

## 고급바이오정보학 -12 강. Structural Genomic Variation 1

### 용어 정리

#### 1. Copy number variation

유전변이의 한 종류로 특정 유전자가 비정상적인 수로 DNA 상에 존재하는 것을 말한다.

#### 2. array comparative genomic hybridization (CGH)

CNV 를 관찰하기 위한 molecular cytogenetic 방법의 하나로 microarray 기술을 이용하여 고 해상도로 전장유전체(genome-wide)를 확인할 수 있다.

#### 3. Haplotype

유전형(allele) 조합을 나타내는 말로 유전적으로 함께 전해지는 chromosome 상의 이웃한 위치를 의미한다. Haplotype 은 특정 한 유전지역 (locus), 몇 개 유전지역 (loci), 혹은 전체 chromosome 을 의미하기도 한다.

#### 4. Recombination

두 가닥의 DNA 가 서로 유전적 정보를 교환하는 것으로 새로운 allele 을 만들어 낸다.

### 내용 정리

#### (학습 1)

1. SNPs: 책에서 스펠링 틀리는 것과 유사

2. CNVs: 논문 소개. 생각보다 많음.

#### (학습 2)

2. Array CGH: chromosome segment 를 single glass slide 에 차례로 놓음=>CGH test 수행: sample 을 green 으로 reference 를 red 로 넣어 라벨 후 다시 슬라이드에 위치=>more green color 는 more copy than reference

control experiments: 정상인 샘플을 reference 와 비교

3. 논문소개-High resolution, genome-wide CNV detection

African 이 more diverse, more CNV #

2 genome 간에 0.78% CNV difference 임이 중요!

전에 생각한 것보다 8 배 많음

Database of Genome Variants=chromosome 별로 catalog 로 볼 수 있음

CNV 크기는 보통 20~40kb 정도까지.

5. Case: AF-14 aCGH Genome Overview

Decipher 에서 기존의 정상인 CNV 영역과 겹치므로 pathogenic 하다고 보기에 어려움

#### (학습 3)

CNV 의 종류: Bi-allelic/multi-allelic(11%정도)

#### (학습 4) CNV Genotyping 과 CNV 생성 매커니즘

##### 1) Genotyping CNVs: Haplotypes

우리가 CNV 를 genotyping 할 때 haplotype 에 대해 잘 아는 것은 중요합니다.

**haplotype** 는 **CNV** 의 구조와, 어떻게 구성되어 있는지, 0 카피인지, 1 카피인지 2 카피인지 등의 정보를 줍니다.

이유를 이해하는 것이 중요한데, 이 슬라이드를 통해 **haplotyping** 의 중요성을 어떻게 **genotyping** 해석의 오류가 발생하는지를 설명하여 보이고자 합니다.

**haplotype** 는 **CNV** 의 구조와, 어떻게 구성되어 있는지, 0 카피인지, 1 카피인지 2 카피인지 등의 정보를 줍니다.

만약 하나 이상의 이 유전자를 가지고 있으면 **RH+**가 되고, 카피가 전혀 없으면 **RH-**가 됩니다.

이 유전자를 따라 아랫쪽으로 가다보면 유사한 유전자인 **RHCE** 유전자가 있습니다. 이것은 **RH+/-**에 전혀 영향을 주지 않는 유전자입니다.

그러나 서열 유사성만이 있지요. 만약에 **RHD** 유전자의 일부를 디자인한 실험을 진행한다면 녹색을 얻게 될 겁니다.

그것은 **RHD** 뿐만 아니라 **RHCE** 유전자에도 혼성화 할 것이기 때문이지요.

만약 여기에 보듯이 **RHD** 유전자 인근인 두 번째 색상을 디자인한다면, **RHD** 유전자를 나타내는 빨간색을 얻게 되고, **RHCE** 는 녹색을 얻게 되겠지요.

여기에 아빠 엄마 아들의 실제 예가 있습니다. 몇 개의 **RHD** 유전자가 있냐는 질문을 받게 되면, **qPCR** 이나 다른 방법을 사용하게 됩니다.

첫 눈에 볼 때는 에러라고 판단하기 쉬운데 계속해서 실험을 반복하게 되겠지요.

왜냐면 엄마에게서의 두 개와 아빠에게서의 두 개 유전자가 아들에게서 하나로 나타나는 것을 이해하기 어려우니까요.

저는 여기에서 엄마의 염색체를 관찰한 **FISH** 실험 자료를 보여드리고자 합니다.

보시면 엄마는 실제로 두 개의 **RHD** 유전자를 갖고 있지 않습니다. 여기보듯이 하나의 카피만 존재하지요. 한 카피나 한 염색체에 다른 한 카피는 다른 염색체에 존재합니다.

제가 말씀드렸듯이 이 실험에서의 우리 전략에 기초하면, **RHCE** 는 **RH+/-**에 영향을 주지 않는다는 것을 확인할 수 있습니다.

여기서 아빠는 **RHD** 유전자 2 카피를 가지고 있습니다. 우리는 **FISH** 실험을 통해 **haplotype** 에 대해서 보다 깊이 알게 되었습니다.

우리가 확인한 것은 아빠는 확실히 2 카피를 가지고 있으나 각각의 염색체에 있지 않지요.

두 카피가 하나의 염색체에 존재합니다.

그리고 **RHCE** 유전자가 여기 있습니다. 다른 염색체에는 0 카피가 보이고요. 여기 **RHCE** 유전자가 보이고요.

그래서 이 시나리오에서 가장 그럴듯한 것은 아빠에게서 이 염색체가 하나 전해지고, 나머지 하나는 엄마에게서 전해진 것이지요.

이것이 바로 아들이 **RHD** 유전자 하나를 갖고 있는 이유지요.

왜냐하면 아빠가 **RHD** 를 가지지 않은 염색체가 있기 때문입니다.

그래서 이것은 왜 이것이 **genotyping error** 가 아닌지 설명해주고 또한 **haplotype** 을 이해하는 것이 **CNV** 를 해석하는데 있어 중요한지를 설명해주고 있습니다.

## 2) Major mechanisms

기작에 대해서 우리가 하는 것은 무엇일까요? 이 **CNV** 들은 어떻게 만들어지는 것일까요?

생성에 관해서 크게 두 가지 범주로 나뉘어질 수 있다는 것에 대한 연구들이 존재합니다. 두 가지 중 하나는 재조합에 기초한 기작이고, 다른 하나는 복제와 관련된 기작입니다. 재조합에 기초한 기작에는 두 종류가 있습니다.

하나는 NAHR (non-allelic homologous recombination) 이고 이것은 duplication, deletion, inversion 등이 나타납니다.

또 하나는 NHEJ (non-homologous end joining) 이며 이것은 duplication 과 deletion 에 의해 나타나면 inversion 은 나타나지 않습니다.

복제와 관련한 기작에는 3 종류가 존재합니다. duplication 과 deletion 을 동반하는 replication slippage,

duplication, deletion, inversion, translocation 을 유발하는 FoSTeS 방법과 MMBIR 방법들이 있습니다.

이 두 가지 기작은 complex CNV 를 만들어내는 것으로 알려져 있습니다.

따라서 complex CNV 가 발견되면, 바로 FoSTeS 난 MMBIR 의 기작을 생각해야 합니다.

#### NAHR [https://en.wikipedia.org/wiki/Non-allelic\\_homologous\\_recombination](https://en.wikipedia.org/wiki/Non-allelic_homologous_recombination)

DNA 복제후에는 DNA 가 여기서 다시 만나게 됩니다.

그러나 DNA 서열상에 유사서열이 있어서 제대로 짝을 못 짓는 경우가 있습니다.

그래서 이부분이 제대로 짝지어야 하는데 여기서 잘못된 부분과 짝짓게 된 것이 보이지요.

이것이 일어나면, 재조합이 발생하게 되어 duplication 이나 deletion 이 발생하게 됩니다.

여기서 파란색과 빨간색 유전자가 유사한 서열을 보인다고 합시다.

보시다시피 여기에 이것은 그대로 유지되게 되며, 이 특정 염색체에서 이것은 duplicated 된 것이 보입니다.

그래서 두 카피가 되게 됩니다. 그리고 이 염색체는 deletion 이 되게 되지요. 그래서 결국 0 카피가 되게 됩니다.

그것은 바로 NAHR 이 어떻게 duplication 이나 deletion 을 만들어 내는지를 나타내고 있습니다.

#### NHEJ [https://en.wikipedia.org/wiki/Non-homologous\\_end\\_joining](https://en.wikipedia.org/wiki/Non-homologous_end_joining)

또 다른 기작은 NHEJ 입니다.

이것은 주로 G1 에 일어나나 DNA 합성 후의 G2 기에서도 종종 발생합니다.

저는 G1 기의 예를 보여 드리겠습니다. G1 기에서는 각 염색체가 단일 chromatid 로 되어 있고 여기에 그것을 표시하고 있습니다.

우리는 4 개의 유전자가 있다고 하죠 여기에 서로 다른 색으로 표시를 했습니다. 이 기작을 시작하게 하는 것은 DNA 끊김입니다. 여기 보시면 하나 두개가 끊어져 있지요.

때로는 DNA 끊김이 있는 경우 세포가 그것을 알고 수정하게 됩니다.

그런데 어떤 경우에는 이것이 제대로 일어나지 않습니다.

예를 들어 이 경우네, 끊김이 겹보기에는 연결된 것처럼 보이지만, B 와 C 유전자는 이 염색체에서 duplicated 되어 있고 이 염색체에서는 deletion 되어 있습니다.

이것이 어떻게 NHEJ 가 CNV 를 만들어 내는지를 보여주고 있습니다.

#### MMBIR <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2621351/pdf/pgen.1000327.pdf>

MMBIR 은 앞서 말씀드렸듯이 **complex CNV** 를 만드는 기작입니다.

보시는 그림은 2009 년 해스팅의 논문에서 따온 그림인데,

여기 두개 염색체에서 나온 딸 염색사가 있습니다.

여기보면 끊긴 팔이 보이고 있는데,

이때 일어나는것은 이 부분에서 발생해야 하는 DNA 합성이 **strand** 침입으로 발생하는 것입니다.

실제로 사용되어야 하는 **template** 가 아닌 침입을 통해 다른 염색체의 **template** 을 사용해 DNA 합성을 이어가게 됩니다.

이 합성이 일어나게 되면, 그림과 같은 이런 구조를 얻게 됩니다.

그러면 끊김이 다시 발생하게 되고, 여기서처럼 다른 **template** 인 **template 2** 를 사용하게 되고, 계속 되지요.

결국 원래 딸염색체로 돌아가죠.

우리가 발견한 것은 MMBIR 은 여기의 원래 DNA 서열을 야기하게 된다는 것인데, 왜냐하면 **micro homology** 를 통해 발생한 끊김이 발생하기 때문입니다.

결국 유전체의 다른 부분에 있는 DNA 가 이 특정 염색체에 복제된 것이지요.

## CNV 의 영향

CNV 의 영향은 무엇일까? 영향은 다양합니다.

이 슬라이드에서 우리는 유전체 CNV 가 미치는 3 가지 영향에 대해서 살펴봅니다.

각각 파란색, 녹색 등으로 표시된 유전자 A B C 가 있는 시나리오를 생각해 봅시다.

CNV 는 **deletion** 의 형태로도 존재하므로 이 유전자 중 하나가 **delete** 되었다고 하지요.

B 가 사라졌다고 하죠. 이 경우 이것은 이 유전자의 생산물의 감소를 야기할 가능성이 높습니다.

복제가 발생할 수도 있는데, 이 경우라면 RNA 와 단백질의 증가를 기대할 수 있겠지요.

실제로 두 유전자 사이에 끊김이 발생하는데, 이렇게 **deletion** 된 경우 새로운 합성유전자를 만들 수도 있습니다.

이런 합성유전자 중 일부는 암에서 매우 중요한 역할을 합니다. 그러나 정상인에게서도 이 합성유전자는 만들어질 수 있습니다.

이 경우에는 B 와 C 에서 끊김이 발생했고, 합성유전자가 만들어졌네요.

보시다시피 CNV 가 유전체의 구조와 하위 기능에 줄 수 있는 영향은 매우 다양합니다.

## Reymond et al., Cur Opin Genet Dev 17: 381, 2007

2007 년 레이몬드의 논문에서 그들은 CNV 가 유전적 영향을 준다는 것을 보였는데요.

여기 보이듯이 그들이 한 것은 파란색으로 보이는 이 유전자가 다른 유전자의 전사를 돕는 단백질을 발현하게 되는데요.

이 부분이 제거되면, 여기 이 두 유전자가 남게 됩니다.

하지만 이 유전자는 제대로 된 전사가 되지 못합니다. 왜냐하면 여기 이 유전자의 생성물이 만들어졌기 때문입니다.

특정한 경우에 여기보여지는 유전자처럼 전체 유전자의 제거가 아니라 일부 엑손이 제거된 경우가 있습니다.

이것은 남은 것들이 결국엔 전사되어 불완전한 단백질이 만들어지는 것을 의미하게 됩니다.

특정 서열지역이 다른 유전자지역과 합쳐져서 효율적인 전사를 야기하는 경우가 발생할 수 있습니다.

그러나 제거가 발생한 경우는 유전자 발현에 타격을 입게 되지요.

마지막으로 헤테로크로마틴화 된 지역이 있다면 이것은 전사의 과정에서 아주 짝짝하게 뭉쳐 있게 됩니다.

따라서 이 예에서는 다이아몬드가 인슐레이터를 의미하는데, 이것이 제거 되었다고 하면, 이 헤테로크로마틴 부분이 이 인슐레이터에 의해 멈추않고 계속해서 염색체를 타고 흐르게 됩니다.

이 유전자가 정상적으로 발현되었을 때는, 인슐레이터가 헤테로크로마틴을 아래로 흐르게하지 못하게 하므로 이 녹색 유전자를 발현시키지 않고, 비활성화 시키게 됩니다.

이러한 예들은 어떻게 CNV 가 발현인자, 유전자 발현, 구조 등에 영향을 주는지에 대한 많은 예들입니다.

### Quiz

**Q1. 다음 중 나머지 하나와 CNV 생성 매커니즘이 다른 하나는 무엇인가?**

**1. Replication slippage**

**2. FoSTeS**

**3. NHEJ**

**4. MMBIR**

정답 : ③

해설 : NHEJ 는 recombination-based 매커니즘이고 나머지는 replication-based 매커니즘이다.

**Q2.다음의 그림이 설명하고 있는 CNV 생성 매커니즘은 무엇인가?**

- NHEJ
- MMBIR : 정답
- NAHR
- FoSTeS

**Q3. 다음 중 CNV 의 설명으로 옳지 않은 것은?**

- CNV 은 copy number variant 의 약자로 유전자 copy 의 수에 변동이 있는 것을 의미한다.
- Array CGH 의 방법을 이용하면 CNV 를 detection 할 수 있다.
- Bi-allelic 과 multi-allelic CNV 가 존재하며 현재까지 알려진 CNV 중에는 Bi-allelic CNV 의 수가 더 많다.
- CNV 는 개인 간 차이가 있으며 array CGH 를 이용하면 reference 없이도 copy number 를 detection 할 수 있다.

정답 : ④

해설 : CNV 는 개인마다 다르게 형성되는 것은 맞으나 aCGH 에서 관심 있는 샘플과 reference 샘플을 동시에 측정하여 그 비율을 통해 CNV 를 확인하기 때문에 reference 가 반드시 필요하다.

요약하기

- 유전변이의 종류를 정리한다.
- Single nucleotide polymorphisms (SNP)
- Copy number variation (CNV)
  - **Array CGH 실험 방법을 정리한다.**
  - Test 하고자 하는 genomic DNA 에 형광인자 1 을 붙여 준비한다.
  - Reference 가 되는 genomic DNA 에 형광인자 2 를 붙여 준비한다.
  - Microarray 기판에 두 시료를 혼성화 시킨다.
  - 형광 1 와 2 의 세기를 측정하여 유전자 copy 의 상대적 수를 비교한다.
  - **2010 년 Nature 에 출판된 논문에 언급된 CNV 의 일반적 특징에 대해 정리한다.**
  - 한 사람이 가지고 있는 CNV 의 수는 유럽인종에서 1,117 개, 아프리카인종에서 1,488 개 이다.
  - CNV 의 크기는 443bp 에서 1,28Mb 의 크기를 가지면 median 길이는 2.9kb 이다.
  - 한 번의 실험으로 확인할 수 있는 CNV 의 49% 정도이다.
  - 발견된 총 CNV 중에서 (51,997) 대략 70%의 CNV 가 novel 한 것으로 확인되었다.
  - **CNV 의 종류와 특징을 간단히 정리한다.**
  - CNV 는 Bi-allelic CNV 와 multi-allelic CNV 로 나뉘며, duplication 이나 deletion 에 의해서 각 chromosome 내의 allele 수가 변화한다.
  - 이때 bi-allelic CNV 는 총 copy 의 수가 정상 2 인 경우에서 각각의 chromosome 에서 duplication 된 경우인 4 개까지 변화하며, multi-allelic CNV 인 경우는 deletion 혹은 duplication 이 몇 번 일어나느냐에 따라 그 수가 매우 다양하다.
  - **CNV 의 생성 mechanism 에 대해서 정리하고 각각이 CNV 에 대해 어떤 역할을 하는지 정리한다.**
  - Recombination-based mechanism
    - 1) NAHR: duplication, deletion, inversion
    - 2) NHEJ: Duplication, deletion
  - Replication-based mechanism
    - 1) Replication slippage: duplication, deletion
    - 2) FoSTeS: Duplication, deletion, inversion, translocation (Complex CNVs)
    - 3) MMBIR: duplication, deletion, inversion, translocation (Complex CNVs)

#### 참고문헌

- Sebat J, et al. (2004). Large-scale copy number polymorphism in the human genome. Science 305 (5683): 525-528.
- lafrate A, et al. (2004). Detection of large-scale variation in the human genome. Nature Genetics 36 (9): 949-51.
- Reymond A, et al. (2007). Side effects of genome strctural changes. Curr Opin Genet Dev 17 (5): 381-6
- Conrad DF. et al. (2010) Origins and functional impact of copy number variation in the human genome. Nature 464 (7289): 704-12