2012. 1. 30 CS4HS

생물 정보학및 암 정보 의학

김선

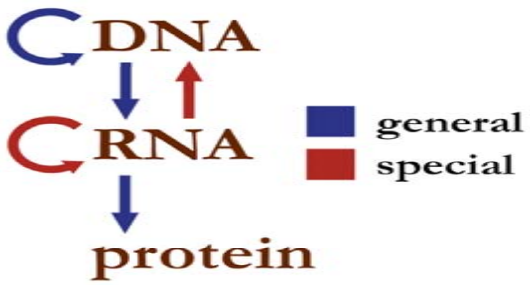
서울대학교
컴퓨터 공학부
생물정보연구소
생물정보학 협동과정

Outline

- 생물정보학
- 맞춤의학과 생물정보학
- 유전체학, 후생 유전체학을 이용한 암연구와 맞춤의학

PART1. 생물정보학

Central Dogma in Biology



http://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology

가장 중요한 질문은 DNA, RNA, and Proteins 을 전체 세포에서 우리가 측정 할수 있느냐 하는 것.

DNA Sequencing

The Nobel Prize in Chemistry 1980
Paul Berg, Walter Gilbert, Frederick Sanger

- The Nobel Prize in Chemistry 1980
- Nobel Prize Award Ceremony
- Paul Berg
- Walter Gilbert
- Frederick Sanger

Paul Berg | Walter Gilbert | Frederick Sanger

The Nobel Prize in Chemistry 1980 was divided, one half awarded to Paul Berg *"for his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA"*, the other half jointly to Walter Gilbert and Frederick Sanger *"for their contributions concerning the determination of base sequences in nucleic acids"*.

The 1st Whole Genome Sequencing

Science 28 July 1995:
Vol. 269 no. 5223 pp. 496-512
DOI: 10.1126/science.7542800

‹ Prev | Table of Contents | Next ›

Whole-genome random sequencing and assembly of Haemophilus influenzae Rd

RD Fleischmann, MD Adams, O White, RA Clayton, EF Kirkness, AR Kerlavage, CJ Bult, JF Tomb, BA Dougherty, JM Merrick and et al.

± Author Affiliations

ABSTRACT

An approach for genome analysis based on sequencing and assembly of unselected pieces of DNA from the whole chromosome has been applied to obtain the complete nucleotide sequence (1,830,137 base pairs) of the genome from the bacterium Haemophilus influenzae Rd. This approach eliminates the need for initial mapping efforts and is therefore applicable to the vast array of microbial species for which genome maps are unavailable. The H. influenzae Rd genome sequence (Genome Sequence DataBase accession number L42023) represents the only complete genome sequence from a free-living organism.

Human Genome Sequencing (2001)

articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

*A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

The Sequence of the Human Genome

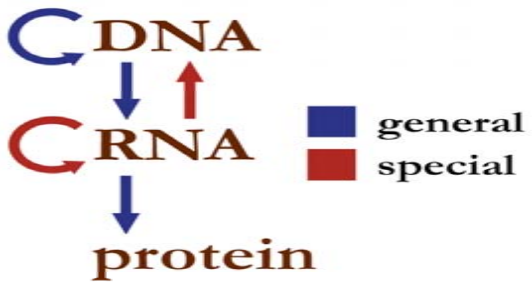
J. Craig Venter,[1*] Mark D. Adams,[1] Eugene W. Myers,[1] Peter W. Li,[1] Richard J. Mural,[1] Granger G. Sutton,[1] Hamilton O. Smith,[1] Mark Yandell,[1] Cheryl A. Evans,[1] Robert A. Holt,[1] Jeannine D. Gocayne,[1] Peter Amanatides,[1] Richard M. Ballew,[1] Daniel H. Huson,[1] Jennifer Russo Wortman,[1] Qing Zhang,[1] Chinnappa D. Kodira,[1] Xiangqun H. Zheng,[1] Lin Chen,[1] Marian Skupski,[1] Gangadharan Subramanian,[1] Paul D. Thomas,[1] Jinghui Zhang,[1] George L. Gabor Miklos,[2] Catherine Nelson,[3] Samuel Broder,[1] Andrew G. Clark,[4] Joe Nadeau,[5] Victor A. McKusick,[6] Norton Zinder,[7] Arnold J. Levine,[7] Richard J. Roberts,[8] Mel Simon,[9] Carolyn Slayman,[10] Michael Hunkapiller,[11] Randall Bolanos,[1] Arthur Delcher,[1] Ian Dew,[1] Daniel Fasulo,[1] Michael Flanigan,[1] Liliana Florea,[1] Aaron Halpern,[1] Sridhar Hannenhalli,[1] Saul Kravitz,[1] Samuel Levy,[1]

Bioinformatics

- Whole genome sequencing은 많은 양의 데이터를 만들었으며, 짧은 DNA 단편을 연결 할 수 있도록 하는 정교한 알고리즘을 필요로 하게 됨.
- whole genome sequencing이 시작되면서 "Bioinformatics"라는 용어를 만들어 짐.
- 생물정보는 게놈프로젝트의 "설계" 단계에서 필요에 의해 시작된 학문. (이전에도 수학, 컴퓨터를 이용한 생물 연구는 많이 되어 있었다 (computational or mathematical biology). 이에 대한 차이는 나중에 논의함).

We have sequences of genomes. Now what?

DNA to RNA

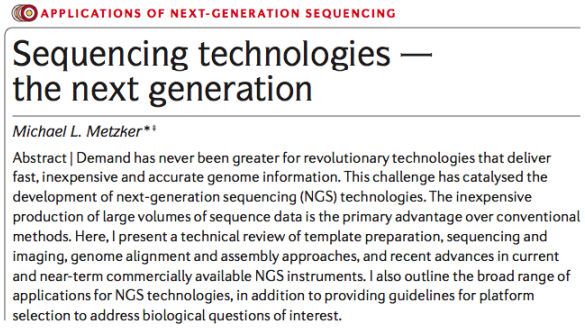


http://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology

Need for Very High Throughput Sequencing Technology

- 다양한 조건에서 RNA 측정하기 위해서는 여러 번 sequencing을 해야 함.
- 많은 사람(유전체 집단)의 서열을 필요로 함.
- 인간게놈프로젝트는 과학사에서 2번째로 많은 비용이 들어간 프로젝트임.
- 이렇게 많은 비용이 들어가는 sequencing을 여러 번 할 수 있을까?

Revolution Again


APPLICATIONS OF NEXT-GENERATION SEQUENCING

Sequencing technologies — the next generation

Michael L. Metzker[*‡]

Abstract | Demand has never been greater for revolutionary technologies that deliver fast, inexpensive and accurate genome information. This challenge has catalysed the development of next-generation sequencing (NGS) technologies. The inexpensive production of large volumes of sequence data is the primary advantage over conventional methods. Here, I present a technical review of template preparation, sequencing and imaging, genome alignment and assembly approaches, and recent advances in current and near-term commercially available NGS instruments. I also outline the broad range of applications for NGS technologies, in addition to providing guidelines for platform selection to address biological questions of interest.

Bioinformatics vs. Computational Biology

Computational Biology — 1960's

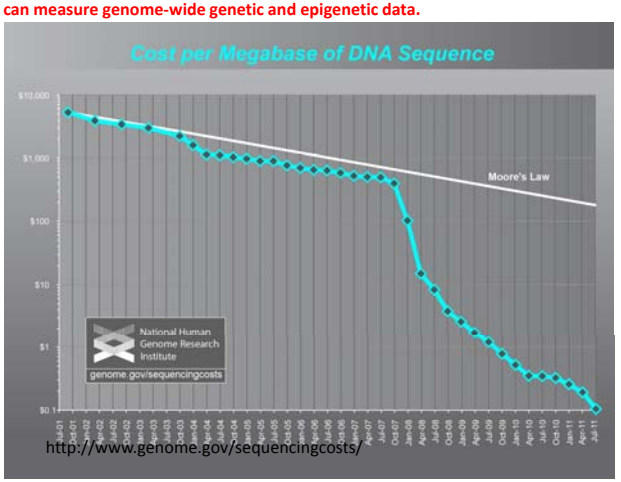
↓

DATA ! **Bioinformatics** — 1990's

↓

Bioinformatics & Computational Biology — 2010's

NGS can measure genome-wide genetic and epigenetic data.

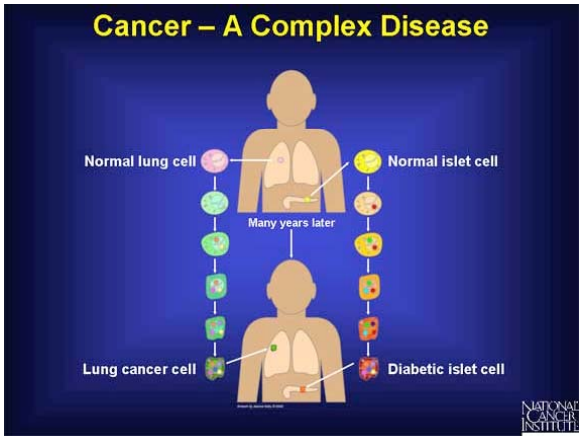


http://www.genome.gov/sequencingcosts/

Availability of Data and Bioinformatics

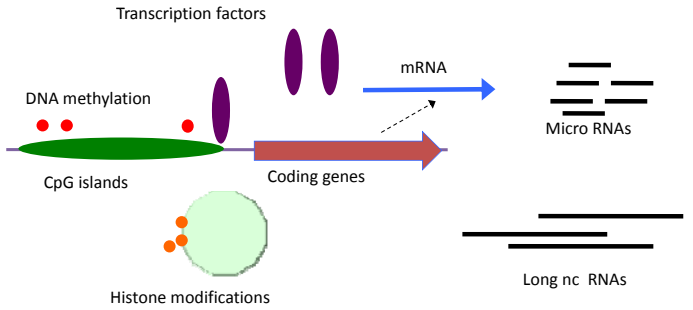
- 차세대 또는 3세대 시퀀싱 기술은 세포 내부의 메커니즘 데이터를 측정할 수 있음.
- 20년 이상 개발되어 온 여러 computational bioinformatics 방법들은 세포 내부의 데이터를 분석하는데 사용될 수 있음.

맞춤의학



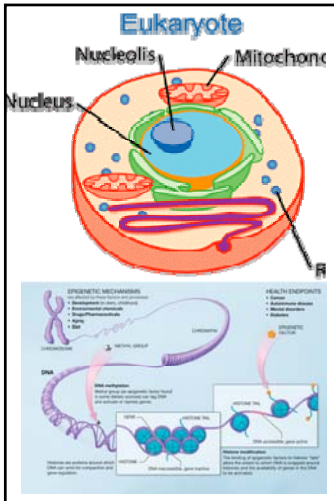
http://www.cancer.gov/cancertopics/understandingcancer/geneticvariation/page2

Genetic and Epigenetic Elements

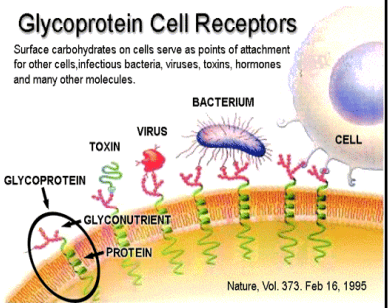


Data Measurement from Cell Surface to DNA

- Is a gene there?
  - → Genome sequencing
- Is the gene disease susceptible?
  - → SNP, GWAS
- Is the gene active?
  - → Epigenomics
- Are proteins made?
  - → Proteomics
- Are proteins functional (or mal-functional)?
  - → PTMs: glycomics and glycoproteomics



Data Can be Measured!

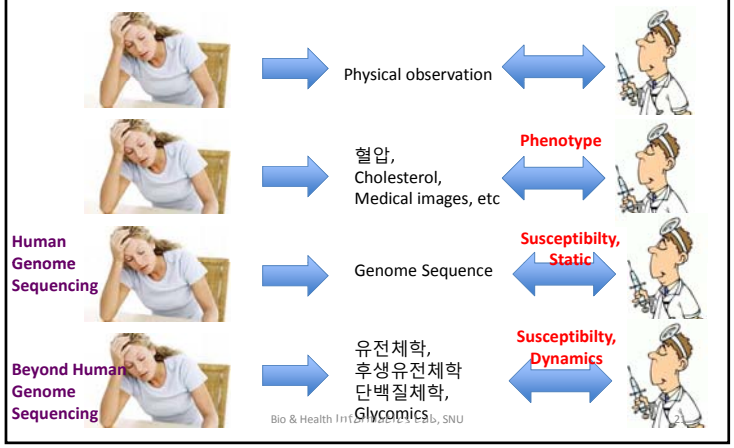


Mass Spectrometry

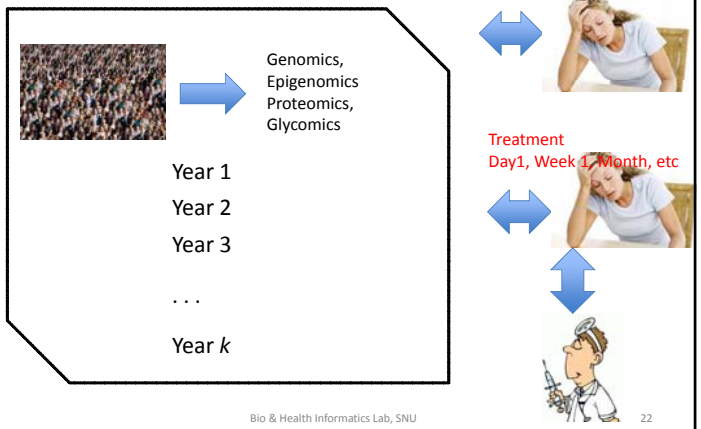
High throughput sequencing technology

Glycan microarray

What Is Happening?



Comparing Patients to DB



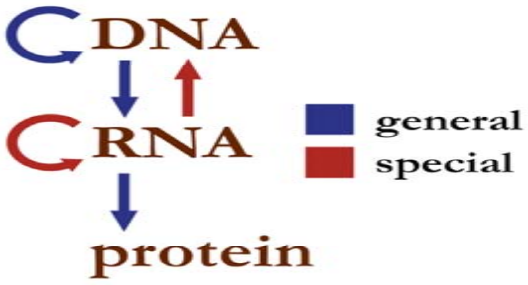
생물정보학 협동과정

http://ipbi.snu.ac.kr

서울대 생물정보 연구소

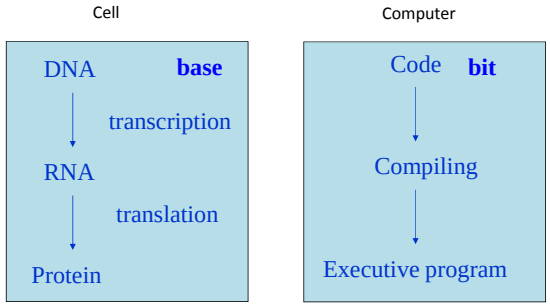
유전학 (genomics) 개요

Central Dogma in Biology



http://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology

Central dogma



Chromosome(염색체)

DNA들의 집합체

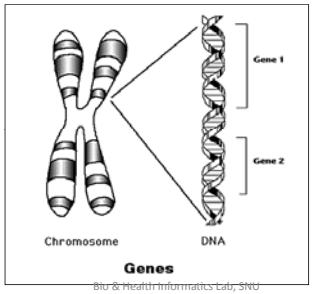
A C G G C A G C G A G C G A C G G C G A G G G

GENE (유전자)

특정 Protein이나 RNA를 encoding 하는 염색체 상의 서열 집합

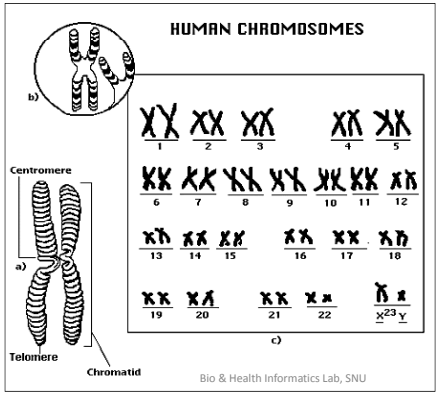
Gene

Chromosome



Genome

개체를 대표하는 chromosome들의 합



Genome Variation

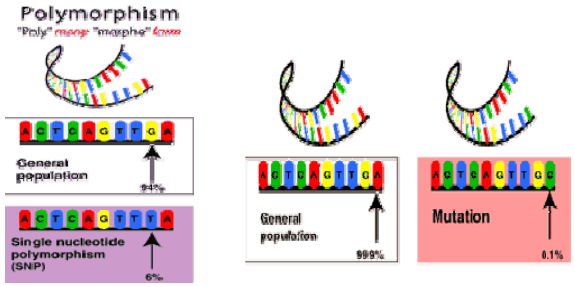
- Genetic variations (SNP, single nucleotide polymorphism)
- Gene fusion
- Alternative splicing
- Genome re-arrangement
- Copy number variations

Genetic Variation(유전자 변이)

Genetic Variations

- 유전자 변이는 사람의 46개 염색체 각각에서 나타날 수 있지만, 모든 염색체에서 고르게 나타나는 것은 아님.
- 유전자 변이는 돌연변이와 다형성(polymorphisms)을 포함
- Human genome variation의 90%가 단일염기다형성(SNPs)의 형태로 나타남

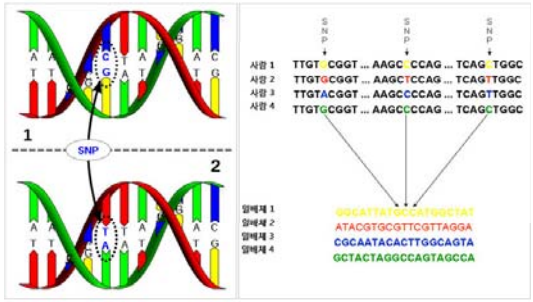
GENOME VARIATIONS



GENOME VARIATIONS

- 단일염기다형성(single nucleotide polymorphisms – SNPs)
  - DNA 염기서열에서 하나의 염기서열(A,T,G,C)의 차이를 보이는 유전적 변이
- 대략 1,000개의 염기마다 1개 꼴로 나타남
  - 전체 DNA의 **0.1%**
- SNP는 **질병과 관련된 유전자 연구**, 의약관련 연구(**개인 맞춤 의약**)의 매우 중요한 도구
  - 암, 심장병, 정신병등 다양한 질병과 관련
  - 특정약물에 대한 개개의 반응성 파악 및 최적의 약물 개발등

GENOME VARIATIONS



http://koreagenome.kobic.re.kr/sub_4.html

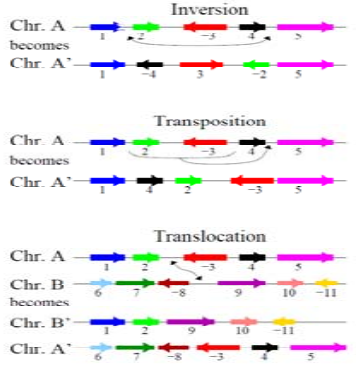
dbSNP

- dbSNP 는 생명체에서 연구되어진 단일염기다형성과 관련된 모든 자료를 저장, 관리하는 데이터베이스
- dbSNP는 임상적으로 의미 있는 인간의 변이 뿐만 아니라 양성 polymorphisms도 포함하며, 연구자들로부터 받은 자료들을 모아 저장하기도 함.
- 다형성의 종류와 대립유전자의 정보 제공
- http://www.ncbi.nlm.nih.gov/projects/SNP/
- *Build 135: in* 1000 genomes,
  - submitted SNP = 57,911,353
  - reference SNP (unique SNP) =39,484,957

Genome re-arrangement

Genome Rearrangement

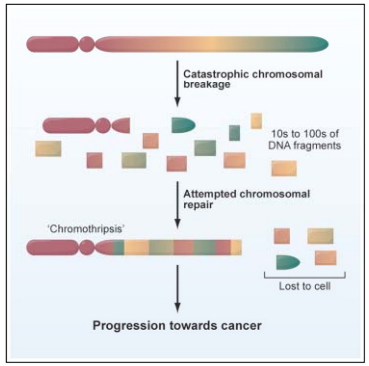
genome rearrangements의 3가지 유형



http://lacim.uqam.ca/~chauve/Enseignement/INF7440/H05/BASE/ICCS-2001.pdf

Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development

- 2%~3%의 암은 특정 유전체 영역에서 10–100s의 재배열을 보임.
- 모든 종양에서 발견됨
  - 특히, bone cancers(최대 25%)
- 암을 유발할 수 있는 genomic lesions을 만들기도 함.



Cell, Volume 144, Issue 1, 27-40, 7 January 2011

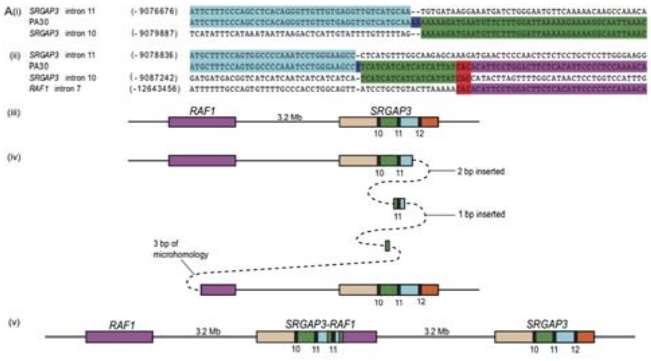
Gene Fusion

Gene Fusion

- 서로 다른 두 개의 유전자가 결합한 것
  - Translocation, interstitial deletion 또는 chromosomal inversion등에 의해 발생할 수 있음.
- Fusion gene은 oncogenes이기도 함.
  - 대부분의 fusion genes은 혈액암, 육종(sarcomas), 전립선 암에서 발견됨.
- Oncogenic fusion genes은 원래의 유전자와 다른 혹은 새로운 기능을 갖는 유전자를 만들기도 함.

http://en.wikipedia.org/wiki/Fusion_gene

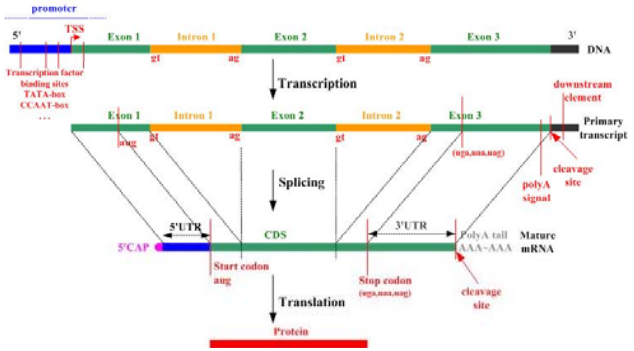
Gene Fusion and Cancer



***RAF* gene fusion breakpoints in pediatric brain tumors are characterized by significant enrichment of sequence microhomology**
Genome Res. 2011. 21: 505-514

Alternative splicing

Eukaryotic Gene



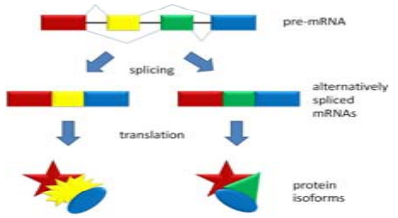
Adapted in part from http://online.itp.ucsb.edu/online/infobio01/burge/

Alternative splicing

- Alternative splicing의 역할
  - 하나의 유전자로부터 다양한 단백질이 만들어질 수 있음
  - Alternative splicing event의 약 80% 이상이 단백질 수준에서의 변화
  - 진화적인 관점에서 보면 Alternative splicing이 진핵 생물체의 표현형적 다양성에 관여
  - 많은 인간 질병이 Alternative splicing에 의해 유발

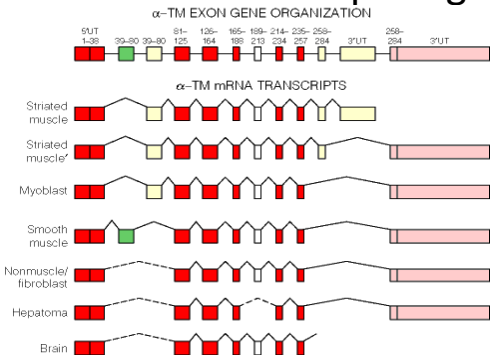
Alternative splicing of gene

Pre-mRNA는 서로 다른 splice 결합을 통해
두 개 이상의 mRNA molecules을 만듦.



사람의 경우
multi-exon gene의 95%에서 alternatively splice가 일어남.

Alternative Splicing



http://www.cs.uni.edu/~fienup/cs188s05/lectures/lec25_4-19-05.htm

Copy number variations
(유전자 복제 수 변이)

Copy Number Variations

- **<u>유전자 복제 수 변이(Copy Number Variations)</u>**
  - Reference 유전체와 비교하여 copy number의 차이를 보이는 1kb의 DNA조각
- 유전자의 삭제(deletions), 중복(duplications), 역위(inversions) 그리고 전좌(translocations)와 같은 유전체의 구조적 재배열에 의해 일어날 수 있음.
- CNVs는 수백 bp~약 1Mb에 이르는 염기 서열이 결실되거나 증폭되는 변이로, 이로 인해 특정 유전자의 숫자가 사람마다 달라지게 됨.
- 각각 다른 사람의 genome은 대략 0.4%의 copy number가 다를것으로 예상됨.
- CNVs는 질병에 대한 직접적인 원인 혹은 감수성(susceptibility) 인자로 작용
  - 알츠하이머병, 크론병, 파킨슨병, 자폐증등
  - **<u>CNVs는 암세포와 관련</u>**

Copy Number Variations

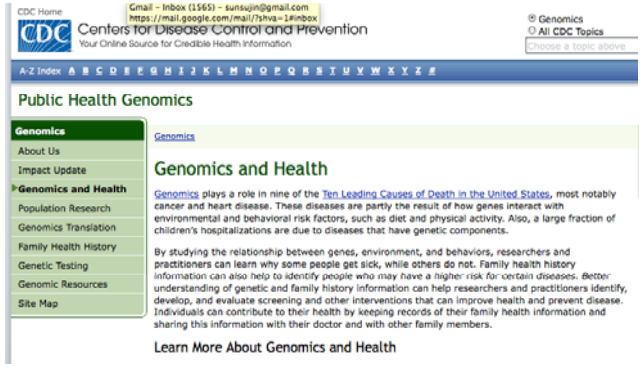
Copy Number Variation and Genetic Disease

By: Ingrid Lobo, Ph.D. (*Write Science Right*) © 2008 Nature Education
Citation: Lobo, I. (2008) Copy number variation and genetic disease. *Nature Education* 1(1)

Did you know that a large number of your genes exist in variable numbers of copies? While they can overlap with disease-related genes, these variants exist in healthy individuals too.

It is well known that errors during mitosis and meiosis can result in duplications and deletions of genes on a chromosomal level, which can lead to disorders. In fact, in the days prior to DNA sequencing, rare changes in gene numbers could actually be detected at the chromosomal level using a microscope. Nonetheless, scientists did not think wide-scale variations in gene copy numbers existed on a subchromosomal scale, nor did they believe that any such variations could lead to disease. Only recently did confusing laboratory results stimulate investigators to ask whether all autosomal genes are present in two copies, with a single allele inherited from each parent. By that time, the Human Genome Project and advances in genotyping provided the tools necessary to investigate variations in gene copy numbers on a subchromosomal scale (Lander *et al.*, 2001; Venter *et al.*, 2001).

Genomics and Disease



http://www.cdc.gov/genomics/public/index.htm

Epigenomics
(후성유전체학)

Epigenomics

- Epi (epi →on; upon) + genomics
- Yes, it is a control mechanism for genomic elements (e.g., genes).
- DNA methylation
- Histone modification
- microRNA, long non-coding RNA

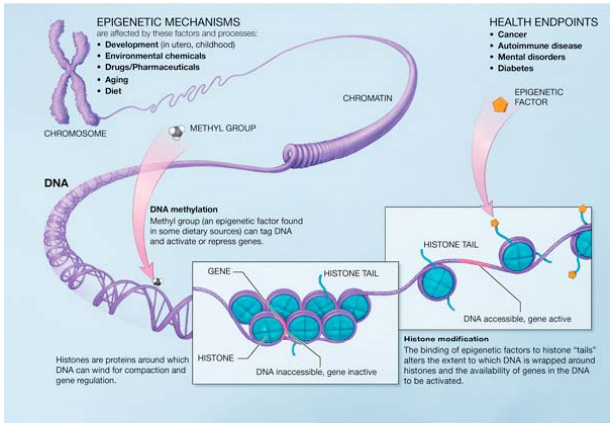
What is Epigenomics?

- Genomics : Hardware
- Epi-genomics : Software

  NOVA Science

  http://www.teachersdomain.org/asset/biot09_vid_epigenetics/

- A group of modifications at genetic level
- Epigenome tells body how to work and when to work



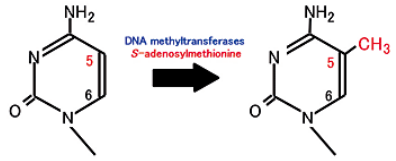
http://nihroadmap.nih.gov/epigenomics/epigeneticmechanisms.asp

DNA methylation

DNA Methylation

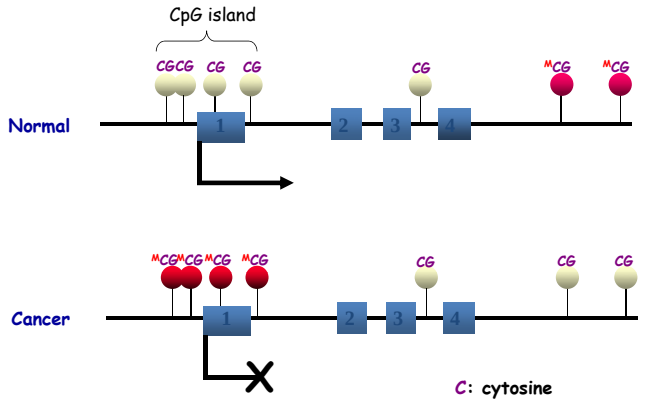
- DNA methylation
  - 고등동물의 정상적인 기관의 발달과 세포분화에 있어서 중요한 부분

Methylation of Cytosine

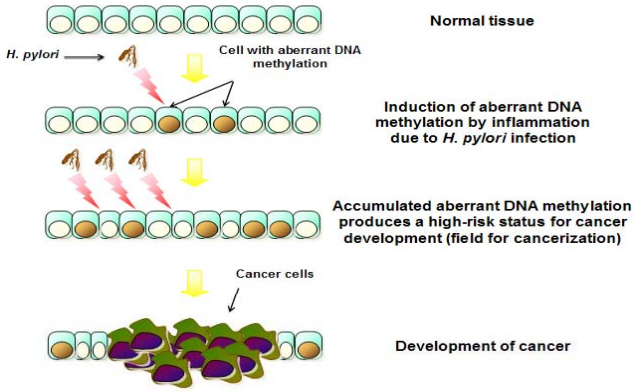


5′–CpG–3′
3′–GpC–5′

DNA Methylation and Gene Silencing in Cancer Cells



DNA Methylation and Cancer

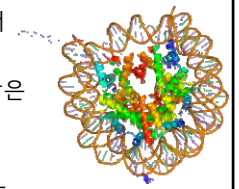


National cancer center institute

Histone modification

Histone and DNA

- 핵 내 DNA와 결합하고 있는 염기성 단백질
  - 실을 실패에 감싸서 실이 엉키지 않도록 보관하고, 바느질 할 때 실패의 실을 풀어서 사용하는 것처럼 30억 bp DNA(실)는 실패(히스톤)에 감겨져 있음.
  - 2m길이의 DNA를 눈에 보이지 않을 만큼 작은 세포 속에 저장가능.
  - 응축된 후에는 5000배 가까이 짧아짐.
- **Chromatin regulation**
  - Histone modifications은 유전자 발현 및 세포 사멸조절, DNA 복제 및 수선, 체세포분열 등과 같은 생물학적 기작에 관여.



http://en.wikipedia.org/wiki/Histone

Histone modifications: Histone Code

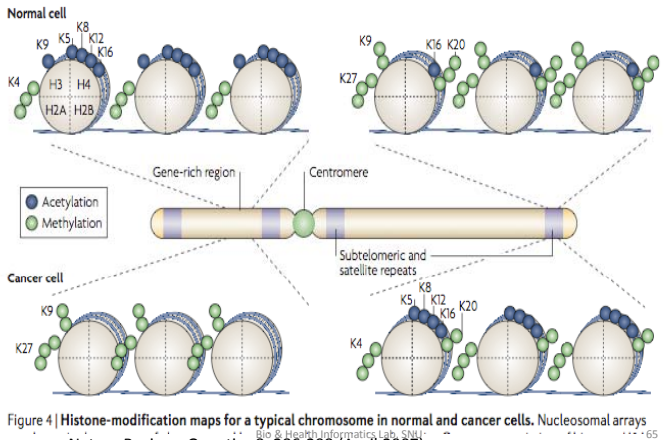
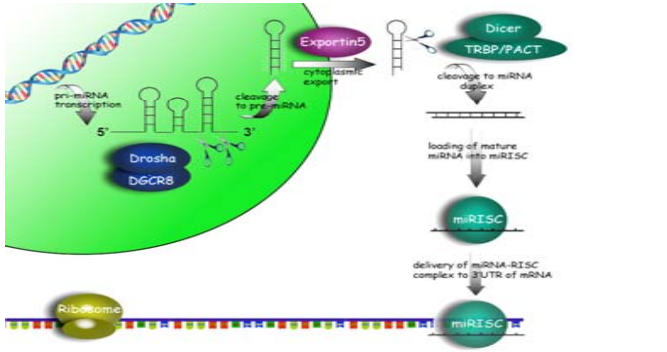


Figure 4 | **Histone-modification maps for a typical chromosome in normal and cancer cells.** Nucleosomal arrays

Nature Reviews Genetics 8, 286-298 (April 2007)

microRNA

microRNA .v.s. protein coding gene



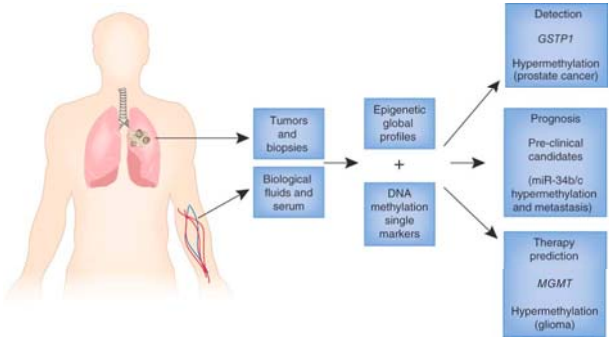
http://www.microrna world.com/intro.htm

Roles of MicroRNA in Cancer

- MicroRNAs as oncogenes(발암유전자)
- MicroRNAs as tumor suppressors(종양억제)
- MicroRNAs as modulators of tumor progression and metastasis(종양 진행 및 전이 조절자)
- Global deregulation of microRNAs in cancer

Ventura and Jacks, Cell. 2009 Feb 20;136(4):586-91

Cancer epigenetics reaches mainstream oncology



Nature Medicine 330–339 (2011)

Acknowledgement

장현숙, 유정현
서울대학교 생물정보연구소