Outline

- Introduction to Genome Biology
- Introduction to Measurement Technology
- Statistics Applied to Genomic Data

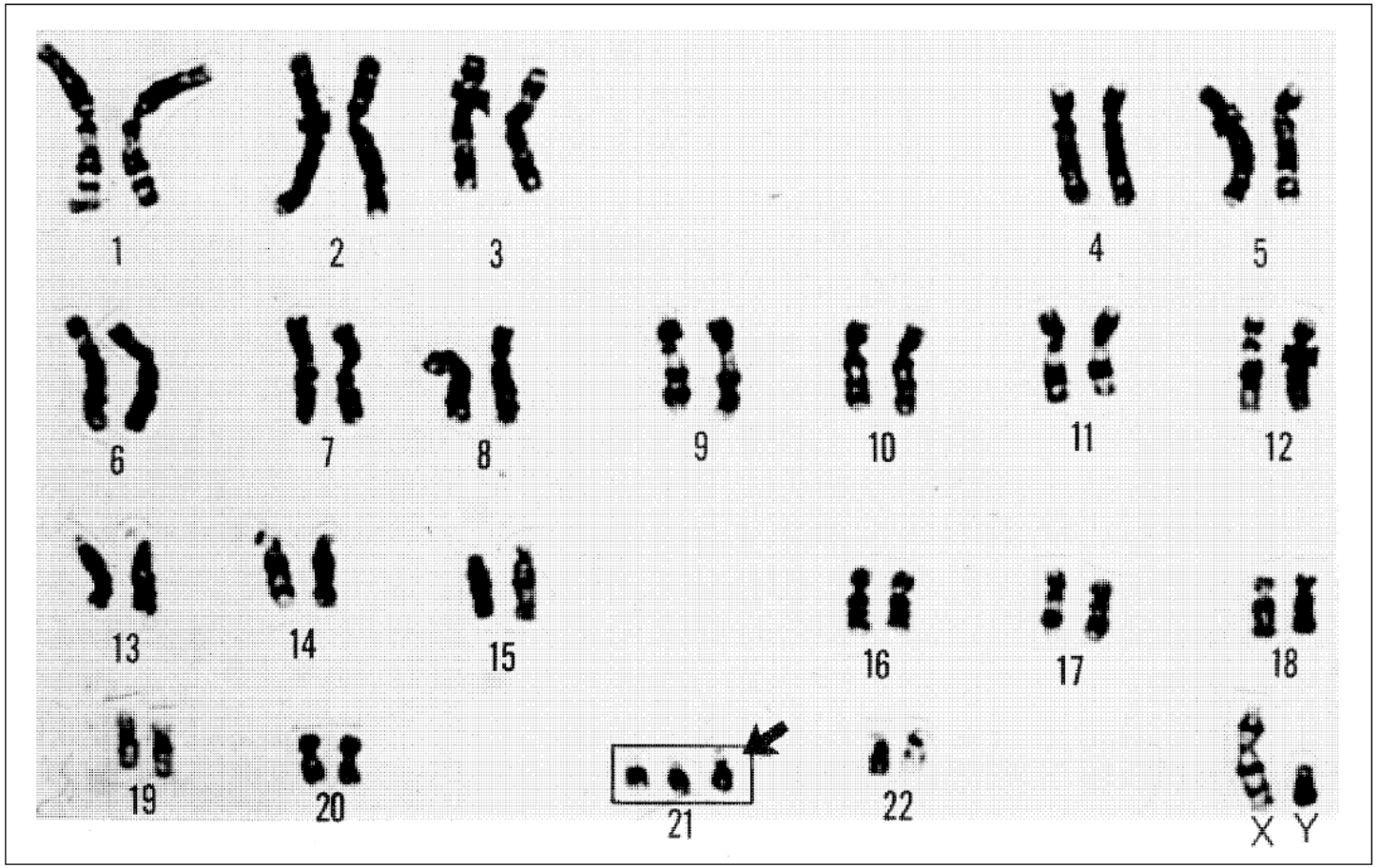
Introduction to Genome Biology

Key terms

- DNA
- Chromosome
- Allele
- Gene
- Transcription or Gene Expression
- RNA
- Translation
- Protein
- Genome
- Strand
- Base-pair (A T C G)


Por qué son tan
marranos mis hijos?

Chromosomes



These are actually human? And for a down syndrome patient

What is Genomics?

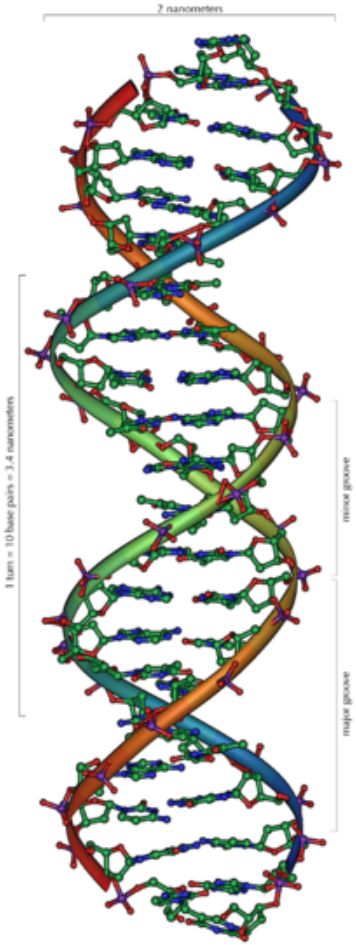
- Each cell contains a complete copy of an organism's genome, or blueprint for all cellular structures and activities.

- The genome is distributed along chromosomes, which are made of compressed and entwined DNA.

- Cells are of many different types (e.g. blood, skin, nerve cells), but all can be traced back to a single cell, the fertilized egg.

- Genomics is the study of molecular information to understand natural human variation and disease.

DNA



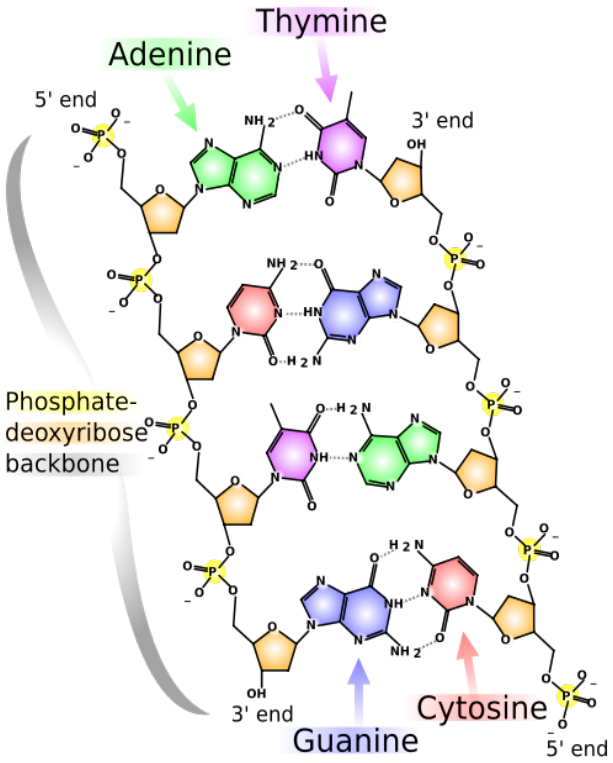
DNAs (Deoxyribonucleic acids) are molecules to store genetic information of a living organism.

DNA consists of two polymers made from four types of nucleotides: adenine (A) guanine (G), cytosine (C) and thymine (T).

Purines: A, G; Pyrimidines: C, T

Two polymers are complementary to each other and from a double-helix structure

```
5′-ACCGTTCGACGGTAA-3′
   ||||||||||||||||
3′-TGGCAAGCTGCCATT-5′
```



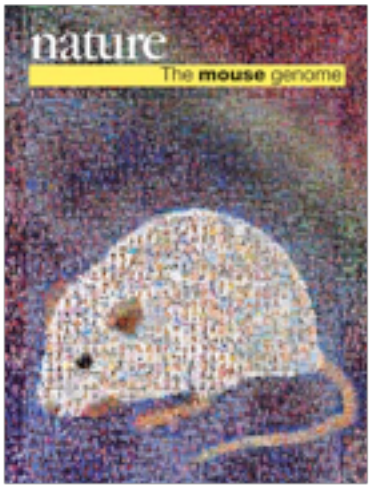
Watson and Crick 1953

Measurement

- For a small enough piece, we can measure the sequence of bases, referred to as *sequencing*
- Human Genome Project

D. melanogaster, *Science*, 2000

H. sapiens, *Nature*, 2000 and *Science*, 2000



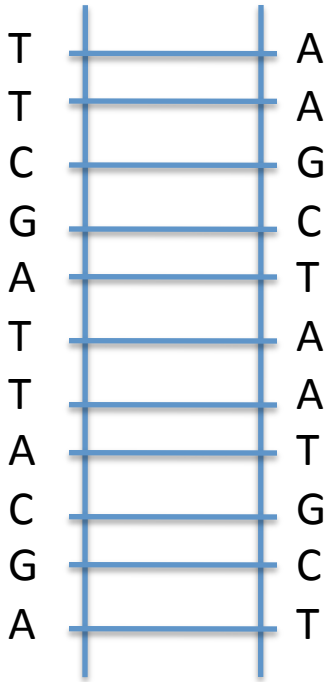
M. musculus, *Nature*, 2002

Genome

TCAGTTGGAGCTGCTCCCCCACGGCCTCTCCTCACATTCCACGTCCTGTAGCTCTATGACCTCCACCTTTGAGTCCCTCCTC
TCACACCTGACATGAAAAGGCACATGAGGATCCTCAAATACCCCGTGATCAGTCTCAGGGTAGCTCTCATAGCCTGGACAGG
GCCCCCCTCGGGGGTTGCGCCCAGGTCCAGGCGGGGGATGCACAGCAACAGTCACCGAAGCAGAAGCCGTCACAGTGGTGAT
GGGCTGGCAGTAGCTGGGCACAGAGCTGCCCATGGCGGTGGACGTTGGGTTCCGAGGGTTGTGAGAACGGGCCCCACGGGGC
CCTGAGCGGTCCCTATTGCTAGGGCCAGAATGCCCTTCAGTAGAAATTTCAAAAGCGTCTCTGCGCGGTCTGTAGGGGGGTG
GCCGCAAGCCTTCTCTAGGGGGATCCCTTCGAGGCTGCTGGCCTTGCCGTCCAGGGGACAAGGAGCCAGAGTCCAGGTGGGG
CTGTTGCCGAGGGGTCAAGGGAGGCTGATGTCTGGAGTCCGGATGGACCACCTGCAGAGGAGAGACATAGGTCAACACAGGG
AGGTAGGATGGTGGTGATGTTCCACCCACAAAAGAAAACCTATTCCTTTAGAAACCTCCAGGATGTGAATCCTGCCTGCACC
TGCACAGCTGGCTGGAGGCATATAGCCACTGCCCATAGATCTCAACTTACCCTCACAACCAACTGCCCCCAGGCCTAAGTTC
TCTGCCTCAAAACTGCCAAGGCCTGGATAGCCAAGAGCCTGGGTGTCTTGGAAATATGCAACCATAAATAGTAGCTTTTAGA
AGTATAAGGCTCCTGTTTCTGGGTCATATTAGTGTTGTTTTCACCTGTCCCCAGCCCTAAGCCAGGTGTGGCCAGAAGCAAA
TGTACTGTAAGAGCAGAGCAAAAACTTCCACACAGATAGTTCTGTTAGGCAATACATCTCTGCCTGACTATTAGGAATCTGG
TTTCTGGGTCCTCTGTACAAAGCTCGGAGCAACACAGTGGCCACATCAATCAAAAGGACCGTGACCAACTTCAAAGTCGGTG
AGCTTGTACCTATTTTTAGGCTCCTGCTGAACAGAACCAGATTCACACTACAGCTCAGCAGGGCATCGTCACGGGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGGGGGGGGGGTGGACAGAGGACGGGGACACAATTCACTGGCCAGCCCTT
CTCTCCTTCAAGGAAGGCTGCTCTAGCCTGGGACTGGAATACACATTTCCTGTAAACATGGTGGGGGCCTCAGGCAAGCCAG
AGTTTTGGAGCCTTCCTTAACTCTTCAAGGTGAGCATCTTGACTTGGAGGGTGGGGGTGCGGGTAAGGAAGGAACCTGTGGA
CTCCTCCCTACAAGACAGAAAAGGAATAAGCCACGAAGACAATAACGATTTTTGTATCAAGCGTCCTCTCCCATTTCAGCTT
ACCTGACAATGAAATCAAATTCGGACCCTGCAAGCATCAGTACACCCAGCAGAGTGGACACAGCACCGTCCAGAACGGGAGC
AAACATGTGCTCCAGAGCGAGCATAGCCCTGTGGTTCTTGTCCCCAATGGCTGTCAGAAAGGCCTGAACAAAGGAGAAAATT
GACACGGTCACATTCTGGGTGTGGTAAAGTGCTCAGCTGTGTCTATACTTGGGTTTTGTAT…

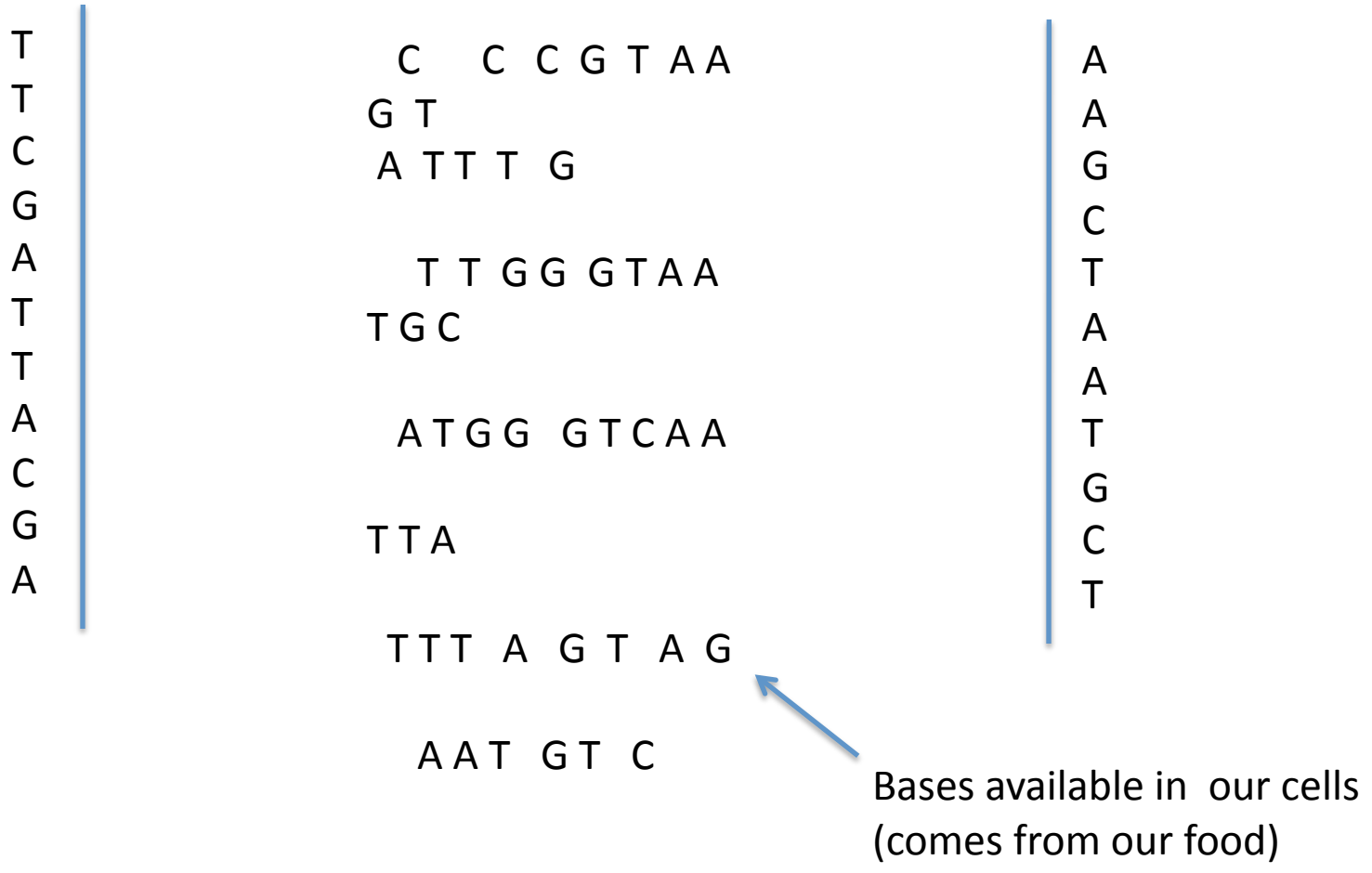
Total amount of DNA in human genome: 3 * 10^9 base pairs (bp)

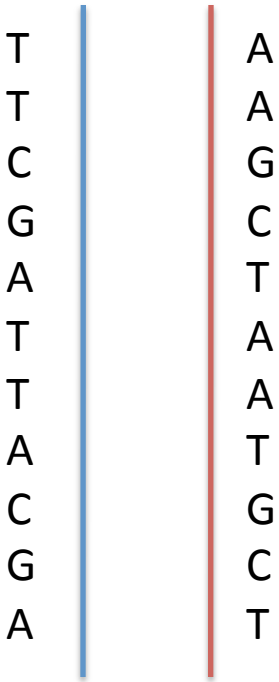
Replication

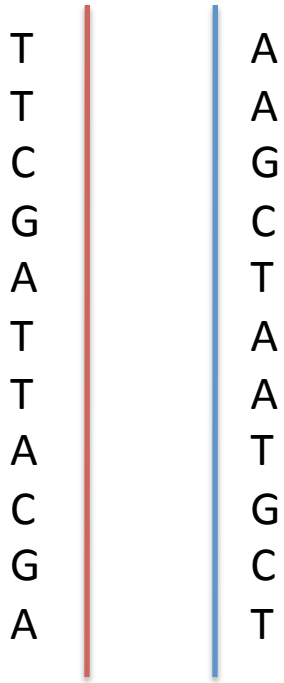


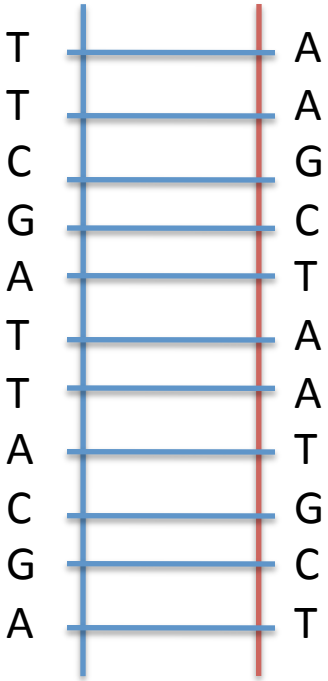
T
T
C
G
A
T
T
A
C
G
A

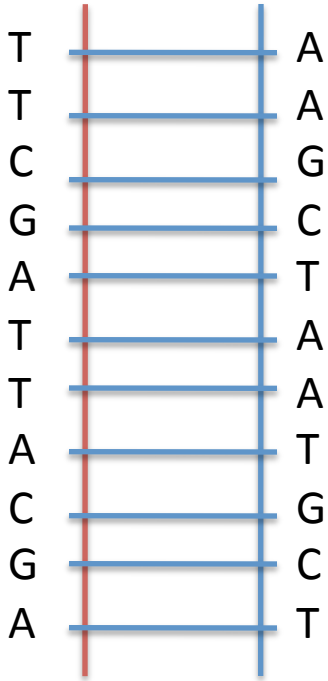
A
A
G
C
T
A
A
T
G
C
T


T
T
C
G
A
T
T
A
C
G
A
C C C G T A A
G T
A T T T G
T T G G G T A A
T G C
A T G G G T C A A
T T A
T T T A G T A G
A A T G T C
A
A
G
C
T
A
A
T
G
C
T
Bases available in our cells
(comes from our food)


T
T
C
G
A
T
T
A
C
G
A
A
A
G
C
T
A
A
T
G
C
T


T
T
C
G
A
T
T
A
C
G
A
A
A
G
C
T
A
A
T
G
C
T


T A
T A
C G
G C
A T
T A
T A
A T
C G
G C
A T

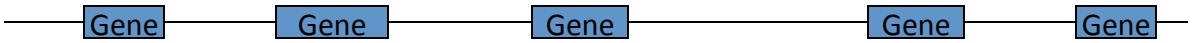

T A
T A
C G
G C
A T
T A
T A
A T
C G
G C
A T

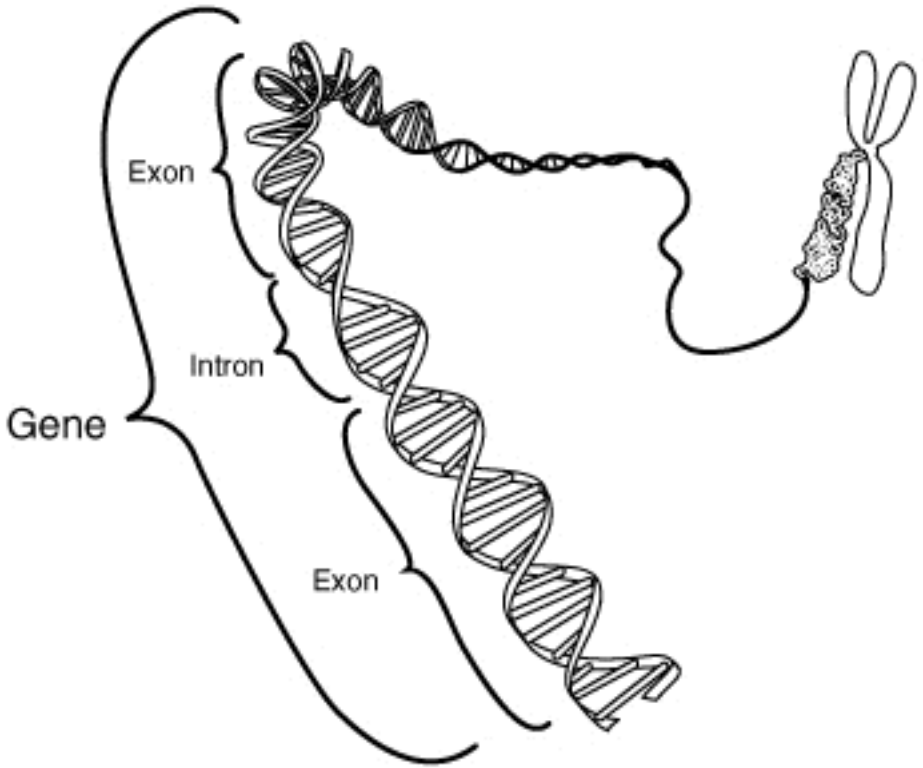
Why are these two different? What explains similarities?

Differences explained by 1-10% difference in genome

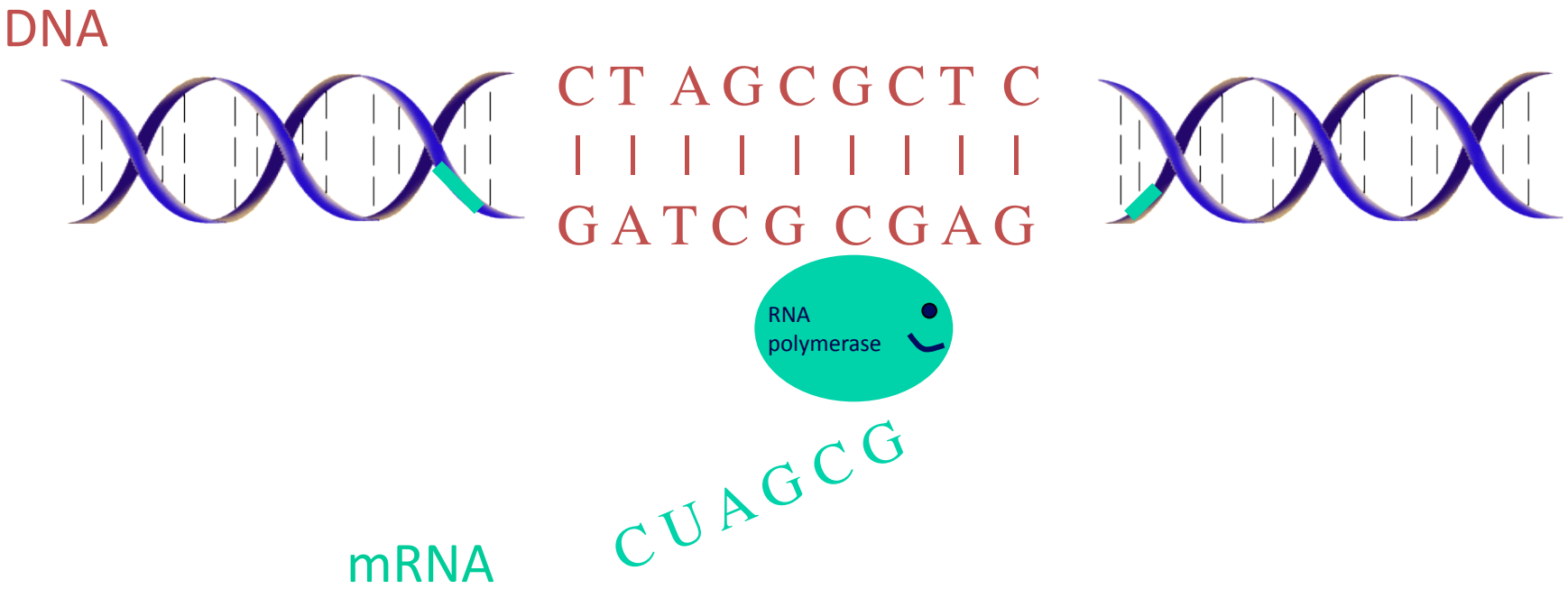
Similarities explained by similar genes

Genes

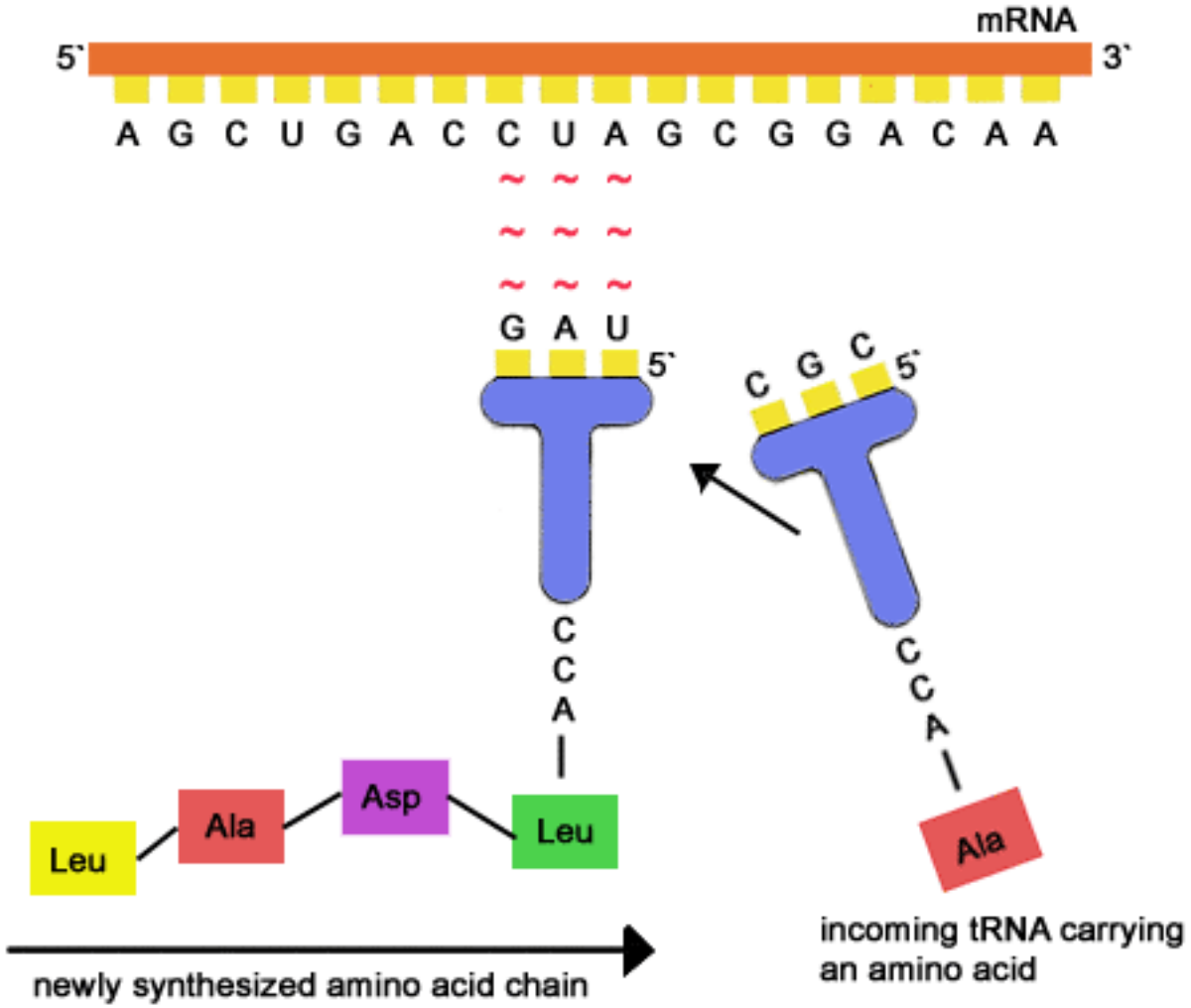




Transcription

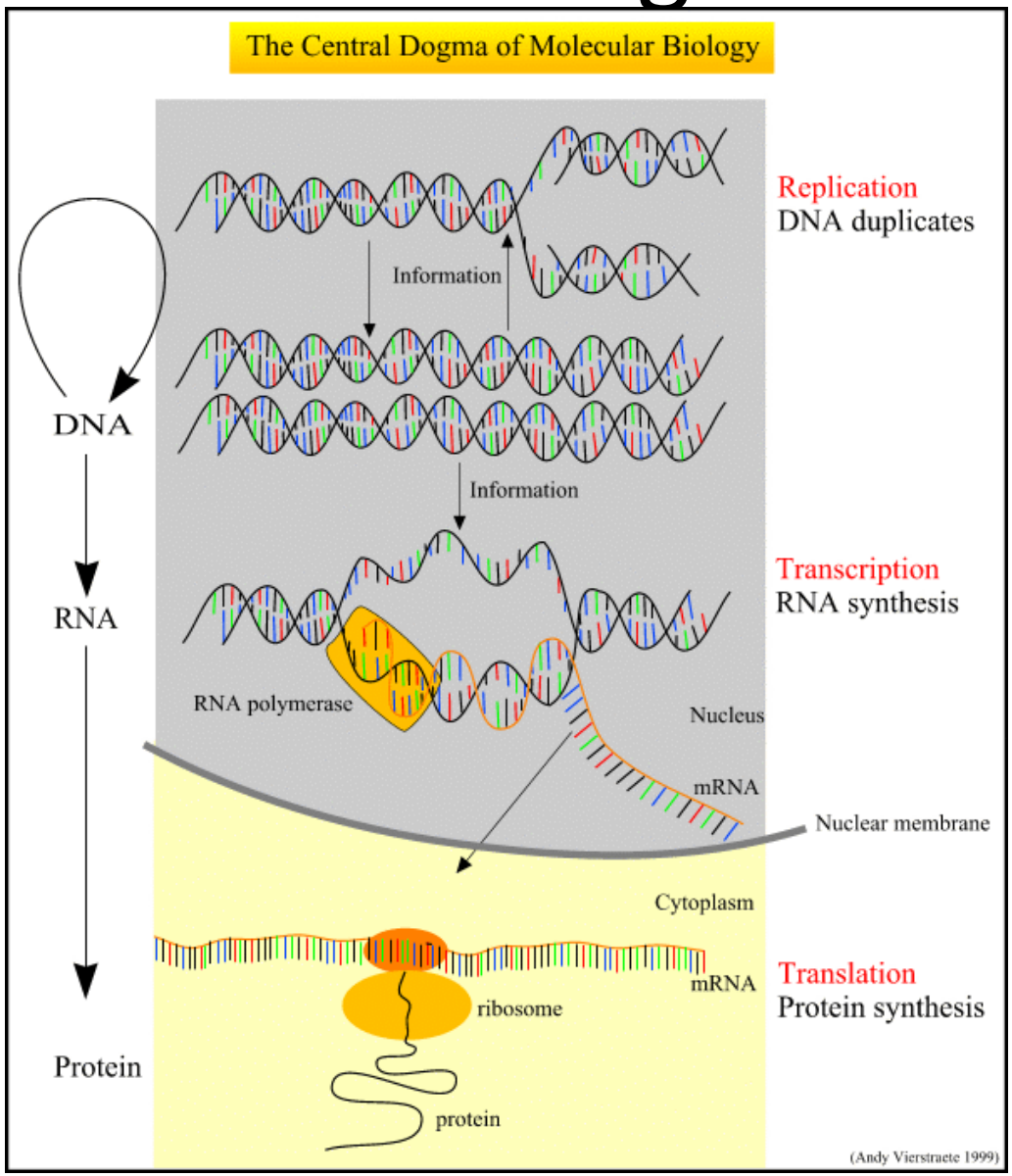


Translation



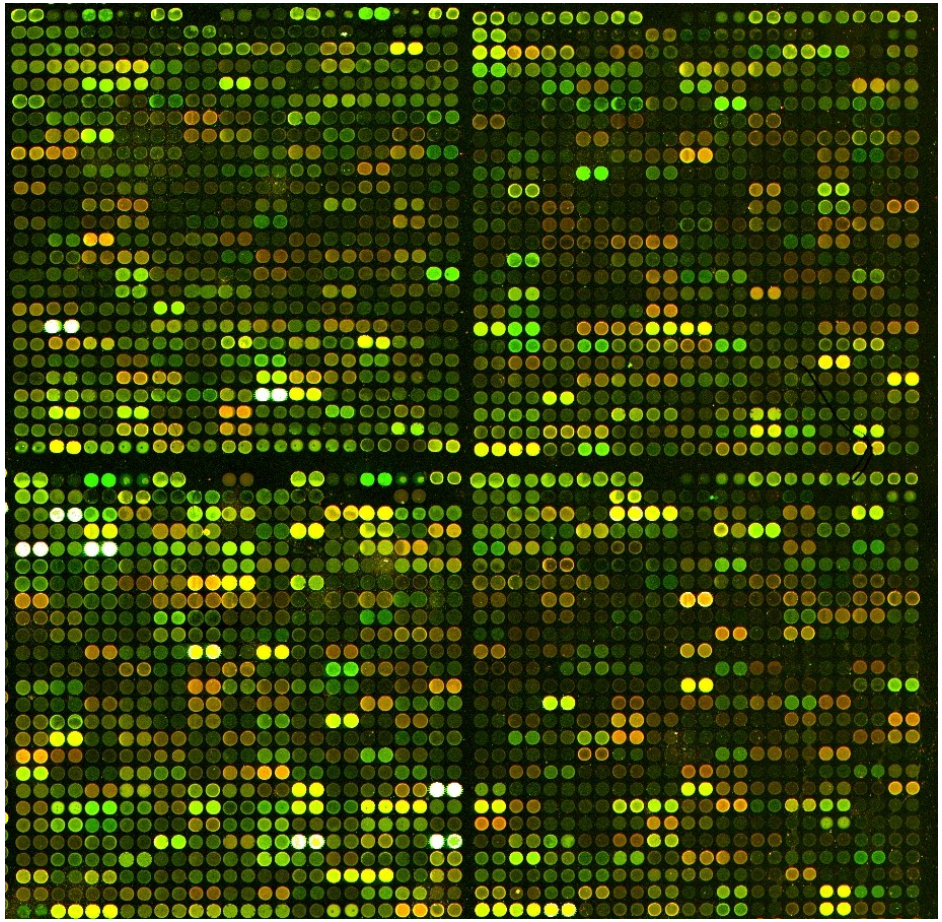
http://gel.ym.edu.tw/~ycl6/sc2005/images/translation.gif

Central Dogma



Measurement: Microarrays

- More on this later



What makes them different?

Much human variation is due to difference in ~ 6 million base pairs (0.1 % of genome) referred to as SNPs

Single Nucleotide Polymorphism (SNP)

Genomic DNA:

SNP

A

TACATAGCCATCGGTANGTACTCAATGATGATA

G

Three genotypes

AA

Madre

TACATAGCCATCGGTAAGTACTCAATGATGATA
ATGTATCGGTAGCCATTCATGAGTTACTACTAT

Padre

TACATAGCCATCGGTAAGTACTCAATGATGATA
ATGTATCGGTAGCCATTCATGAGTTACTACTAT

AG

Mama

TACATAGCCATCGGTAAGTACTCAATGATGATA
ATGTATCGGTAGCCATTCATGAGTTACTACTAT

Papa

TACATAGCCATCGGTAGGTACTCAATGATGATA
ATGTATCGGTAGCCATCCATGAGTTACTACTAT

GG

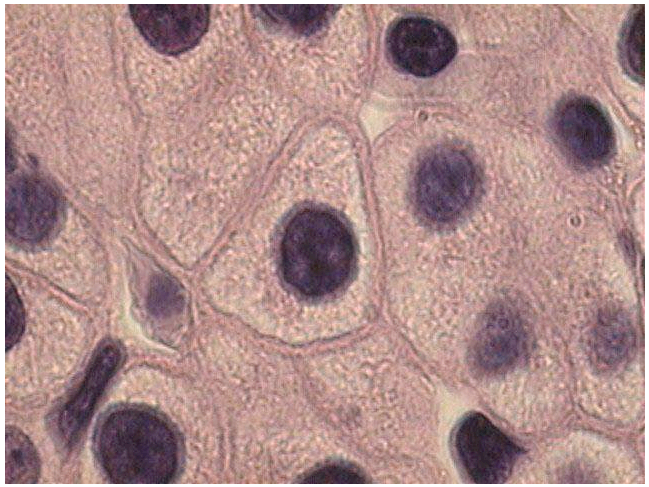
Mama

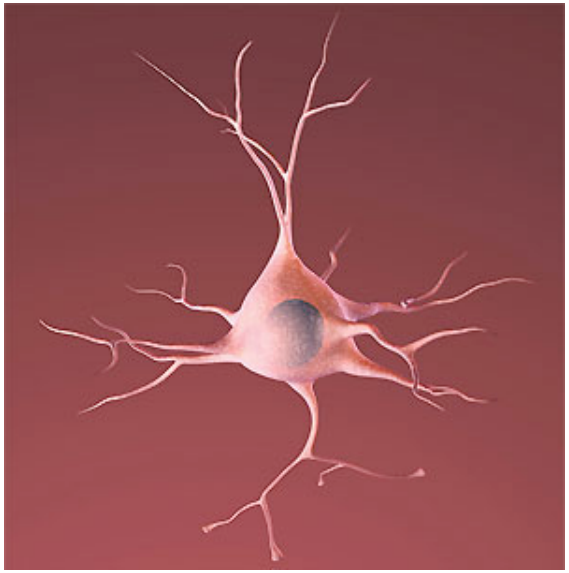
TACATAGCCATCGGTAGGTACTCAATGATGATA
ATGTATCGGTAGCCATCCATGAGTTACTACTAT

Papa

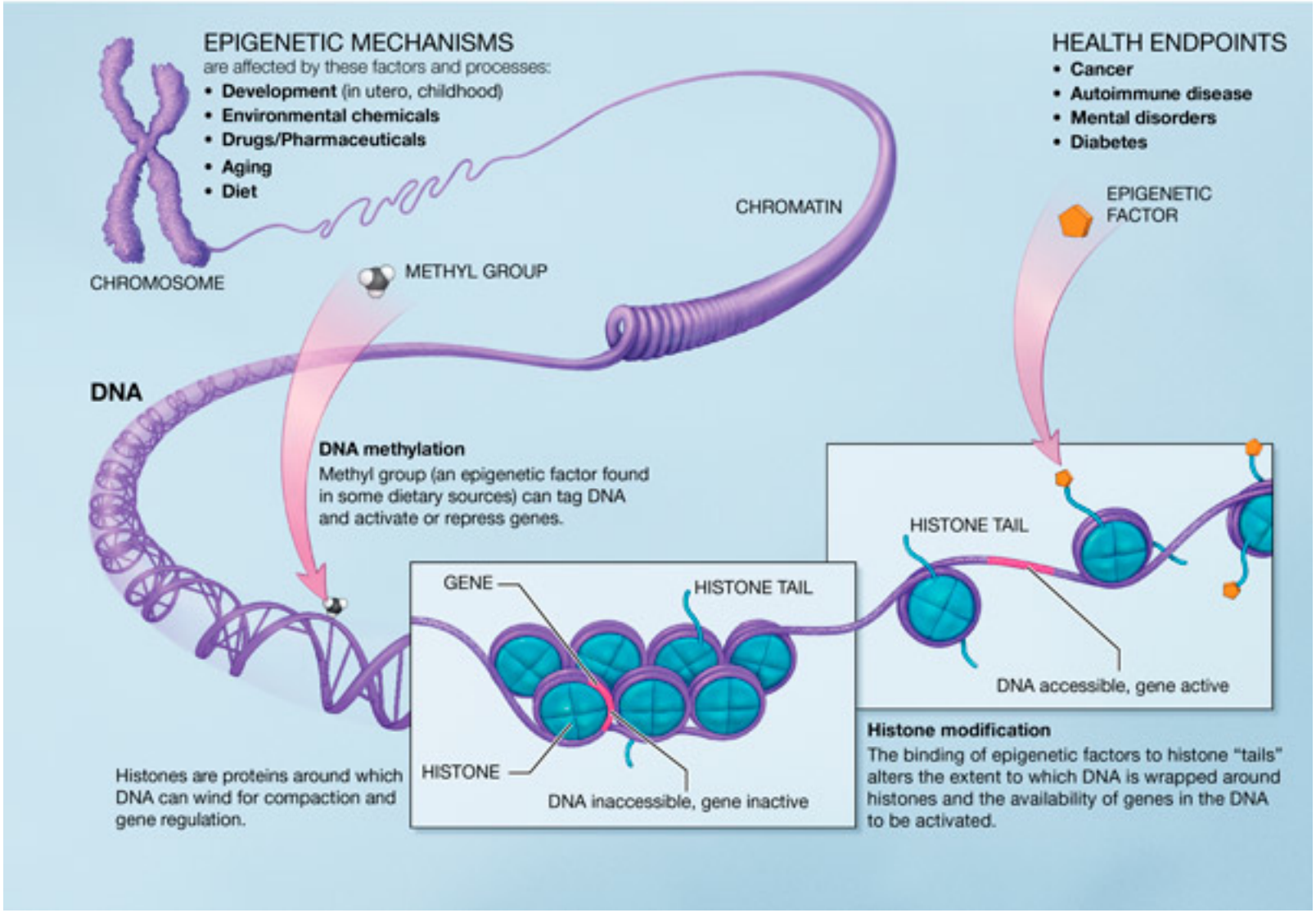
TACATAGCCATCGGTAGGTACTCAATGATGATA
ATGTATCGGTAGCCATCCATGAGTTACTACTAT

How many basepair differences?



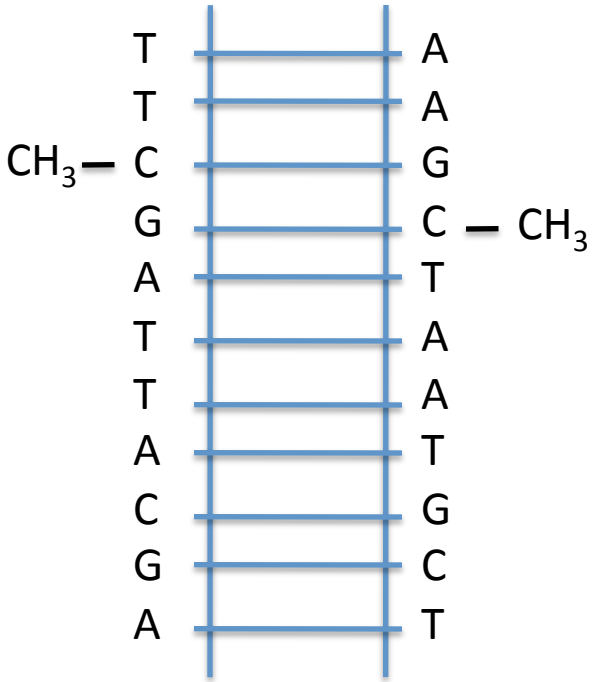


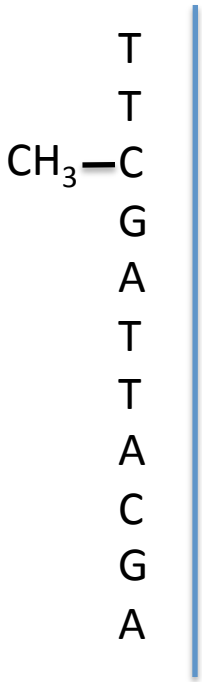
Epigenetics

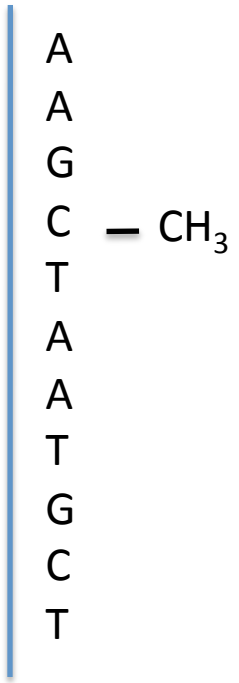


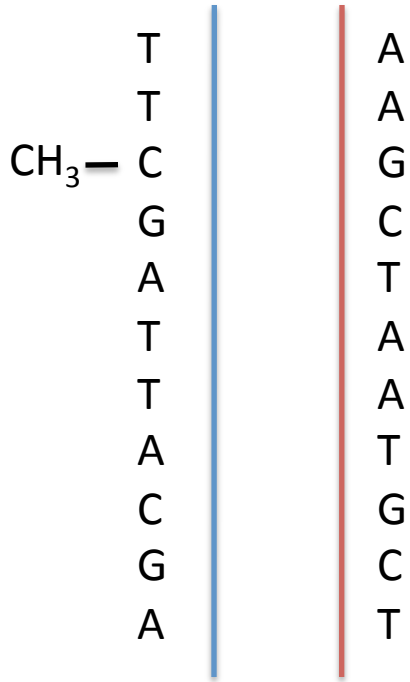
http://nihroadmap.nih.gov/EPIGENOMICS/images/epigeneticmechanisms.jpg

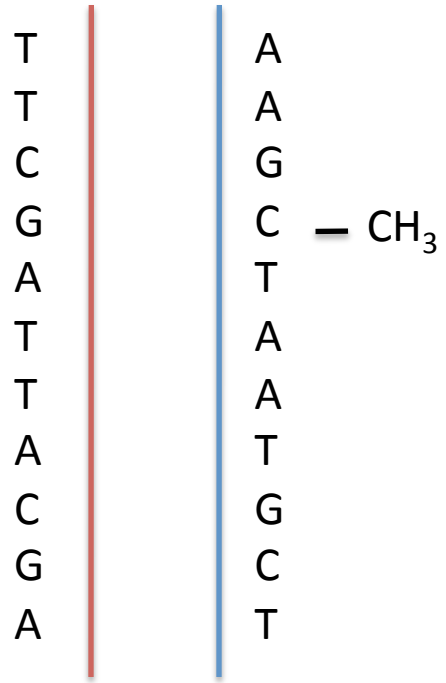
DNA Methylation

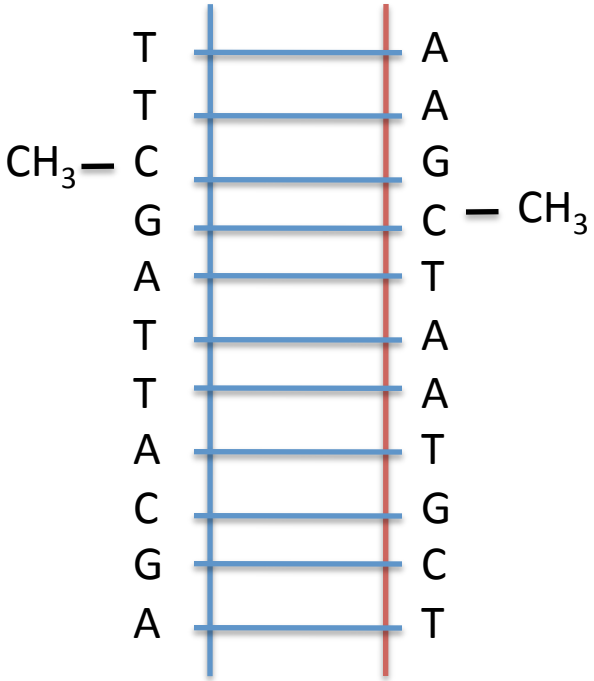


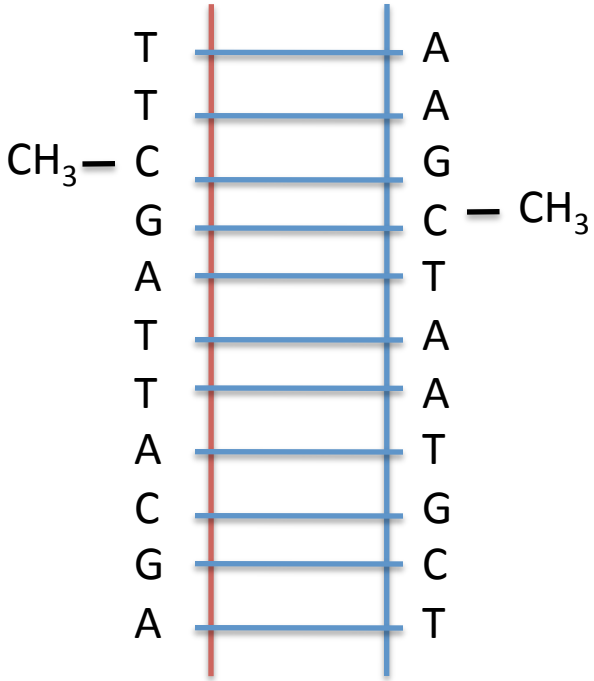

T
T
CH3—C
G
A
T
T
A
C
G
A


A
A
G
C — CH3
T
A
A
T
G
C
T

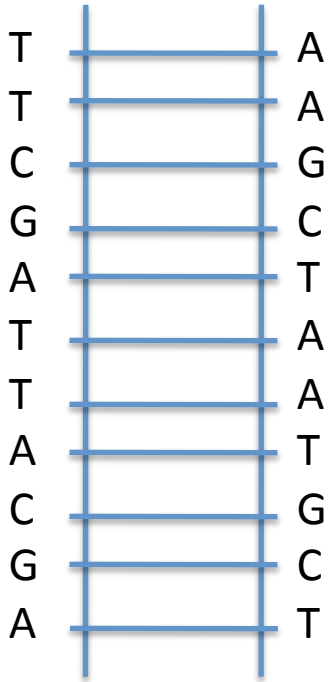

CH3
T
T
C
G
A
T
T
A
C
G
A
A
A
G
C
T
A
A
T
G
C
T


T
T
C
G
A
T
T
A
C
G
A
A
A
G
C
CH3
T
A
A
T
G
C
T

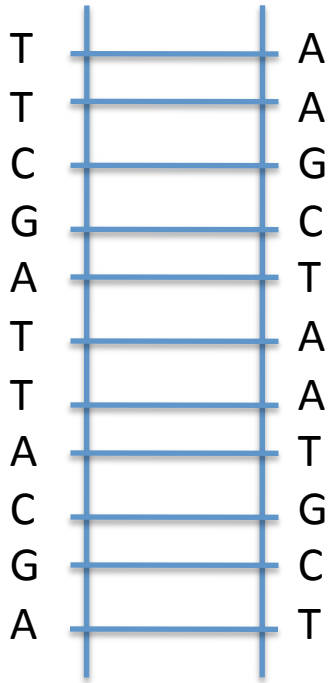

T A
T A
CH3 C G
G C CH3
A T
T A
T A
A T
C G
G C
A T


T A
T A
CH3 C G
G C CH3
A T
T A
T A
A T
C G
G C
A T

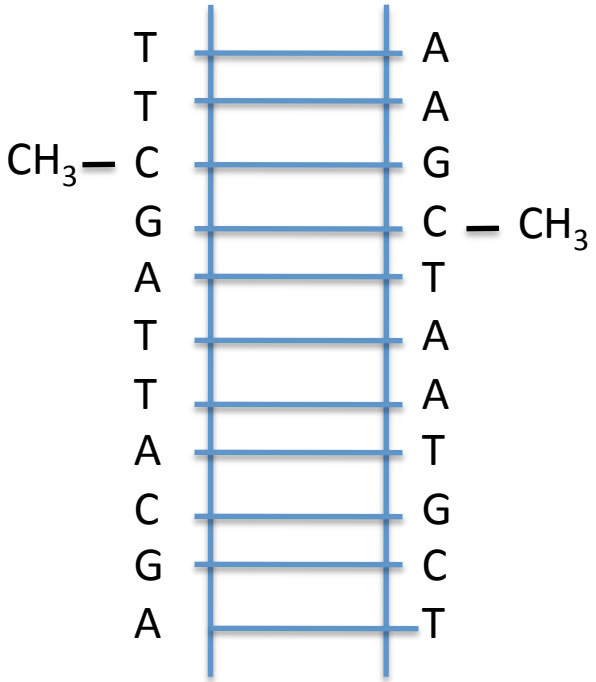
Liver



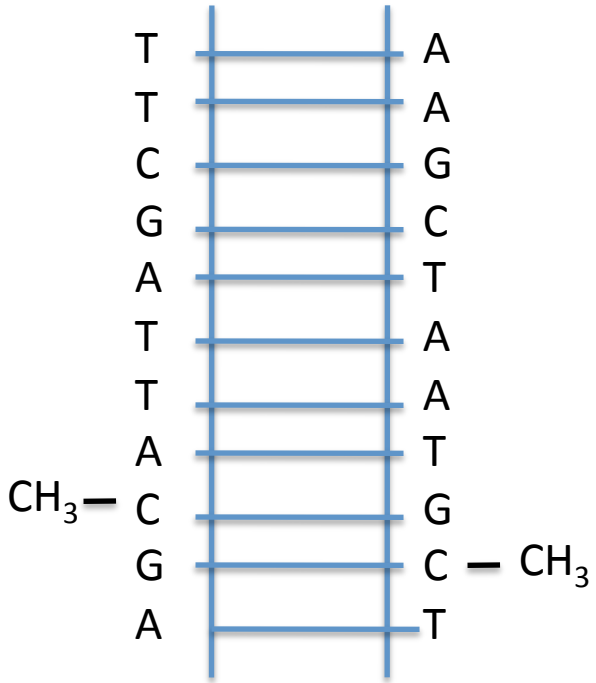
Brain



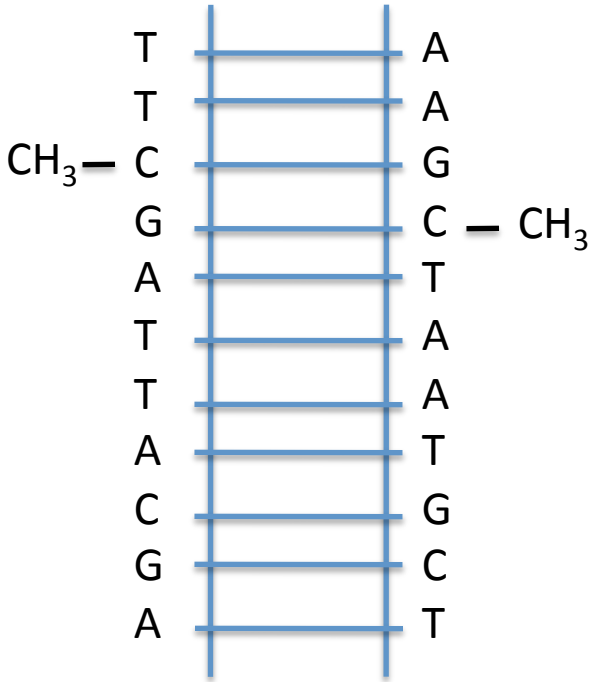
Liver

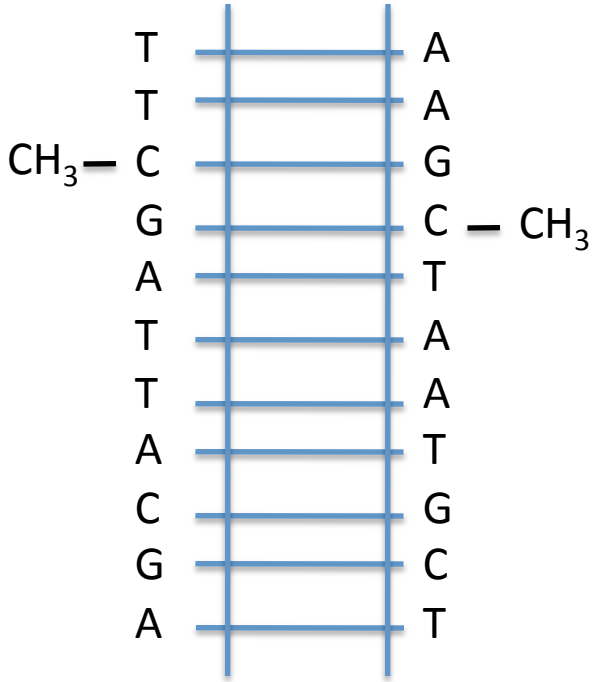

T A
T A
CH3 C G
G C CH3
A T
T A
T A
A T
C G
G C
A T

Brain

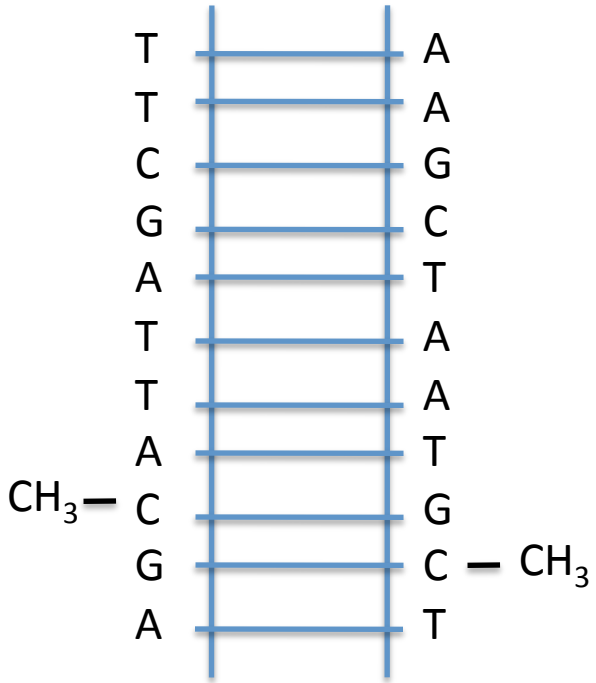

T A
T A
C G
G C
A T
T A
T A
A T
CH3 C G
G C CH3
A T

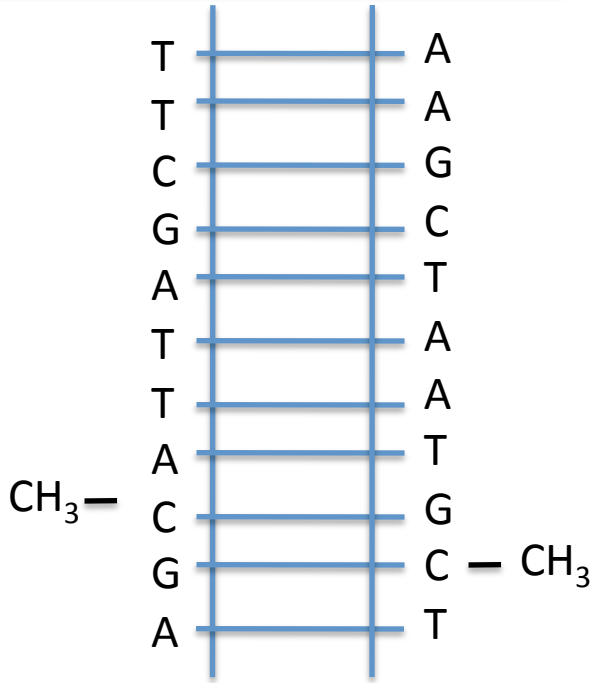
Liver


T
T
CH3 – C
G
A
T
T
A
C
G
A
A
A
G
C – CH3
T
A
A
T
G
C
T


T
T
CH3 – C
G
A
T
T
A
C
G
A
A
A
G
C – CH3
T
A
A
T
G
C
T

Brain


T
T
C
G
A
T
T
A
CH3 – C
G
A
A
A
G
C
T
A
A
T
G
C – CH3
T


T
T
C
G
A
T
T
A
CH3 –
C
G
A
A
A
G
C
T
A
A
T
G
C – CH3
T