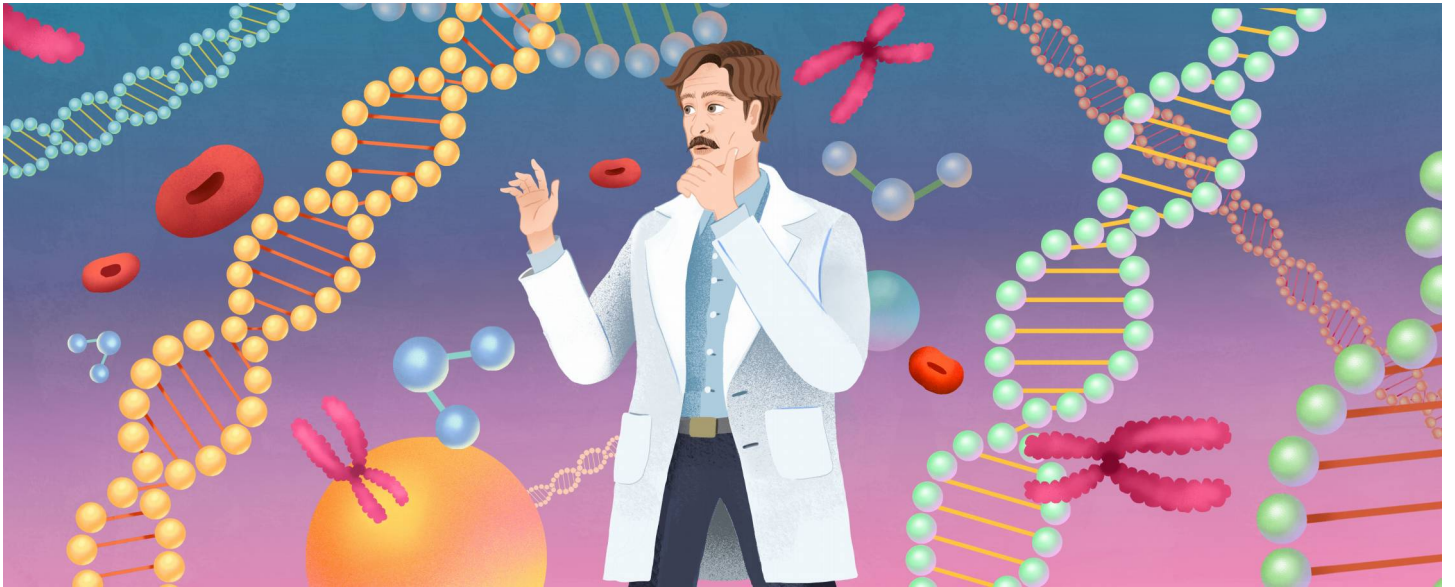


물리학자가 본 생명현상의 숨은 조력자, 정크 DNA

2018년 9월 27일

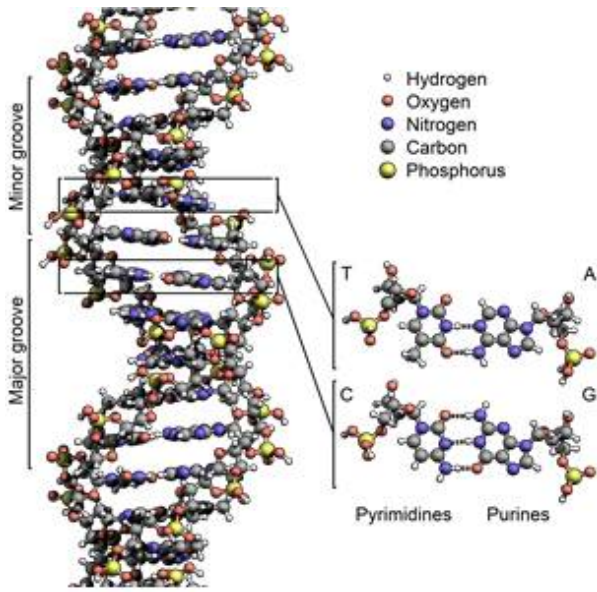
홍석철



생명현상은 대단히 신비롭다. 물리학 전공자로서 생명의 신비에 이끌려 생물물리를 연구하고 있는 필자에게는 생명현상 자체가 대단히 매혹적이다. 의도된 행동, 목적지향적 선택, 이타적 희생, 정치, 종교 행위를 가능하게 하는 의식 및 정신은 무생물계나 하등생명체에서는 볼 수 없는, 고차원적인 생명의 놀라운 특징이다. 신경과학의 발전에 힘입어 이러한 현상들에 대한 지식이 하루가 다르게 축적되고 있으나, 물리학자들이 범접하여 이해에 도달하기에는 아직 멀리 떨어져 있는 것으로 생각된다.

사실 이러한 정신적인 특징을 제쳐두고 보다 낮은 수준에서 바라보더라도 생명현상은 대단히 놀랍다. 가령 미소한 나노기계들이 에너지를 소비하면서 엔트로피의 강한 압박을 뚫고 일견 매우 질서정연하게 생명현상을 유지하고 구동해 나가는 과정을 살펴보자. 열에너지가 나름의 역할을 할 수 있을 만큼 미소한 세계에서 다종 다수의 분자들이 교묘한 균형, 즉 분자 하나하나의 구조적 평형 상태와 다종 다수의 분자들 간의 아슬아슬한 상호작용이 평형 상태를 이루며 한 방향으로 가는 듯한 질서 있는 물리적 사건을 이끌어낸다. 적절한 조건-온도, 염도 등 환경조건, 에너지 조건, 필요한 구성 인자들의 존재 유무와 같은 '사회적' 조건-이 갖춰지면 놀랍게도 질서 있는 반응이 일사천리로 자발적으로 일어난다. 어떻게 원하는, 바람직한 반응이 '저절로' 일어날까? 여기서 '저절로'는 에너지 공급 하에서 또는 전체의 엔트로피는 늘리면서 일어나는 것이어서 물리 법칙을 거스르는 것은 아니다. '저절로' 일어나는 현상에 대한 가장 좋은 답은 아마도 진화론에 바탕을 둔, (최적까지는 아닌) 적합화 과정의 산물로 보는 것이다. 각설하고 생명현상은 분자의 상호작용이든, 거대한 생태계의 거동이든 적응의 산물이라는 것은 확실해 보인다.

여기서 주제를 좁혀 구체적인 사례에 대해서 얘기해보고자 한다. 세포 안에 있는 유전 물질, 즉 DNA의 사용과 관리에 대해서 얘기하고자 한다. 어떻게 이러한 생명체를 이루게 되었는지, 어떻게 진화를 거쳐 현재와 같은 생명체가 되었는지, 신경이나 사고와 기억이 무엇인지와 같은 것들은 현재 물리학이 설명하기 어려운 것이니 그냥 무시하기로 한다. 이미 많은 생물학자, 생화학자, 생물리학자들의 노력으로, 적어도 몇몇 간단한 모델 시스템에서는, 유전자가 DNA로부터 시작하여 RNA를 거쳐 단백질로 발현된다는 사실과 그에 대한 기계적이고 물리적인 과정이 잘 밝혀져 있다. 그런데 한 발짝 멀리 떨어져서 세포 내 전체 DNA를 들여다보면 많은 중요한 문제가 대부분 해결되지 않은 채 남아 있음을 알 수 있다. 그래서 물리학자의 관점에서 물리적으로 중요한 문제에 대해서 얘기해보고자 한다.



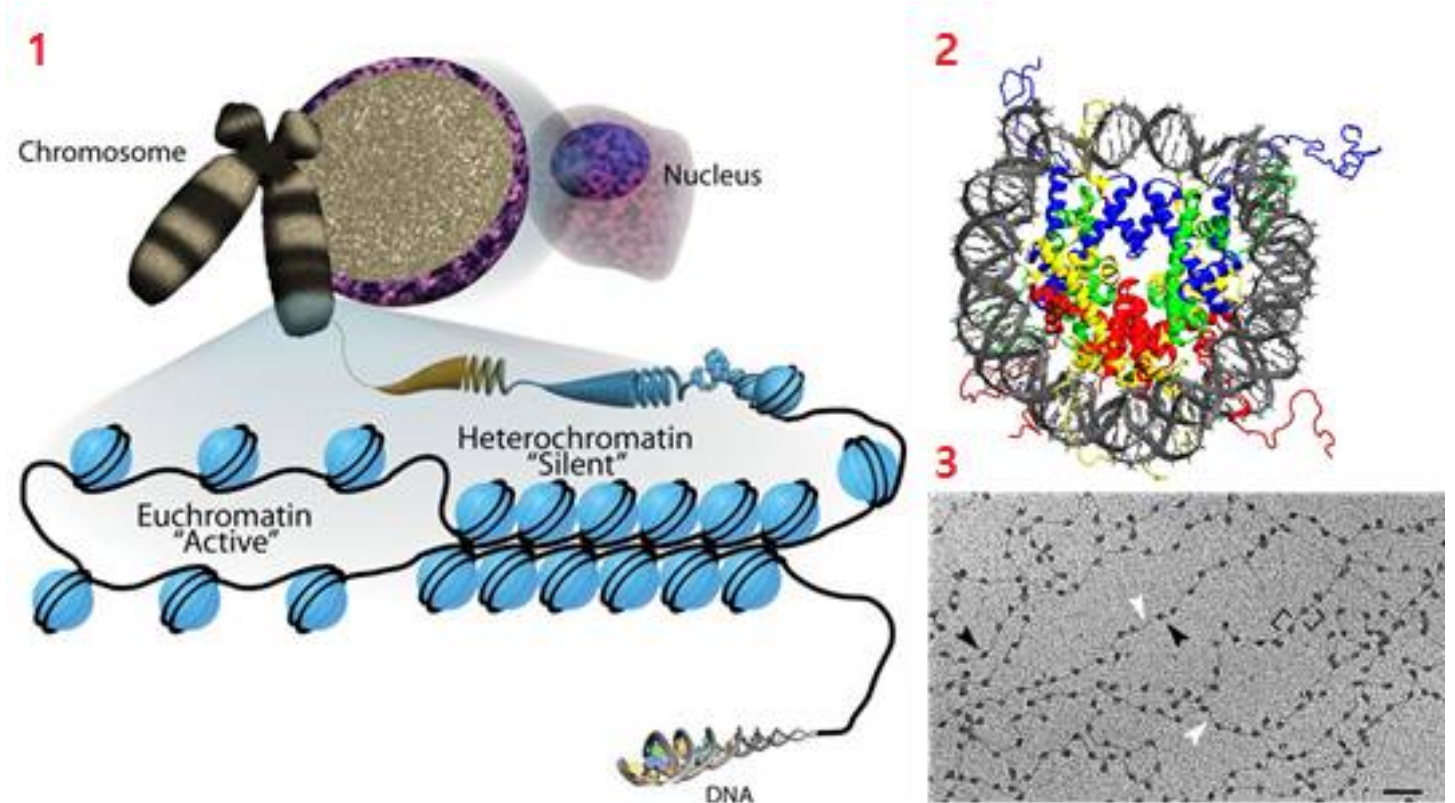
DNA의 구조

인간은 약 100조 개의 세포로 이루어진 거대한 생명체이다. 줄기세포를 제외한 다수 다종의 세포들은 전문적인 세포들로 분화되어 각자 기능을 수행하고 긴밀히 소통하며 하나의 생명체의 생존과 발전에 이바지한다. 겉으로 보기에 매우 다른 세포들이지만 이들은 모두 같은 DNA 분자를 가지고 있다. DNA는 deoxy-ribo-nucleic acid의 약자로 핵산의 일종이며 왓슨-크릭이 결정한 바와 같이 오른나사 방향의 이중나선 모양을 갖는다. 각 사슬은 네 종류(A: 아데닌, G: 구아닌, T: 티민, C: 시토신)의 염기- 이들 모두 아민^{amine}(-NH₃⁺)기가 있어 염기성을 띠 -로 구성된 선형 폴리머이며, 아데닌은 티민과, 구아닌은 시토신과 결합하는 상보성 규칙에 따라 두 사슬이 지퍼가 잠기듯 꽉 채워진 형태를 갖는다. 이들 염기가 의미유전자를 전달하는 알파벳이라고 할 수 있다. 전기 문명이 ON-OFF의 두 가지 신호로 정보를 저장하는 데 반해 유전자 정보는 4개의 알파벳을 통해 정보를 저장하고 있다. DNA의 물성 및 분자적 특질을 놓고도 재미있는 얘기를 할 수 있지만 여기서는 DNA를 보다 큰 규모에서 실처럼 생각하고 이들이 조직화되는 방식을 생각해볼도록 하자. 앞서 얘기한 바와 같이 모든 세포는 같은 DNA를 가지고 있다. 염기쌍의 수가 무려 30억 개이다. 따라서 46개의 염색체로 나뉜 전체 DNA를 모두 이어붙이면 그 길이가 약 1m에 달한다. 그러니 한 개체의 모든 세포에 있는 DNA를 이어붙이면 10¹⁴m가 될 것이다. 지구와 태양 사이의 거리가 1억 5천만km(1.5×10¹¹m)임을 생각하면 실로 엄청난 길이라고 할 수 있다. 만약 모든 생존 인류 60억의 DNA를 붙이면 6×10²³m나 된다.

1m나 되는 DNA가 세포 속, 그것도 크기가 수십 마이크로론에 불과한 핵 속에 구겨 놓여져 있다. 하지만 DNA는 뜨개질을 하다가 만 실타래가 아니다. DNA는 유전정보가 들어있는 가장 중요한 물질, 분자이며, 세포가 자주 펴봐야 하는 참고서이자 설계도이다. 어질러진 서랍에서 원하는 물건을 찾는 일이나 헝클어진 실타래를 가지런히 펴는 일이 얼마나

성가신 줄 안다면, 기다란 실타래와도 같은 DNA가 세포핵 내부에 아무렇게나 보관되어 있을 때 얼마나 심각한 문제가 발생할지 상상할 수 있을 것이다. 실제로 DNA는 히스톤이라 불리는 단백질 4종(H2A, H2B, H3, H4), 각 한 쌍, 8개의 조각이 구슬처럼 조립된 구조 위에 1.7 바퀴 감겨 있고(이를 염주구조beads-on-a-string라고 한다), 이 염주구조가 다시 규칙적으로 돌돌 말려 있다고 한다.[1] 그리고 이들 굵은 다발이 다시 접혀 말려 있다고 한다. 히스톤에 감겨서 생긴 크로마틴 구조도 마찬가지로지만, 상위 구조에 대한 우리의 지식은 매우 제한적이다.

이렇게 긴 DNA가 있고 이것이 유전정보를 가지고 있다는 것을 알고 있을 때, 과학자와 인류가 궁금했던 것은 도대체 어떤 단백질들이 여기에 저장되어 있고, 어디에 있는지에 대한 것이었다. 2003년에 인간 게놈 프로젝트가 첫 결과를 발표했을 때, 불과 1.5%의 DNA만이 유전자를 담고 있다는 사실에 대부분의 사람들은 매우 놀랐다. 왜냐하면 반대로 말하자면 98.5%의 DNA가 단백질을 만드는 어떤 정보도 담고 있지 않다는 뜻이기 때문이다.[2] 인간 단백질유전자는 겨우 3만 개 미만인 것으로 밝혀졌는데, 이는 약 5백만 개 염기쌍을 갖는 대장균이 만드는 2천여 개와 비교할 때 겨우 10배에 불과한 것이었다.



1. DNA의 단계적 집적 2. 146개 염기쌍이 여덟 개의 히스톤 단백질을 1.7바퀴 감고 있는 모습(적: H2B 황: H2A, 녹: H4, 청: H3) 3. 염주구조의 전자현미경 사진

나머지 부분은 무엇인가? 사람들은 이 유전자를 담고 있지 않은 DNA 부분을 쓰레기, 즉 정크 DNA^{junk DNA}라고 부르기 시작하였다.[3] 필자의 포스닥 지도교수였던 코자렐리 교수는 이를 가리켜 유전자 주변에 고비사막이 펼쳐져 있는 것 같다고 하였다. 반드시 없어질 필요는 없겠으나 불필요한 것은 쉽게 사라질 수 있는 진화의 엄청난 압력에도 불구하고 왜 98.5%의 DNA를 끌고 다니는가? 이에 답하기 위해서 인간 게놈 과제의 후속 과제로서 ENCODE^{ENCyclopedia Of DNA Elements}라는 연구 컨소시엄이 형성돼서 RNA로 전사가 되는 DNA의 영역과 비율을 밝혀내고자 하였다. 2007년 파일럿 과제의 결과와, 2010년에 공표된 1단계 과제의 결과에서, 약 80%의 DNA가 일단 RNA로 전사가 된다는 점이

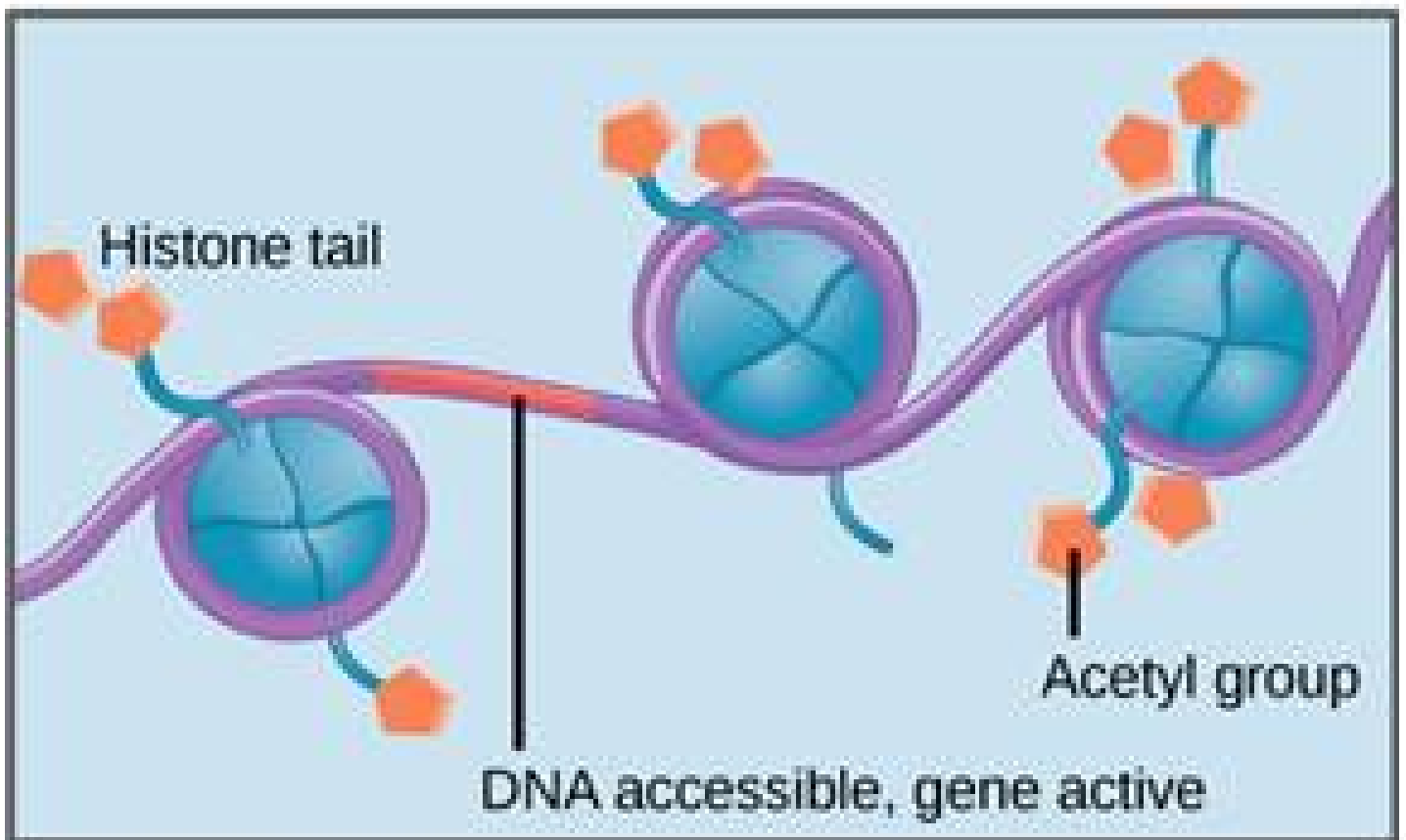
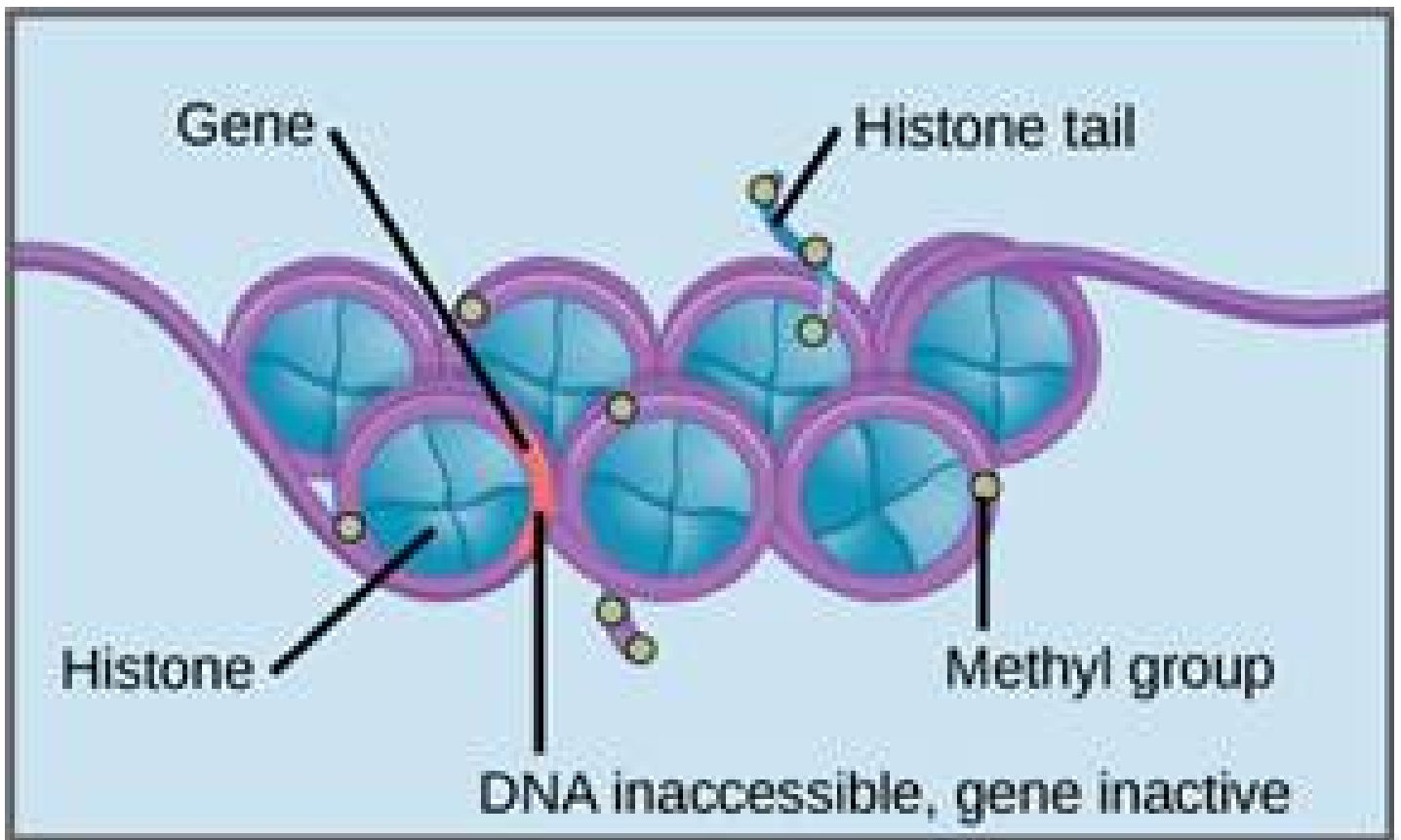
밝혀지면서 또 한 번 이 분야 연구자들을 놀라게 하였고, 수많은 논란과 논쟁을 불러일으켰다.[4] 그럼 DNA는 왜 단백질로 번역되지 않는 다량의 RNA를 일단 만드는가? 이들은 쓸모없는 부산물에 지나지 않는가? 아니면 어떤 중요한 기능이 있는가? 현재 연구자들은 정크 DNA의 기능(또는 무기능)을 밝히는 데 노력하고 있다.

한 가지 흥미로운 것은 유전자를 담고 있는 DNA의 길이나 유전자 개수는 우리가 흔히 생각하는 생명체의 고등한 정도에 따라 크게 다르지 않지만, 정크 DNA의 비중은 하등생명과 고등생명체 사이에 상당한 차이를 보인다는 점이다. [5] 이로부터 정크 DNA(또는 여기서 비롯된 RNA)는 고등생명체에서 필요한 조절과 제어에 관여할 것이라고 추측할 수 있다. 정크 DNA의 기능을 구조적 역할과 섬세한 제어, 조절로 보는 것이 상당 부분 사실에 부합할 것이라고 생각된다.

그럼 그 역할은 구체적으로 무엇일까? 프로모터(promoter), 인핸서(enhancer)라고 불리는 DNA 지역은 아미노산으로 '번역'되지 않으나(단백질의 일부가 되지 않음) 단백질 발현의 중요한 제어부 역할을 한다는 점은, 이미 생화학이나 분자생물학 교과서에 수록된 잘 알려진 사실이다. 중심체(centromere, 세포 분열기에 염색체를 딸세포로 각각 가져가기 위해서 방추사라는 단백질 튜브 빗줄이 와서 붙는 부분, 텔로미어(telomere, 염색체 말단부로 줄기세포나 암세포에서는 그 길이가 유지되는 부분)라고 불리는 염색체의 부분은 유전자를 담는 대신 구조적 역할을 갖는 DNA이다. 하지만 이것이 전부 아닐 것이다.

세포마다 발현하는 단백질이 달라야 하는 문제를 생각해보자. 아마 단세포에서 출발해 군집을 이루다 역할이 분화되며 거대한 개체가 된 다세포 생명체에게 이러한 상황은 필연적일 수 있다. 하지만 어떻게 세포마다 다른 유전자를 발현하고 딸세포에 특정 발현 패턴을 전달하는 것은 중요하고도 어려운 문제이다.

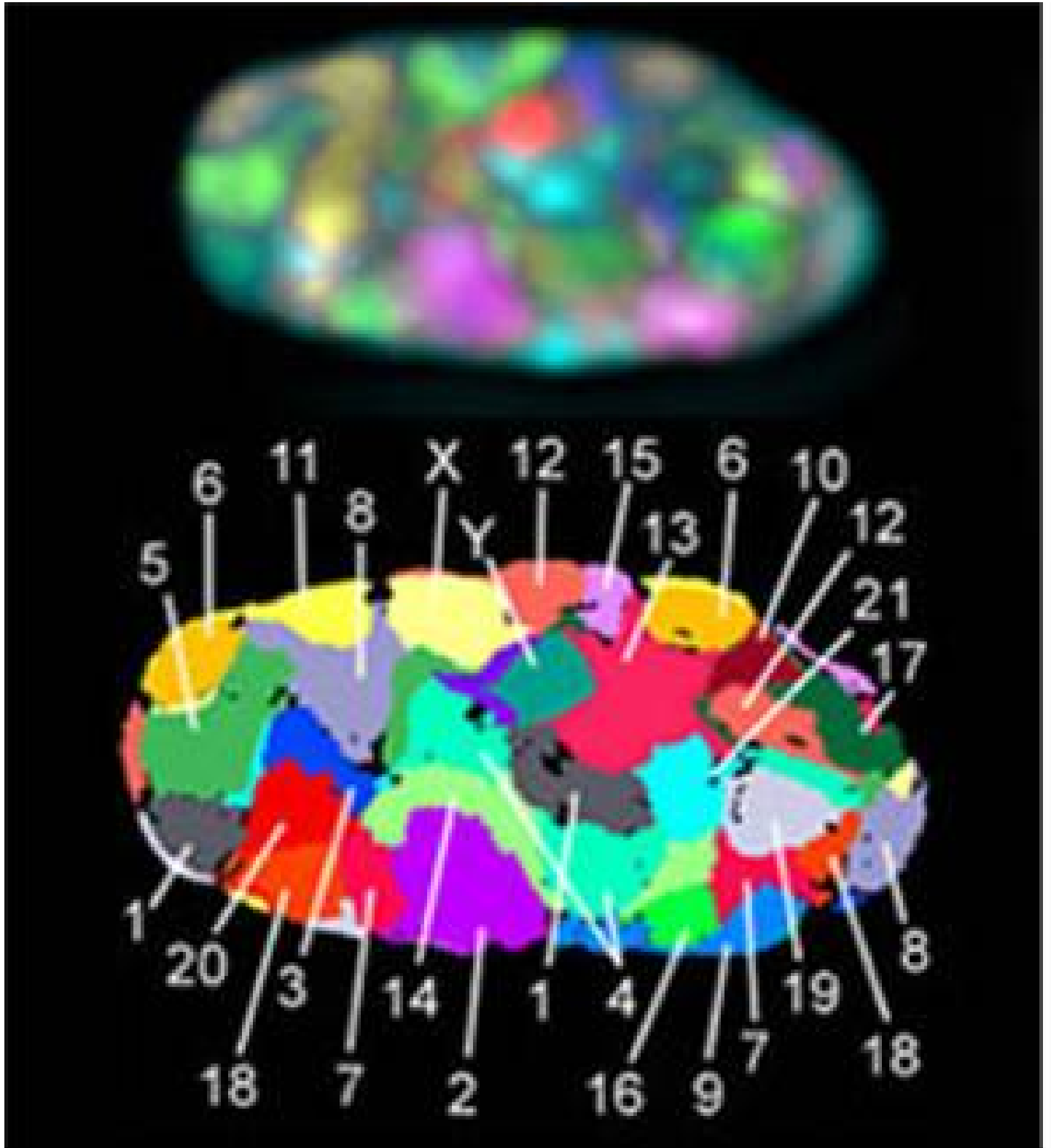
최근 대두되는 생물 연구 분야 중, DNA의 염기 중 하나인 시토신에 메틸기(-CH₃)가 추가됐을 때 크로마틴의 응축이 유도되고 이를 통해 유전자 발현이 억제되는 현상을 연구하는 후성유전학(epi-genetics, 에피지네틱스)이라는 분야가 있다.[6] 즉 세포마다 그 특징이 다른 이유는 이들이 발현하는 단백질의 종류가 달라서인데, 그 이유는 세포마다 각 유전자의 후성유전적인 처리가 달라서 켜지고 꺼지는 유전자가 다르기 때문이라는 것이다. 이 원리는 상당히 보편적으로 받아들여지고 있다. 여기서 바로 추측할 수 있는 것은 돌연변이 없이 DNA 서열이 멀쩡해도 후성유전적 처리의 오류가 병으로 이어질 수 있다는 것이다. 이에 대한 증거나 사례가 이미 많이 알려져 있고, 지금도 계속 밝혀지고 있다. 이미 암, 성장 발육, 대사질환 등이 후성유전 현상과 매우 긴밀함이 알려져 있다.[6]



DNA의 메틸화 정도와 히스톤의 표지상태에 따라 유전자의 활성 상태가 달라진다.

그럼 어떻게 길고도 긴 DNA가 세포 특징에 맞게 후성처리되고, 관련 유전자들이 효율적으로 조직되어 발현되는가? (최적화까지는 아니더라도 유전정보를 긴 분자 위에서 원활히 접근하여 활용하려면 '효율적'이어야만 할 것 같다.) 최근 염색체의 공간적 조직화에 대한 연구에 따르면, 46개의 염색체는 그 개체마다 일정 영역을 차지하고 있고[7] 각각의 염색체도 완전히 무질서하게 헝클어져 있다기보다 공간적 질서를 일부 이루고 있다.[8] 또한 유전자 발현도 유전자 개수만큼 별개의 장소에서 일어나기보다는 세포핵 내부의 특정 지역에서 집중적으로 일어나는 것처럼 보인다는 게 밝

혀졌다.[9] 4C, 5C, Hi-C 등으로 불리는 게놈 차원 DNA 분자들의 공간적 근접성에 대한 연구들은 세포핵 내부에서 염색체가 어떻게 공간적으로 분포, 위치하는가를 보여준다.[8][10] 세포핵 내부에서 DNA는 구불구불한 1차원 실이 접힌 상태로 존재하기 때문에, 공간적으로 근접한 두 지점이 서로 다른 염색체에 속해 있을 수 있고, 동일 염색체 사슬 상에서 멀리 떨어져 있는 경우에도 실제 공간에서는 근접해 있을 수 있다. 혹시 세포 특이적으로 발현이 필요한 유전자들이 선택적으로 후성유전적으로 활성화되고, 효과적인 발현을 위해서 공간적으로 근접하게 되어 발현되는 것은 아닐까? 유전자 발현에 필요한 분자들의 확산^{random walk}에 의해 접근하여 결합한 후 반응이 시작된다면 너무나 요행을 바라는 것이 아닐까 생각된다. 혹시 순전히 통계물리적, 열역학적 효과에 의해 이미 효과적인 유전자 발현에 적합하게 큰 틀에서 염색체가 조직화되어 있지는 않을까?



염색체가 영역을 형성하고 있다.

왓슨-크릭이 주목하였듯이 DNA는 유전 정보를 담고 있고 그 복제의 원리는 분자의 상호 인식에 대한 특별한 정보를 담은 구조적 상보성에서 비롯된다. 고등생명체에 차고 넘치는 정크 RNA^{단백질로 발현되지는 않는 정크 DNA에서 유래한 RNA}들은 모두 특정 서열을 담고 있다. 이들은 유전체의 다른 부분들과 유사한 서열을 갖는 경우가 많아서 DNA 부분들을 잘 인식하여 서로 다른 부분의 DNA를 연결하고 다리를 놓을 수 있고, 동시에 아미노산과도 잘 결합한다. 이들이 만들어내는 구조가 유전자 발현과 억제 조절에 필요한 단백질 인자들에 대한 교묘한 교통정리에 관여하는지도 모른다. 여기서 부터는 ‘~ 일 수도 있다’가 너무 많다. 생명체는 굳어지기 없이 가장 효율적으로 작동할 필요는 없지만 지나치게 운에 맡겨 돌아가지는 않을 것 같다. 하나하나의 과정은 확률적, 통계적인 요소가 있어도 결국 회로가 틀림없이 돌아가기 위해서 2차, 3차의 안전장치(대안의 반응 경로)를 통해서라도 확고한 결과가 나오게 할 것 같다. 반응, 사건이 웬만큼 잘 일어나기 위한 확실한 단계들이 준비되어 있을 것이고 그 분자적 도구가 정크 DNA에서 비롯된 것이 아닌가 생각된다. 게놈에서 쓸데없어 보였던 것, 존재감이 드러나지 않아 게놈의 암흑물질이라 불릴 만한 정크 DNA가 생명현상의 분자적 원리를 제공하는 빛과 같은 존재가 될 수 있다는 것이다. 무용으로 보였던 것이 참으로 유용한 경우라고 할 수 있겠다.

아직 이러한 복잡한 회로의 핵심을 담는 기본적인 모델이나 물리적 이론이 없다. 물리학자들은 일반적인 것, 전체를 아우르는 것, 직관적인 것을 좋아한다. 생명이 영위되는 데 새롭고 대단한 원리들이 있을 것으로 생각된다. 물리가 큰 밑그림을 그리면 화학 반응이 채색을 하고 생물학이 디테일을 완성하는 것이 아닐까? 여기에 물리학자들의 이론적, 실험적 노력이 필요하고 이로써 생명현상이 원리적으로 이해될 수 있을 것으로 생각된다.

그럼에도 불구하고 여전히 질문은 끊이지 않는다. 그럼 왜 정크 DNA에서 비롯된 무수한 인자들은 게놈 속에 존재하는가? 아직은, 아마도 오랜 진화의 과정에서 쓸모없는 물건의 유용한 사용처를 저절로 터득했다고 궁색한 변명이 떠오를 수밖에 없다.

참고문헌

1. B. Alberts, et al, "The Molecular Biology of the Cell" 4th ed. Garland Science, pp. 207-216.
2. Ibid, Table I, p. 202.
3. Ibid, p. 200. 한편, "Junk DNA" by N. Carey는 이 분야에 대한 훌륭한 교양서이다.
4. The ENCODE project consortium, "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project," Nature, 447, pp. 799-816 (2007); M. B. Gerstein, et al, "Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project," Science, 330, 1775-1787 (2010); The modENCODE consortium, "Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE," Science, 330, 1787-1797, (2010).
5. J. S. Mattick, "A New Paradigm for Developmental Biology," J. Ex. Med., 210, pp.1526-1547, (2007).
6. C. D. Allis, et al., "The Molecular Hallmarks of Epigenetic Control," Nat. Rev. Genet., 17, pp. 487-500, (2016).
7. A. Bolzer, et al, "Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes," PLoS Biol. 3, pp. 826-842, (2005).

8. E. Liberman-Aiden, et al, "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome," *Science*, 326, pp. 289-293, (2009).
9. D. A. Jackson, et al, "Visualization of focal sites of transcription within humannuclei," *EMBO J*, 12, pp. 1059-1065, (1993).
10. J.-M. Belton, et al, "Hi-C: A comprehensive technique to capture the conformation of genomes," *Methods*, 58, pp 268-276, (2012).