

# 유전자 발현의 변화를 발생시키는 분자적, 진화적 과정

조지훈

성균관대학교 생명과학과 진화유전체학연구실

E-mail: jojihoon98@gmail.com

## 요약문

유전자 발현(gene expression)의 유전 가능한 변화는 종 내, 종 간에서 흔하게 나타난다. 이 변화는 분자 수준에서 유전자 조절네트워크의 구조나 기능을 변화시키는 돌연변이(mutation)에 의해 유발되고, 자연선택(natural selection)에 따라서 조절된다. 오늘날의 다양한 고해상도 방법론(High-throughput methods)은 돌연변이의 도입과 유전자 발현(특히, 전사)에 영향을 미치는 이 돌연변이의 시스-(cis-), 트랜스-(trans-) 조절의 특성을 연구할 수 있게 해 주었고, 다양한 분자 메커니즘이 어떻게 이 변화를 발생시키는지 밝히는데 큰 역할을 하고 있다. 또한, 이 변화에 따른 돌연변이의 영향을 자연에서 관찰되는 것과 비교하는 연구는 중립적(neutral), 비-중립적(non-neutral) 진화과정의 역할을 구분해주고 있다. 이러한 분자생물학, 진화생물학의 통합은 오늘날 우리가 보고 있는 유전자 발현의 변화가 어떻게 일어났는지 이해하고, 미래에는 어떻게 진화할 가능성이 가장 높은 지에 대해서 예측할 수 있게 해 줄 것이다.

**Key Words:** 유전자 발현의 진화, 진화생물학, 분자생물학, 돌연변이, 시스-조절 변이, 트랜스-조절 변이

본 자료는 Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.* 22, 203–215 (2021). 의 논문을 한글로 번역, 요약한 자료입니다.

## 목 차

1. 서론
2. 시스-(cis-), 트랜스-(trans-)조절 변화
3. 시스-조절 변화의 메커니즘
  - 3.1. 코어프로모터
  - 3.2. 인핸서

- 3.3. 전사인자 결합부위(transcription factor binding sites)
- 3.4. 염색질 접근성(chromatin accessibility)
- 3.5. 번역 이후(post-transcriptional) 과정에서의 시스-조절 변화의 원인
4. 트랜스-조절 변화의 메커니즘
  - 4.1. 단백질 암호화, 비암호화 서열(coding and non-coding sequences)
  - 4.2. 전사인자
  - 4.3. 전사인자 이외의 트랜스-조절 변화를 유발하는 원인들
  - 4.4. 트랜스-조절 돌연변이의 영향에 대한 측정
5. 진화적 변화의 메커니즘
6. 미래의 방향

## 1. 서론

유전자 발현조절은 유전형에 따른 표현형을 나타내는데 있어 가장 중요한 과정으로 이 조절과정에 대한 변화는 다양한 형질을 나타내는 것에 기여하며, 종 간에서도 흔하게 나타난다. 예를 들어, 동식물의 색소침착, 생쥐의 몸 크기 차이, 효모의 포자형성, 인간의 질병과 같은 외형적, 생리적, 행동적 특성에 기여한다. 이러한 이유로 이 조절과정의 변화가 어떻게 일어나고 진화하는지에 대해 이해하는 것은 다양한 생물학 분야에서 중요하게 다뤄지고 있다.

유전적 변이는 유전자 발현의 변화에 기초가 되는 **조절네트워크(regulatory network)**의 활성화에 영향을 주게 되는데, 이 네트워크는 단백질, RNA, DNA 서열 간의 상호작용을 포함한다. 예를 들어, 인핸서(enhancer)와 프로모터(promoter) 부위에서 일어나는 전사인자 단백질과 DNA 서열의 상호관계는 이 조절네트워크를 연구할 때 가장 자주 고려되는 요소이고, 이외에도 단백질 간의 상호작용, 세포신호전달, 대사상태 등도 이 네트워크의 활성화에 영향을 준다. 예시에서 언급한 모든 요소들을 변화시킬 수 있는 **돌연변이(Mutations)**는 유전자 발현의 변화를 유발한다. 돌연변이는 **시스-작용(cis-acting)** 또는 **트랜스-작용(trans-acting)** 돌연변이 두 개로 분류 될 수 있다. 시스-작용 돌연변이는 같은 염색체 상에 존재하는 유전자나 가까이 위치하는 유전자의 발현에 변화를 유발하는 것을 의미하고, 트랜스-작용 돌연변이는 RNA와 단백질 같이 이동 가능한 분자들에 의해서 유전자 발현이 변화하는 것으로 유전체 상에 어느 곳이나 존재할 수 있다. 두 분류의 돌연변이 모두 유전자 발현의 변화에 기여하지만, 각 분자 메커니즘의 차이는 진화적 시간에 따라서 서로 다르게 작용해왔다.

유전자 발현의 변화와 이와 연관된 시스-, 트랜스- 작용의 변화를 설명하기 위해 이미 다양한 동물, 식물, 미생물에서 유전체 수준의 연구들이 진행되었다. 모든 형질이 그렇듯이, 이 변화는 돌연변이에 의한 새로운 유전적 변이의 도입과 유전적 부동(genetic drift)의 영향을 받는 자연선택에 따른 변이들의 필터링 과정들을 반영한다. 하지만, 이러한 각 과정들이 어떻게 변이들을

유발하는지 알아내는 것은 어렵다. 예를 들어, 어떤 한 유전자의 발현량이 다른 유전자들보다 더 변화를 보인다면 이는, 그 유전자 발현이 선택적 제약(selective constraint)이 적은 것 이거나, 새로운 돌연변이의 대부분이 발현량을 변화시키는 것이다. 이러한 변이를 만들어내는 과정에서 자연선택의 역할에 대한 연구는 지금까지 새로운 돌연변이의 영향일 것이라는 추정이 대부분이었다. 하지만, 최근 발전된 DNA 합성, 유전체 교정, 고해상도 발현량 분석 등의 기술로 조절 돌연변이가 발생하는 것과 그 특성을 규명하는 것을 대규모로 연구하여 더욱 자세히 이해 할 수 있게 되었다.

이 리뷰논문에서는 유전자 발현의 변화를 발생시키는 분자적, 진화적 과정들에 대해 현재 우리가 이해하고 있는 내용들에 대해서 다루고자 한다. 이 중에서도 활용 가능한 데이터가 가장 많고 단백질 수준보다 기술적으로 더 쉽게 연구할 수 있는 RNA 수준의 발현 변화에 대해 중점을 두고자 한다. 도입부에서는 **시스-, 트랜스-조절 변화(cis- and trans-regulatory variation)**가 유전자 발현의 변화에 어떤 기여를 하는지에 대한 연구들을 간략히 리뷰하고자 한다. 다음으로 이 조절변화 서열 상의 돌연변이의 영향과 종 내, 종 간의 발현 차이에서의 작용 등을 포함하여 이에 대한 분자 수준의 원인에 대해서 논하고자 한다. 마지막으로, 새로운 돌연변이 영향과 자연집단에서 분리되는 유전적 변이의 영향을 비교하는 것이 유전자 발현의 진화적 과정들을 어떻게 밝혀주는지를 보여주며 마무리 할 것이다.

## 2. 시스-(cis-), 트랜스-(trans-)조절 변화

시스-, 트랜스-조절 변화를 구분하는 것은 돌연변이와 유전자 발현에 대한 영향 간의 관계를 이해 할 수 있게 해 주며, 각 조절변화의 영향을 구분하여 연구하기 위해 주로 다음 두 가지의 연구 방법론이 활용되고 있다. 첫 번째 방법론은 F1 잡종(F1 hybrid)의 **대립유전자-특이적 유전자 발현(ASE: allele-specific expression, 이하 "ASE")**을 활용하여 공통의 트랜스-조절 배경에서의 시스-조절 대립유전자의 활성을 부모 세대와 비교하는 것이고, 두 번째는 **유전자 발현에 영향을 미치는 정량적 형질위치(eQTLs: expression quantitative trait loci, 이하 "eQTLs")**를 확인하기 위해 유전적 변이와 유전자 발현 간의 통계적 연관 관계를 연구하는 것이다. 이 두 방법 모두 시스-, 트랜스-조절 변화에 대해 전반적인 영향과 개별적인 영향을 포함한 정보를 제공한다.

유전자 발현에 시스-, 트랜스-조절 변화가 미치는 영향을 연구하기 위한 ASE 연구는 이미 다양한 동식물에서 수행되었다. 이 연구들에서는 비근교계(outbred) 집단에서의 개체 간, 한 종에서 더욱 변이된 유형(isolated strain) 간, 그리고 종간에서의 유전자 발현 분석을 다루는데, 각 비교는 전체 진화과정에서 유전자 발현의 진화가 어떻게 진행되었는지 보여준다. 먼저, 종 내에서의 연구 결과는 시스-조절 변화보다 트랜스-조절 변화가 유전자 발현 변화에 더 큰 기여를 하고 있음을 보여준다. 이 패턴으로 트랜스-조절 변화가 더욱 큰 돌연변이를 표적으로 한다는 것을 알 수 있는데, 이는 즉, 시스- 보다 트랜스-조절 변화가 유전체 내에서 더 많은 부위에서 유전자 발현 변화에 영향을 주는 돌연변이를 유발시킨다는 것을 의미한다. 반면에, 시스-조절변화는 서로 다른 종 간에서 유전자 발현의 변화에 비슷하거나 더 높은 수준의 기여를 한다. 이 시스-조절 변화와 트랜스-조절

변화의 영향 간의 차이를 직접적으로 비교한 연구 결과들은, 시스-조절변화가 트랜스- 보다 발산시간(divergence time)에 따라서 상대적으로 더 증가했음을 암시한다. 이 상대적 차이는 시스-조절 변화가 트랜스-조절 변화보다 진화적으로 더욱 이롭게 (혹은 덜 해롭게) 작용하고 있다고 해석할 수 있다. 이는 아마도 각 조절 변화의 **다면발현(pleiotropy)**의 차이에서 기인한 것이라 할 수 있는데, 이에 대한 내용은 아래의 섹션에서 계속해서 다룰 것이다.

유전자 발현의 변화에 영향을 주는 eQTLs에 대한 연구 또한, 다양한 분류군에서 연구가 수행되었다. 이 연구들은 유전체 내에서 조절변이들(regulatory variants)의 수, 위치, 영향에 대해서 알 수 있게 해주었고, 유전자 발현의 변화는 대부분 여러 개의 변이가 함께 영향을 주는 다유전자성(polygenic)이라는 것을 보여주었다. 또한, eQTLs는 유전체 상에서 비교적 큰 부분을 차지하고 여러 개의 유전적 변이들을 가지고 있을 수 있기 때문에, 유전자 발현에 원인이 되는 변이를 확인하는 것은 매우 어렵다. 이 때문에, 많은 재조합 중단지점(recombination break points)을 알 수 있는 방법일수록 이를 분석하는데 유리하다. 예를 들어, 특정 표현형을 가지는 개체들에 대해서 한 세대 이상에서의 재조합 중단지점을 함께 연구한 eQTL 확인 실험에서는 더 많은 eQTLs를 알 수 있게 해 주었고, 하나의 유전자(TDH3 gene in *S. cerevisiae*)에 100개 이상의 eQTLs가 영향을 주기도 하는 것을 밝혀내기도 했다.

eQTLs은 영향을 미치는 유전자와 유전체 상에서 근접해있으면(=proximal) 이는 시스-작용을 하는 것이고, 멀리 떨어져 있으면(=distal) 이는 트랜스-작용을 하는 것으로 간주된다. 이 가정에 맞게, 근접한 eQTLs는 유전자의 발현에 대립유전자-특이적으로 영향을 준다. 실제로 인간의 세포 내에서의 유전자 발현을 대규모로 연구한 GTEx (Genotype-Tissue Expression) 컨소시엄의 연구에서는 시스-작용을 하는 eQTLs이 대립유전자-특이적 발현과 강한 상관관계가 있음을 보여주었다. 또 다른 연구에서는 유전가능한 유전자 발현변화의 대부분은 트랜스-작용을 하는 eQTLs로 설명이 가능하다는 것을 보여주기도 했고, 이 eQTLs 중 많은 유전자에 영향을 주는 것들은 '**핫스팟(Hot spot)**'으로써 잘 알려져 있다. 이 두 가지 eQTLs를 GTEx의 연구에서 발표한 숫자에 따라서 살펴보면, 거의 95%의 단백질암호화유전자(protein-coding genes)는 적어도 하나의 시스-작용 eQTLs를 가지는 것으로 확인된 반면, 트랜스-작용 eQTLs는 단 121개의 단백질암호화유전자에서만 확인되었다. 이렇게 각 eQTLs가 서로 크게 다르게 확인되는 것은 eQTLs를 비교하고자 하는 연구에서 반드시 고려되어야 하며, 트랜스-조절 변화를 연구할 때에는 시스-조절 변화보다 더욱 많은 통계적 검증과 다중 검정이 필요하다. 이 때문에 일부 연구에서는 시스-작용 eQTLs를 확인하는 데에만 집중하기도 한다.

ASE, eQTL 연구 모두 시스-, 트랜스-조절 변화에 대한 이해에 많은 기여를 했지만, 이 변화에 의해서 달라지는 특이적 유전적 변화나 분자적인 메커니즘들에 대해서는 큰 통찰을 주지는 못하였다. 이러한 연구들이 단일변이 수준(single-variant resolution)까지 도달하게 된다면, 오늘날 우리가 관찰할 수 있는 조절변화의 패턴을 이해하는 데 있어 매우 중요한 아이디어를 제공할 수 있을 것이다. 이제 이어지는 다음 두 개의 섹션에서는 시스-, 트랜스-조절 변화를 유발하는 분자적인 메커니즘에 대해서 더 자세히 살펴보고자 한다.

### 3. 시스-조절 변화의 메커니즘

시스-조절 변화는 한 유전자의 특정 대립유전자의 발현을 조절하는 서열에 영향을 주는 유전적 변화에 의해 유발된다. 이 변화는 유전자의 **코어프로모터(core promoter)**나 **인핸서(enhancer)**에 영향을 주고 이 두 가지 모두 **전사인자(transcription factors)**에 대한 결합 부위, 전사인자가 DNA에 결합하는 것에 영향을 주는 염색체구조 등을 포함한다. 다음으로는 시스-조절 변화의 원천이 되는 여러 구성요소들에 대해서 살펴보고자 한다.

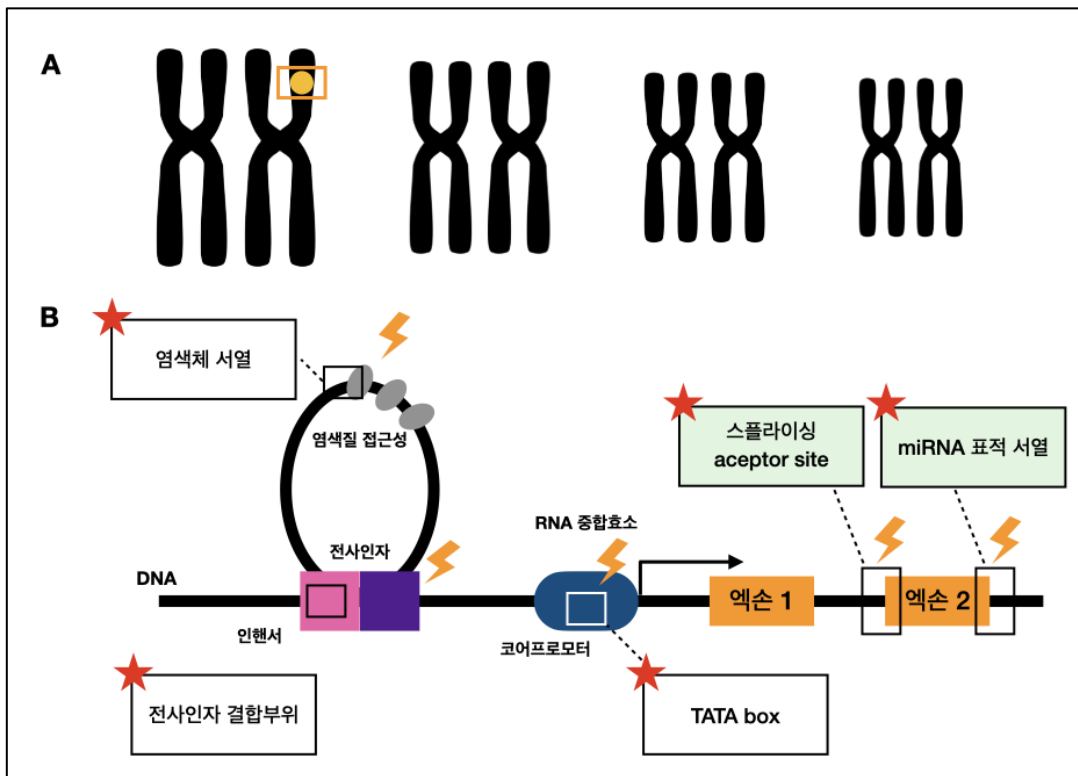


그림 1. 진핵생물에서 시스-조절 변화의 원인.

A. 시스-조절 변화 원인 돌연변이 위치의 염색체 모식도. 시스-조절 변화의 원인이 되는 돌연변이(주황색 네모 테두리)는 영향을 주는 유전자(주황색 동그라미)의 근처에서 일어난다. B. 시스-조절 변화의 원인이 되는 유전체 내 돌연변이 위치들(주황색 번개 모양 = 돌연변이, 빨간색 별 모양 = 염기서열 돌연변이, 흰색 바탕 = DNA상의 돌연변이, 초록색 바탕 = RNA상의 돌연변이).

#### 3.1. 코어프로모터

근접한 위치에서 일어나는 유전자의 발현은 전사과정에 필수적인 전사인자 결합 부위를 가지는 코어프로모터 서열에 의해서 조절된다 (그림 1). 코어프로모터 서열은 일반적으로 전사시작위치(transcription start site)와 가깝게 위치한다. 일부 코어프로모터들은 일정한 위치와 함께 TATA box와 같은 별개의 서열을 가지고 있는 반면, 나머지는 더 넓은 위치에 분포하는 CpG islands와 같은 서열 모티프들을 가지고 있다.

고해상도 돌연변이유발(high-throughput mutagenesis) 연구는 코어프로모터의 활성을 변화시키는 수천 개의 단일염기변이를 확인할 수 있게 해 주고, 이 변이가 종 내와 종 간에서 어떻게 유전자 조절 변화에 영향을 미치는 지를 보여준다. 이에 대한 최초의 연구에서는 영향력이 큰 대부분의 돌연변이들은 TATA box와 initiator region (=전사시작위치와 겹치는 TATA box 이 후의 코어프로모터 서열)에 위치한다는 것과, 이 두 핵심모티프 이외에서의 돌연변이는 통계적으로 유의미한 영향을 나타내지는 못한다는 것을 확인하였다. 하지만, K562 세포주를 활용하여 진행된 최근의 연구에서는 두 핵심 모티프 이외의 돌연변이들 또한 코어프로모터의 활성에 영향을 줄 수 있음을 확인하였다. 즉, 요약하면 코어프로모터의 서열 모든 부위에서의 돌연변이는 유전자 발현에 영향을 줄 수 있다는 것이다.

이러한 코어프로모터가 유전자 발현 변화에 미치는 중요성에도 불구하고, 코어프로모터의 서열, 히스톤 표지(histone marks), 기능 등은 종 간에서 매우 보존되어 있다. 이러한 보존성은 아마도 유전자를 발현시키기 위한 기능적 프로모터의 필요성과 이러한 서열에 결합하는 단백질의 강한 기능적 제약(functional constraints)에 의해 주도되는 것으로 보인다. 실제로, TATA box와 같이 일반적인 전사인자들이 결합하기 위한 서열인 프로모터 내의 서열들은 포유류의 코어프로모터에서 대부분 매우 보존되어 있다. 그러나, 인간과 붉은털원숭이(rhesus macaque) 사이의 코어프로모터 서열을 비교한 한 연구는 소수 유전자의 프로모터는 양성선택(positive selection)에 의해서 분기될 수 있음을 보여주었고, 다른 연구에서는 코어프로모터의 획득과 손실(gain and loss)이 생쥐와 인간 사이에서의 유전자 발현 차이에 기여했음을 보여주었다. 더욱이, 코어프로모터 자체의 변이는 유전자 발현 차이의 원인이 아닐지라도, 코어프로모터 구조의 변화는 유전자 발현 차이를 만들어 낼 수 있는데, 이는 TATA box의 유무, 코어프로모터의 뉴클레오솜 상의 위치 등이 효모의 유전자 발현 차이에 미치는 영향을 연구한 결과에서 확인되었다.

### 3.2. 인핸서

코어프로모터와 비교해서, 인핸서는 일반적으로 전사시작위치의 앞부분, 뒷부분 또는 인트론 부위에 멀리 떨어져서 위치하고 (그림 1), 유전자 발현에 영향을 주는 시스-조절 변화에 보다 더 자주 그 원인으로 지목된다. 왜냐하면, 인핸서는 코어프로모터에 비해 더 시간 특이적, 조직 특이적, 환경 특이적인 방식으로 유전자의 발현을 조절하기 때문이다. 또한, 인핸서는 다면발현의 특성을 나타내기 때문에 기능적인 제약을 덜 받고, 그래서 더 진화가능하다. 실제 한 예로, 포유류 종에서 일반적으로 인핸서와 연관된 히스톤 표지는 코어프로모터에서보다 훨씬 더 다양성을 보인다.

### 3.3. 전사인자 결합부위(transcription factor binding sites, 이하 "TFBSs")

모든 시스-조절 DNA 서열이 가지고 있는 기본적인 기능적 요소는 전사과정을 촉진하거나, 억제시킬 수 있는 TFBSs이다. 이 서열들은 짧고, 잘 변하여 상대적으로 빠르게 진화한다. TFBSs에 다양한 변화를 유발하는 돌연변이는 시스-조절 활성을 달라지게 한다. 앞선 코어프로모터와



인핸서와 마찬가지로 TFBSs 에서의 돌연변이 또한 유전자 발현에 변화를 일으키는데, TFBSs의 돌연변이는 이 세 가지 중에 가장 큰 영향을 준다. TFBSs는 인핸서 내에서 가장 보존적인 서열이지만, 종 내, 종 간에서의 유전자 발현의 변화에 원인이 되는 유전적 변이를 포함할 수 있다. 하지만, 인핸서 또는 비슷한 시스-조절 서열에서 확인되는 기능적 변화의 대부분에서 어떤 특이적 유전적 변화가 그 변화를 일으키는지는 아직도 밝혀지지 않았다.

### 3.4. 염색질 접근성(chromatin accessibility)

한 TFBS가 한 유전자의 발현을 조절하기 위해서는 전사인자가 그 DNA 서열에 결합할 수 있어야 한다. 진핵생물에서 DNA는 뉴클레오솜(nucleosome)이라는 히스톤 단백질 복합체에 의해서 감싸져서 염색질을 이루게 되며, 이 뉴클레오솜은 염색질에 대한 접근을 막을 수 있다 (그림 1). 시스-조절 DNA 서열에서의 뉴클레오솜과 전사인자 사이의 정상적인 상호작용은 유전자 발현에 영향을 줄 수 있기 때문에, 염색질 구조에 영향을 미치는 유전적 변화는 또 하나의 시스-조절 변화의 원인이다. 실제로, 프로모터에서의 뉴클레오솜 분포의 다양한 패턴은 유전자 발현의 변화, 종 수준의 유전자 발현 다양성, 유전자 발현에 영향을 주는 돌연변이의 영향 등과 상관관계가 있음이 많은 연구에서 증명되었다. 이 연구결과들은 뉴클레오솜 분포의 다양한 패턴이 유전자 발현의 진화적 궤적을 연구하는데 중요한 역할을 할 수 있다는 것을 의미한다.

하지만, 염색질 구조의 변화가 유전자 발현의 진화에 기여를 한다는 것에 대한 직접적인 증거는 아직 많이 부족하고, 이제 막 정보가 쌓이기 시작하고 있는 단계이다. 초파리에서의 연구를 예로 들면, 염색질 접근성과 TFBSs에 대한 정보를 함께 보았을 때, TFBS만을 가지고 보았을 때 보다 유전자 발현의 다양성을 더 잘 설명할 수 있었다. 또한, 효모(Yeast)를 활용한 연구에서도 서로 다른 염색질 구조가 유전자 발현과 상관관계가 있음을 보여주었다. 그러나, 몇몇 연구에서는 염색질 구조의 변화가 유전자 발현 변화의 직접적인 원인이 아니라, 구조변화에 따라 노출된 TFBSs의 변화에 의한 결과임을 보여주기도 하였다. 그러므로, 염색질 구조변화와 유전자 발현 간의 직접적인 인과관계에 대해서는 더 많은 연구가 필요하다.

### 3.5. 번역 이후(post-transcriptional) 과정에서의 시스-조절 변화의 원인

코어프로모터, 인핸서, 염색질 접근성 등은 시스-조절 변화의 가장 주된 요인들로 거론되고 있지만, 유전자 발현에서 대립유전자 특이적 변화를 유발하는 것은 이 요인들에 의해서만 나타나지는 않는다. 예를 들어, 스플라이스 부위(splice sites)는 microRNA에 의한 조절과 마찬가지로 mRNA 스플라이싱, mRNA의 안정성에 영향을 주는 폴리아데닐화(polyadenylation)의 변화, 세포 내에서의 위치와 단백질 번역, mRNA 분해와 관련된 3'비번역부위(3' untranslated region) 등에 대립유전자-특이적으로 영향을 미칠 수 있다. 또한, mRNA 내에서의 서열 변이도 리보솜의 점유율과 단백질 번역효율 등에 영향을 준다. 미래의 연구들에서는 이러한 번역 이후 과정들이 유전자 발현 조절의 진화에 어떠한 기여를 하는지에 대해 더욱 면밀히 확인해 보아야 할 것이다.

#### 4. 트랜스-조절 변화의 메커니즘

영향을 주는 유전자와 근접한 위치에 작용하는 시스-조절 변화와는 달리 트랜스-조절 변화는 유전체 상 어느 곳에서나 유전자 발현에 영향을 줄 수 있다. 그러므로 이 트랜스-조절 변이의 잠재적인 위치는 단백질 암호화 서열(coding sequence), 비암호화 서열(non-coding sequence) 모두가 될 수 있으며, 이는 유전자의 발현이나 유전자가 만들어내는 산물의 활성을 직접적으로 또는 간접적으로 모두 영향을 준다 (그림 2). 트랜스-조절 변이의 이러한 광범위한 분포 가능성 때문에 어떤 부분이 트랜스-조절 변이인지 확인하는 것은 매우 어렵다. 그러므로, 이를 확인하기 위해 전유전체(genome-wide) 수준의 돌연변이유발(mutagenesis) 실험 또는 맵핑전략(mapping strategies)이 필수적이며, 때로는 연결된 여러 부위 중에 실제 트랜스-조절 변화에 원인이 되는 변이를 구분하기 위한 실험 검증 과정이 추가적으로 필요하다.

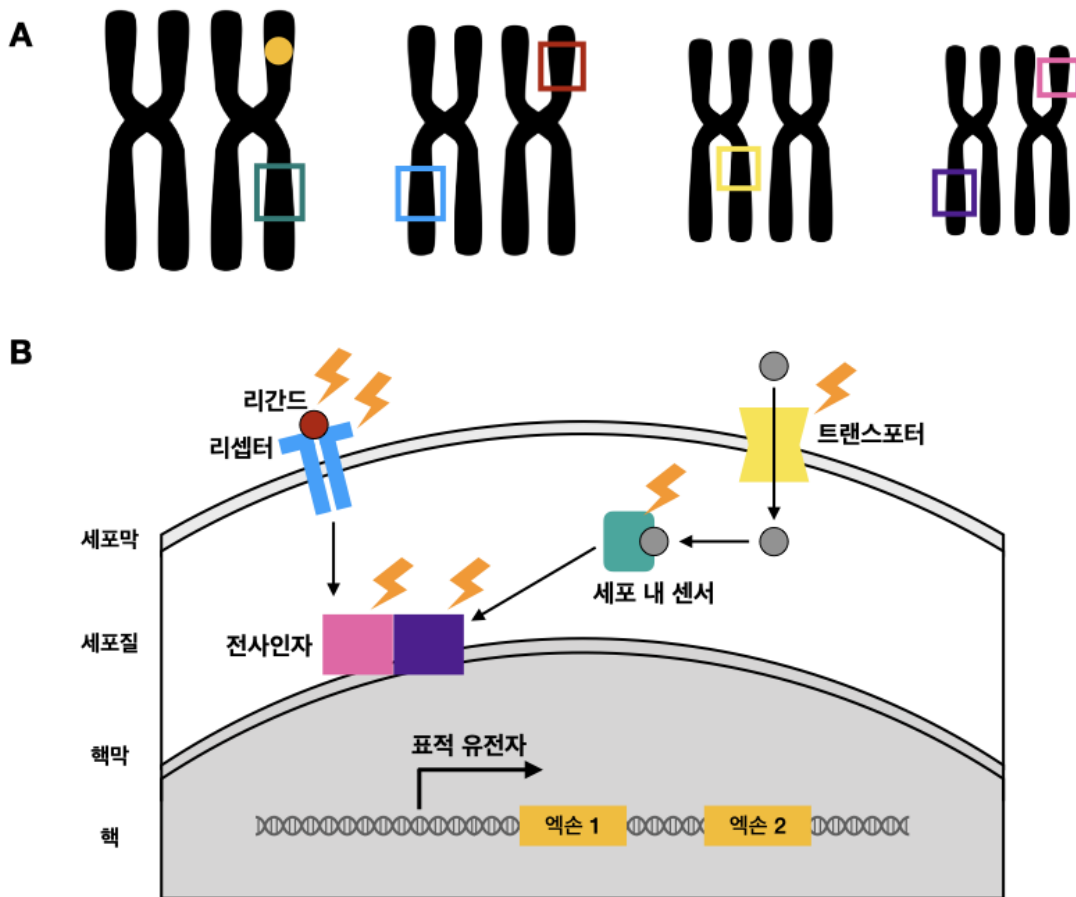


그림 2. 트랜스-조절 변화의 원인.

A. 트랜스-조절 변화 원인 돌연변이 위치의 염색체 모식도. 트랜스-조절 변화의 원인이 되는 돌연변이(네모 테두리들)는 유전체 어디에서나 일어나서 유전자(주황색 동그라미)에 영향을 준다. B. 트랜스-조절 변화의 원인이 되는 돌연변이(주황색 번개 모양: 돌연변이).



#### 4.1. 단백질 암호화, 비암호화 서열(coding and non-coding sequences)

비록 트랜스-조절 변이의 영향은 RNA나 단백질과 같이 확산 가능한 분자들에 의해 매개되어 작용하지만, 인간의 조절 변이에 대한 연구들은 대부분의 트랜스-작용 변이들이 이 분자들을 암호화하는 DNA 서열에 위치하지 않는다는 것을 보여준다. 실제로, 여러 대규모 전유전체상관분석연구(GWAS: genome-wide association study)에서 트랜스-조절 변이들의 대부분은 단백질 비암호화 서열에서 발견된다는 것과, 시스-작용 서열들이 그 유전자 발현을 조절하는 것을 확인했다. 이는 즉, 어떤 유전자가 시스-에 따른 영향을 받아 발현량이 변화하면, 그 변화가 다른 유전자들의 발현을 트랜스-로써 영향을 줄 수 있다는 것이다. 이러한 내용은 monocyte와 adipose tissue 등에서 이루어진 여러 eQTL 연구들에서도 확인이 되었다.

하지만, 효모(*S. cerevisiae*)에서의 연구들은 이 종의 단백질 암호화, 비암호화 서열 상의 트랜스-조절 변이의 분포가 또 다른 패턴을 보여준다는 것을 확인하였다. 인간에서 많은 유전자들에 영향을 미치는 트랜스-조절 eQTLs를 가지는 핫스팟 유전자들은 지역적이고(local) 잠재적으로 시스-작용 eQTLs를 가질 확률이 높지만, 효모에서 확인된 트랜스-조절 변이는 주로 단백질 암호화 서열 안에서 확인되었다. 즉, 효모는 단백질 암호화 영역에서 인간보다 더 높은 비율의 트랜스-조절변이를 가지고 있는데, 이는 인간에서 약 97%에 달하는 단백질 비암호화 영역이 효모에서는 27% 정도 밖에 되지 않기 때문이거나, 실험실에서 활용되는 효모균주들이 야생형 균주에서는 존재하지 않는 많은 변이들을 가지고 있기 때문일 수 있다. 효모를 포함한 다양한 종들에서 트랜스-작용 변화를 일으키는 단백질 암호화, 비암호화 영역에서의 변이의 상대적인 기여에 대해 정확하게 확인하려면 더 광범위한 연구와, 이 결과에 대한 야생형 집단(natural populations)에서의 실험적인 검증 또한 필요할 것이다.

만약, 트랜스-조절 변이들이 단백질 암호화 서열보다 비암호화 서열에서 일반적으로 더 많이 확인 된다면, 이는 비암호화 서열에서의 돌연변이는 다면발현성이 덜(less pleiotropic)한 경향이 있기 때문일 것이다. 예를 들어, 암호화 서열에서의 돌연변이는 그 유전자의 서열을 바꾸지만, 비암호화 서열에서의 돌연변이는 조직-특이적(tissue-specific) 인핸서들의 활성화만 영향을 미치기 때문에, 보다 더 적은 형질에 영향을 주게 된다. 실제로 인간의 단백질 비암호화 서열에서 확인되는 트랜스-작용 eQTLs는 코어프로모터 보다 인핸서에 영향을 주는 경우가 대부분이고, 종종 조직-특이적으로 작용한다. 그 이유를 생각해 보자면, 다면발현성이 더 높은 곳 에서의 돌연변이들은 덜한 곳보다 일반적으로 더 해로울 것이고, 단백질 암호화 영역에서의 돌연변이는 비암호화 영역에서의 돌연변이 보다 더 강하게 선택되어 자연 집단 내에서 그 빈도를 줄이고자 할 것이기 때문이다. 하지만, 이러한 패러다임은 오히려 시스-조절 서열이 더 많은 다면발현성을(more pleiotropic) 가진다는 것과 단백질 서열은 더 모듈화(more modular)되어있다는 데이터들에 의해서 도전받고 있다.

#### 4.2. 전사인자

전사인자는 가장 가능성이 높은 트랜스-조절 변화의 원인으로 여겨지는데, 특히 핫스팟 eQTLs에서 더 그렇다. 그 이유는 대부분의 전사인자는 하나가 아닌 여러 타겟 유전자들의 발현량을 조절하기 때문이다. 실제로 전사인자는 인간과 효모에 걸쳐서 핫스팟 eQTLs에 가장 큰 원인이 된다.

하지만, 전사인자가 여러 타겟 유전자들의 발현량에 영향을 주는 능력은 전사인자가 진화하는 데에 기능적 제약이 되는 결과를 낳기도 한다. 이 때문에 전사인자들의 단백질 암호화 서열의 DNA-결합 특성과 일반적인 생리적 기능은 오랜 진화과정동안 보존적이다. 이러한 보존적인 특성에도 불구하고 전사인자는 자신의 DNA 결합 특이성(DNA binding specificity)에 영향을 주어 기능적으로 분기(diverge)가 가능한데, 이는 효모의 교배형(mating type) 조절, 식물의 개화와 세포분열 조절, 곤충의 체절패턴 조절 등의 다양한 연구에서 보고되었다.

#### 4.3. 전사인자 이외에 트랜스-조절 변화를 유발하는 원인들

전사인자를 암호화하지 않는 유전자들에 영향을 미치는 변이들 또한 트랜스-조절 변화의 중요한 원인이 된다. 예를 들어, 염색질 조절자(chromatin regulators)는 유전자 발현에 광범위한 영향을 주고, 효모에서의 한 eQTL 연구에서는 이 유전자들이 많은 유전자들의 발현에 영향을 주는 트랜스-작용 eQTLs으로 작용한다는 것을 보여주었다. 효모에서의 또 하나의 기능적 연구에서는 공동인자(cofactors)에서 변이의 트랜스-조절 효과가 전사인자의 활성화와 글루코스 리셉터 RGT2, 막단백질 SSY1과 같이 대사과정에 영향을 미치는 유전자들의 활성을 조절한다는 것을 증명하였다. 인간에서는 트랜스-eQTLs이 전사인자가 아닌 체계에서 담즙산(bile acids)의 수송을 변화시키는 SLCO1A6 유전자에서 확인 되는 것을 보여주었다. 이 연구들과 다른 연구들에서 보여주고 있는 트랜스-조절 변화의 다양한 원인은 전사과정, 구조, 신호전달, 대사네트워크 등의 상호 연결성에서 비롯되었을 것이며, 이 트랜스-조절 변이들을 예측하고 규명해내는 연구가 시스템생물학적인 이해와 함께 매우 중요하다는 것을 강조해주고 있다. 또한, 발현하는 모든 유전자가 모든 형질에 대해서 영향을 미칠 수 있는 가능성이 있다는 유전학의 '옴니제닉(omnigenic)' 모델과 일치 한다. 궁극적으로 트랜스-조절 변화의 원인에 대해서 온전히 이해하기 위해서는 트랜스-조절 변이의 후보군들에 대해서 더욱 활발한 기능적 검증 연구가 필요할 것이다.

#### 4.4. 트랜스-조절 돌연변이의 영향에 대한 측정

시스-조절 돌연변이의 영향을 확인하기 위한 전략인 표적 돌연변이유발(targeted mutagenesis)은 트랜스-조절 돌연변이에 대해서 균일한 측정을 하는 데에 활용할 수 없다. 왜냐하면, 트랜스-조절 돌연변이는 유전체의 어느 곳이나 존재하기 때문이다. 그러므로, 트랜스-조절 돌연변이를 제대로 확인하기 위한 최적의 방법은 유전체 내에 돌연변이를 무작위로 도입한 뒤 이에 따른 유전자 발현 변화에 대한 영향을 확인하는 것이다. 일반적으로 **돌연변이 축적(mutation accumulation)**과 **무작위 돌연변이유발(random mutagenesis)**라는 두 가지 전략이 이에 활용된다. 두 방법 모두 돌연변이가 시스-작용을 하는지 트랜스-작용을 하는지는 구별해 낼 수 없는데, 한 중점 유전자(focal gene)의 발현에 영향을 미치는 무작위로 도입된 돌연변이의 대부분은 트랜스-작용을 할 것으로 예상된다. 즉, 이러한 연구에서 확인된 시스-조절 돌연변이는 무시할 수 있을 정도라는 것이다. 실제로 효모에서의 TDH3 유전자에 대한 연구는 한 무작위 돌연변이가 이 유전자의 발현에 시스-가 아닌 트랜스-로 영향을 미칠 확률이 최소 265배 더 높다는 것을 예측하였다.

돌연변이 축적 연구는 일반적으로 **돌연변이변화(mutational variance, 이하 " $V_m$ ")**를 추정하여 유전자 발현에 대한 새로운 돌연변이의 효과를 개괄적으로 보여주는데,  $V_m$ 은 각 세대별로 새로운 돌연변이로 인해 야기되는 발현량 변화의 증가를 설명해 준다. 지금까지의 연구에서  $V_m$  파라미터는 효모, 초파리, 선충 등에서 전유전체 수준에서 예측되었고, 이 데이터는 새로운 돌연변이가 유전자 발현량에 광범위한 영향을 준다는 것을 제시했다. 한 예로, 200세대에 걸친 돌연변이 축적 실험을 진행한 초파리에서의 연구에서는 12개의 독립적인 각 strains에서 360개의 돌연변이를 관찰하였고, 약 ~39%의 유전자에서 유의미한 유전자 발현 변화를 발견하였다. 이 예시와 같이, 일반적으로 돌연변이 축적 연구는 많은 돌연변이들이 여러 유전자의 발현에 영향을 준다는 것과 이는 트랜스-조절 효과를 보이는 것과 일치하는 것을 보여준다.

단일 유전자의 발현량에 영향을 주는 돌연변이 세트를 특이적으로 관찰하는 돌연변이유발 연구는 돌연변이 축적 연구에 중요한 보충연구이다. 유전자의 발현량에 영향을 주는 트랜스-조절 돌연변이에 대해서 더 심층적인 샘플을 제공해 주기 때문이다. 지금까지 돌연변이유발 접근법은 효모의 TDH3 유전자의 프로모터에 의해 발현량이 변화하는 트랜스-조절 돌연변이에 대한 돌연변이 영향의 분포를 연구하기 위해 가장 많이 활용되었다. 이 연구는 TDH3 유전자는 효모의 유전체 상에서 가장 발현량이 높은 유전자 중 하나이지만, 이 유전자의 발현량을 증가시키는 돌연변이들은 이를 감소시키는 돌연변이만큼 흔하다는 것을 보여주었다. 또한, 이와 같은 접근법은 효모의 다른 9개의 유전자 프로모터 작용을 변화시켜서 유전자 발현에 영향을 주는 수천 개의 돌연변이들의 특성을 확인하는 데에도 활용되었다.

이렇게 특정 유전자의 발현에 영향을 주는 트랜스-작용 돌연변이에 중점을 두는 연구는 같은 유전자에 대해 시스-와 트랜스-작용 돌연변이의 영향을 서로 직접적으로 비교할 수 있게 해 준다. 예를 들어, 효모의 TDH3 프로모터에서 235개의 시스-조절 돌연변이의 영향과 유전체 전체에 퍼져있는 ~47,000개의 돌연변이의 영향을 비교한 한 연구에서는 시스-조절 돌연변이가 트랜스-조절 돌연변이보다 TDH3 프로모터의 작용에 의한 유전자 발현에 더 큰 영향을 미치는 경향이 있음을 보여주었다. 또한, 이 시스-조절 돌연변이들은 트랜스-조절 돌연변이보다 더 유전자의 발현을 감소시키며, 이배체(diploid) 세포에서 우세적인 영향을 보였다. 지금까지 알려진 정보에 따르면, TDH3 유전자는 시스-, 트랜스-조절 돌연변이를 비교할 수 있는 현존하는 유일한 유전자이다. 하지만, 만약 다른 유전자들이 비슷한 경향을 보인다면 시스-, 트랜스-조절 돌연변이 간의 이 차이는 예상되는 다면발현의 차이와 결합되어 있을 것이고, 이는 유전자 발현의 진화에 시스-, 트랜스-조절 돌연변이의 서로 다른 기여를 설명해 줄 수 있을 것이다.

## 5. 진화적 변화의 메커니즘

새로운 돌연변이가 유전자 발현의 변화를 어떻게 만들어내는지 이해하는 것은 유전자 발현이 어떻게 진화하는지 이해하는 데 있어 매우 중요하다. 왜냐하면, 이는 **중립적인 과정(neutral processes)**에 따라 서로 다른 진화 상의 시간이 지난 뒤에만 우리가 확인할 수 있는 유전자 발현의 변화를 예측할 수 있게 해 주기 때문이다. 즉, 유전자의 발현량이 중립적으로 진화할 때, 돌연변이는

이 발현에 영향을 줄 수 있는 새로운 변이들을 도입하게 되고, 유전적 부동(genetic drift)은 우연히 생긴 이 변이들을 수정하거나 제거하고 돌연변이의 영향을 효과적으로 무작위 샘플링 할 수 있게 해준다. 하지만, 자연선택이 한 유전자의 발현에 작용할 때, 특정 조절 변이들은 그 영향에 따라서 다른 변이들에 비해 수정되거나, 제거될 확률이 높고, 돌연변이 영향의 분포가

**다형성(polymorphism)**이 관찰된 영향의 분포와 다른 것의 원인이 된다 (그림 3). 그러므로 돌연변이의 영향을 다형성 부위(polymorphic sites) 또는/그리고 발산 부위(divergent sites)의 영향과 함께 비교하는 것은 자연선택의 영향을 추론하는 가장 효율적인 방법이다. 이 일반적인 전략은 돌연변이 추적 연구에서 추론된 돌연변이의 영향을 활용하고, 특정 유전자의 발현에 더 심층적인 영향을 미치는 시스-, 트랜스-조절 돌연변이 연구에서 종 내, 종 간에서 유전자 발현의 변화 생성에 자연선택의 역할을 추론하는데 널리 사용되었다.

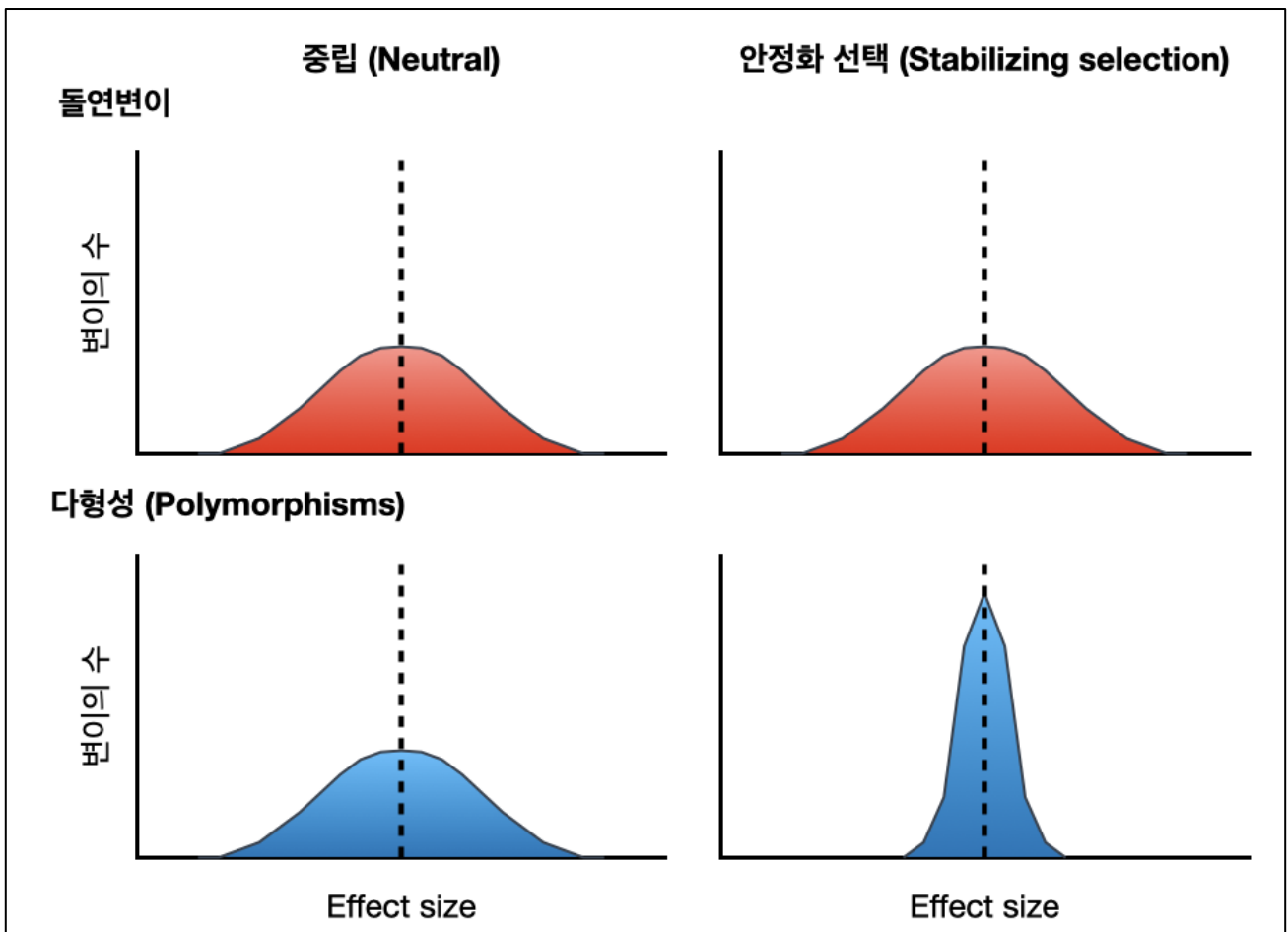


그림 3. 돌연변이의 영향으로부터 추론 될 수 있는 자연선택의 작용.

유전자 발현이 중립적으로 진화할 경우(왼쪽 패널) 무작위 샘플링에 의한 돌연변이의 영향과 다형성의 영향 사이에는 차이가 없음. 하지만, 유전자 발현이 안정화 선택 하에서 진화 할 경우(오른쪽 패널) 다형성의 영향은 돌연변이의 영향보다 더 적은 분산을 가진다(모든 그림에서의 점선은 Effect size가 0인 것을 의미하고, 이는 유전자 발현이 변하지 않는 것을 의미함).

위에서 설명하였듯이, 돌연변이 축적 연구에서는 일반적으로 유전자 발현에 미치는 돌연변이의 영향을 ' $V_m$ '이라는 파라미터로 측정한다. 초파리인 *Drosophila* spp.에서의 한 연구 예시로, 각 세대에서 어떻게 유전자 발현 변화가 증가하는지 알아보기 위해 이 파라미터가 세 쌍의 *Drosophila* 종에서 mutation-drift equilibrium 하에서 진화할 것으로 예상되는 유전자 발현의 변화를 계산하는 데 활용 되었다. 이 세 종들 간의 발현량 차이를 중립적인 조건에서의 기대치와 비교한 결과, 이 차이는 중립 모델에서 예측된 값보다 현저히 낮았는데, 이를 통해 **안정화 선택(stabilizing selection)**이 이 종들 간의 유전자 발현의 변화를 낮추는데 작용을 했음을 추론 할 수 있었다. 이외의 다른 초파리에서의 연구와 280세대에 걸친 선충에서의 연구 또한, 이와 같은 결과를 보여주어, 안정화 선택이 유전체 수준에서 유전자 발현의 변화가 생기는 것을 제한하고 있다는 것은 정설로 자리매김했다.

돌연변이 영향의 유전자-특이적 분포를 연구하는 것은 이에 대한 분석들을 개선하기 시작하여 유전자 발현의 변화에서 선택(selection)의 영향에 대해 보다 구체적인 질문을 제기할 수 있도록 해 준다. 예를 들어, 인간과 쥐의 인핸서의 돌연변이의 영향은 다른 설치류와 영장류 계통의 발산부위(divergent sites)의 영향을 예측하는데 활용되었고, 각 인핸서에 다른 유형의 선택이 작용한다는 증거를 제공했다. 또한, 돌연변이의 영향과 야생종에서 분석된 다형성 사이의 직접적인 비교연구도 효모의 TDH3 유전자에서 수행되어 안정화 선택이 유전자 발현 변화에 작용하고 있다는 증거들을 제시 하였다.

## 6. 미래의 방향

분자생물학은 새로운 돌연변이가 유전자 발현의 변화를 어떻게 만들어내는지 설명한다. 반면에 집단유전학(population genetics)은 이러한 돌연변이가 일단 발생하면 진화적 발산(evolutionary divergence)에 어떻게 기여하는지를 설명한다. 이 두 가지 관점은 우리가 왜 자연적으로 발생하는 유전자 발현의 변화를 확인 수 있는것인지 이해하기 위해서 항상 함께 고려되어야 한다. 그리고 앞으로는, 이 연구분야가 더 성장하기 위해서 다음 세 가지 방향이 매우 중요할 것이다.

첫 번째, 시스-, 트랜스-조절 돌연변이에 대해 이해하기 위해, 돌연변이 영향의 유전자-특이적 분포(gene-specific distribution)에 대한 연구가 더욱 필요하다. 왜냐하면, 새로운 돌연변이가 유전자 발현에 영향을 줄 것으로 예상되는 것은 유전자에 따라서 매우 다르고 시스-, 트랜스-작용 돌연변이에 따라서도 다르기 때문이다. 그러나, 우리는 이제 겨우 그 이러한 차이의 범위들을 알아가고 있는 단계이며, 진화적 다형성과 발산을 정확히 예측하기 위해 돌연변이의 어떤 영향이 가장 중요한지 아직 알지 못한다. 이제 앞으로는 포화 돌연변이유발(saturation mutagenesis)과 대규모 유전체 편집(massively parallel genome editing) 등의 새로운 연구기술이 이에 큰 도움을 줄 것이다.

두 번째 중요한 방향은 조절네트워크(regulatory network)의 특성이 어떻게 돌연변이 영향의 분포를 형성하는지 이해하는 것이다. 유전자의 발현에 대한 새로운 돌연변이의 영향은 그것이 조절네트워크의 구조나 기능에 어떻게 영향을 미치는지에 따라 결정되기 때문이다. 실제로, 초파리와 효모에서 이 조절네트워크의 구조와 함께 유전자의 발현 다형성과 발산을 비교하기 위한 많은 연구들이 있었지만, 이에 대한 일반적인 결론을 내리지는 못하여 조절네트워크의 구조와 기능에 대해 더 많은 연구가 필요하다. 또한, 이 조절네트워크는 세포 타입, 유전적 배경, 성별과 환경 등에 따라서 다르게 작용하기 때문에 연구에 더 어려움을 주고 있다. 하지만, 최근 단일세포 RNA 시퀀싱과 같은 기술의 발전이 조절네트워크가 새로운 돌연변이에 의해 어떻게 영향을 받는지에 대한 연구에 힘을 더해줄 것이다.

일단 유전자 발현에 대한 새로운 돌연변이의 영향을 알 수 있거나 예측할 수 있게 되면, 마지막 중요 방향은 이러한 돌연변이에 의해서 변화된 유전자 발현과 **피트니스(fitness)\***를 연결 짓는 것과 기존의 집단유전학에서의 체계를 활용해서 서로 다른 조절 돌연변이의 진화적 과정을 예측하는 것이다 \*[피트니스(fitness): 군집 또는 종이 주어진 환경에 적응하여 건강하게 성장하고(healthy growth) 생식하는(reproductive success) 정도를 수치화한 것으로, 유전학적으로는 주어진 유전형질이 다음 세대로 전달되는 정도를 나타냄]. 유전자 발현과 상대적 피트니스 사이의 연관성을 설명해주는 피트니스 곡선(fitness curves)은 아직 효모의 소수 유전자에서만 적용 가능할 뿐, 다른 대부분의 종에서는 알지 못해서 이를 다양한 종에서 이해할 수 있는 효과적인 방법이 반드시 필요하다.

이 세 영역에서의 도전과제를 해결하는 것은 서로 다른 분야인 분자생물학과 진화생물학을 연결해 주는데 있어 매우 중요할 것이며, 현재와 미래에 존재할 생물학 세계에 대해 이해하는 데 있어 필수적이다.

The views and opinions expressed by its writers do not necessarily reflect those of the Biological Research Information Center.

조지훈(2021). 유전자 발현의 변화를 발생시키는 분자적, 진화적 과정. BRIC View 2021-R19  
Available from <https://www.ibric.org/myboard/read.php?Board=report&id=3784> (May 27, 2021)

Email: [member@ibric.org](mailto:member@ibric.org)