The Basic Genome
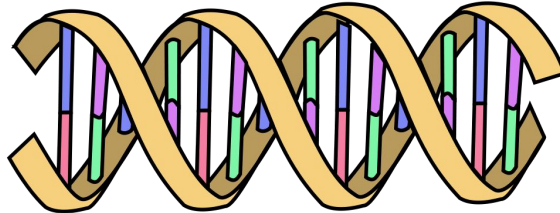
Robert Brown

Thanks UCLA Computational Medicine!

Human genetics
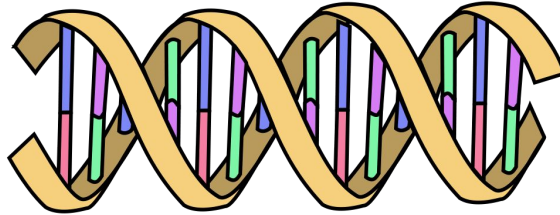
DNA:

...CTCGTCACTTCACGCATG...
|||||||||||||||||||||
...GAGCAGTGAAGT??????...

- 3 billion base pairs long
- <u>Nucleotides</u>
    - Adenine
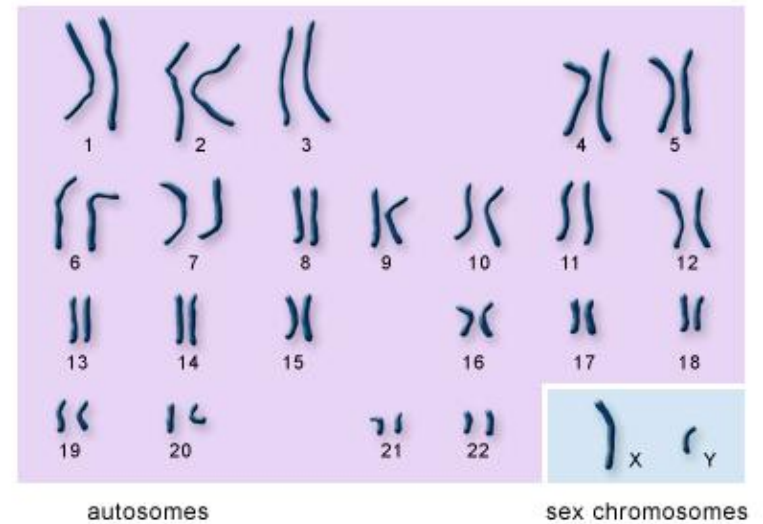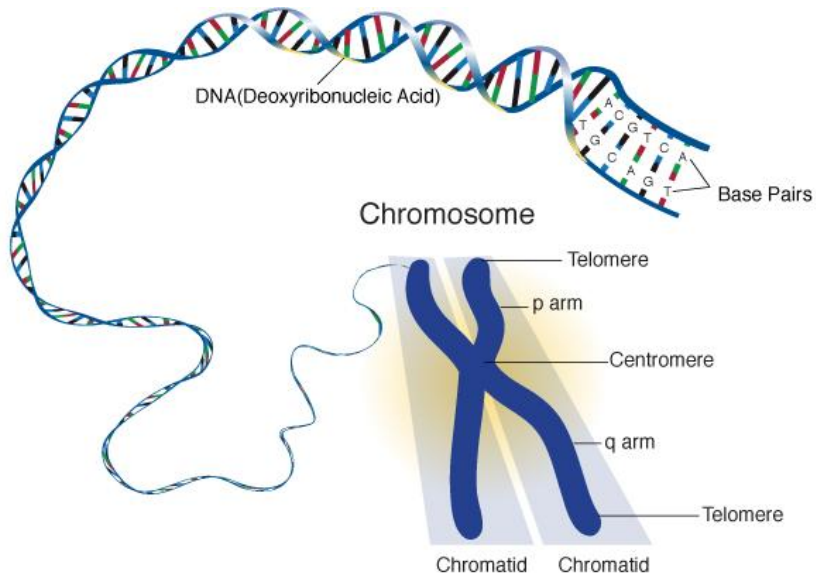   - Thymine
    - Guanine
    - Cytosine

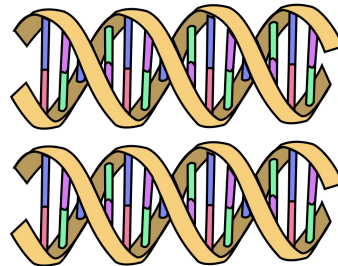# Human genetics

DNA:

...CTCGTCACTTCACGCATG...

- 3 billion base pairs long
- <u>Nucleotides</u>
  - Adenine
  - Thymine
  - Guanine
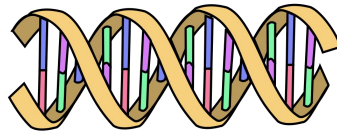 - Cytosine

Chromosomes

Source: National Human Genome Research Institute

Source: National Library of Medicine

Human genetics

DNA:

...CTCGTCACTTCACGCATG...

- 3 billion bases long
- Nucleotides
  - Adenine
  - Thymine
  - Guanine
  - Cytosine

...CTCGTCACATCACGTATG...
|||||||||||||||||||
...GAGCAGTGTAGTGCATAC...

...CTCCTCACTTCACGTATG...
|||||||||||||||||||
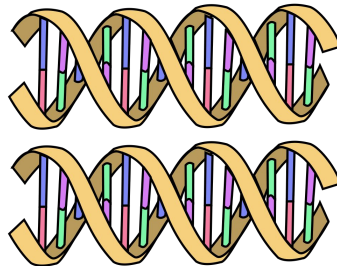...GAGCAGTGAAGTGCATAC...

We are diploid!!

Human genetics

DNA:

...CTCGTCACTTCACGCATG...

- 3 billion bases long
- Nucleotides
    - Adenine
    - Thymine
    - Guanine
    - Cytosine

...CTCGTCACATCACGTATG...
|||||||||||||||||||
...GAGCAGTGTAGTGCATAC...

...CTCCTCACTTCACGTATG...
|||||||||||||||||||
...GAGCAGTGAAGTGCATAC...

We are diploid!!

Human genetics

DNA:

...CTCGTCACTTCACGCATG...

- 3 billion bases long
- Nucleotides
  - Adenine
  - Thymine
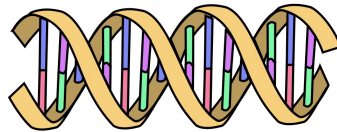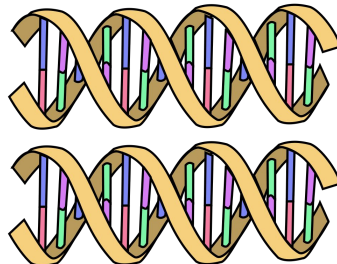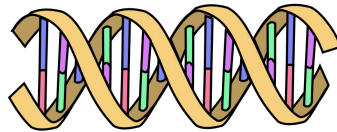 - Guanine
  - Cytosine

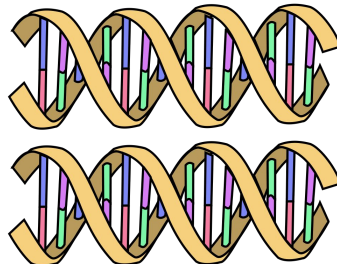...CTCGTCACATCACGTATG...
...CTCCTCACTTCACGTATG...

Human genetics

DNA:

...CTCGTCACTTCACGCATG...

- 3 billion bases long
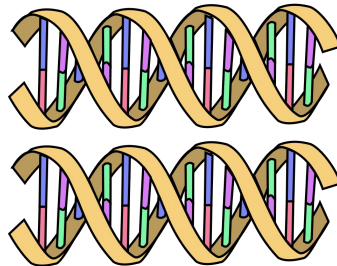- Nucleotides
  - Adenine
  - Thymine
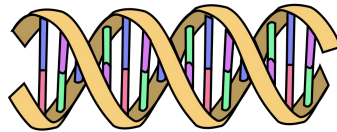 - Guanine
  - Cytosine

...CTCGTCAC**A**TCACG**T**ATG...
...CTC**C**TCACTTCACG**T**ATG...

Human genetics

DNA:

...CTCGTCACTTCACGCATG...

- 3 billion bases long
- Nucleotides
  - Adenine
  - Thymine
  - Guanine
  - Cytosine

...CTCGTCAC**A**TCACG**T**ATG...
...CTC**C**TCACTTCACG**T**ATG...

Single Nucleotide Polymorphisms (SNPs)

Human genetics

DNA:

...CTCGTCACTTCACGCATG...

- 3 billion bases long
- Nucleotides
  - Adenine
  - Thymine
  - Guanine
  - Cytosine

Haplotype 1

...CTCGTCAC**A**TCACG**T**ATG...

...CTC**C**TCACTTCACG**T**ATG...

Haplotype 2

Human genetics

DNA:

...CTCGTCACTTCACGCATG...

- 3 billion bases long
- Nucleotides
    - Adenine
    - Thymine
    - Guanine

Haplotype 1, Chromosome A

...CTCGTCAC**A**TCACG**T**ATG...
...CTC**C**TCACTTCACG**T**ATG...

Haplotype 2, Chromosome B

Human genetics

DNA:

...CTCGTCACTTCACGCATG...

- 3 billion bases long
- Nucleotides
  - Adenine
  - Thymine
  - Guanine
  - Cytosine
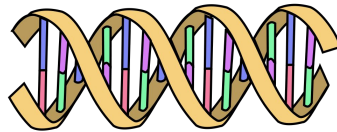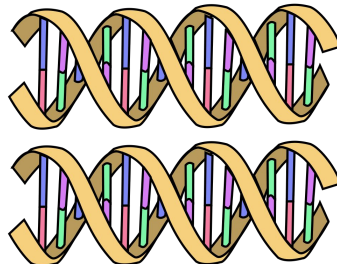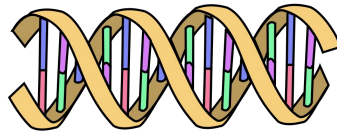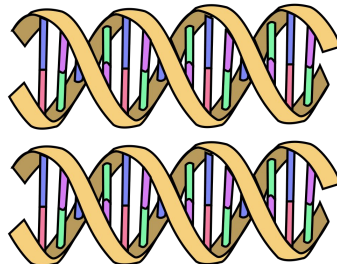
...CTCGTCAC**A**TCACG**T**ATG...
...CTC**C**TCACTTCACG**T**ATG...

What is a genotype?
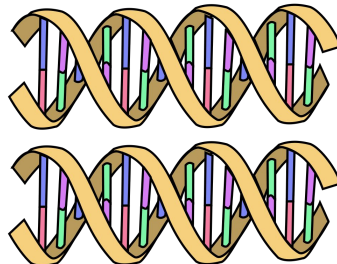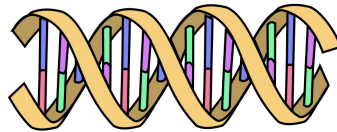
Human genetics

...CTCGTCAC**A**TCACG**T**ATG...
...CTC**C**TCACTTCACG**T**ATG...

...CTCGTCACTTCACG**T**ATG...
...CTC**C**TCACTTCA**A**GCATG...

Human genetics

99.9% the same!

...CTCGTCAC**A**TCACG**T**ATG...
...CTC**C**TCACTTCACG**T**ATG...

...CTCGTCACTTCACG**T**ATG...
...CTC**C**TCACTTCA**A**GCATG...

Genetic variation

99.9% the same!
0.1% different

...CTCGTCAC**A**TCACG**T**ATG...
...CTC**C**TCACTTCACG**T**ATG...

...CTCGTCACTTCACG**T**ATG...
...CTC**C**TCACTTCA**A**GCATG...

Genetic differences drive traits and disease susceptibility

Genetic variation

...C**A**CGTCACTTCAC**GTA**TG...
...CTC**C**TCACTTCAC---TG...

↑   ↑

SNP   Indel

...CTCGTCACTTCAC**GTA**TG...
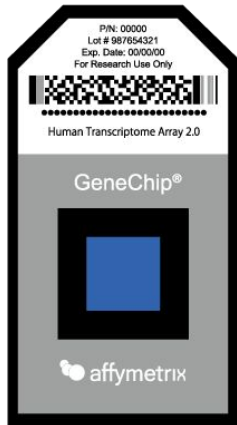...CTC**C**TCA**TTC**CAC**GTA**TG...

↑

"structural variation"

Words to confuse you!

"Variant" "Mutation" "Allele"

Genotype and Sequencing

Genotyping
- Cheap
- Probes limited number of known variants
- Common variants

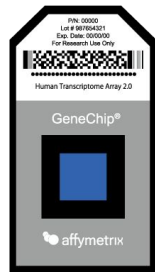**Imputation**
"Fills in the blanks" of known variants

Sequencing
- Becoming less expensive
- Reads entire genome or portion of genome
- Finds unknown variants
- Common or rare variants

The problem of phasing

...C**A**CGTCAC**T**TCAC**GTA**TG...
...CTC**C**TCAC**T**TCAC---TG...

Unphased
C/C
T/A
C/C
G/C
T/T
C/C
A/A
C/C
T/T
T/T
C/C
A/A
C/C
GTA/---
T/T
G/G

?

Phased
C|C
A|T
C|C
G|C
T|T
C|C
A|A
C|C
T|T
T|T
C|C
A|A
C|C
GTA|---
T|T
G|G

Phased vs Unphased and Inheritance

…CTCGTCAC**T**TCAC---TG…
…C**A**CGTCAC**T**TCAC**GTA**TG…

…CTC**C**TCAC**T**TCAC---TG…
…CTCGTCACGTCAC---TG…

Dad
T/A
G/G
T/T
---/GTA

Mom
T/T
G/C
G/T
---/---

…C**A**CGTCAC**T**TCAC**GTA**TG…
…CTC**C**TCAC**T**TCAC---TG…

Unphased
T/A
G/C
T/T
---/GTA

→ Family Data
or
Computers!
(Computational Phasing) →

Phased
A|T
G|C
T|T
GTA|---