다.

6500만년동안, *Alu* 서열은 RNA-중합효소 III에 의한 전사체를 통해 증폭해왔고, 영장류 계놈 내에 약 140만 복사의 수에 도달되었다. 그들은 가동성 인자 중에서 가장 큰 집단이며, 인간 계놈의 10%를 구성한다. *Alu* 서열이 유전적으로 기능이 없다고 생각되었지만, 최근 많은 연구자들이 새로운 기능 및 질병과의 관련성을 증명해왔다. 이들 *Alu* 서열은 삽입 돌연변이, *Alu*-매개 재조합, 유전자 발현에 대해 유전자 전환 그리고 스플라이싱 사이트를 유발하고, 유전자 구조, 단백질 서열, 스플라이싱 모티프와 발현 양상에 영향을 준다. 우리는 *Alu*의 구조와 기원, 이들 패밀리의 컨센서스 서열, *Alu*의 진화와 분포 그리고 그들의 기능에 대하여 요약 정리하였다. 또한 영장류의 진화과정에 있어 질병과 관련하여 *Alu* 패밀리의 새로운 연구방향을 제시하였다.

참 고 논 문

- Batzer, M. A., P. L. Deininger, U. Hellmann-Blumberg, J. Jurka, D. Labuda, C. M. Rubin, C. W. Schmid, E. Zietkiewicz and E. Zuckerkandl. 1996. Standardized nomenclature for *Alu* repeats. *J. Mol. Evol.* **42**, 3-6.
- Batzer, M. A. and P. L. Deininger. 2002. *Alu* repeats and human genomic diversity. *Nat. Rev. Genet.* **3**, 370-379.
- Bernardi, G. 2001. Misunderstandings about isochores. *Gene* **276**, 3-13.
- Boeke, J. D. 1997. LINEs and *Alus* the polyA connection. *Nature Genet.* **16**, 6-7.
- Britten, R. J. 1994. Evidence that most human *Alu* sequences were inserted in a process that ceased about 30 million years ago. *Proc. Natl. Acad. Sci. USA* **91**, 6148-6150.
- Brookfield, J. F. 2001. Selection on *Alu* sequences? *Curr. Biol.* **11**, 900-901.
- Chae, J. J., Y. B. Park, S. H. Kim, S. S. Hong, G. J. Song, K. H. Han, Y. Namkoong, H. S. Kim and C. C. Lee, 1997. Two partial deletion mutations involving the same *Alu* sequence within intron 8 of the LDL receptor gene in Korean patients with familial hypercholesterolemia. *Hum. Genet.* **99**, 155-163.
- Chen, C. A., J. Gentles, J. Jurka and S. Karlin. 2002. Genes, pseudogenes, and *Alu* sequence organization across human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. USA* **99**, 2930-2935.
- Dagan, T, R. Sorek, E. Sharon, G. Ast and D. Graur. 2004. *AluGene*: a database of *Alu* elements incorporated within protein-coding genes. *Nucleic Acids Res.* **32**, D489-492.
- Deininger, P. L. and V. K. Slagel, 1988. Recently amplified *Alu* family members share a common parental *Alu* sequence. *Mol. Cell. Biol.* **8**, 4566-4569.
- Deininger, P. L., M. A. Batzer, C. A. Hutchison and M. H. Edgell. 1992. Master genes in mammalian repetitive DNA amplification. *Trends Genet.* **8**, 307-311.

12. Deininger, P. L. and M. A. Batzer. 1999. *Alu* repeats and human disease. *Mol. Genet. Metab.* **67**, 183-193.
13. Feng, Q., J. V. Moran, H. H. Kazazian, Jr and J. D. Boeke. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905-916.
14. Flint, J., J. Rochette, C. F. Craddock, C. Dode, B. Vignes, S. W. Horsley, L. Kearney, V. J. Buckle, H. Ayyub and D. R. Higgs. 1996. Chromosomal stabilisation by a subtelomeric rearrangement involving two closely related *Alu* elements. *Hum. Mol. Genet.* **5**, 1163-1169.
15. Ganguly, A., T. Dunbar, P. Chen, L. Godmilow and T. Ganguly. 2003. Exon skipping caused by an intronic insertion of a young *Alu* Yb9 element leads to severe hemophilia A. *Hum. Genet.* **113**, 348-352.
16. Grover, D., M. Mukerji, P. Bhatnagar, K. Kannan and S. K. Brahmachari. 2004. *Alu* repeat analysis in the complete human genome: trends and variations with respect to genomic composition. *Bioinformatics* **20**, 813-817.
17. Hilgard, P., T. Huang, A. W. Wolkoff and R. J. Stockert. 2002. Translated *Alu* sequence determines nuclear localization of a novel catalytic subunit of casein kinase 2. *Am. J. Physiol. Cell Physiol.* **283**, C472-C483.
18. Houck, C. M., F. P. Rinehart and C. W. Schmid. 1979. A ubiquitous family of repeated DNA sequences in the human genome. *J. Mol. Biol.* **132**, 289-306.
19. Hutchinson, G. B., S. E. Andrew, H. McDonald, Y. P. Goldberg, R. Graham, J. M. Rommens and M. R. Hayden. 1993. An *Alu* element retroposition in two families with Huntington disease defines a new active *Alu* subfamily. *Nucleic Acids Res.* **21**, 3379-3383.
20. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. 2001. An assembly and annotation of the first draft sequence of the entire human genome that includes a comprehensive analysis of repeated DNA sequences. *Nature* **409**, 860-921.
21. Jurka, J. and T. Smith. 1988. A fundamental division in the *Alu* family of repeated sequences. *Proc. Natl. Acad. Sci. U S A.* **85**, 4775-4778.
22. Jurka, J. 1993. A new subfamily of recently retroposed *Alu* repeats. *Nucleic Acids Res.* **21**, 2252.
23. Jurka, J. and P. Klonowski. 1996. Integration of retroposable elements in mammals: selection of target sites. *J. Mol. Evol.* **43**, 685-689.
24. Jurka, J. 2000. Repbase update: A database and an electronic journal of repetitive elements, *Trends Genet.* **16**, 418-420.
25. Jurka, J., M. Krnjajic, V. V. Kapitonov, J. E. Stenger and O. Kohany. 2002. Active *Alu* elements are passed primarily through paternal germlines. *Theor. Popul. Biol.* **61**, 519-530.
26. Jurka, J., O. Kohany, A. Pavlicek, V. V. Kapitonov and M. V. Jurka. 2004. Duplication, coclustering, and selection of human *Alu* retrotransposons. *Proc. Natl. Acad. Sci. U S A.* **101**, 268-272.
27. Kapitonov, V. and J. Jurka. 1996. The age of *Alu* subfamilies. *J. Mol. Evol.* **42**, 59-65.
28. Knebelmann, B., L. Forestier, L. Drouot, S. Quinones, C. Chuet, F. Benessy, J. Saus and C. Antignac. 1995. Splice-mediated insertion of an *Alu* sequence in the COL4A3 mRNA causing autosomal recessive Alport syndrome. *Hum. Mol. Genet.* **4**, 675-679.
29. Kolomietz, E., M. S. Meyn, A. Pandita and J. A. Squire. 2002. The role of *Alu* repeat clusters as mediators of recurrent chromosomal aberrations in tumors. *Genes Chromosomes Cancer* **35**, 97-112.
30. Korenberg, J. R. and M. C. Rykowski. 1988. Human genome organization: *Alu*, lines, and the molecular structure of metaphase chromosome bands. *Cell* **53**, 391-400.
31. Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
32. Le Goff, W., M. Guerin, M. J. Chapman and J. Thillet. 2003. A CYP7A promoter binding factor site and *Alu* repeat in the distal promoter region are implicated in regulation of human CETP gene expression. *J. Lipid Res.* **44**, 902-910.
33. Lev-Maor, G., R. Sorek, N. Shomron and G. Ast. 2003. The birth of an alternatively spliced exon: 3' splice-site selection in *Alu* exons. *Science* **300**, 1288-1291.
34. Li, L. and P. F. Bray. 1993. Homologous recombination among three intragene *Alu* sequences causes an inversion-deletion resulting in the hereditary bleeding disorder Glanzmann thrombasthenia. *Am. J. Hum. Genet.* **53**, 140-149.
35. Li, T. H. and C. W. Schmid. 2004. *Alu*'s dimeric consensus sequence destabilizes its transcripts. *Gene* **324**, 191-200.
36. Li, W. H., Z. Gu, H. Wang and A. Nekrutenko. 2001. Evolutionary analyses of the human genome. *Nature* **409**, 847-849.
37. Makalowski, W., G. A. Mitchell and D. Lauda. 1994. *Alu* sequences in the coding regions of mRNA: A source of protein variability. *Trends Genet.* **10**, 188-193.
38. Matera, A. G., U. Hellmann and C. W. Schmid. 1990. A transpositionally and transcriptionally competent *Alu* subfamily. *Mol. Cell. Biol.* **10**, 5424-5432.
39. Mathias, S. L., A. F. Scott, H. H. Jr. Kazazian, J. D. Boeke and A. Gabriel. 1991. Reverse transcriptase encoded by a human transposable element. *Science* **254**, 1808-1810.
40. Mazzarella, R. and D. Schlessinger. 1997. Duplication and distribution of repetitive elements and non-unique regions in the human genome. *Gene* **205**, 29-38.
41. Mighell, A. J., A. F. Markham and P. A. Robinson. 1997. *Alu* sequences. *FEBS Lett.* **417**, 1-5.
42. Mitchell, G. A., D. Labuda, G. Fontaine, J. M. Saudubray, J. P. Bonnefont, S. Lyonnet, L. C. Brody, G. Steel, C. Obie and D. Valle. 1991. Splice-mediated insertion of an *Alu* sequence inactivates ornithine delta-aminotransferase: a role for *Alu* elements in human mutation. *Proc. Natl. Acad. Sci. USA* **88**, 815-819.

43. Nekrutenko, A. and W. H. Li. 2001. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* **17**, 619-621.
44. Norris, J., D. Fan, C. Aleman, J. R. Marks, P. A. Futreal, R. W. Wiseman, J. D. Iglehart, P. L. Deininger and D. P. McDonnell. 1995. Identification of a new subclass of *Alu* DNA repeats which can function as estrogen receptordependent transcriptional enhancers. *J. Biol. Chem.* **270**, 22777-22782.
45. Pavlicek, A., K. Jabbari, J. Paces, V. Paces, J. V. Hejnar and G. Bernardi. 2001. Similar integration but different stability of *Alus* and *LINEs* in the human genome. *Gene* **276**, 39-45.
46. Ricci, V., S. Regis, M. Di Duca and M. Filocamo. 2003. An *Alu* mediated rearrangement as cause of exon skipping in Hunter disease. *Hum. Genet.* **112**, 419-425.
47. Roy, A. M., M. L. Carroll, S. V. Nguyen, A. H. Salem, M. Oldridge, A. O. M. Wilkie, M. A. Batzer and P. L. Deininger. 2000. Potential gene conversion and source genes for recently integrated *Alu* elements. *Genome Res.* **10**, 1485-1495.
48. Roy-Engel, A. M., M. L. Carrol, E. Vogel, R. K. Garber, S. V. Nguyen, A. H. Salem, M. A. Batzer and P. L. Deininger. 2001. *Alu* insertion polymorphisms for the study of human genomic diversit. *Genetics* **159**, 279-290.
49. Roy-Engel, A. M., M. L. Carroll, M. El-Sawy, A. H. Salem, R. K. Garber, S. V. Nguyen, P. L. Deininger and M. A. Batzer. 2002. Non-traditional *Alu* evolution and primate genomic diversity. *J. Mol. Biol.* **316**, 1033-1040.
50. Roy-Engel, A. M., A. H. Salem, O. O. Oyeniran , P. L. Deininger, D. J. Hedges, G. E. Kilroy, M. A. Batzer and P. L. Deininger. 2002. Active *Alu* element "A-tails": size does matter. *Genome Res.* **12**, 1333-1344.
51. Salem, A. H., G. E. Kilroy, W. S. Watkins , L. B. Jorde and M. A. Batzer. 2003. Recently integrated *Alu* elements and human genomic diversity. *Mol. Biol. Evol.* **20**, 1349-1361.
52. Schmid, C. W. 1996. *Alu*: structure, origin, evolution, significance and function of one- tenth of human DNA. *Prog. Nucleic Acid Res. Mol. Biol.* **53**, 283-319.
53. Singer, M. F. 1982. *SINEs* and *LINEs*: highly repeated short and long interspersed sequences in mammalian genomes. *Cell* **28**, 433-434.
54. Sorek, R., G. Ast and D. Graur. 2002. *Alu*-containing exons are alternatively spliced. *Genome Res.* **12**, 1060-1067.
55. Ullu, E. and C. Tschudi. 1984. *Alu* sequences are processed 7SL RNA genes. *Nature* **312**, 171-172.
56. Vansant, G. and W. F. Reynolds. 1995. The consensus sequence of a major *Alu* subfamily contains a functional retinoic acid response element. *Proc. Natl. Acad. Sci. USA* **92**, 8229-8233.
57. Vervoort, R., R. Gitzelmann, W. Lissens and I. Liebaers. 1998. A mutation (IVS8+0.6kbpdelTC) creating a new donor splice site activates a single nucleotide polymorphisms within *Alus*, as well as a cryptic exon in an *Alu*-element in intron 8 of the human b-glucuronidase gene. *Hum. Genet.* **103**, 686-693.
58. Vidaud, D., M. Vidaud, B. R. Bahnak, V. Siguret, S. Sanchez, Y. Laurian, D. Meyer, M. Goosens and J. M. Lavergne. 1993. Haemophilia B due to a de novo insertion of a human-specific *Alu* subfamily member within the coding region of the factor IX gene. *Eur. J. Hum. Genet.* **1**, 30-36.