

## 인간 게놈의 Copy Number Variation과 유전자 질환

오정환 · Ichiro Nishimura\*

경희대학교 치의학전문대학원 구강악안면외과학교실

\*미국 UCLA 치과대학 보철학교실

### Abstract

#### UNDERSTANDING OF EPIGENETICS AND DNA METHYLATION

Jung-Hwan Oh, Ichiro Nishimura\*

*Department of Oral and Maxillofacial Surgery, School of Dentistry, Kyung-Hee University, Seoul, Korea*

*\*The Jane and Jerry Weintraub Center for Reconstructive Biotechnology,*

*Department of Advanced Prosthodontics, School of Dentistry, UCLA, USA*

Genetic variation in the human genome occurs on various levels; from the single nucleotide polymorphism to large, microscopically visible chromosome anomalies. It can be present in many forms, including variable number of tandem repeat (VNTRs; e.g., mini- and microsatellites), presence/absence of transposable elements (e.g., Alu elements), single nucleotide polymorphisms, and structural alterations (e.g., copy number variation, segmental duplication, inversion, translocation).

Until recently SNPs were thought to be the main source of genetic and phenotypic human variation. However, the use of methods such as array comparative genomic hybridization (array CGH) and fluorescence in situ hybridization (FISH) have revealed the presence of copy number variations (CNVs) ranging from kilobases (kb) to megabases (Mb) in the human genome. There is great interest in the possibility that CNVs play a role in the etiology of common disease such as HIV-1/AIDS, diabetes, autoimmune disease, heart disease and cancer. The discovery of widespread copy number variation in human provides insights into genetic variability among populations and provides a foundation for studies of the contribution of CNVs to evolution and disease.

**Key words:** Genetic variation, Single nucleotide polymorphism (SNP), Copy number variation (CNV)

### I. 서론

2001년도<sup>1)</sup> 인간게놈 프로젝트에 의해 인간 DNA의 30억 개 염기서열이 밝혀진 후, 많은 연구자들은 유전 정보가 어떻게 단백질, 조직, 인체의 구조와 기능에 영향을 미치는지를 이해하기 위하여 유전자의 구성에 관심을 갖기 시작하였고, 개인 유전자의 단일 염기 수준에서 변이형태인 단일염기다형성 (single nucleotide polymorphism, SNP)를 발견하였다<sup>2-5)</sup>.

단일염기다형성은 특정 염기 쌍 부위에서 하나의 염기가 다른 것으로 대체된 두 유전자 서열의 단일염기 교환을 의미하는 용어이다. 예를 들면, cytosine이 thymine으로, adenine이 guanine으로 치환된 유전변이이며, 지금까지 인간 게놈에 약 1000 만개의 단일염기다형성이 존재하는 것으로 알려져 있다.

단일염기다형성은 인간의 유전형과 표현형의 다양성을 설명하는 가장 주된 유전적 변이로 알려졌으나, 최근에는 특정 유전좌위 연구보다 발전된 유전체 전체적인 접근법의 개

발로 kilobase에서 megabase 단위의 큰 copy number variation이 발견되면서 인간 게놈의 유전적 변이가 단순한 단일염기다형성 뿐만 아니라 구조적 변형까지 다양한 형태와 다양한 크기로 존재하는 것으로 밝혀졌다<sup>6-9)</sup>.

인간 게놈의 다양성은 단일염기다형성에서부터 현미경으로 관찰할 수 있는 염색체수준까지 여러 가지 형태로 이루어져 있다. 여기에는 다양한 크기의 variable number of tandem repeat (VNTRs; e.g., mini- and microsatellites), Alu elements, 단일염기다형성, 그리고 구조적 변이 (예를 들면, copy number variations, 절편복제, 역위, 전좌) 등이 포함된다 (Table 1)<sup>10-12)</sup>.

Simple sequence length polymorphism (SSLP)는 VNTRs, microsatellites, 그리고 tandem repeats를 포함한, 개인마다 다양한 수로 존재하는 tandem repeats를 구성하는 특정 DNA부위에 사용할 수 있는 용어이다. 인간은 65%정도만 유일한 DNA로 구성되어 있고, 나머지 부분은

한 가지 또는 다른 형태의 반복적인 서열로 이루어져 있다. 수백 또는 수천의 copies 형태의 서열을 “moderately repetitive sequence”라고 한다. 인간 DNA의 약 25% 정도가 이 카테고리에 속한다. Moderately repetitive non-coding DNA의 대부분은 “Long INterpersed Elements (LINEs)”를 형성한다. 이것은 retrovirus-like ancestors에서 유래된 것으로 생각된다. 나머지 10%정도의 인간 DNA는 10만에서 100만의 copies 형태의 서열로 존재한다. 대부분의 이 highly repetitive DNA는 “Short INterpersed Elements (SINEs)”. 이러한 서열들은 대부분 비기능성이며 대표적인 SINE로는 300 bp 크기의 Alu element가 있다. 인간 DNA에는 약 300,000 ~ 500,000 개의 Alu element가 산재되어 있다. 이것은 인간 유전정보의 약 6-8% 정도를 구성한다.

인간 게놈 전체에 산재되어 있는 LINEs and SINEs와 달리 long clusters of tandem repeats 형태로 발견되는

**Table 1.** Classification of genetic variation in the human genome (by Feuk et. al. 2006)<sup>12)</sup>

Variation type	Definition	Frequency in the human genome
SNP	Single base pair variation found in >1% of Chromosome	~10 million SNPs in the human population
Insertion/deletion Variation (InDel)	Deletion or insertion of a segment of DNA includes small polymorphic changes and large chromosomal aberrations. InDels >1 kb in size are often also called CNVs	~1 million InDels polymorphisms >1 bp in the human genome
Microsatellite (e.g. CA repeats)	Sequence containing variable numbers of 1-6 bp repeats totaling <200 bp in length	>1 million microsatellites in the human genome, accounting for ~3% of the sequence.
Minisatellite and VNTRs	Polymorphic sequence containing 20-50 copies of 6-100 bp repeats	~150,000 minisatellites, of which 20% are polymorphic
Multisite variant (MSV)	Single nucleotide variant with complex characteristics due to CNV or gene conversion	The number of MSVs is currently unknown
Intermediate-sized structural variant	Gain or loss of a DNA sequence >8 kb in size also includes inversion breakpoints	297 ISVs were identified using a fosmid library from a single genome
Copy number polymorphism; (large-scale CNV)	Copy number change >1 kb. If the frequency is >1%, it is called a CNP. LCVs are CNVs ~50 kb in size or greater	The frequency of CNVs in the human genome is unknown. Estimates of larger CNVs (>50 kb)
Inversion	Rearrangement causing a segment of DNA to be present in reverse orientation	Estimates of microscopically detectable inversion frequencies are 0.12-0.7% (pericentric) and 0.1-0.5% (paracentric)
Translocation	Rearrangement in which a DNA fragment is attached to different chromosome	1/500 is heterozygous for a reciprocal translocation and 1/1000 for Robertsonian translocations
Unbalanced Rearrangements	Rearrangements which lead to a net gain or loss of DNA are referred to as unbalanced	Unbalanced rearrangements occur in ~1/1500 live births

매우 반복적인 DNA가 있는데, 이를 “satellite DNA”라고 하고, tandem은 바로 옆에 존재하는 DNA와 사이에 틈이 없이 반복되는 서열을 의미한다. Satellite DNA는 영구적으로 단단하게 “heterochromatin”이라는 코일형태를 이루고 있다.

Satellite보다 적은 수이면서 짧은 tandem repeats를 “mini-satellites” 또는 “Variable Number of Tandem Repeats (VNTRs)”라고 한다. 일반적으로 한 개의 VNTR에는 5~50개의 tandem repeats가 존재한다.

## II. Copy Number Variation의 개념과 의미

인간 게놈은 부모 각각으로부터 받은 23쌍의 염색체 안에 DNA구조의 약 60억 개의 뉴클레오타이드, 3만 개의 유전자로 구성되어 있다. 인간 게놈에서 유전자는 일반적으로 두 개의 copies로 이루어져 있다. Copy number는 단일 숙주 세포 안에서 발견되는 하나의 유전자 또는 플라스미드의 copy 갯수를 의미한다. 고등생물일수록 각 염색체에 두 개의 copy number를 갖는다. 동일한 염색체의 각 쌍은 동일한 선형배열로 구성된 같은 유전자의 copy를 갖는다. 실제로는 간혹 같은 유전자가 서로 다른 대립형질(allele)을 가지면, 동일한 염색체이지만 완벽하게 일치하지 않다. 두 개의 동일한 대립형질을 가진 유전자를 “동형접합자(homozygous)”, 서로 다른 대립형질을 가진 유전자를 “이형접합자(heterozygous)”라고 한다. 염색체 내에 두 개의 “homologous copies”를 가진 세포나 유기체를 “이배체(diploid)”라고 하고, 각 염색체에 오직 한 개의 copy만 있는 경우를 “반수체(haploid)”라고 한다.

인간의 유전자는 일반적으로 두 개의 copies를 갖지만 최근 연구에 의하면 유전자에 따라 하나, 둘, 또는 세 개의 copies를 갖는 경우가 보고되고, 간혹 유전자 전체가 결손 경우도 발견되는데, 이를 “copy number variation (CNV)”라고 한다. CNV는 복제(duplication)나 결손(deletion)에 의하여 DNA 절편의 1 Kb(kilobase)이상의 copy number 변화를 의미한다(Fig. 1)<sup>6-9,12,19)</sup>. CNV는 tandem duplication과 같이 단순한 구조일 수도 있고, 게놈의 다양한 부위에서 상동서열의 복합적인 복제나 상실에 의해서도 발생할 수도 있다. CNV라는 용어는 전이인자에 의해 반복되는 삽입과 결손을 제외한 copy number polymorphism, large-scale copy number variants, inter-mediated-sized structural variations 등을 포괄하는 개념이다. CNV는 tandem duplication과 같은 단순구조일 수도 있고 게놈의 여러 부위에서 homologous sequences

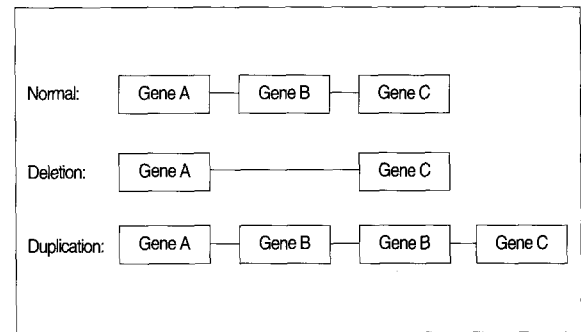


Fig. 1. Copy number variation : deletion and duplication.

Table 2. Summary of genome-wide studies of structural variation (by Eichler EE, 2006)<sup>18)</sup>

Study	Assay	Variant types	Size range	Number of Variants	Median size (kb)
Iafrate	BAC-based Array CGH	Deletion/insertion	>50 kb	255	~150
Sebat	ROMA	Deletion/insertion	>100 kb	76	222
Sharp	BAC-based Array CGH	Deletion/insertion	>50 kb	160	~150
Tuzun	Paired-end sequence	Deletion/insertion	>6-1,900 kb	297	15.2
Conrad	Mendelian errors	Deletion	300 bp-1,200 kb	586	8.5-10
McCarroll	Clustered genotype Errors and mendelian errors	Deletion	1 kb-745 kb	541	7.0
Hinds	High-density Oligonucleotide Hybridization	Deletion	70 bp-7 kb	215	0.75

의 복합적인 획득 또는 상실에 의해 생길 수도 있다<sup>19)</sup>. "Copy number polymorphism"이라는 용어는 모든 CNV에 적용될 수 있지만 인구 1% 이상에서 발생하는 경우에 사용된다<sup>7,12)</sup>.

최근 array-comparative genome hybridization (array-CGH)법 등 새로운 분석법의 개발로 특정 유전좌위 연구에서 유전체 전체 분석법으로 변화되면서 삽입 또는 결손과 연관된 CNV 연구들이 많이 보고되었는데 (Table 2), Iafrate 등<sup>6)</sup>은 39명의 건강한 사람을 대상으로 array-based comparative genomic hybridization 방법을 이용하여 유전적 불균형을 포함하고 있는 255개의 genomic clone을 발견하고, 한 개체 이상에서 발생하는 102개의 large-scale copy number variations (LCVs)를 보고하였다. 102개의 LCVs 중 26개가 기존의 절편복제 부위와 중첩된다고 하였다. 대부분의 LCVs의 길이는 약 150 kb에서 425 kb였으며, 염색체 1p13.3 부위의 아밀라제 알파 1a와 알파 2a locus를 포함하고 있었다. 255개의 다형성 클론 중 142개가 코딩영역이 연관되어 있었다. 67개의 클론은 하나 또는 그 이상 전체 유전자를 포함하고 있었다. 이것은 LCVs가 intergenic 또는 intronic region에만 국한되는 것이 아니라는 것을 의미하며, 14개의 LCVs는 인간 유전증후군 또는 암과 연관된 부위에서 발견되었다.

Scherer 등<sup>13)</sup>은 CNV를 1000 개 또는 50만개 이상의 뉴클레오타이드에 영향을 미치는 초현미경적 염색체 변이라고 하였다. 이들은 4개의 다른 지리적 조상을 가진 270명의 건강한 개체에서 1447개의 CNV를 보고하였다. 이러한 상대적으로 많은 CNVs 숫자는 3억 6000만 뉴클레오타이드, 인간 게놈의 약 12%에 해당되는 수치이며, 이 연구는 기존의 단일염기다형성 연구에서 두 인간은 유전적으로 99.7%가 동일하다는 생각을 바꾸어 놓게 되었다.

Sebat 등<sup>7)</sup>은 20명의 개놈에서 추출한 BgIII-절편을 증폭하고 올리고 뉴클레오타이드 microarray를 이용하여 평균 크기 465 kb의 76개 CNV를 발견했다. Tuzun 등<sup>14)</sup>은 인간 포스미드 DNA genomic library에서 크기가 8 kb에서 40 kb정도되는 139개의 삽입, 102개의 결손, 그리고 56개 역위의 미세크기의 구조적 변이를 보고하였는데, 기존의 Sebat와 Iafrate 등이 발표한 large-scale variation과 결과를 비교하여 297개의 부위 중 252개의 부위가 기존에 발견되지 않은 새로운 부위라고 하고, 기존의 larger copy number polymorphisms과 구분하기 위하여 삽입, 결손, 역위, intermediated-sized structural variants (ISVs)라는 용어를 사용하였다.

Haploinsufficiency는 dosage-sensitive gene의 한 copy가 결손되어 발육성 지연이나 결손을 야기하는 경우를 의미하는 반면에, Haplosufficiency는 발육성 지연이나 결손을 야기하지 않는 유전체 결손을 나타내는 용어이다.

최근 결손과 연관된 대표적인 CNV 연구를 살펴보면, Conrad 등<sup>14)</sup>은 The international HapMap Project의 SNP genotype data와 comparative genome hybridization법을 이용하여 멘델 유전법칙에 불일치한 대상으로 300 bp-12 Mb 크기의 586개 결손 변이를 보고하였는데, 결손 CNV는 상대적으로 유전자가 드문 부위에서 발견되었다고 하였다. McCarroll 등<sup>15)</sup>도 The international HapMap Project의 단일염기다형성 자료를 이용하여 크기가 1 kb-745 kb 정도되는 541개의 결손 변이를 보고하였다. 541개 중 120개가 두 개의 copies 모두가 결손된 상동 결손(homologous deletion)이었으며, 10개 정도의 상동결손 부위는 스테로이드 대사, 후각, 그리고 약물대사와 관련된 유전자의 엑손 부위였다고 보고하였다.

Hinds 등<sup>16)</sup>은 24명의 연관성이 없는 개체의 DNA와 고밀도 올리고 뉴클레오타이드 분석법을 이용한 연구를 통해 크기 70 bp-10 kb 크기의 215개의 결손변이를 발표하였다. 그 중 41개의 결손은 24명에서 10%이상의 대립인자 빈도를 보였다고 하였다. 대부분의 결손이 연관 불균형(linkage disequilibrium) 상태로 결손변이와 단일염기다형성이 유사한 진화과정을 가지고 있다고 하였다. 이것은 대부분의 결손변이가 근처 단일염기다형성과 연관 불균형 관계였다는 McCarroll 등의 보고(참고문헌 요망)와 유사한 결과이다.

현재까지 The Database of Genomic Variants (<http://projects.tcag.ca/variation>)에 143 Mb genome sequence의 1237개의 CNV가 등록되어 있는데<sup>17)</sup>, 등록된 모든 CNV의 평균 크기가 ~118 kb이지만 중간 크기는 ~18 kb밖에 되지 않는 것은 최근 결손과 연관된 CNV연구 결과 때문인 것으로 생각된다. Redon 등<sup>19)</sup>은 복제에 비해 훨씬 낮은 비율의 결손을 관찰하였는데, 결손은 복제에 비해 유전자에서 떨어져 있는 경향을 가지고 있다는 것을 의미한다.

### III. Copy Number Variation 형성 기전

흔히 CNVs는 큰 크기의 homologous repeats 또는 절편복제(segmental duplication)와 인접되어 있거나 이를 포함하고 있는데 절편복제 (또는 low-copy repeats)는 반수체 유전체당 두 개 또는 그 이상의 copies를 가지고, 서로 다른 copies의 염기서열이 90%이상 동일하며, 크기가 1 kb 이상인 DNA 절편으로 정의된다<sup>12)</sup>.

절편 복제는 DNA 절편의 tandem repetition, duplicative transposition-like process, 유전체 절편의 복사과정을 통해 발생할 수 있다. 절편복제와 연관된 CNVs는 non-allelic homologous recombination (NAHR)을 통한 구조적 염색체 재배열(chromosomal rearrangements)에 민감

하다. NAHR은 동일한 염색체에서 절편복제가 절편복제 부위의 copy number 변화를 용이하게 해준다. 정상인에서의 CNVs 형성 이외에도, NAHR은 게놈의 불안정성, 조기 발현, 또는 고도의 장애의 원인이 되는 큰 구조적 다형성과 염색체 재배열을 초래한다.

Tuzun 등<sup>14)</sup>은 구조적 변이의 약 55% (163 of 297)가, Redon 등<sup>19)</sup>은 1,447 CNVs의 24%가 게놈상의 절편복제와 관련되어 있다고 하였다. 모든 CNVs가 절편복제와 관련된 것은 아니며, 일부는 non-homology-based mutational mechanism에 의해 발생할 수도 있다<sup>21)</sup>. 어떤 CNVs는 left-handed DNA 그리고 십자형(cruciforms)를 포함한 right-handed  $\beta$ -helical duplex와 구조가 다른 DNA부위인 non- $\beta$  DNA 구조와 연관성이 있다. 이러한 DNA구조는 염색체 재배열을 촉진시키는 것으로 알려져 있다. CNV의 크기와 변이기전 사이에 연관성이 있는데, 보다 큰 CNV는 절편복제와 연관되고 작은 크기의 CNV는 non-homology-based mutational mechanism이 우세한 것으로 알려져 있다.

#### IV. CNV와 연관된 질병 연구

큰 크기의 복제와 결손은 dosage-sensitive gene과 관련된 copy numbers의 변화를 야기하여 특정 유전질환을 유발할 수 있는 것으로 알려져 있다 (Table 3). 또한 CNV는 유전자, 전사 또는 번역 수준에서의 변이를 가져오고 결국 표현형이 변화될 수 있다. 예를 들어, dosage-sensitive gene의 1개의 copy가 결손 되었을 경우 회복될 수 없는 기능결손을 가져올 수 있다. 또한 CNV가 여러 유전자의 긴 길이에 걸쳐 존재한다면 다양한 질환을 유발할 수 있다. Redon 등<sup>19)</sup>은 질환과 연관된 유전자의 14.5%와 중첩되는 1400여 부위에서 CNV를 발견하였다 (www.ncbi.nlm.nih.gov/omim).

최근 HIV-1/AIDS, 홍반성 낭창, 심장질환 및 암과 연관된 CNVs에 관한 연구를 살펴보면, Gonzalez 등<sup>21)</sup>은 강력한 human immunodeficiency virus-1 (HIV1)-suppressive chemok인이며 HIV coreceptor CCR5의 ligand인 유전자 코딩 CCL3L1의 CNV와 후천성 면역결핍증에 대

**Table 3.** Examples of disorders caused by structural variations (by Feuk et. al, 2006)<sup>11)</sup>

Indication	Locus	Gene involved	Type of rearrangement	Effect on gene
Severe speech and language disorder	7q31.1	FOXP2	Balanced translocation	Postulated down-regulation
Blepharophimosis Syndrome (BPES)	3q22.3	FOXL2	Four different microdeletion (126 kb to 1.9 Mb in size)	Postulated down-regulation
Campomelic dysplasia	17q24.3	SOX9	Two balanced translocations and one complex balanced translocation	Postulated down-regulation
Peter's anomaly	1q41	TGFB2	Balanced translocation	Postulated down-regulation
Ptack-Shaffer syndrome	11p11.2	ALX4	~1.37 Mb deletion	Postulated down-regulation
Short stature	Xp22.33	SHOX	Ring(X) with deletion of 700-900 kb of Xp and Xq pseudoautosomal regions	Postulated down-regulation
Spastic paraplegia Type II with axonal neuropathy	Xq22.2	PLP1	100-150 kb duplication	Postulated down-regulation
Townes-Brocks syndrome	16q12.1	SALL1	Balanced translocation	Postulated down-regulation
X-linked recessive Hypoparathyroidism	Xq27.1	SOX3	Deletion of ~25 kb with insertion of ~340 kb	Postulated down-regulation

해 연구하였는데, HIV/AIDS에 대한 감수성이 높을수록 개체의 평균보다 낮은 CCL3L1의 copy number를 가지며, 후천성 면역결핍증의 이환뿐만 아니라 진행속도와도 연관이 있는 것으로 보고하였다. 일반인보다 유전자 양이 낮은 성인 후천성 면역결핍증 환자는 질병이 급속히 진행되고 빨리 사망에 이르는 것으로 알려졌다. 그러나 어린이 후천성 면역결핍증 환자에서는 질병에 대한 CCL3L1의 copy number의 영향이 발견되지 않았다.

Yang 등<sup>22)</sup>은 인간 전신 홍반성 낭창과 관련된 보체 요소 C4 (with isotopes C4A and C4B)의 변이와 CNV에 대한 연구에서 전체 C4의 두 개의 copies를 가진 경우 SLE 질환에 대한 감수성이 증가하지만 5 copies 이상을 가진 경우 감소하였다. 0-1개 copy의 C4A를 가진 경우 SLE의 감수성이 증가하지만 3 copies 이상인 경우 오히려 감소하였다.

Sebat 등<sup>23)</sup>은 자폐증 환자 14명, 대조군 2명, 총 16명에서 17개의 CNVs를 발견하였다. 자폐증 환자 14명 중 12명이 결손과 연관되었으나 대조군 2명은 복제와 연관되었다고 하였다.

Christiansen 등<sup>24)</sup>은 염색체 1q12.1 부위 유전자의 결손이 선천성 심장질환과 관련성을 보고하였고, Gal 등<sup>25)</sup>은 DNA 회복효소 유전자 XRCC3 241 Met의 다형성이 이차 종양 유발가능성을 높이는 반면, XRCCq 399Gln 유전자의 다형성은 종양에 의한 사망률(mortality)을 낮추는 것으로 보고하였다.

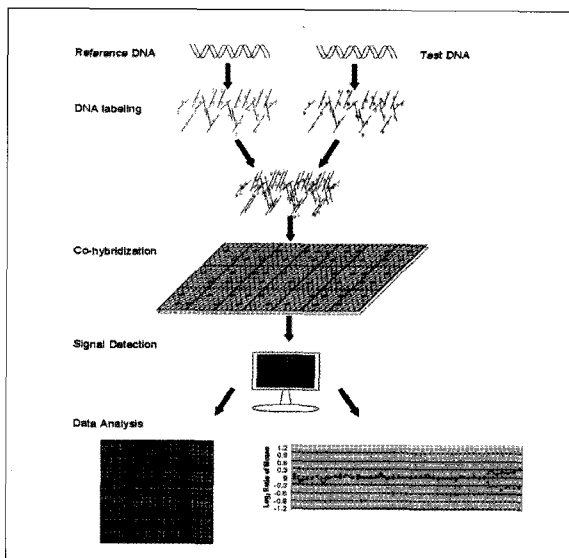


Fig. 2. Schematic representation of array-CGH technology (Bejjani et. al. 2006)<sup>27)</sup>.

## V. CNV의 분석방법

최근 인간 게놈의 다양한 CNV를 연구하기 위하여 Microarray Comparative Genomic Hybridization (array-CGH)<sup>26,27)</sup>과 Fluorescent In Situ Hybridization (FISH) 방법<sup>28)</sup>이 많이 사용되고 있다.

### 1. MicroArray Comparative Genomic Hybridization (Array-CGH)

Array-CGH법은 인간 게놈의 CNV와 다양한 암 연구에 아주 유용한 방법으로, 이 방법의 개발로 최초로 게놈 전체 CNV를 연구할 수 있게 되었다. 다른 두 개의 fluorophores로 표시된 대조군의 total genomic DNA와 실험군의 total genomic DNA를 metaphase chromosome 또는 DNA microarray에 hybridization시킨다. 대조군과 실험군 게놈에 존재하는 서열의 상대적인 CNV에 비례하여 상대적인 hybridization intensity가 나타나게 된다(Fig. 2). 만약 대조군이 정상이면 실험군 게놈의 DNA copy number variation을 의미하는 신호의 강도가 증가되거나 감소된다. 만약 서로 다른 색으로 표식될 수 있다면 두 개 이상의 genome DNA를 검사할 수도 있다. 결과는 linear scale상의 1.0값 또는 logarithmic scale상의 0.0값의 표준치로 나타낸다(Fig. 3)<sup>28)</sup>.

Hybridization되는 표적으로는 플라스미드, 코스미드, BACs, 또는 P1 인공 염색체에 의하여 클로닝된 올리고 뉴클레오타이드, cDNA, 또는 유전체 절편 등이 될 수 있다. Array-CGH의 선예도(resolution)는 nucleic acid tar-

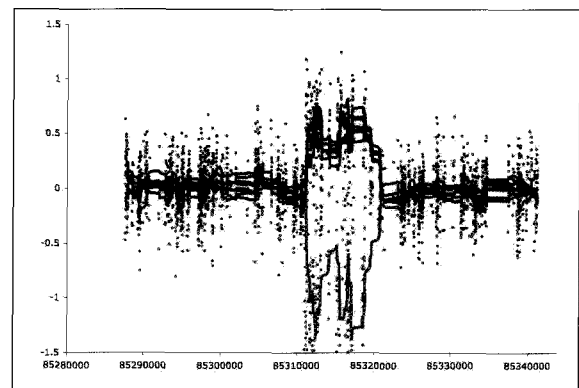


Fig. 3. Log2 Ratio of array-CGH: a ~11kb deletion on chromosome 8 revealed by ultra-high resolution CGH. Blue lines: individuals with two copies. Red line: individual with Zero copies (<http://www.sanger.ac.uk/humgen/cnv/>).

gets의 크기와 유전체 상의 coverage density에 의하여 좌우되는데, 핵산 표적의 크기가 작을수록, 고유 염색체 상의 표적에 더 많이 접촉할수록 선예도가 증가하게 된다. 이 방법의 경제적이고 주어진 플랫폼에 많은 개체의 빠른 스크리닝이 가능하다는 장점이 있지만, 표적의 크기와 수에 의하여 선예도가 좌우되는 단점이 있다.

## 2. Fluorescent in Situ Hybridization (FISH)

FISH법은 염색체 상에서 특정 DNA서열의 존재 유무와 위치를 검사할 수 있는 세포적 cytogenic technique이다. 이 방법은 염색체의 특정부위의 염기 서열과 매우 유사한 fluorescent probe를 사용하고, 형광 현미경으로 염색체 상에서 probe가 결합하는 위치를 찾아낸다 (Fig. 4, 5).

표적에 특이적으로 결합하는 probe에 nick translation 또는 PCR using tagged nucleotides 방법을 이용하여 fluorophores를 부착시킨다. 간기 또는 세포분열 중기 염색체를 일반적으로 유리로 만들어진 기질에 부착시킨다. Probe를 염색체 DNA에 적용시키고 hybridizing을 위해 12시간 정도 배양시킨다. 세척과정을 통해 unhybridized 또는 partially hybridized probe를 제거한다. 현미경을 이용하여 형광신호를 관측하고 이미지를 기록한다.

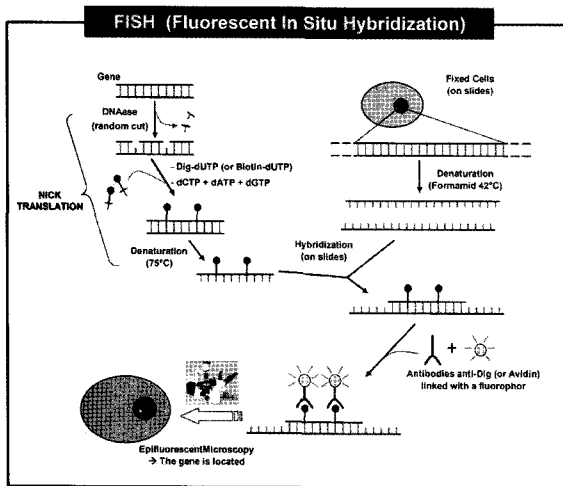


Fig. 4. Scheme of the principle of the FISH experiment to localize a gene in the nucleus ([http://en.wikipedia.org/wiki/Fluorescent\\_in\\_situ\\_hybridization](http://en.wikipedia.org/wiki/Fluorescent_in_situ_hybridization)).

## VI. 요약

인간 게놈의 DNA서열의 차이는 개개인의 특이성을 의미하기 때문에 염기서열의 변화는 질병에 대한 감수성, 약물에 대한 반응 등 개인의 성향에 큰 영향을 미치게 된다. 인간 게놈에는 여러 가지 형태의 유전적 변이가 존재하지만 그 중 단일염기다형성이 인간의 유전적, 표현형의 다양성을 설명하는 주된 유전적 변이로 생각되었으나 최근 유전체 전체 분석법의 발전으로 1 kb 이상 크기의 CNV의 발견으로 개체간의 유전적 다양성에 대한 더 많은 이해가 가능하게 되었고, 진화와 유전 질환에 대한 CNV의 역할을 조사하는 연구의 기초를 제공하게 되었다.

현재 인간게놈의 CNV를 찾아내고 특성화 작업을 목표로 하는 The Copy Number Variation Project를 위해 The Wellcome Trust Institute (Hinxton, United Kingdom), Hospital for Sick Children (Toronto), University of Tokyo (Tokyo), Affymetrix (Santa Clara, CA), 그리고 Harvard Medical School/Brigham and Women's Hospital (Boston, MA) 등이 참여하는 international consortium이 구성되어 보다 심도 있는 연구가 진행되고, 또한 향후 진보된 DNA microarray-based technology와 서열화 기술의 개발로 인간 게놈 상의 모든 유전적 변이를 발견하게 되고 포괄적인 CNV 지도를 완성하고 인간 유전자 다양성, 인간의 진화, 유전적 질환, 개인 맞춤형 의학에 대한 새로운 이해와 연구가 가능하게 될 것으로 기대된다.

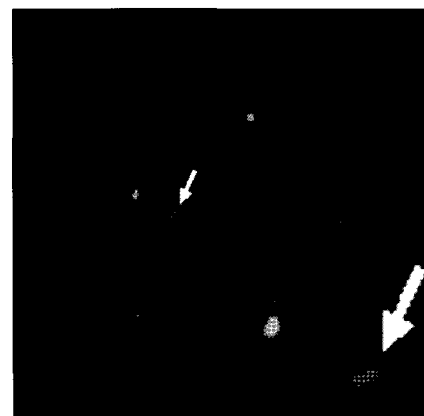


Fig. 5. Fluorescence in situ hybridization (FISH) image of CNV 14q12 (red) and RP11-79B13 (green), a control region on chromosome 14 that did show a deviation from normal copy number according to the array-CGH experiments. The arrow marks the pair of FISH signals at CNV 14q12. Quantitative-FISH showed that the marked probe was indeed due to hybridization at 2 adjacent loci rather than a split signal (Braude et. al. 2006)<sup>29)</sup>.

참고문헌

1. Venter JC, Adams MD, Myers EW et al : The sequence of the human genome. *Science* 291 : 1304, 2001.
2. Thorisson GA, Stein LD: The SNP consortium website : past, present and future. *Nucleic Acids Res* 31(1) : 124, 2003.
3. The International HapMap Consortium. The International HapMap Project. *Nature* 426 : 789, 2003.
4. The International HapMap Consortium. A haplotype map of the human genome. *Nature* 437 : 1299, 2005.
5. The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409 : 928, 2001.
6. Iafrate AJ, Feuk L, Rivera MN et al : Detection of large-scale variation in the human genome. *Nat Genet* 36 : 949, 2004.
7. Sebat J, Lakshmi B, Troge J et al : Large-scale copy number polymorphism I the human genome. *Science* 305 : 525, 2004
8. Sharp AJ, Locke DP, McGrath SD et al : Segmental duplications and copy-number variation in the hman genome. *Am J Hum Genet* 77 : 78,2005
9. Tuzun E, Sharp AJ, Bailey JA et al : Fine-scale structural variation of the human genome. *Nat Genet* 37 : 727, 2005.
10. Clark P. David. *Molecular biology: Understanding the genetic revolution*. 1st Ed. Burlington : Elsevier Inc. 2005.
11. Freeman JL, Perry GH, Feuk L et al: Copy number variation : New insights in genome diversity. *Genome Res* 949, 2007.
12. Feuk L, Carson AR, Scherer SW: Structural variation in the human genome. *Nat Rev Genet* 7:85, 2006.
13. Sherer SW, Cheung J, MacDonald JR et al : Human chromosome 7: DNA sequence and biology. *Science* 300: 767, 2003.
14. Conrad DF, Andrews TD, Carter NP et al : A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* 38 : 75, 2006.
15. McCarroll SA, Hadnott TN, Perry GH et al : Common deletion polymorphisms in the human genome. *Nat Genet* 38 : 86, 2006.
16. Hinds DA, Klock AP, Jen M et al : Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nat Genet* 38 : 82, 2006.
17. Nadeau JH, Lee C: Copies count. *Nature* 439 : 798, 2006.
18. Eichler EE: Widening the spectrum of human genetic variation. *Nat Genet* 38(1) : 9, 2006.
19. Redon R, Ishikawa S, Fitch KR et al : Global variation in copy number in the human genome. *Nature* 444 : 444, 2006.
20. Shaw CJ, Lupski JR: Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum Mol Genet.* 13(1) : 57, 2004..
21. Gonzalez E, Quinones MP, Barnshad MJ et al : The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307 : 1434, 2005.
22. Yang Y, Chung EK, Savelli SL et al : Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE) : low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 80(6) : 1037, 2007.
23. Sebat J, Lakshmi B, Malhotra D et al : Strong association of de novo copy number mutations with autism. *Science* 20(316) : 445, 2007.
24. Christiansen J, Dyck JD, Elyas BG et al : Chromosome 1q21.1 contiguous gene deletion is associated with congenital heart disease. *Circ Res* 92 : 1492, 2004.
25. Gal TJ, Huang WY, Chen C et al : DNA repair gene polymorphisms and risk of second primary neoplasms and mortality in oral cancer patients. *Laryngoscope* 115 : 2221, 2005.
26. Albertson DG, Pinkel D: Genomic microarrays in human genetic disease and Cancer. *Hum Mol Genet* 12(2) : 145, 2003.
27. Bejjani B, Shaffer LG: Application of array-based comparative genomic hybridization to clinical diagnostics. *J Mol Diagnostics* 8(5) : 528, 2006.
28. Pinkel D, Albertson DG: Array comparative genomic hybridization and its applications in cancer. *Nat Genet* 37:11, 2005.
29. Braude I, Vukovic B, Prasad M et al: Large scale copy number variation (CNV) at 14q12 is associated with the presence of genomic abnormalities in neoplasia. *BMC Genomics* 7 : 138, 2006.

저자 연락처

우편번호 130-702  
 서울특별시 동대문구 회기동  
 경희대학교 치의학전문대학원 구강악안면외과학교실  
**오정환**

원고 접수일 2007년 12월 3일  
 게재 확정일 2008년 3월 5일

Reprint Requests

**Jung-Hwan Oh**  
 Dept. of OMFS, Kyung-Hee University Dental School  
 Hoegi-dong, Dongdaemun-gu, Seoul, 130-702, Korea  
 Tel: 82-2-958-9440 Fax: 82-2-966-4572  
 E-mail: omsojh@khu.ac.kr

Paper received 3 December 2007  
 Paper accepted 5 March 2008