The molecular basis of life

Sarah Graham
Language and Genetics Department
Max Planck Institute for Psycholinguistics

LSA Linguistic Institute
University of Michigan, Ann Arbor
29 June 2013

Universality and Variability

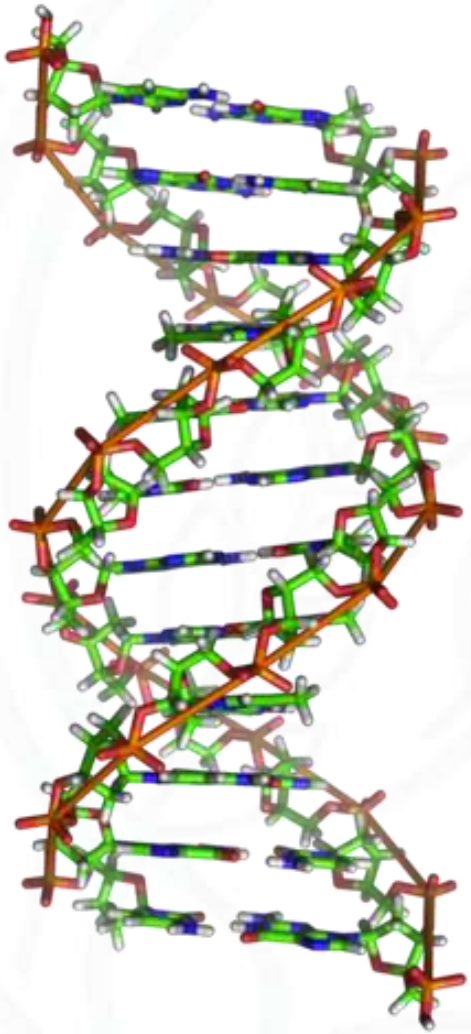
Part I: Universality

The core molecular principles underpinning living things.

Part II: Variability

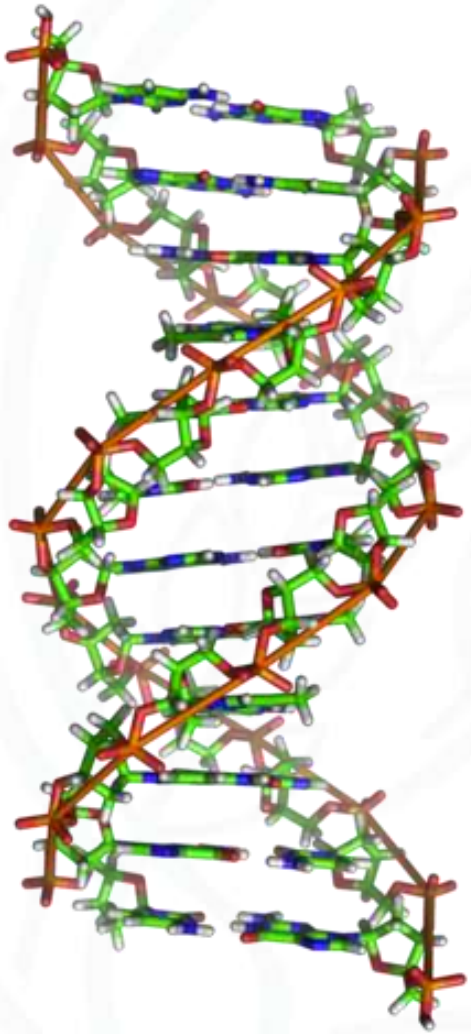
How humans differ from other species and from each other.

DNA



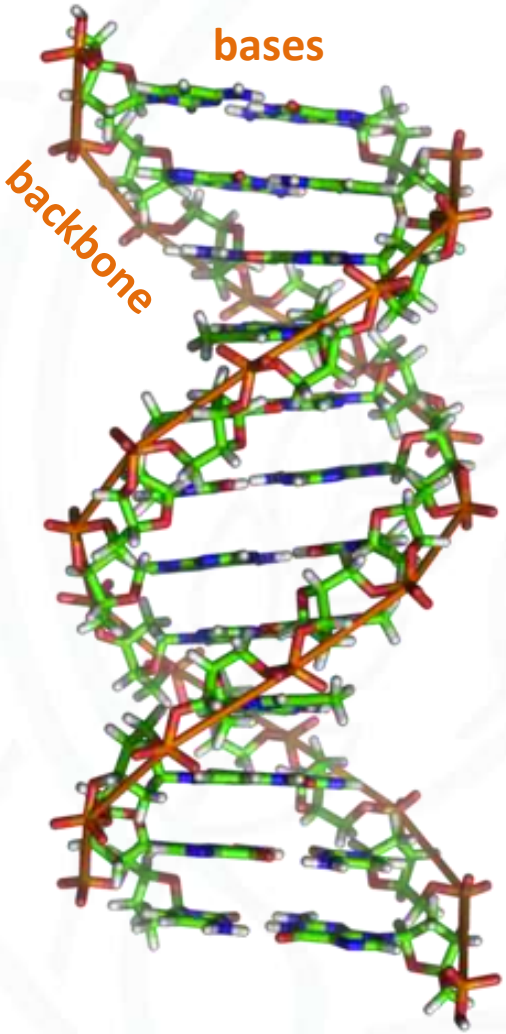
- **60 years** since this structure was worked out.
- What have we learnt?

DNA



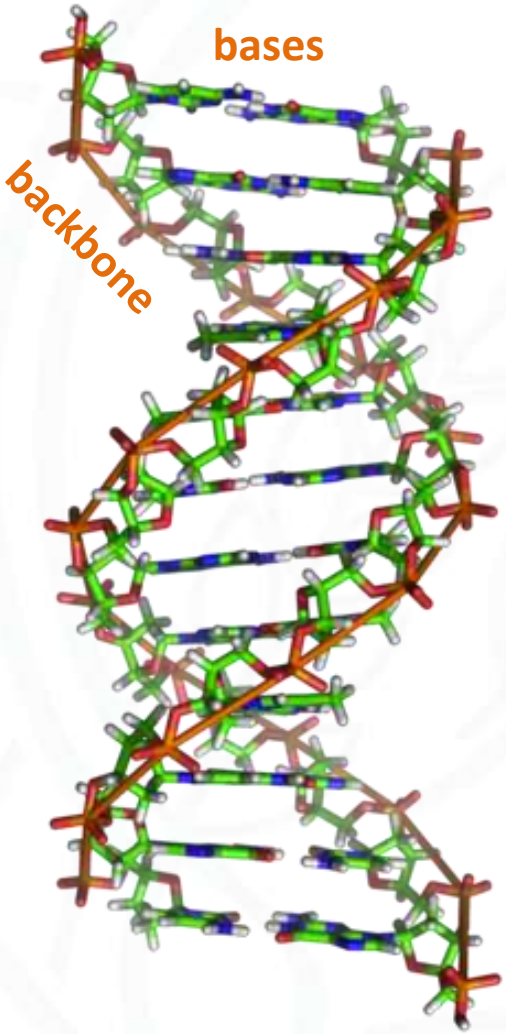
- **DNA** is the **genetic material** for all life on Earth.
- It stores the **information** which is passed on from parent to offspring.
- An organism's **genome** is all of its DNA.

DNA

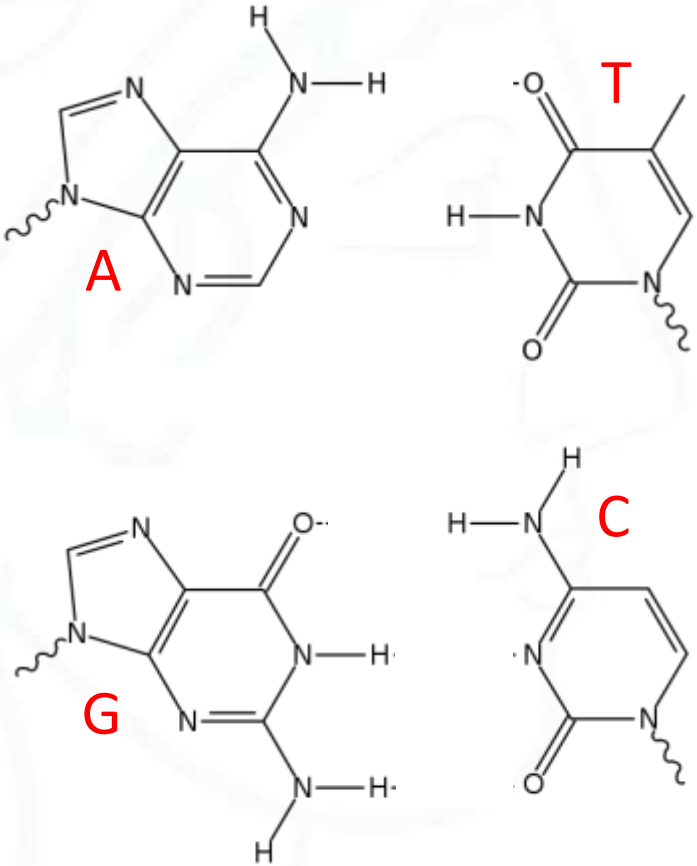


- DNA consists of two **strands** joined together as a **helix**.
- Each strand has a **backbone** supporting a string of **bases**.
- A base attached to a backbone is called a **nucleotide**.
- The backbone is for **structure**.
- The bases carry the **information**.

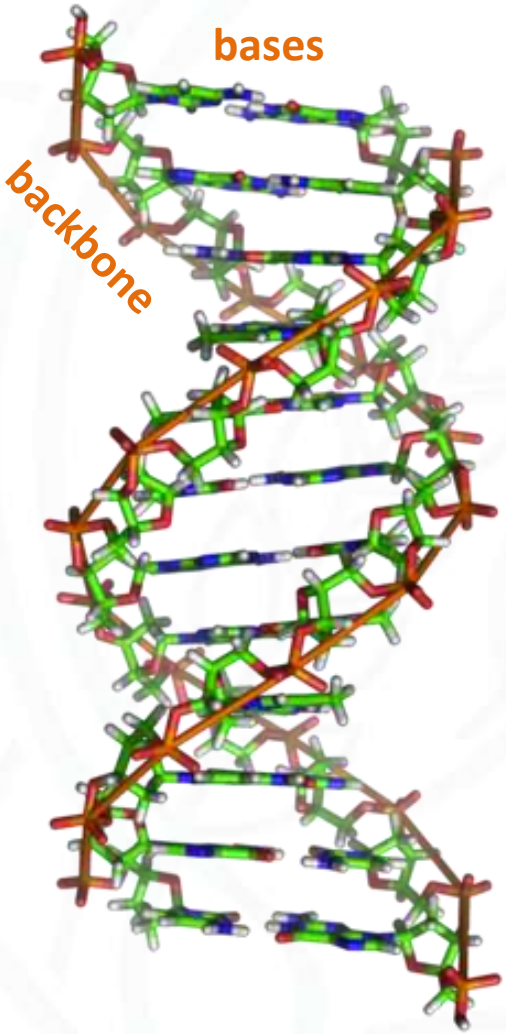
DNA



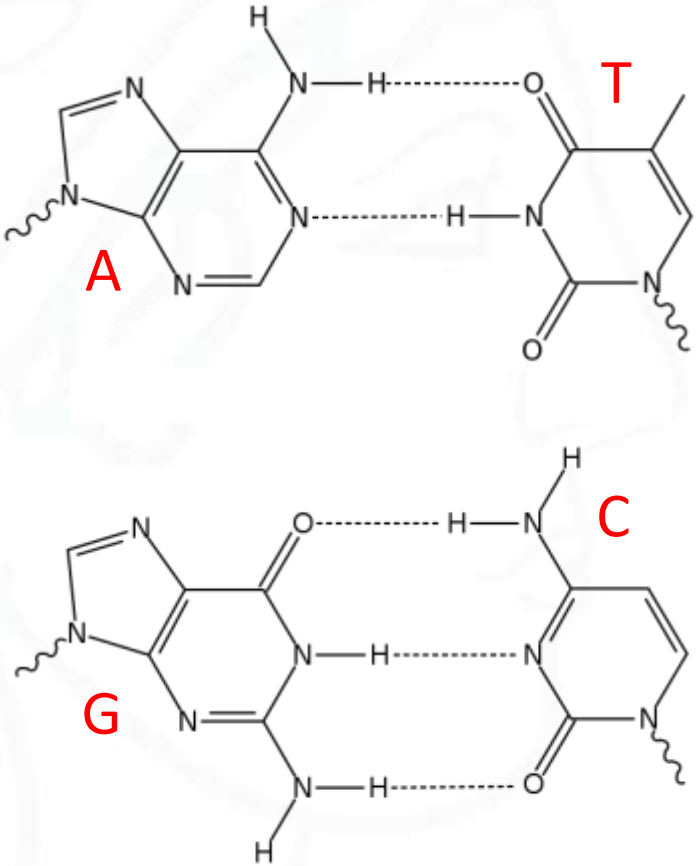
4 kinds of **base**:



DNA



4 kinds of **base**:

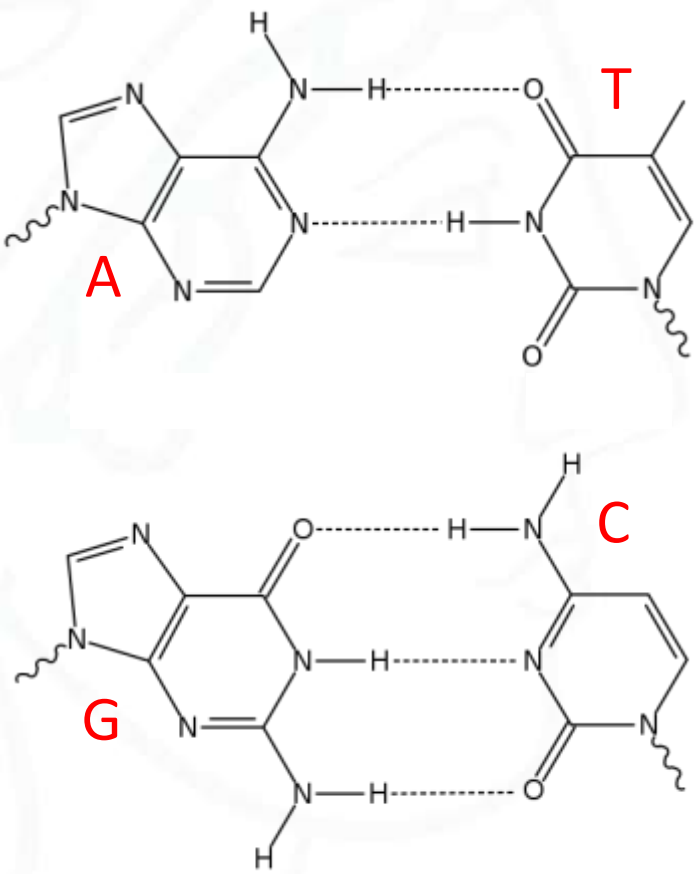


DNA

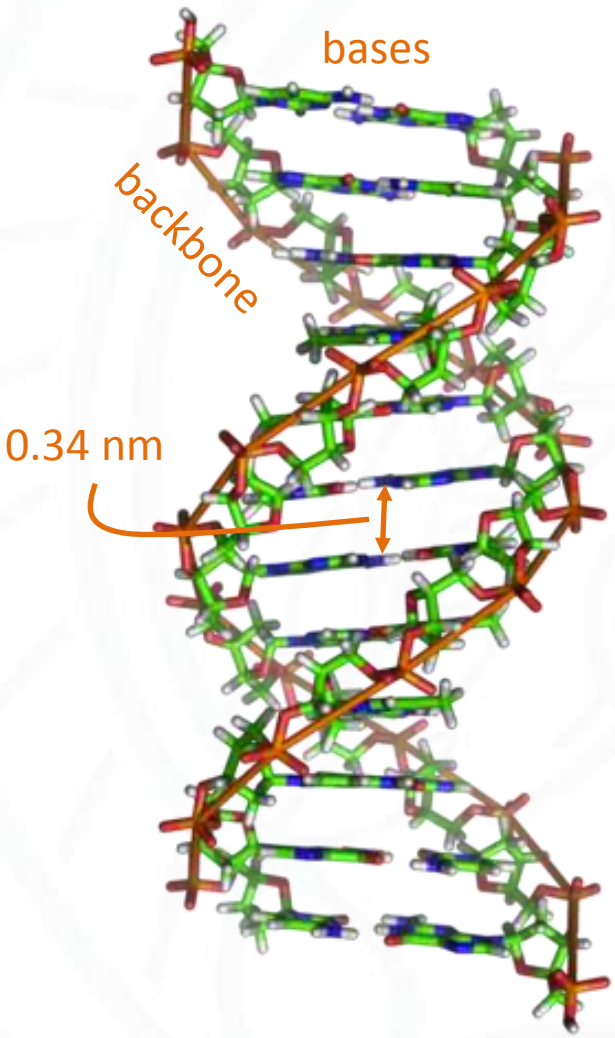
LANGUAGE AND GENETICS DEPARTMENT

- **Base pairing** provides the mechanism for the replication and use of DNA.
- The order of the bases is called the **sequence** of the DNA.
- If we know the sequence of one strand, we also know the sequence of the **opposite** strand.

4 kinds of base:

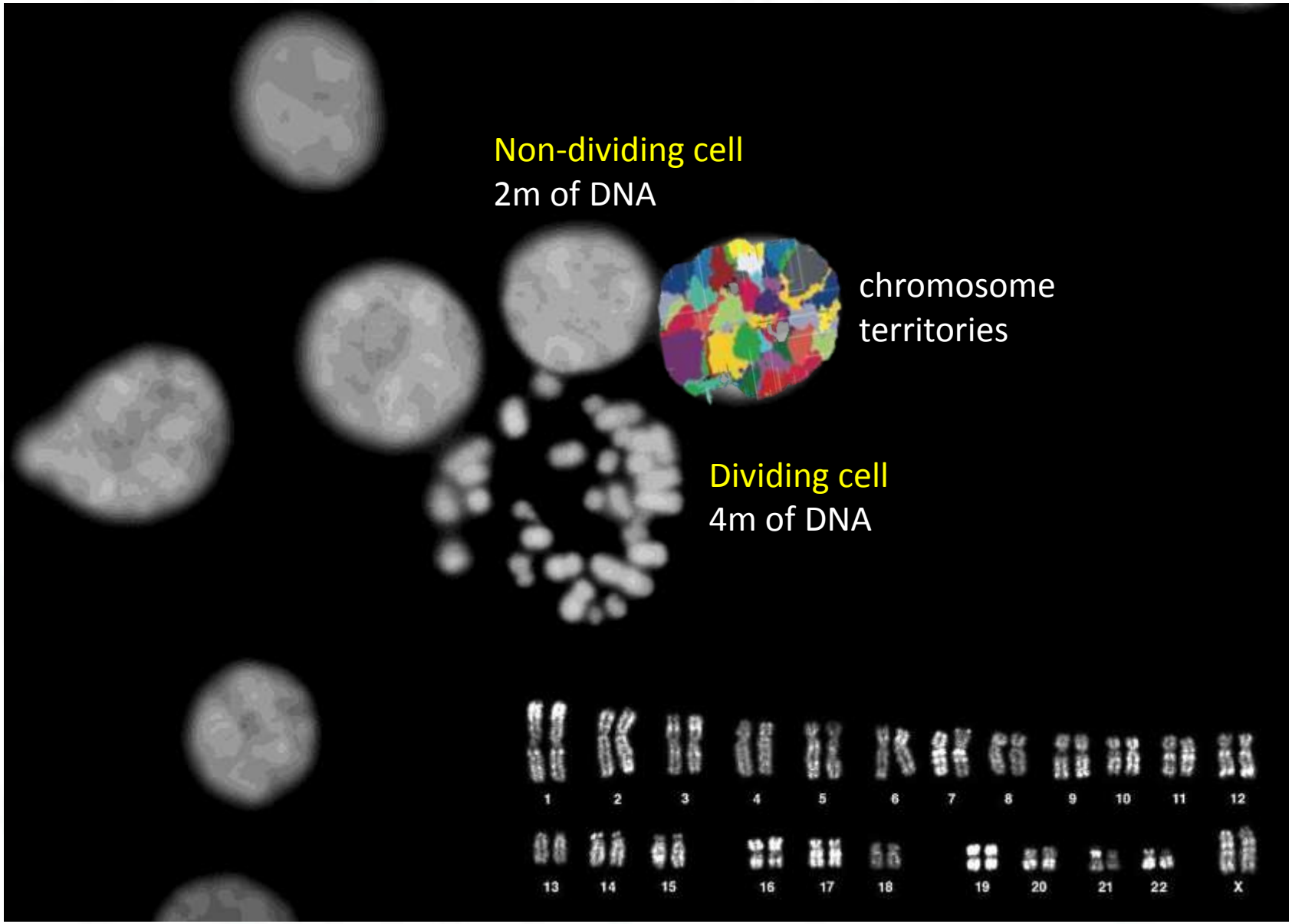


The human genome



- 3 billion bases
- 1 metre of DNA in 23 pieces (**chromosomes**)
- 2 copies of the genome in every cell - one from each parent
- 23 **pairs** of **chromosomes**
  - 1 pair of **sex chromosomes**
  - 22 pairs of **autosomes**

DNA content of human cells



A bit of the human genome

LANGUAGE AND GENETICS DEPARTMENT

TTCTGTGTGTGTCACCATGTCTTTGTGCTTCTGGGTGCTTCTGTGTTTGTTTCTGGCCGCGTTTCTGTGTTGGACAGGGGTGACTTTGTGCCGGATGGCTTCTGTGTGAGAGCGCGCGCG
AGTGTGCATGTCGGTGAGCTGGGAGGGTGTGTCTCAGTGTCTATGGCTGTGGTTCGGTATAAGTCTGAGCATGTCTGCCAGGGTGTATTTGTGCCTGTATGTGCGTGCCTCGGTGGGCAC
TCTCGTTTCCTTCCGAATGTGGGGCAGTGCCGGTGTGCTGCCCTCTGCCTTGAGACCTCAAGCCGCGCAGGCGCCCAGGGCAGGCAGGTAGCGGCCACAGAAGAGCCAAAAGCTCCCGGG
TTGGCTGGTAAGGACACCACCTCCAGCTTTAGCCCTCTGGGGCCAGCCAGGGTAGCCGGGAAGCAGTGGTGGCCCGCCCTCCAGGGAGCAGTTGGGCCCCGCCCGGGCCAGCCCCAGGAG
AAGGAGGGCGAGGGGAGGGGAGGGAAAGGGGAGGAGTGCCTCGCCCCTTCGCGGCTGCCGGCGTGCCATTGGCCGAAAGTTCCCGTACGTCACGGCGAGGGCAGTTCCCCTAAAGTCCTG
TGCACATAACGGGCAGAACGCACTGCGAAGCGGCTTCTTCAGAGCACGGGCTGGAACTGGCAGGCACCGCGAGCCCCTAGCACCCGACAAGCTGAGTGTGCAGGACGAGTCCCCACCACA
CCCACACCACAGCCGCTGAATGAGGCTTCCAGGCGTCCGCTCGCGGCCCGCAGAGCCCCGCCGTGGGTCCGCCCGCTGAGGCGCCCCCAGCCAGTGCGCTCACCTGCCAGACTGCGCGCC
ATGGGGCAACCCGGGAACGGCAGCGCCTTCTTGCTGGCACCCAATGGAAGCCATGCGCCGGACCACGACGTCACGCAGGAAAGGGACGAGGTGTGGGTGGTGGGCATGGGCATCGTCATG
TCTCTCATCGTCCTGGCCATCGTGTTTGGCAATGTGCTGGTCATCACAGCCATTGCCAAGTTCGAGCGTCTGCAGACGGTCACCAACTACTTCATCACTTCACTGGCCTGTGCTGATCTG
GTCATGGGCCTGGCAGTGGTGCCCTTTGGGGCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGAGTTTTGGACTTCCATTGATGTGCTGTGCGTCACGGCCAGC
ATTGAGACCCTGTGCGTGATCGCAGTGGATCGCTACTTTGCCATTACTTCACCTTTCAAGTACCAGAGCCTGCTGACCAAGAATAAGGCCCGGGTGATCATTCTGATGGTGTGGATTGTG
TCAGGCCTTACCTCCTTCTTGCCCATTCAGATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGCTGTGACTTCTTCACGAACCAAGCCTATGCC
ATTGCCTCTTCCATCGTGTCCTTCTACGTTCCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCAGCTCCAGAAGATTGACAAATCTGAGGGCCGCTTC
CATGTCCAGAACCTTAGCCAGGTGGAGCAGGATGGGCGGACGGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAAGGAGCACAAAGCCCTCAAGACGTTAGGCATCATCATGGGC
ACTTTCACCCTCTGCTGGCTGCCCTTCTTCATCGTTAACATTGTGCATGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTCCTAAATTGGATAGGCTATGTCAATTCTGGT
TTCAATCCCCTTATCTACTGCCGGAGCCCAGATTTCAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTATGGGAATGGCTACTCCAGCAACGGCAACACA
GGGGAGCAGAGTGGATATCACGTGGAACAGGAGAAAGAAAATAAACTGCTGTGTGAAGACCTCCCAGGCACGGAAGACTTTGTGGGCCATCAAGGTACTGTGCCTAGCGATAACATTGAT
TCACAAGGGAGGAATTGTAGTACAAATGACTCACTGCTGTAAAGCAGTTTTTCTACTTTTAAAGACCCCCCCCCCCAACAGAACACTAAACAGACTATTTAACTTGAGGGTAATAAACTT
AGAATAAAATTGTAAAATTGTATAGAGATATGCAGAAGGAAGGGCATCCTTCTGCCTTTTTTATTTTTTTAAGCTGTAAAAAGAGAGAAAACTTATTTGAGTGATTATTTGTTATTTGTA
CAGTTCAGTTCCTCTTTGCATGGAATTTGTAAGTTTATGTCTAAAGAGCTTTAGTCCTAGAGGACCTGAGTCTGCTATATTTTCATGACTTTTCCATGTATCTACCTCACTATTCAAGTA
TTAGGGGTAATATATTGCTGCTGGTAATTTGTATCTGAAGGAGATTTTCCTTCCTACACCCTTGGACTTGAGGATTTTGAGTATCTCGGACCTTTCAGCTGTGAACATGGACTCTTCCCC
CACTCCTCTTATTTGCTCACACGGGGTATTTTAGGCAGGGATTTGAGGAGCAGCTTCAGTTGTTTTCCCGAGCAAAGTCTAAAGTTTACAGTAAATAAATTGTTTGACCATGCCTTCATT
GCACCTGTTTCTCCAAAACCCCTTGACTGGAGTGCTGTTGCCTCCCCCACTGGAAACCGCAGGTAACTACTTGTAATTACTGCCCATGACTTAATGTAGAATGATACAAGAATGACATGC
ACAGATTGCTTAACCCTTTCATTTGCCTTTGAGTCTGCTGCTGCAAAGCTGCATCTCTCCTGACACTTGTGCCCCAAATCAGTTCTGCCTGCTCTTAGTATAGCTCAACTCTCCCTATGG
TTATTGTTCTGTGTTGTTACCTCAGAAACACTGACTCACAGAAGCGGAGTTAAGGGGATATGTTTTTTTCTCTCCACGTGCACCCACCACCCACCTTCCAGTTCTACTTGTTTCAAAACT
GTTTATATTTCTGTCTTGGCCATGTGTTACAGTGGAGCTCTTTGTACTGCATCAGGGCTTGGCATTTTAGGGATAAGGAAGATGTTCTTATGAGGAAGCTACTCAGACATGGCCCCGTAA
TTCTGAGGGAAAATTCAAAAGGCATTGGTCATGGGGAGAAAAGCTGGAGAACACATAACTGATGGATACCTCATGAACTAGAAACAGAATTTTAACCCCTTTTCCTTCTTTCCTTTGGTC

A gene

LANGUAGE AND GENETICS DEPARTMENT

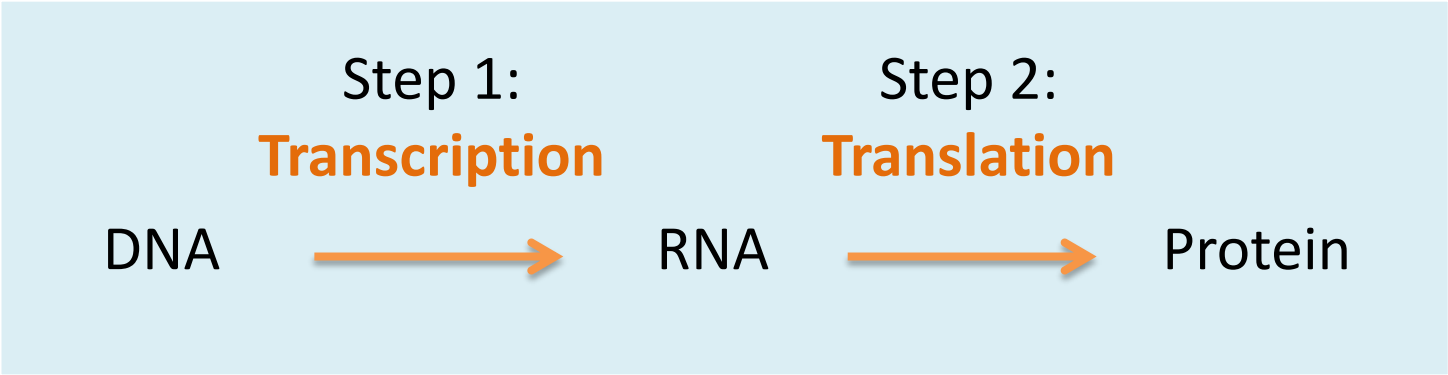
TTCTGTGTGTGTCACCATGTCTTTGTGCTTCTGGGTGCTTCTGTGTTTGTTTCTGGCCGCGTTTCTGTGTTGGACAGGGGTGACTTTGTGCCGGATGGCTTCTGTGTGAGAGCGCGCGCG
AGTGTGCATGTCGGTGAGCTGGGAGGGTGTGTCTCAGTGTCTATGGCTGTGGTTCGGTATAAGTCTGAGCATGTCTGCCAGGGTGTATTTGTGCCTGTATGTGCGTGCCTCGGTGGGCAC
TCTCGTTTCCTTCCGAATGTGGGGCAGTGCCGGTGTGCTGCCCTCTGCCTTGAGACCTCAAGCCGCGCAGGCGCCCAGGGCAGGCAGGTAGCGGCCACAGAAGAGCCAAAAGCTCCCGGG
TTGGCTGGTAAGGACACCACCTCCAGCTTTAGCCCTCTGGGGCCAGCCAGGGTAGCCGGGAAGCAGTGGTGGCCCGCCCTCCAGGGAGCAGTTGGGCCCCGCCCGGGCCAGCCCCAGGAG
AAGGAGGGCGAGGGGAGGGGAGGGAAAGGGGAGGAGTGCCTCGCCCCTTCGCGGCTGCCGGCGTGCCATTGGCCGAAAGTTCCCGTACGTCACGGCGAGGGCAGTTCCCCTAAAGTCCTG
TGCACATAACGGGCAGAACGCACTGCGAAGCGGCTTCTTCAGAGCACGGGCTGGAACTGGCAGGCACCGCGAGCCCCTAGCACCCGACAAGCTGAGTGTGCAGGACGAGTCCCCACCACA
CCCACACCACAGCCGCTGAATGAGGCTTCCAGGCGTCCGCTCGCGGCCCGCAGAGCCCCGCCGTGGGTCCGCCCGCTGAGGCGCCCCCAGCCAGTGCGCTCACCTGCCAGACTGCGCGCC
ATGGGGCAACCCGGGAACGGCAGCGCCTTCTTGCTGGCACCCAATGGAAGCCATGCGCCGGACCACGACGTCACGCAGGAAAGGGACGAGGTGTGGGTGGTGGGCATGGGCATCGTCATG
TCTCTCATCGTCCTGGCCATCGTGTTTGGCAATGTGCTGGTCATCACAGCCATTGCCAAGTTCGAGCGTCTGCAGACGGTCACCAACTACTTCATCACTTCACTGGCCTGTGCTGATCTG
GTCATGGGCCTGGCAGTGGTGCCCTTTGGGGCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGAGTTTTGGACTTCCATTGATGTGCTGTGCGTCACGGCCAGC
ATTGAGACCCTGTGCGTGATCGCAGTGGATCGCTACTTTGCCATTACTTCACCTTTCAAGTACCAGAGCCTGCTGACCAAGAATAAGGCCCGGGTGATCATTCTGATGGTGTGGATTGTG
TCAGGCCTTACCTCCTTCTTGCCCATTCAGATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGCTGTGACTTCTTCACGAACCAAGCCTATGCC
ATTGCCTCTTCCATCGTGTCCTTCTACGTTCCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCAGCTCCAGAAGATTGACAAATCTGAGGGCCGCTTC
CATGTCCAGAACCTTAGCCAGGTGGAGCAGGATGGGCGGACGGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAAGGAGCACAAAGCCCTCAAGACGTTAGGCATCATCATGGGC
ACTTTCACCCTCTGCTGGCTGCCCTTCTTCATCGTTAACATTGTGCATGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTCCTAAATTGGATAGGCTATGTCAATTCTGGT
TTCAATCCCCTTATCTACTGCCGGAGCCCAGATTTCAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTATGGGAATGGCTACTCCAGCAACGGCAACACA
GGGGAGCAGAGTGGATATCACGTGGAACAGGAGAAAGAAAATAAACTGCTGTGTGAAGACCTCCCAGGCACGGAAGACTTTGTGGGCCATCAAGGTACTGTGCCTAGCGATAACATTGAT
TCACAAGGGAGGAATTGTAGTACAAATGACTCACTGCTGTAAAGCAGTTTTTCTACTTTTAAAGACCCCCCCCCCCAACAGAACACTAAACAGACTATTTAACTTGAGGGTAATAAACTT
AGAATAAAATTGTAAAATTGTATAGAGATATGCAGAAGGAAGGGCATCCTTCTGCCTTTTTTATTTTTTTAAGCTGTAAAAAGAGAGAAAACTTATTTGAGTGATTATTTGTTATTTGTA
CAGTTCAGTTCCTCTTTGCATGGAATTTGTAAGTTTATGTCTAAAGAGCTTTAGTCCTAGAGGACCTGAGTCTGCTATATTTTCATGACTTTTCCATGTATCTACCTCACTATTCAAGTA
TTAGGGGTAATATATTGCTGCTGGTAATTTGTATCTGAAGGAGATTTTCCTTCCTACACCCTTGGACTTGAGGATTTTGAGTATCTCGGACCTTTCAGCTGTGAACATGGACTCTTCCCC
CACTCCTCTTATTTGCTCACACGGGGTATTTTAGGCAGGGATTTGAGGAGCAGCTTCAGTTGTTTTCCCGAGCAAAGTCTAAAGTTTACAGTAAATAAATTGTTTGACCATGCCTTCATT
GCACCTGTTTCTCCAAAACCCCTTGACTGGAGTGCTGTTGCCTCCCCCACTGGAAACCGCAGGTAACTACTTGTAATTACTGCCCATGACTTAATGTAGAATGATACAAGAATGACATGC
ACAGATTGCTTAACCCTTTCATTTGCCTTTGAGTCTGCTGCTGCAAAGCTGCATCTCTCCTGACACTTGTGCCCCAAATCAGTTCTGCCTGCTCTTAGTATAGCTCAACTCTCCCTATGG
TTATTGTTCTGTGTTGTTACCTCAGAAACACTGACTCACAGAAGCGGAGTTAAGGGGATATGTTTTTTTCTCTCCACGTGCACCCACCACCCACCTTCCAGTTCTACTTGTTTCAAAACT
GTTTATATTTCTGTCTTGGCCATGTGTTACAGTGGAGCTCTTTGTACTGCATCAGGGCTTGGCATTTTAGGGATAAGGAAGATGTTCTTATGAGGAAGCTACTCAGACATGGCCCCGTAA
TTCTGAGGGAAAATTCAAAAGGCATTGGTCATGGGGAGAAAAGCTGGAGAACACATAACTGATGGATACCTCATGAACTAGAAACAGAATTTTAACCCCTTTTCCTTCTTTCCTTTGGTC

What is a gene?

- A **gene** is a region of DNA that gives the instructions for making a **protein**.
- DNA stores **information** – proteins do **work**.

What is a gene?

- A **gene** is a region of DNA that gives the instructions for making a **protein**.
- DNA stores **information** – proteins do **work**.



Step 1: Transcription

LANGUAGE AND GENETICS DEPARTMENT

DNA

TTCTGTGTGTGTCACCATGTCTTTGTGCTTCTGGGTGCTTCTGTGTTTGTTTCTGGCCGCGTTTCTGTGTTGGACAGGGGTGACTTTGTGCCGGATGGCTTCTGTGTGAGAGCGCGCGCG
AGTGTGCATGTCGGTGAGCTGGGAGGGTGTGTCTCAGTGTCTATGGCTGTGGTTCGGTATAAGTCTGAGCATGTCTGCCAGGGTGTATTTGTGCCTGTATGTGCGTGCCTCGGTGGGCAC
TCTCGTTTCCTTCCGAATGTGGGGCAGTGCCGGTGTGCTGCCCTCTGCCTTGAGACCTCAAGCCGCGCAGGCGCCCAGGGCAGGCAGGTAGCGGCCACAGAAGAGCCAAAAGCTCCCGGG
TTGGCTGGTAAGGACACCACCTCCAGCTTTAGCCCTCTGGGGCCAGCCAGGGTAGCCGGGAAGCAGTGGTGGCCCGCCCTCCAGGGAGCAGTTGGGCCCCGCCCGGGCCAGCCCCAGGAG
AAGGAGGGCGAGGGGAGGGGAGGGAAAGGGGAGGAGTGCCTCGCCCCTTCGCGGCTGCCGGCGTGCCATTGGCCGAAAGTTCCCGTACGTCACGGCGAGGGCAGTTCCCCTAAAGTCCTG
TGCACATAACGGGCAGAACGCACTGCGAAGCGGCTTCTTCAGAGCACGGGCTGGAACTGGCAGGCACCGCGAGCCCCTAGCACCCGACAAGCTGAGTGTGCAGGACGAGTCCCCACCACA
CCCACACCACAGCCGCTGAATGAGGCTTCCAGGCGTCCGCTCGCGGCCCGCAGAGCCCCGCCGTGGGTCCGCCCGCTGAGGCGCCCCCAGCCAGTGCGCTCACCTGCCAGACTGCGCGCC
ATGGGGCAACCCGGGAACGGCAGCGCCTTCTTGCTGGCACCCAATGGAAGCCATGCGCCGGACCACGACGTCACGCAGGAAAGGGACGAGGTGTGGGTGGTGGGCATGGGCATCGTCATG
TCTCTCATCGTCCTGGCCATCGTGTTTGGCAATGTGCTGGTCATCACAGCCATTGCCAAGTTCGAGCGTCTGCAGACGGTCACCAACTACTTCATCACTTCACTGGCCTGTGCTGATCTG
GTCATGGGCCTGGCAGTGGTGCCCTTTGGGGCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGAGTTTTGGACTTCCATTGATGTGCTGTGCGTCACGGCCAGC
ATTGAGACCCTGTGCGTGATCGCAGTGGATCGCTACTTTGCCATTACTTCACCTTTCAAGTACCAGAGCCTGCTGACCAAGAATAAGGCCCGGGTGATCATTCTGATGGTGTGGATTGTG
TCAGGCCTTACCTCCTTCTTGCCCATTCAGATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGCTGTGACTTCTTCACGAACCAAGCCTATGCC
ATTGCCTCTTCCATCGTGTCCTTCTACGTTCCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCAGCTCCAGAAGATTGACAAATCTGAGGGCCGCTTC
CATGTCCAGAACCTTAGCCAGGTGGAGCAGGATGGGCGGACGGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAAGGAGCACAAAGCCCTCAAGACGTTAGGCATCATCATGGGC
ACTTTCACCCTCTGCTGGCTGCCCTTCTTCATCGTTAACATTGTGCATGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTCCTAAATTGGATAGGCTATGTCAATTCTGGT
TTCAATCCCCTTATCTACTGCCGGAGCCCAGATTTCAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTATGGGAATGGCTACTCCAGCAACGGCAACACA
GGGGAGCAGAGTGGATATCACGTGGAACAGGAGAAAGAAAATAAACTGCTGTGTGAAGACCTCCCAGGCACGGAAGACTTTGTGGGCCATCAAGGTACTGTGCCTAGCGATAACATTGAT
TCACAAGGGAGGAATTGTAGTACAAATGACTCACTGCTGTAAAGCAGTTTTTCTACTTTTAAAGACCCCCCCCCCCAACAGAACACTAAACAGACTATTTAACTTGAGGGTAATAAACTT
AGAATAAAATTGTAAAATTGTATAGAGATATGCAGAAGGAAGGGCATCCTTCTGCCTTTTTTATTTTTTTAAGCTGTAAAAAGAGAGAAAACTTATTTGAGTGATTATTTGTTATTTGTA
CAGTTCAGTTCCTCTTTGCATGGAATTTGTAAGTTTATGTCTAAAGAGCTTTAGTCCTAGAGGACCTGAGTCTGCTATATTTTCATGACTTTTCCATGTATCTACCTCACTATTCAAGTA
TTAGGGGTAATATATTGCTGCTGGTAATTTGTATCTGAAGGAGATTTTCCTTCCTACACCCTTGGACTTGAGGATTTTGAGTATCTCGGACCTTTCAGCTGTGAACATGGACTCTTCCCC
CACTCCTCTTATTTGCTCACACGGGGTATTTTAGGCAGGGATTTGAGGAGCAGCTTCAGTTGTTTTCCCGAGCAAAGTCTAAAGTTTACAGTAAATAAATTGTTTGACCATGCCTTCATT
GCACCTGTTTCTCCAAAACCCCTTGACTGGAGTGCTGTTGCCTCCCCCACTGGAAACCGCAGGTAACTACTTGTAATTACTGCCCATGACTTAATGTAGAATGATACAAGAATGACATGC
ACAGATTGCTTAACCCTTTCATTTGCCTTTGAGTCTGCTGCTGCAAAGCTGCATCTCTCCTGACACTTGTGCCCCAAATCAGTTCTGCCTGCTCTTAGTATAGCTCAACTCTCCCTATGG
TTATTGTTCTGTGTTGTTACCTCAGAAACACTGACTCACAGAAGCGGAGTTAAGGGGATATGTTTTTTTCTCTCCACGTGCACCCACCACCCACCTTCCAGTTCTACTTGTTTCAAAACT
GTTTATATTTCTGTCTTGGCCATGTGTTACAGTGGAGCTCTTTGTACTGCATCAGGGCTTGGCATTTTAGGGATAAGGAAGATGTTCTTATGAGGAAGCTACTCAGACATGGCCCCGTAA
TTCTGAGGGAAAATTCAAAAGGCATTGGTCATGGGGAGAAAAGCTGGAGAACACATAACTGATGGATACCTCATGAACTAGAAACAGAATTTTAACCCCTTTTCCTTCTTTCCTTTGGTC

Step 1: Transcription

LANGUAGE AND GENETICS DEPARTMENT

RNA

GCACAUAACGGGCAGAACGCACUGCGAAGCGGCUUCUUCAGAGCACGGGCUGGAACUGGCAGGCACCGCGAGCCCCUAGCACCCGACAAGCUGAGUGUGCAGGACGAGUCCCCACCACA
CCCACACCACAGCCGCUGAAUGAGGCUUCCAGGCGUCCGCUCGCGGCCCGCAGAGCCCCGCCGUGGGUCCGCCCGCUGAGGCGCCCCCAGCCAGUGCGCUCACCUGCCAGACUGCGCGCC
AUGGGGCAACCCGGGAACGGCAGCGCCUUCUUGCUGGCACCCAAUGGAAGCCAUGCGCCGGACCACGACGUCACGCAGGAAAGGGACGAGGUGUGGGUGGUGGGCAUGGGCAUCGUCAUG
UCUCUCAUCGUCCUGGCCAUCGUGUUUGGCAAUGUGCUGGUCAUCACAGCCAUUGCCAAGUUCGAGCGUCUGCAGACGGUCACCAACUACUUCAUCACUUCACUGGCCUGUGCUGAUCUG
GUCAUGGGCCUGGCAGUGGUGCCCUUUGGGGCCGCCCAUAUUCUUAUGAAAAUGUGGACUUUUGGCAACUUCUGGUGCGAGUUUUGGACUUCCAUUGAUGUGCUGUGCGUCACGGCCAGC
AUUGAGACCCUGUGCGUGAUCGCAGUGGAUCGCUACUUUGCCAUUACUUCACCUUUCAAGUACCAGAGCCUGCUGACCAAGAAUAAGGCCCGGGUGAUCAUUCUGAUGGUGUGGAUUGUG
UCAGGCCUUACCUCCUUCUUGCCCAUUCAGAUGCACUGGUACCGGGCCACCCACCAGGAAGCCAUCAACUGCUAUGCCAAUGAGACCUGCUGUGACUUCUUCACGAACCAAGCCUAUGCC
AUUGCCUCUUCCAUCGUGUCCUUCUACGUUCCCCUGGUGAUCAUGGUCUUCGUCUACUCCAGGGUCUUUCAGGAGGCCAAAAGGCAGCUCCAGAAGAUUGACAAAUCUGAGGGCCGCUUC
CAUGUCCAGAACCUUAGCCAGGUGGAGCAGGAUGGGCGGACGGGGCAUGGACUCCGCAGAUCUUCCAAGUUCUGCUUGAAGGAGCACAAAGCCCUCAAGACGUUAGGCAUCAUCAUGGGC
ACUUUCACCCUCUGCUGGCUGCCCUUCUUCAUCGUUAACAUUGUGCAUGUGAUCCAGGAUAACCUCAUCCGUAAGGAAGUUUACAUCCUCCUAAAUUGGAUAGGCUAUGUCAAUUCUGGU
UUCAAUCCCCUUAUCUACUGCCGGAGCCCAGAUUUCAGGAUUGCCUUCCAGGAGCUUCUGUGCCUGCGCAGGUCUUCUUUGAAGGCCUAUGGGAAUGGCUACUCCAGCAACGGCAACACA
GGGGAGCAGAGUGGAUAUCACGUGGAACAGGAGAAAGAAAAUAAACUGCUGUGUGAAGACCUCCCAGGCACGGAAGACUUUGUGGGCCAUCAAGGUACUGUGCCUAGCGAUAACAUUGAU
UCACAAGGGAGGAAUUGUAGUACAAAUGACUCACUGCUGUAAAGCAGUUUUUCUACUUUUAAAGACCCCCCCCCCCAACAGAACACUAAACAGACUAUUUAACUUGAGGGUAAUAAACUU
AGAAUAAAAUUGUAAAAUUGUAUAGAGAUAUGCAGAAGGAAGGGCAUCCUUCUGCCUUUUUUAUUUUUUUAAGCUGUAAAAAGAGAGAAAACUUAUUUGAGUGAUUAUUUGUUAUUUGUA
CAGUUCAGUUCCUCUUUGCAUGGAAUUUGUAAGUUUAUGUCUAAAGAGCUUUAGUCCUAGAGGACCUGAGUCUGCUAUAUUUUCAUGACUUUUCCAUGUAUCUACCUCACUAUUCAAGUA
UUAGGGGUAAUAUAUUGCUGCUGGUAAUUUGUAUCUGAAGGAGAUUUUCCUUCCUACACCCUUGGACUUGAGGAUUUUGAGUAUCUCGGACCUUUCAGCUGUGAACAUGGACUCUUCCCC
CACUCCUCUUAUUUGCUCACACGGGGUAUUUUAGGCAGGGAUUUGAGGAGCAGCUUCAGUUGUUUUCCCGAGCAAAGUCUAAAGUUUACAGUAAAUAAAUUGUUUGACCAUGCCUUCAUU
GC

RNA vs DNA

- Single strand not double helix.
- Slightly different backbone.
- Uses the base **U** instead of **T**.

Step 2: Translation

RNA	➔	Protein
string of **nucleotides**		string of **amino acids**

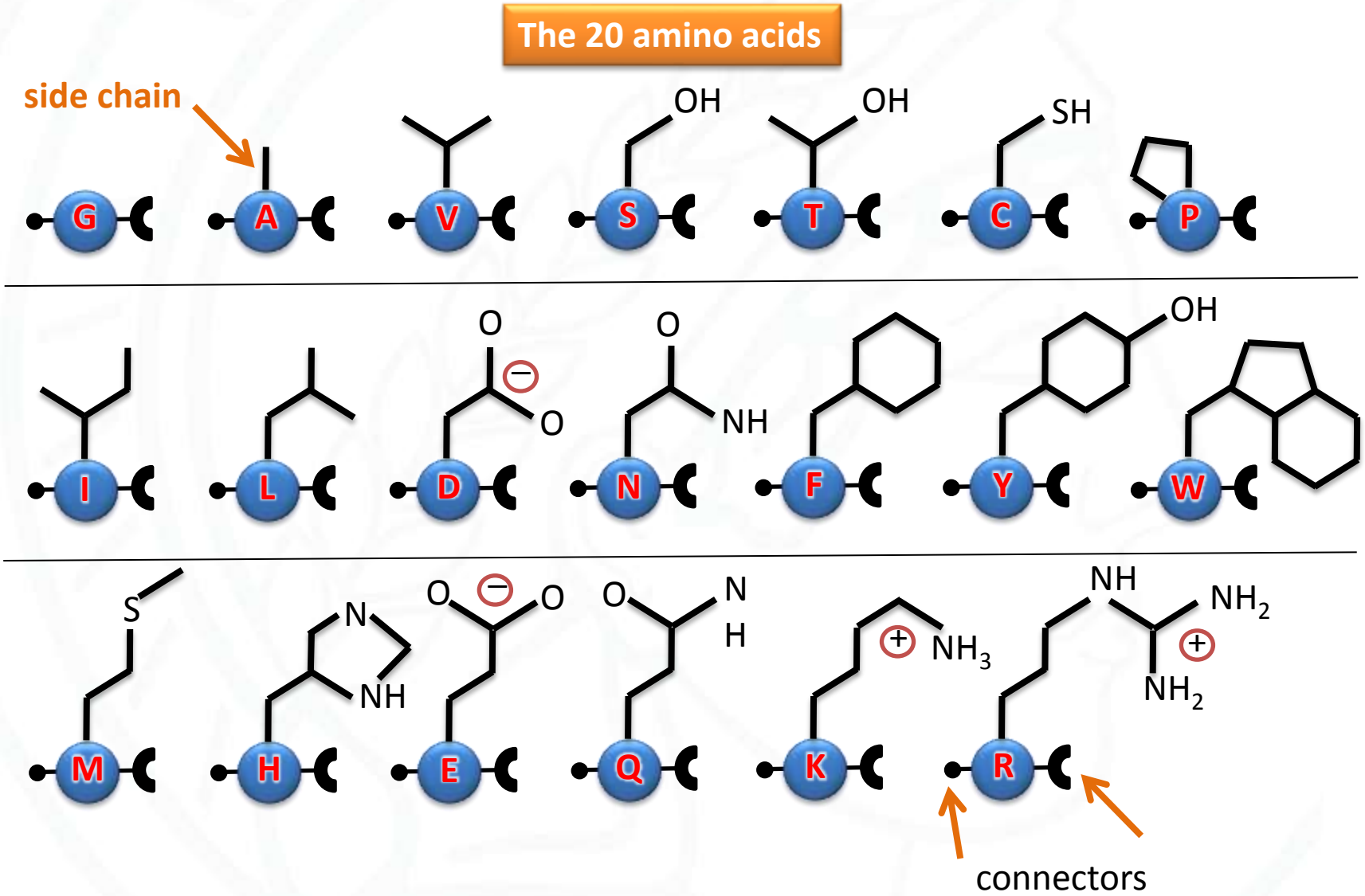
Step 2: Translation

LANGUAGE AND GENETICS DEPARTMENT

The 20 amino acids

G	A	V	S	T	C	P
Glycine	Alanine	Valine	Serine	Threonine	Cysteine	Proline
I	L	D	N	F	Y	W
Isoleucine	Leucine	Aspartate	Asparagine	Phenylalanine	Tyrosine	Tryptophan
M	H	E	Q	K	R	
Methionine	Histidine	Glutamate	Glutamine	Lysine	Arginine	

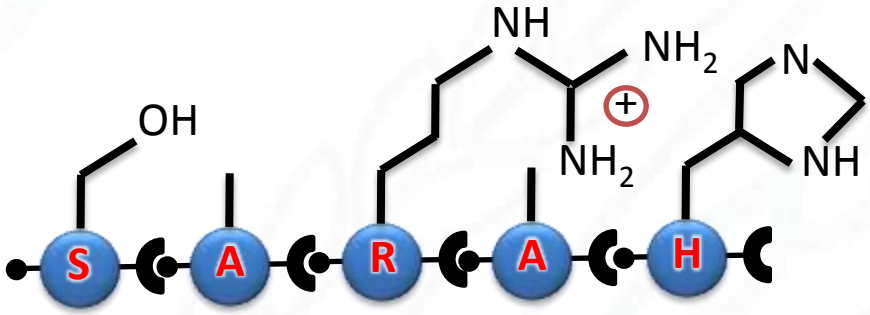
Step 2: Translation



Step 2: Translation

Side chains

Backbone

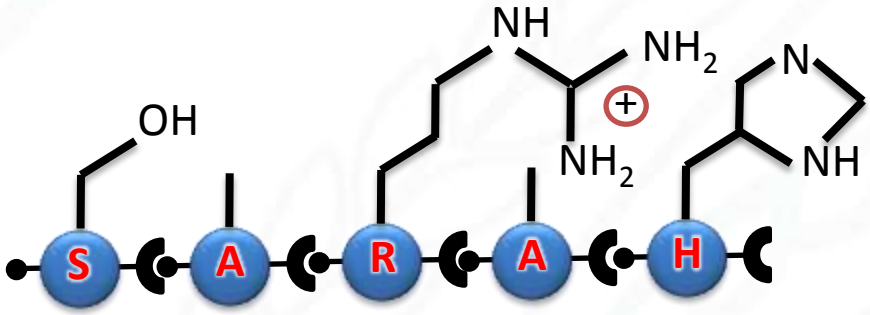


- **Amino acids** can be connected **in any order** (**sequence**) to make a **protein**.
- Proteins are typically 200-2000 amino acids.

Step 2: Translation

Side chains

Backbone



- **Amino acids** can be connected **in any order** (**sequence**) to make a **protein**.
- Proteins are typically 200-2000 amino acids.

RNA	→	Protein
string of **nucleotides**		string of **amino acids**

Step 2: Translation

Each amino acid is encoded by a set of 3 consecutive nucleotides – a **codon** – following the **genetic code**.

UUU	F
UUC	
UUA	
UUG	
CUU	L
CUC	
CUA	
CUG	
AUU	
AUC	I
AUA	
AUG	M
GUU	
GUC	V
GUA	
GUG	

UCU	
UCC	S
UCA	
UCG	
CCU	
CCC	P
CCA	
CCG	
ACU	
ACC	T
ACA	
ACG	
GCU	
GCC	A
GCA	
GCG	

UAU	Y
UAC	
UAA	STOP
UAG	
CAU	H
CAC	
CAA	Q
CAG	
AAU	N
AAC	
AAA	K
AAG	
GAU	D
GAC	
GAA	E
GAG	

UGU	C
UGC	
UGA	STOP
UGG	W
CGU	
CGC	
CGA	R
CGG	
AGA	
AGG	
AGU	S
AGC	
GGU	
GGC	G
GGA	
GGG	

If we know the code, we can ‘read the genome’.

Step 2: Translation

LANGUAGE AND GENETICS DEPARTMENT

RNA

GCACAUAACGGGCAGAACGCACUGCGAAGCGGCUUCUUCAGAGCACGGGCUGGAACUGGCAGGCACCGCGAGCCCCUAGCACCCGACAAGCUGAGUGUGCAGGACGAGUCCCCACCACA
CCCACACCACAGCCGCUGAAUGAGGCUUCCAGGCGUCCGCUCGCGGCCCGCAGAGCCCCGCCGUGGGUCCGCCCGCUGAGGCGCCCCCAGCCAGUGCGCUCACCUGCCAGACUGCGCGCC
AUGGGGCAACCCGGGAACGGCAGCGCCUUCUUGCUGGCACCCAAUGGAAGCCAUGCGCCGGACCACGACGUCACGCAGGAAAGGGACGAGGUGUGGGUGGUGGGCAUGGGCAUCGUCAUG
UCUCUCAUCGUCCUGGCCAUCGUGUUUGGCAAUGUGCUGGUCAUCACAGCCAUUGCCAAGUUCGAGCGUCUGCAGACGGUCACCAACUACUUCAUCACUUCACUGGCCUGUGCUGAUCUG
GUCAUGGGCCUGGCAGUGGUGCCCUUUGGGGCCGCCCAUAUUCUUAUGAAAAUGUGGACUUUUGGCAACUUCUGGUGCGAGUUUUGGACUUCCAUUGAUGUGCUGUGCGUCACGGCCAGC
AUUGAGACCCUGUGCGUGAUCGCAGUGGAUCGCUACUUUGCCAUUACUUCACCUUUCAAGUACCAGAGCCUGCUGACCAAGAAUAAGGCCCGGGUGAUCAUUCUGAUGGUGUGGAUUGUG
UCAGGCCUUACCUCCUUCUUGCCCAUUCAGAUGCACUGGUACCGGGCCACCCACCAGGAAGCCAUCAACUGCUAUGCCAAUGAGACCUGCUGUGACUUCUUCACGAACCAAGCCUAUGCC
AUUGCCUCUUCCAUCGUGUCCUUCUACGUUCCCCUGGUGAUCAUGGUCUUCGUCUACUCCAGGGUCUUUCAGGAGGCCAAAAGGCAGCUCCAGAAGAUUGACAAAUCUGAGGGCCGCUUC
CAUGUCCAGAACCUUAGCCAGGUGGAGCAGGAUGGGCGGACGGGGCAUGGACUCCGCAGAUCUUCCAAGUUCUGCUUGAAGGAGCACAAAGCCCUCAAGACGUUAGGCAUCAUCAUGGGC
ACUUUCACCCUCUGCUGGCUGCCCUUCUUCAUCGUUAACAUUGUGCAUGUGAUCCAGGAUAACCUCAUCCGUAAGGAAGUUUACAUCCUCCUAAAUUGGAUAGGCUAUGUCAAUUCUGGU
UUCAAUCCCCUUAUCUACUGCCGGAGCCCAGAUUUCAGGAUUGCCUUCCAGGAGCUUCUGUGCCUGCGCAGGUCUUCUUUGAAGGCCUAUGGGAAUGGCUACUCCAGCAACGGCAACACA
GGGGAGCAGAGUGGAUAUCACGUGGAACAGGAGAAAGAAAAUAAACUGCUGUGUGAAGACCUCCCAGGCACGGAAGACUUUGUGGGCCAUCAAGGUACUGUGCCUAGCGAUAACAUUGAU
UCACAAGGGAGGAAUUGUAGUACAAAUGACUCACUGCUGUAAAGCAGUUUUUCUACUUUUAAAGACCCCCCCCCCCAACAGAACACUAAACAGACUAUUUAACUUGAGGGUAAUAAACUU
AGAAUAAAAUUGUAAAAUUGUAUAGAGAUAUGCAGAAGGAAGGGCAUCCUUCUGCCUUUUUUAUUUUUUUAAGCUGUAAAAAGAGAGAAAACUUAUUUGAGUGAUUAUUUGUUAUUUGUA
CAGUUCAGUUCCUCUUUGCAUGGAAUUUGUAAGUUUAUGUCUAAAGAGCUUUAGUCCUAGAGGACCUGAGUCUGCUAUAUUUUCAUGACUUUUCCAUGUAUCUACCUCACUAUUCAAGUA
UUAGGGGUAAUAUAUUGCUGCUGGUAAUUUGUAUCUGAAGGAGAUUUUCCUUCCUACACCCUUGGACUUGAGGAUUUUGAGUAUCUCGGACCUUUCAGCUGUGAACAUGGACUCUUCCCC
CACUCCUCUUAUUUGCUCACACGGGGUAUUUUAGGCAGGGAUUUGAGGAGCAGCUUCAGUUGUUUUCCCGAGCAAAGUCUAAAGUUUACAGUAAAUAAAUUGUUUGACCAUGCCUUCAUU
GC

Step 2: Translation

LANGUAGE AND GENETICS DEPARTMENT

GCACAUAACGGGCAGAACGCACUGCGAAGCGGCUUCUUCAGAGCACGGGCUGGAACUGGCAGGCACCGCGAGCCCCUAGCACCCGACAAGCUGAGUGUGCAGGACGAGUCCCCACCACA

CCCACACCACAGCCGCUGAAUGAGGCUUCCAGGCGUCCGCUCGCGGCCCGCAGAGCCCCGCCGUGGGUCCGCCCGCUGAGGCGCCCCCAGCCAGUGCGCUCACCUGCCAGACUGCGCGCC

AUGGGGCAACCCGGGAACGGCAGCGCCUUCUUGCUGGCACCCAAUGGAAGCCAUGCGCCGGACCACGACGUCACGCAGGAAAGGGACGAGGUGUGGGUGGUGGGCAUGGGCAUCGUCAUG

.M.

UCUCUCAUCGUCCUGGCCAUCGUGUUUGGCAAUGUGCUGGUCAUCACAGCCAUUGCCAAGUUCGAGCGUCUGCAGACGGUCACCAACUACUUCAUCACUUCACUGGCCUGUGCUGAUCUG

GUCAUGGGCCUGGCAGUGGUGCCCUUUGGGGCCGCCCAUAUUCUUAUGAAAAUGUGGACUUUUGGCAACUUCUGGUGCGAGUUUUGGACUUCCAUUGAUGUGCUGUGCGUCACGGCCAGC

AUUGAGACCCUGUGCGUGAUCGCAGUGGAUCGCUACUUUGCCAUUACUUCACCUUUCAAGUACCAGAGCCUGCUGACCAAGAAUAAGGCCCGGGUGAUCAUUCUGAUGGUGUGGAUUGUG

UCAGGCCUUACCUCCUUCUUGCCCAUUCAGAUGCACUGGUACCGGGCCACCCACCAGGAAGCCAUCAACUGCUAUGCCAAUGAGACCUGCUGUGACUUCUUCACGAACCAAGCCUAUGCC

AUUGCCUCUUCCAUCGUGUCCUUCUACGUUCCCCUGGUGAUCAUGGUCUUCGUCUACUCCAGGGUCUUUCAGGAGGCCAAAAGGCAGCUCCAGAAGAUUGACAAAUCUGAGGGCCGCUUC

CAUGUCCAGAACCUUAGCCAGGUGGAGCAGGAUGGGCGGACGGGGCAUGGACUCCGCAGAUCUUCCAAGUUCUGCUUGAAGGAGCACAAAGCCCUCAAGACGUUAGGCAUCAUCAUGGGC

ACUUUCACCCUCUGCUGGCUGCCCUUCUUCAUCGUUAACAUUGUGCAUGUGAUCCAGGAUAACCUCAUCCGUAAGGAAGUUUACAUCCUCCUAAAUUGGAUAGGCUAUGUCAAUUCUGGU

UUCAAUCCCCUUAUCUACUGCCGGAGCCCAGAUUUCAGGAUUGCCUUCCAGGAGCUUCUGUGCCUGCGCAGGUCUUCUUUGAAGGCCUAUGGGAAUGGCUACUCCAGCAACGGCAACACA

GGGGAGCAGAGUGGAUAUCACGUGGAACAGGAGAAAGAAAAUAAACUGCUGUGUGAAGACCUCCCAGGCACGGAAGACUUUGUGGGCCAUCAAGGUACUGUGCCUAGCGAUAACAUUGAU

UCACAAGGGAGGAAUUGUAGUACAAAUGACUCACUGCUGUAAAGCAGUUUUUCUACUUUUAAAGACCCCCCCCCCCAACAGAACACUAAACAGACUAUUUAACUUGAGGGUAAUAAACUU

AGAAUAAAAUUGUAAAAUUGUAUAGAGAUAUGCAGAAGGAAGGGCAUCCUUCUGCCUUUUUUAUUUUUUUAAGCUGUAAAAAGAGAGAAAACUUAUUUGAGUGAUUAUUUGUUAUUUGUA

CAGUUCAGUUCCUCUUUGCAUGGAAUUUGUAAGUUUAUGUCUAAAGAGCUUUAGUCCUAGAGGACCUGAGUCUGCUAUAUUUUCAUGACUUUUCCAUGUAUCUACCUCACUAUUCAAGUA

UUAGGGGUAAUAUAUUGCUGCUGGUAAUUUGUAUCUGAAGGAGAUUUUCCUUCCUACACCCUUGGACUUGAGGAUUUUGAGUAUCUCGGACCUUUCAGCUGUGAACAUGGACUCUUCCCC

CACUCCUCUUAUUUGCUCACACGGGGUAUUUUAGGCAGGGAUUUGAGGAGCAGCUUCAGUUGUUUUCCCGAGCAAAGUCUAAAGUUUACAGUAAAUAAAUUGUUUGACCAUGCCUUCAUU

GC

- AUG is the **start** signal and encodes amino acid M.

Step 2: Translation

LANGUAGE AND GENETICS DEPARTMENT

GCACAUAACGGGCAGAACGCACUGCGAAGCGGCUUCUUCAGAGCACGGGCUGGAACUGGCAGGCACCGCGAGCCCCUAGCACCCGACAAGCUGAGUGUGCAGGACGAGUCCCCACCACA

CCCACACCACAGCCGCUGAAUGAGGCUUCCAGGCGUCCGCUCGCGGCCCGCAGAGCCCCGCCGUGGGUCCGCCCGCUGAGGCGCCCCCAGCCAGUGCGCUCACCUGCCAGACUGCGCGCC

AUGGGGCAACCCGGGAACGGCAGCGCCUUCUUGCUGGCACCCAAUGGAAGCCAUGCGCCGGACCACGACGUCACGCAGGAAAGGGACGAGGUGUGGGUGGUGGGCAUGGGCAUCGUCAUG
.M..G..Q..P..G..N..G..S..A..F..L..L..A..P..N..G..S..H..A..P..D..H..D..V..T..Q..E..R..D..E..V..W..V..V..G..M..G..I..V..M.
UCUCUCAUCGUCCUGGCCAUCGUGUUUGGCAAUGUGCUGGUCAUCACAGCCAUUGCCAAGUUCGAGCGUCUGCAGACGGUCACCAACUACUUCAUCACUUCACUGGCCUGUGCUGAUCUG
.S..L..I..V..L..A..I..V..F..G..N..V..L..V..I..T..A..I..A..K..F..E..R..L..Q..T..V..T..N..Y..F..I..T..S..L..A..C..A..D..L.
GUCAUGGGCCUGGCAGUGGUGCCCUUUGGGGCCGCCCAUAUUCUUAUGAAAAUGUGGACUUUUGGCAACUUCUGGUGCGAGUUUUGGACUUCCAUUGAUGUGCUGUGCGUCACGGCCAGC
.V..M..G..L..A..V..V..P..F..G..A..A..H..I..L..M..K..M..W..T..F..G..N..F..W..C..E..F..W..T..S..I..D..V..L..C..V..T..A..S.
AUUGAGACCCUGUGCGUGAUCGCAGUGGAUCGCUACUUUGCCAUUACUUCACCUUUCAAGUACCAGAGCCUGCUGACCAAGAAUAAGGCCCGGGUGAUCAUUCUGAUGGUGUGGAUUGUG
.I..E..T..L..C..V..I..A..V..D..R..Y..F..A..I..T..S..P..F..K..Y..Q..S..L..L..T..K..N..K..A..R..V..I..I..L..M..V..W..I..V.
UCAGGCCUUACCUCCUUCUUGCCCAUUCAGAUGCACUGGUACCGGGCCACCCACCAGGAAGCCAUCAACUGCUAUGCCAAUGAGACCUGCUGUGACUUCUUCACGAACCAAGCCUAUGCC
.S..G..L..T..S..F..L..P..I..Q..M..H..W..Y..R..A..T..H..Q..E..A..I..N..C..Y..A..N..E..T..C..C..D..F..F..T..N..Q..A..Y..A.
AUUGCCUCUUCCAUCGUGUCCUUCUACGUUCCCCUGGUGAUCAUGGUCUUCGUCUACUCCAGGGUCUUUCAGGAGGCCAAAAGGCAGCUCCAGAAGAUUGACAAAUCUGAGGGCCGCUUC
.I..A..S..S..I..V..S..F..Y..V..P..L..V..I..M..V..F..V..Y..S..R..V..F..Q..E..A..K..R..Q..L..Q..K..I..D..K..S..E..G..R..F.
CAUGUCCAGAACCUUAGCCAGGUGGAGCAGGAUGGGCGGACGGGGCAUGGACUCCGCAGAUCUUCCAAGUUCUGCUUGAAGGAGCACAAAGCCCUCAAGACGUUAGGCAUCAUCAUGGGC
.H..V..Q..N..L..S..Q..V..E..Q..D..G..R..T..G..H..G..L..R..R..S..S..K..F..C..L..K..E..H..K..A..L..K..T..L..G..I..I..M..G.
ACUUUCACCCUCUGCUGGCUGCCCUUCUUCAUCGUUAACAUUGUGCAUGUGAUCCAGGAUAACCUCAUCCGUAAGGAAGUUUACAUCCUCCUAAAUUGGAUAGGCUAUGUCAAUUCUGGU
.T..F..T..L..C..W..L..P..F..F..I..V..N..I..V..H..V..I..Q..D..N..L..I..R..K..E..V..Y..I..L..L..N..W..I..G..Y..V..N..S..G.
UUCAAUCCCCUUAUCUACUGCCGGAGCCCAGAUUUCAGGAUUGCCUUCCAGGAGCUUCUGUGCCUGCGCAGGUCUUCUUUGAAGGCCUAUGGGAAUGGCUACUCCAGCAACGGCAACACA
.F..N..P..L..I..Y..C..R..S..P..D..F..R..I..A..F..Q..E..L..L..C..L..R..R..S..S..L..K..A..Y..G..N..G..Y..S..S..N..G..N..T.
GGGGAGCAGAGUGGAUAUCACGUGGAACAGGAGAAAGAAAAUAAACUGCUGUGUGAAGACCUCCCAGGCACGGAAGACUUUGUGGGCCAUCAAGGUACUGUGCCUAGCGAUAACAUUGAU
.G..E..Q..S..G..Y..H..V..E..Q..E..K..E..N..K..L..L..C..E..D..L..P..G..T..E..D..F..V..G..H..Q..G..T..V..P..S..D..N..I..D.
UCACAAGGGAGGAAUUGUAGUACAAAUGACUCACUGCUGUAAAGCAGUUUUUCUACUUUUAAAGACCCCCCCCCCCAACAGAACACUAAACAGACUAUUUAACUUGAGGGUAAUAAACUU
.S..Q..G..R..N..C..S..T..N..D..S..L..L.
AGAAUAAAAUUGUAAAAUUGUAUAGAGAUAUGCAGAAGGAAGGGCAUCCUUCUGCCUUUUUUAUUUUUUUAAGCUGUAAAAAGAGAGAAAACUUAUUUGAGUGAUUAUUUGUUAUUUGUA

CAGUUCAGUUCCUCUUUGCAUGGAAUUUGUAAGUUUAUGUCUAAAGAGCUUUAGUCCUAGAGGACCUGAGUCUGCUAUAUUUUCAUGACUUUUCCAUGUAUCUACCUCACUAUUCAAGUA

UUAGGGGUAAUAUAUUGCUGCUGGUAAUUUGUAUCUGAAGGAGAUUUUCCUUCCUACACCCUUGGACUUGAGGAUUUUGAGUAUCUCGGACCUUUCAGCUGUGAACAUGGACUCUUCCCC

CACUCCUCUUAUUUGCUCACACGGGGUAUUUUAGGCAGGGAUUUGAGGAGCAGCUUCAGUUGUUUUCCCGAGCAAAGUCUAAAGUUUACAGUAAAUAAAUUGUUUGACCAUGCCUUCAUU

GC

- UAA, UAG and UGA are **stop** signals.

A protein

.M..G..Q..P..G..N..G..S..A..F..L..L..A..P..N..G..S..H..A..P..D..H..D..V..T..Q..E..R..D..E..V..W..V..V..G..M..G..I..V..M.

.S..L..I..V..L..A..I..V..F..G..N..V..L..V..I..T..A..I..A..K..F..E..R..L..Q..T..V..T..N..Y..F..I..T..S..L..A..C..A..D..L.

.V..M..G..L..A..V..V..P..F..G..A..A..H..I..L..M..K..M..W..T..F..G..N..F..W..C..E..F..W..T..S..I..D..V..L..C..V..T..A..S.

.I..E..T..L..C..V..I..A..V..D..R..Y..F..A..I..T..S..P..F..K..Y..Q..S..L..L..T..K..N..K..A..R..V..I..I..L..M..V..W..I..V.

.S..G..L..T..S..F..L..P..I..Q..M..H..W..Y..R..A..T..H..Q..E..A..I..N..C..Y..A..N..E..T..C..C..D..F..F..T..N..Q..A..Y..A.

.I..A..S..S..I..V..S..F..Y..V..P..L..V..I..M..V..F..V..Y..S..R..V..F..Q..E..A..K..R..Q..L..Q..K..I..D..K..S..E..G..R..F.

.H..V..Q..N..L..S..Q..V..E..Q..D..G..R..T..G..H..G..L..R..R..S..S..K..F..C..L..K..E..H..K..A..L..K..T..L..G..I..I..M..G.

.T..F..T..L..C..W..L..P..F..F..I..V..N..I..V..H..V..I..Q..D..N..L..I..R..K..E..V..Y..I..L..L..N..W..I..G..Y..V..N..S..G.

.F..N..P..L..I..Y..C..R..S..P..D..F..R..I..A..F..Q..E..L..L..C..L..R..R..S..S..L..K..A..Y..G..N..G..Y..S..S..N..G..N..T.

.G..E..Q..S..G..Y..H..V..E..Q..E..K..E..N..K..L..L..C..E..D..L..P..G..T..E..D..F..V..G..H..Q..G..T..V..P..S..D..N..I..D.

.S..Q..G..R..N..C..S..T..N..D..S..L..L.

A protein

.M..G..Q..P..G..N..G..S..A..F..L..L..A..P..N..G..S..H..A..P..D..H..D..V..T..Q..E..R..D..E..V..W..V..V..G..M..G..I..V..M.

.S..L..I..V..L..A..I..V..F..G..N..V..L..V..I..T..A..I..A..K..F..E..R..L..Q..T..V..T..N..Y..F..I..T..S..L..A..C..A..D..L.

.V..M..G..L..A..V..V..P..F..G..A..A..H..I..L..M..K..M..W..T..F..G..N..F..W..C..E..F..W..T..S..I..D..V..L..C..V..T..A..S.

.I..E..T..L..C..V..I..A..V..D..R..Y..F..A..I..T..S..P..F..K..Y..Q..S..L..L..T..K..N..K..A..R..V..I..I..L..M..V..W..I..V.

.S..G..L..T..S..F..L..P..I..Q..M..H..W..Y..R..A..T..H..Q..E..A..I..N..C..Y..A..N..E..T..C..C..D..F..F..T..N..Q..A..Y..A.

.I..A..S..S..I..V..S..F..Y..V..P..L..V..I..M..V..F..V..Y..S..R..V..F..Q..E..A..K..R..Q..L..Q..K..I..D..K..S..E..G..R..F.

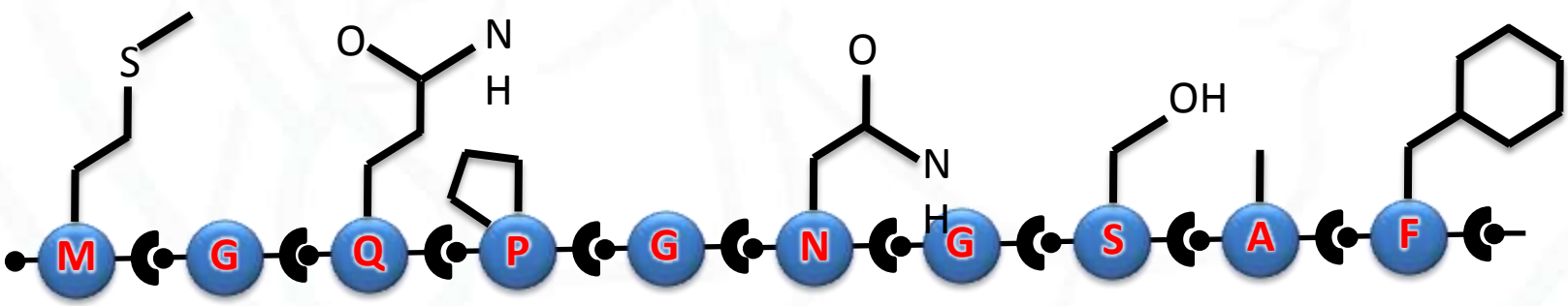
.H..V..Q..N..L..S..Q..V..E..Q..D..G..R..T..G..H..G..L..R..R..S..S..K..F..C..L..K..E..H..K..A..L..K..T..L..G..I..I..M..G.

.T..F..T..L..C..W..L..P..F..F..I..V..N..I..V..H..V..I..Q..D..N..L..I..R..K..E..V..Y..I..L..L..N..W..I..G..Y..V..N..S..G.

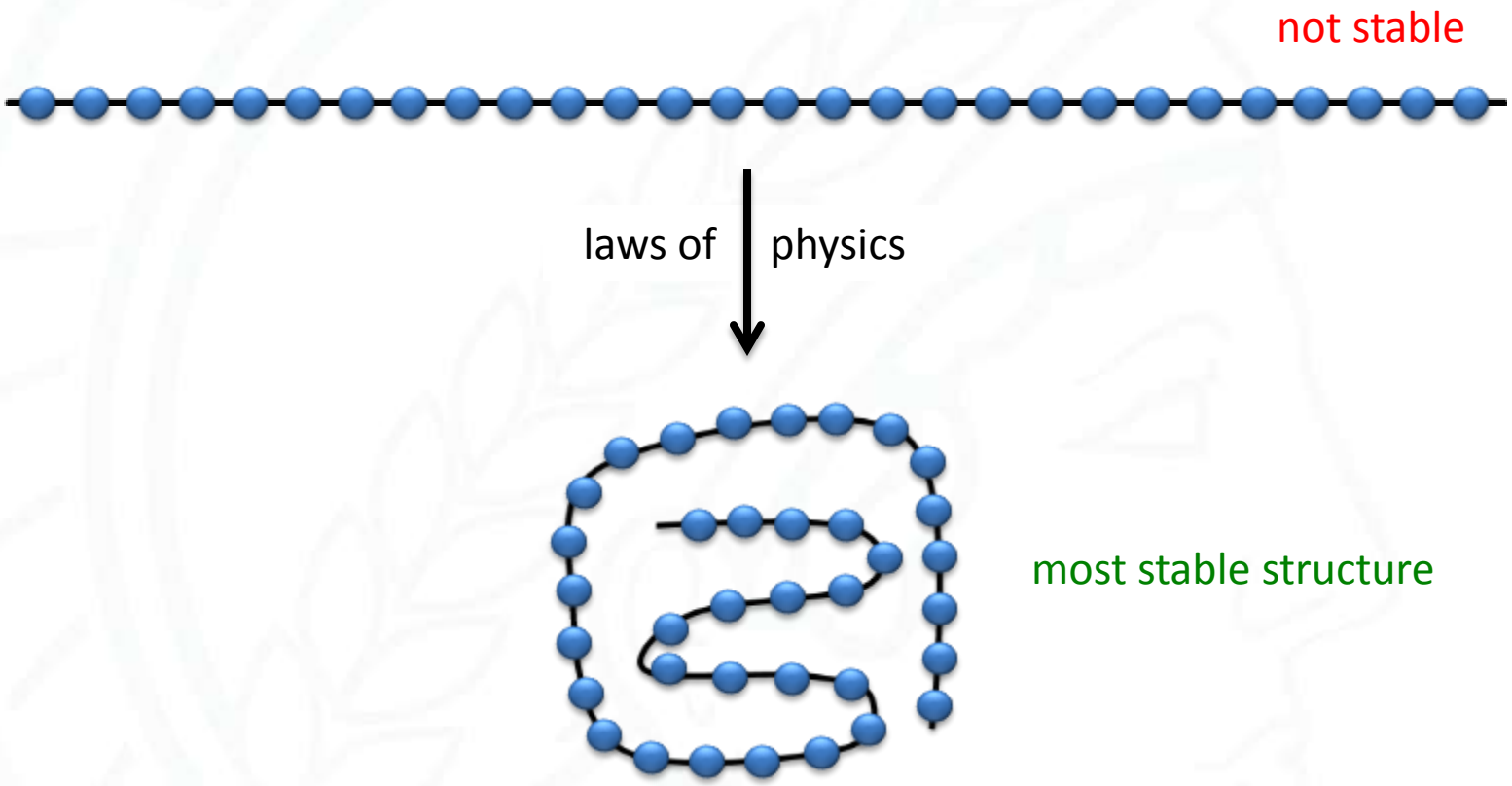
.F..N..P..L..I..Y..C..R..S..P..D..F..R..I..A..F..Q..E..L..L..C..L..R..R..S..S..L..K..A..Y..G..N..G..Y..S..S..N..G..N..T.

.G..E..Q..S..G..Y..H..V..E..Q..E..K..E..N..K..L..L..C..E..D..L..P..G..T..E..D..F..V..G..H..Q..G..T..V..P..S..D..N..I..D.

.S..Q..G..R..N..C..S..T..N..D..S..L..L.



Protein folding



- Every protein **folds** into a unique **structure** (shape).
- A protein's **structure** is determined **entirely** by its amino acid **sequence**.

Flow of information

DNA

↓ Transcription

RNA

↓ Translation

protein

↓ Folding

three-dimensional structure

From sequence to structure

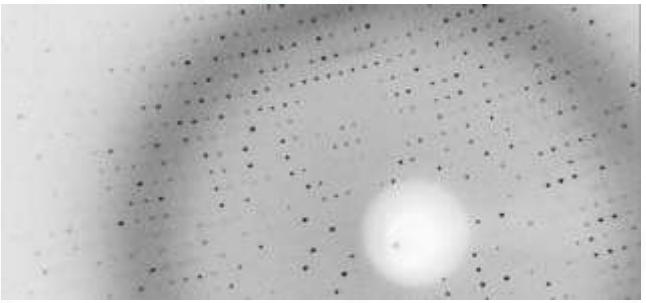
We **cannot yet** predict a protein's **structure** from its amino acid **sequence**…
…but we **can** determine it experimentally.

Protein **crystals**

+

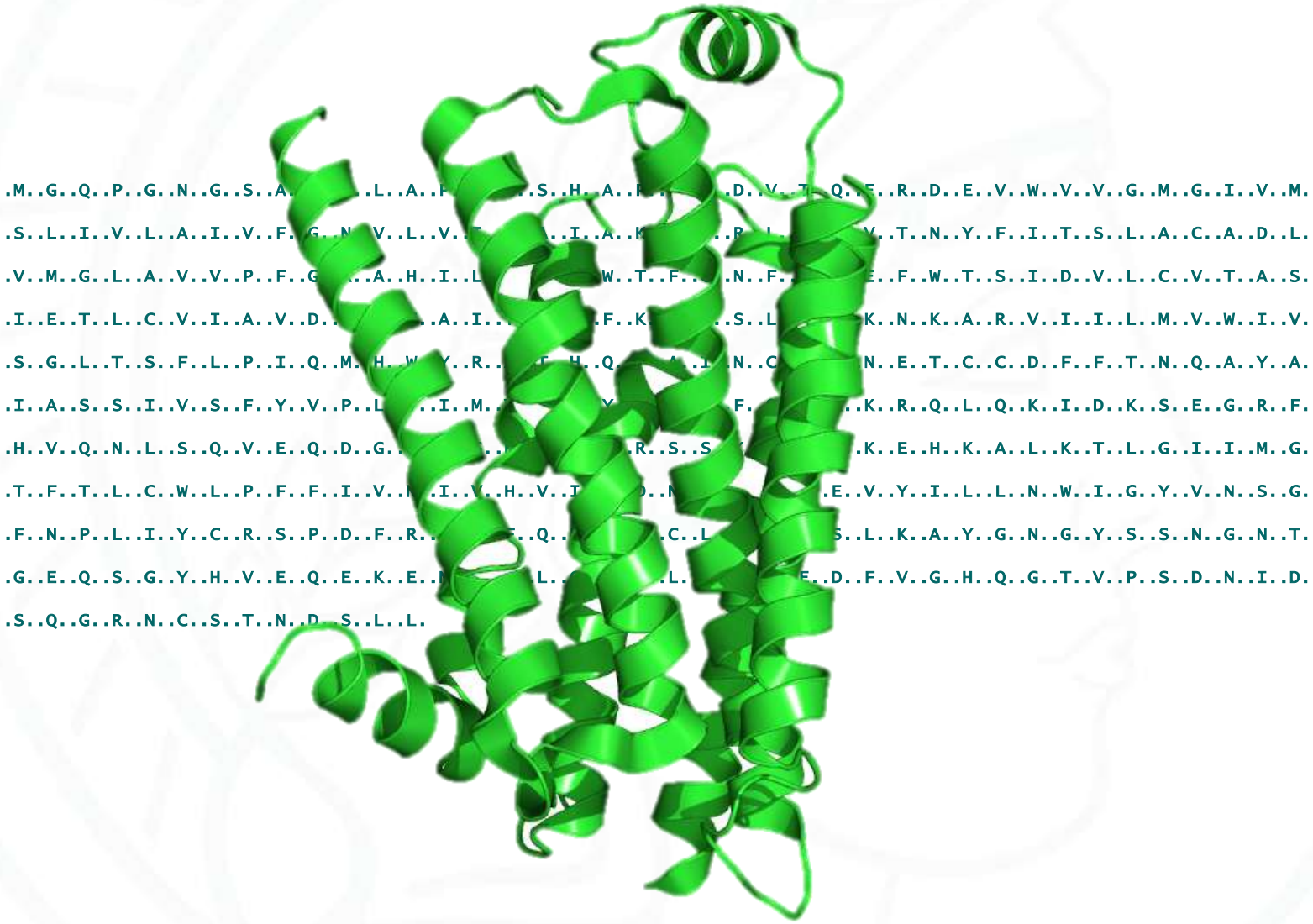
Synchrotron

=

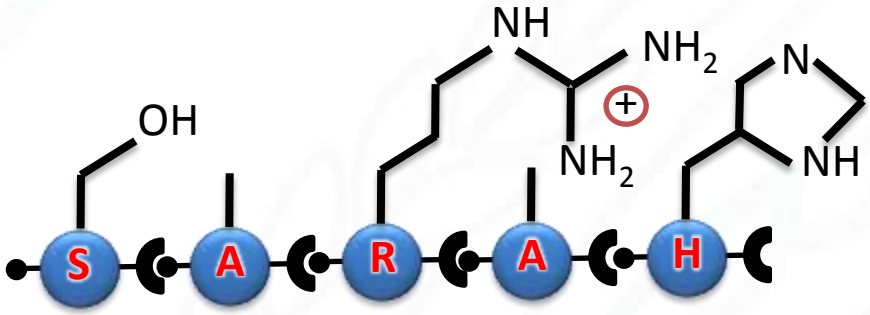


Diffraction pattern

From sequence to structure

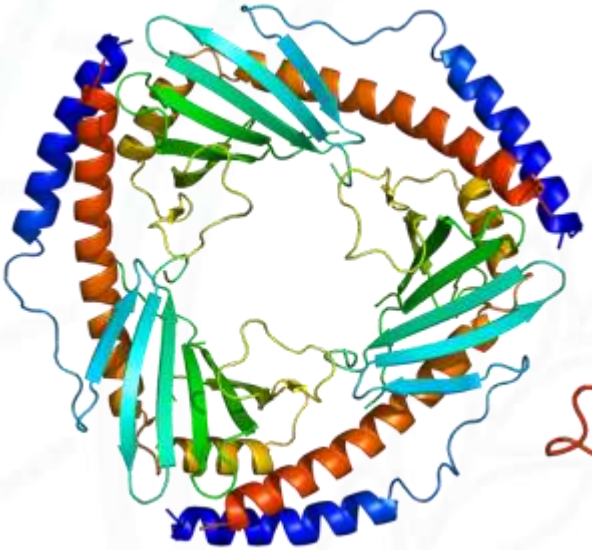


From sequence to structure



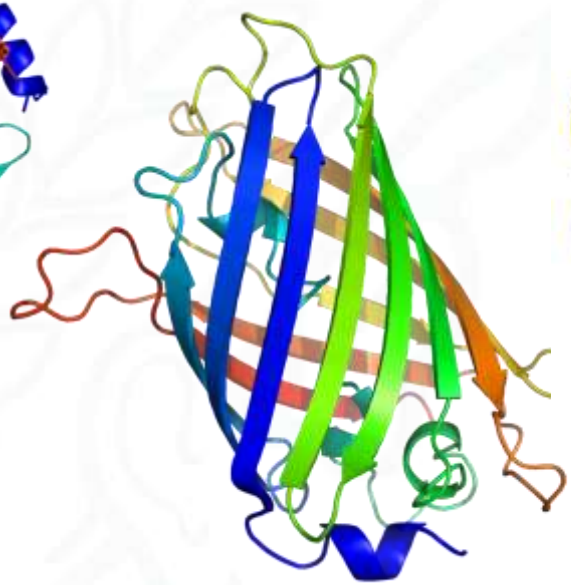
- The number of possible unique proteins is essentially **infinite**.
- For a **5-amino acid** protein: 20^5 unique sequences = **3.2 million**.
- For a **500-amino acid** protein: 20^{500} unique sequences = $\mathbf{10^{650}}$
- Diversity in protein **sequence** = diversity in protein **structure**.

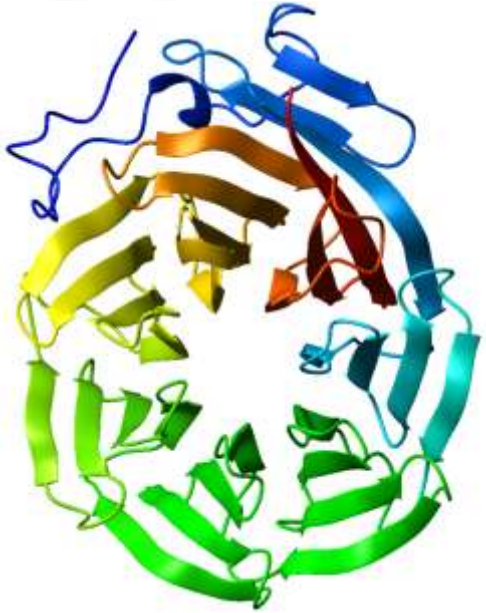
From sequence to structure



C1QBP

GFP





Tup1

- Proteins come in **all shapes and sizes**.
- Each protein has a **unique** shape that determines how it interacts with other molecules, and thus its biological **function**.

Flow of information

DNA

↓ Transcription

RNA

↓ Translation

protein

↓ Folding

three-dimensional structure

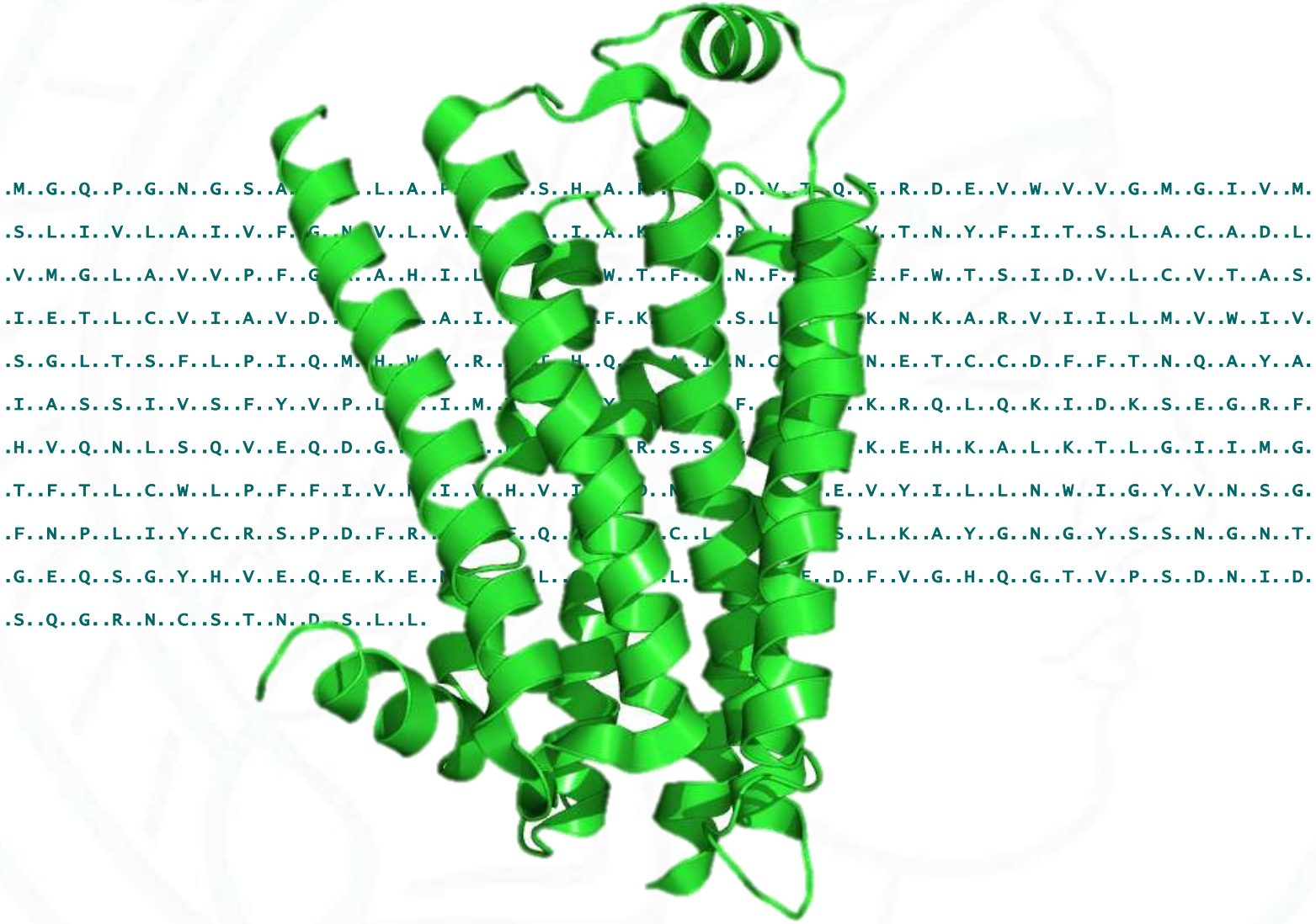
↓

biological function

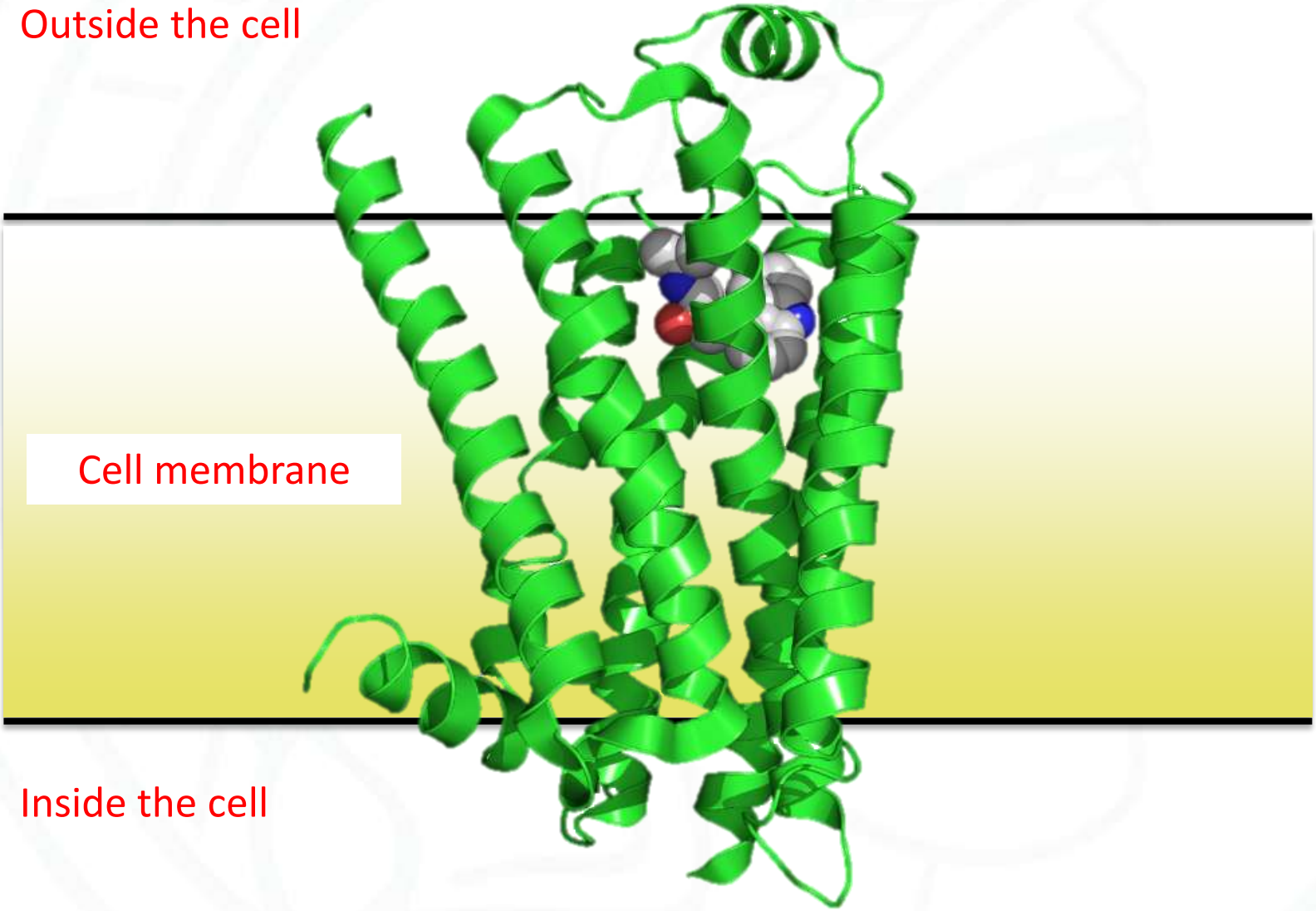
Diverse amino acid **sequences** = diverse **structures** = diverse **functions**

From structure to function

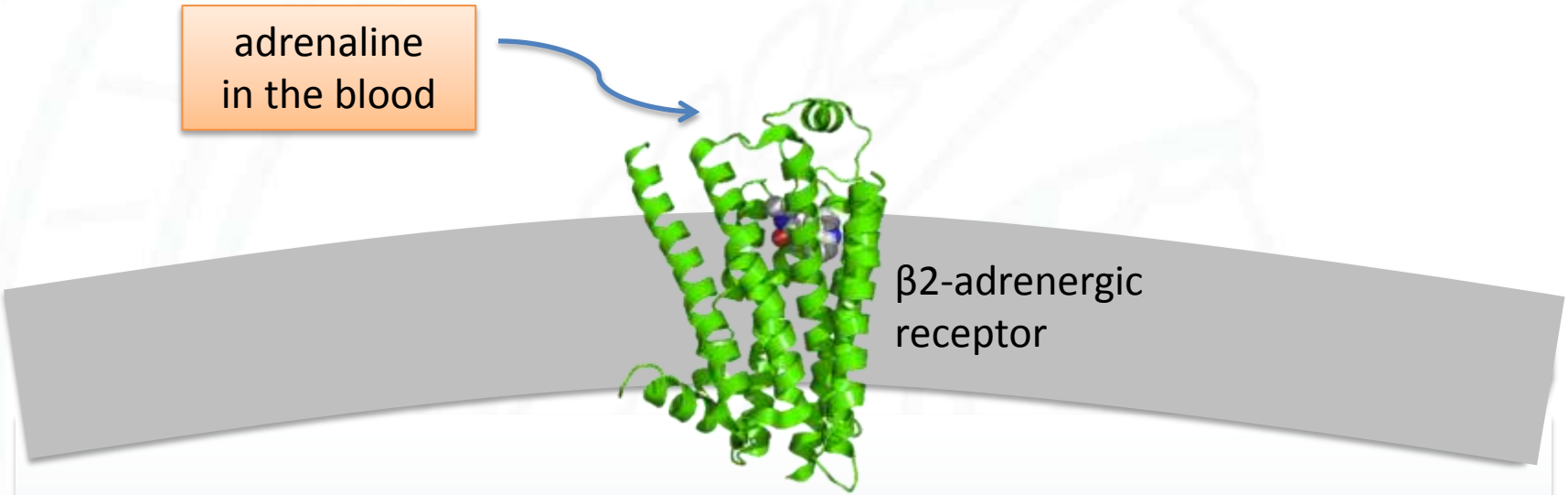
LANGUAGE AND GENETICS DEPARTMENT



From structure to function



From structure to function



↓

Heart rate increases, blood vessels widen, energy stores mobilized.

↓

Muscles can work harder.

↓

Increased chance of survival.

From DNA to function

LANGUAGE AND GENETICS DEPARTMENT

TTCTGTGTGTGTCACCATGTCTTTGTGCTTCTGGGTGCTTCTGTGTTTGTTTCTGGCCGCGTTTCTGTGTTGGACAGGGGTGACTTTGTGCCGGATGGCTTCTGTGTGAGAGCGCGCGCG
AGTGTGCATGTCGGTGAGCTGGGAGGGTGTGTCTCAGTGTCTATGGCTGTGGTTCGGTATAAGTCTGAGCATGTCTGCCAGGGTGTATTTGTGCCTGTATGTGCGTGCCTCGGTGGGCAC
TCTCGTTTCCTTCCGAATGTGGGGCAGTGCCGGTGTGCTGCCCTCTGCCTTGAGACCTCAAGCCGCGCAGGCGCCCAGGGCAGGCAGGTAGCGGCCACAGAAGAGCCAAAAGCTCCCGGG
TTGGCTGGTAAGGACACCACCTCCAGCTTTAGCCCTCTGGGGCCAGCCAGGGTAGCCGGGAAGCAGTGGTGGCCCGCCCTCCAGGGAGCAGTTGGGCCCCGCCCGGGCCAGCCCCAGGAG
AAGGAGGGCGAGGGGAGGGGAGGGAAAGGGGAGGAGTGCCTCGCCCCTTCGCGGCTGCCGGCGTGCCATTGGCCGAAAGTTCCCGTACGTCACGGCGAGGGCAGTTCCCCTAAAGTCCTG
TGCACATAACGGGCAGAACGCACTGCGAAGCGGCTTCTTCAGAGCACGGGCTGGAACTGGCAGGCACCGCGAGCCCCTAGCACCCGACAAGCTGAGTGTGCAGGACGAGTCCCCACCACA
CCCACACCACAGCCGCTGAATGAGGCTTCCAGGCGTCCGCTCGCGGCCCGCAGAGCCCCGCCGTGGGTCCGCCCGCTGAGGCGCCCCCAGCCAGTGCGCTCACCTGCCAGACTGCGCGCC
ATGGGGCAACCCGGGAACGGCAGCGCCTTCTTGCTGGCACCCAATGGAAGCCATGCGCCGGACCACGACGTCACGCAGGAAAGGGACGAGGTGTGGGTGGTGGGCATGGGCATCGTCATG
TCTCTCATCGTCCTGGCCATCGTGTTTGGCAATGTGCTGGTCATCACAGCCATTGCCAAGTTCGAGCGTCTGCAGACGGTCACCAACTACTTCATCACTTCACTGGCCTGTGCTGATCTG
GTCATGGGCCTGGCAGTGGTGCCCTTTGGGGCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGAGTTTTGGACTTCCATTGATGTGCTGTGCGTCACGGCCAGC
ATTGAGACCCTGTGCGTGATCGCAGTGGATCGCTACTTTGCCATTACTTCACCTTTCAAGTACCAGAGCCTGCTGACCAAGAATAAGGCCCGGGTGATCATTCTGATGGTGTGGATTGTG
TCAGGCCTTACCTCCTTCTTGCCCATTCAGATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGCTGTGACTTCTTCACGAACCAAGCCTATGCC
ATTGCCTCTTCCATCGTGTCCTTCTACGTTCCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCAGCTCCAGAAGATTGACAAATCTGAGGGCCGCTTC
CATGTCCAGAACCTTAGCCAGGTGGAGCAGGATGGGCGGACGGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAAGGAGCACAAAGCCCTCAAGACGTTAGGCATCATCATGGGC
ACTTTCACCCTCTGCTGGCTGCCCTTCTTCATCGTTAACATTGTGCATGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTCCTAAATTGGATAGGCTATGTCAATTCTGGT
TTCAATCCCCTTATCTACTGCCGGAGCCCAGATTTCAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTATGGGAATGGCTACTCCAGCAACGGCAACACA
GGGGAGCAGAGTGGATATCACGTGGAACAGGAGAAAGAAAATAAACTGCTGTGTGAAGACCTCCCAGGCACGGAAGACTTTGTGGGCCATCAAGGTACTGTGCCTAGCGATAACATTGAT
TCACAAGGGAGGAATTGTAGTACAAATGACTCACTGCTGTAAAGCAGTTTTTCTACTTTTAAAGACCCCCCCCCCCAACAGAACACTAAACAGACTATTTAACTTGAGGGTAATAAACTT
AGAATAAAATTGTAAAATTGTATAGAGATATGCAGAAGGAAGGGCATCCTTCTGCCTTTTTTATTTTTTTAAGCTGTAAAAAGAGAGAAAACTTATTTGAGTGATTATTTGTTATTTGTA
CAGTTCAGTTCCTCTTTGCATGGAATTTGTAAGTTTATGTCTAAAGAGCTTTAGTCCTAGAGGACCTGAGTCTGCTATATTTTCATGACTTTTCCATGTATCTACCTCACTATTCAAGTA
TTAGGGGTAATATATTGCTGCTGGTAATTTGTATCTGAAGGAGATTTTCCTTCCTACACCCTTGGACTTGAGGATTTTGAGTATCTCGGACCTTTCAGCTGTGAACATGGACTCTTCCCC
CACTCCTCTTATTTGCTCACACGGGGTATTTTAGGCAGGGATTTGAGGAGCAGCTTCAGTTGTTTTCCCGAGCAAAGTCTAAAGTTTACAGTAAATAAATTGTTTGACCATGCCTTCATT
GCACCTGTTTCTCCAAAACCCCTTGACTGGAGTGCTGTTGCCTCCCCCACTGGAAACCGCAGGTAACTACTTGTAATTACTGCCCATGACTTAATGTAGAATGATACAAGAATGACATGC
ACAGATTGCTTAACCCTTTCATTTGCCTTTGAGTCTGCTGCTGCAAAGCTGCATCTCTCCTGACACTTGTGCCCCAAATCAGTTCTGCCTGCTCTTAGTATAGCTCAACTCTCCCTATGG
TTATTGTTCTGTGTTGTTACCTCAGAAACACTGACTCACAGAAGCGGAGTTAAGGGGATATGTTTTTTTCTCTCCACGTGCACCCACCACCCACCTTCCAGTTCTACTTGTTTCAAAACT
GTTTATATTTCTGTCTTGGCCATGTGTTACAGTGGAGCTCTTTGTACTGCATCAGGGCTTGGCATTTTAGGGATAAGGAAGATGTTCTTATGAGGAAGCTACTCAGACATGGCCCCGTAA
TTCTGAGGGAAAATTCAAAAGGCATTGGTCATGGGGAGAAAAGCTGGAGAACACATAACTGATGGATACCTCATGAACTAGAAACAGAATTTTAACCCCTTTTCCTTCTTTCCTTTGGTC

= increased chance of survival

Flow of information

DNA

↓ Transcription

RNA

↓ Translation

protein

↓ Folding

three-dimensional structure

↓

biological function

↓

integration of function into cell/organism

↓

observed **phenotypes**

Flow of information

DNA

↓ Transcription

RNA

↓ Translation

protein

↓ Folding

Dynamic response → three-dimensional structure

↓

biological function

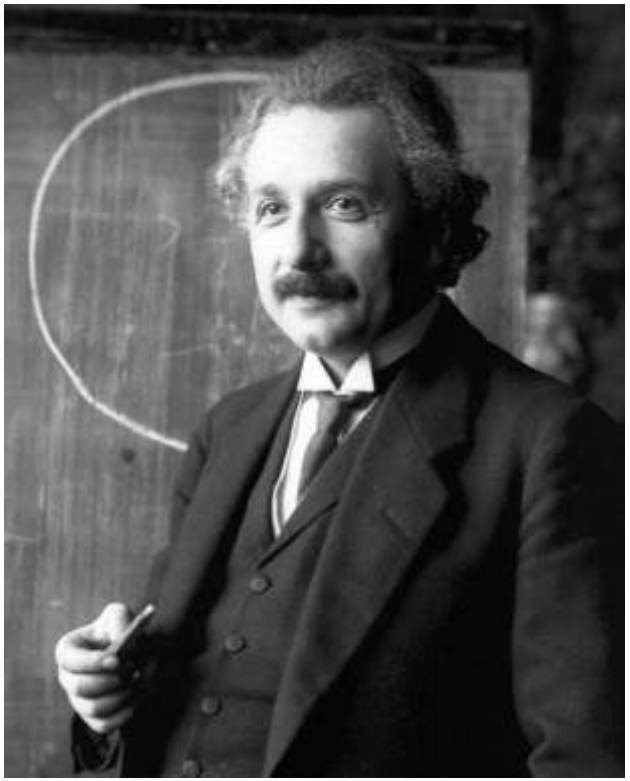
↓

integration of function into cell/organism

↓

observed **phenotypes**

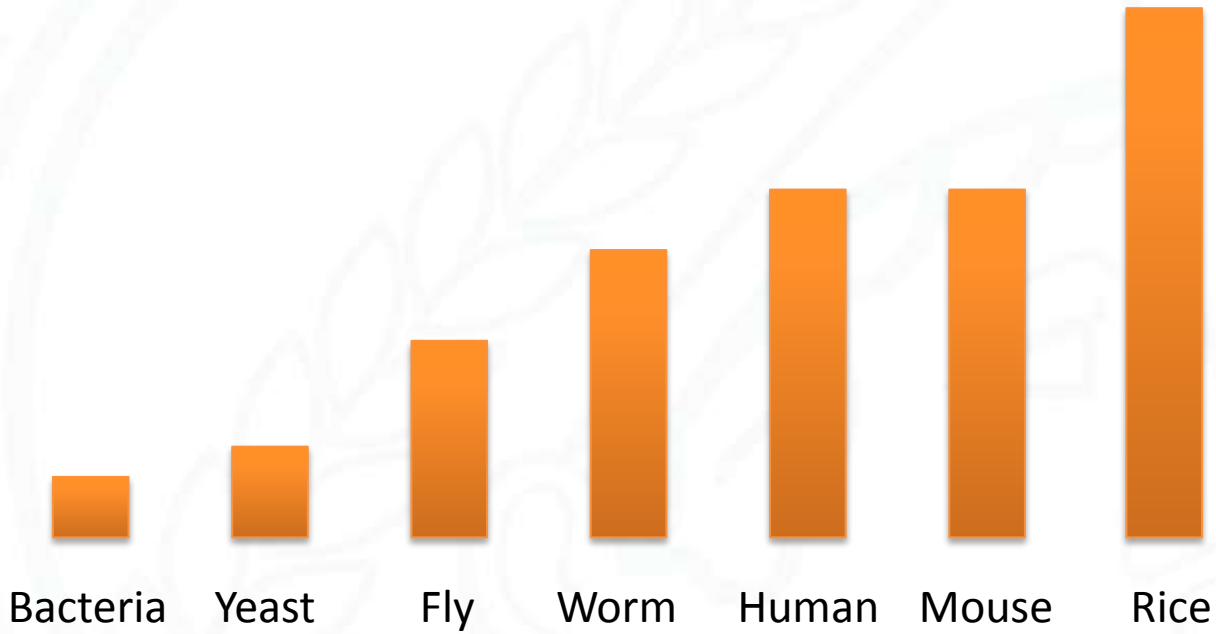
But that's not the end of the story…



Homo sapiens
23000 genes

Arabidopsis thaliana
26000 genes

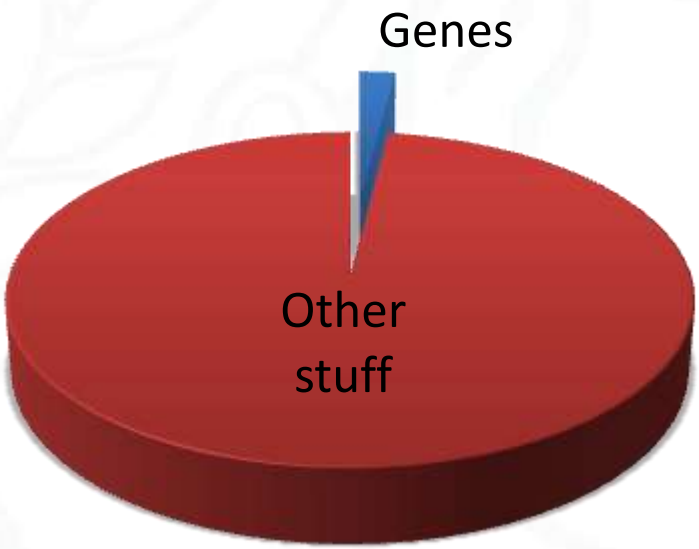
But that's not the end of the story…



- Humans have the same number of genes as the **average mammal**.
- There must be **something else** in the genome that accounts for our complex brain.

Contents of the human genome

- Less than 2% of the genome is genes that code for protein – the **exome**.
- So why do we focus so much on the protein-coding DNA rather than the 98% of the genome that is **non-coding**?



Contents of the human genome

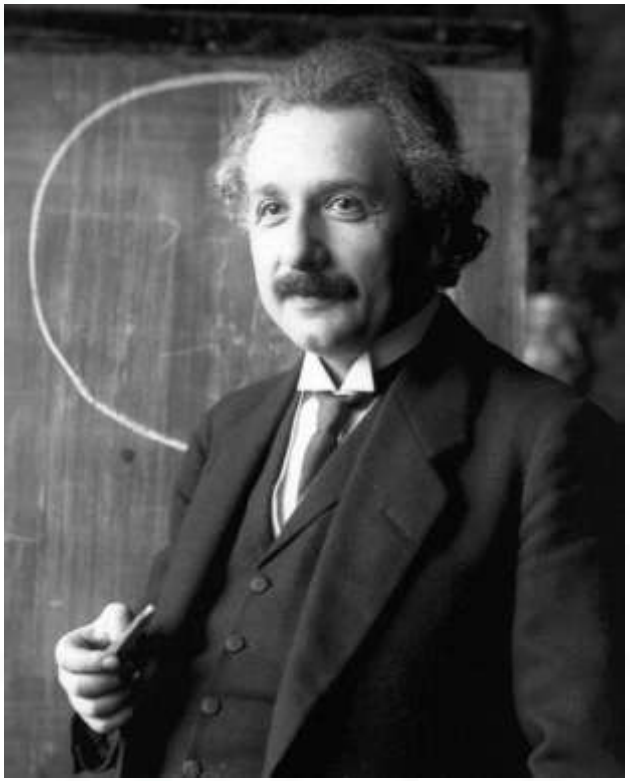
- Less than 2% of the genome is genes that code for protein – the **exome**.
- So why do we focus so much on the protein-coding DNA rather than the 98% of the genome that is **non-coding**?

Reason 1: **proteins are very important** and studying them has been enormously successful in explaining biology and treating disease.

Reason 2: our vast **knowledge and experience** of protein-coding DNA means that it is easier to study than non-coding DNA.

Reason 3: it is likely that only a **minor proportion** of the non-coding DNA is truly **important** – and the important bits are difficult to find.

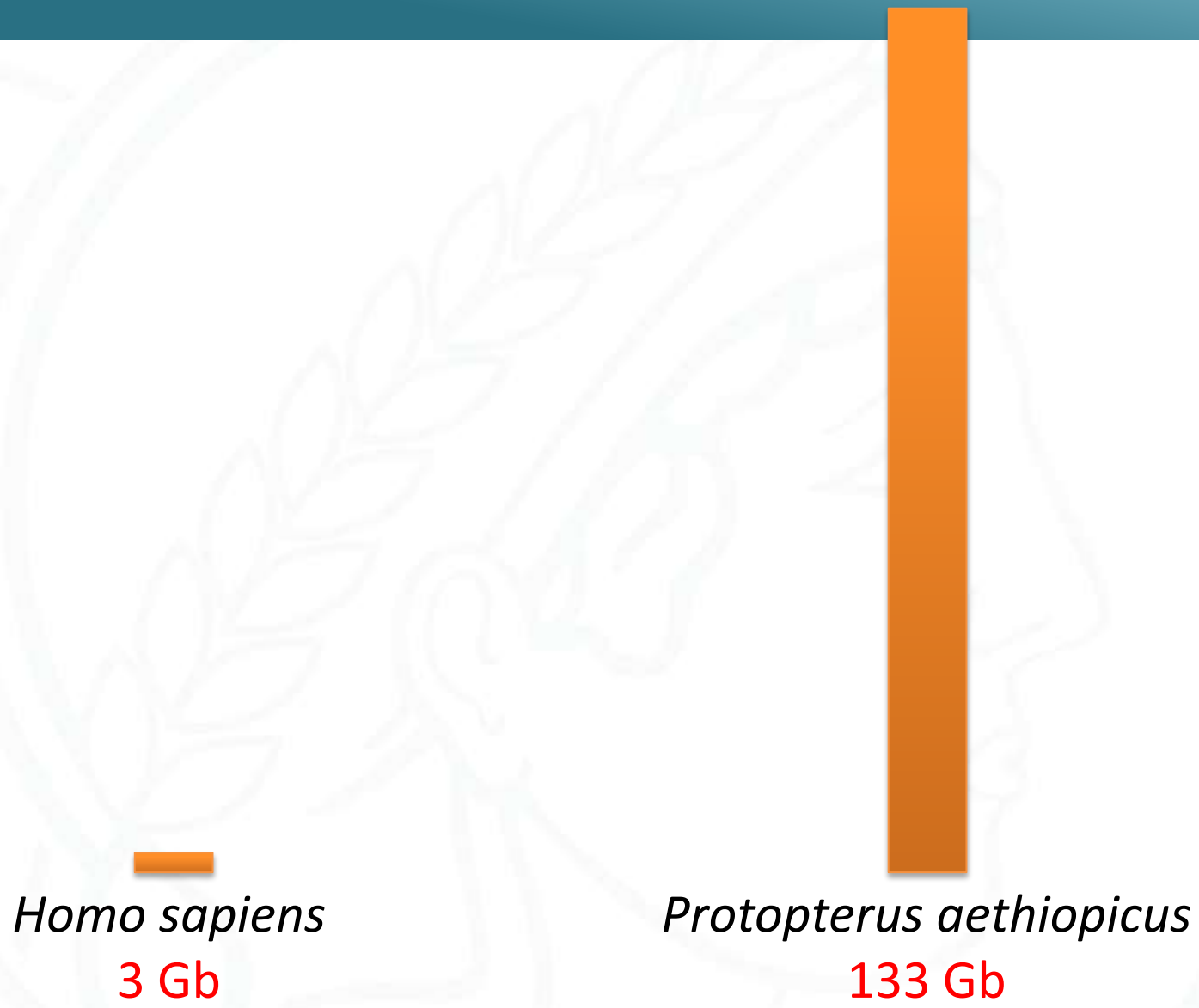
Why not all non-coding DNA matters



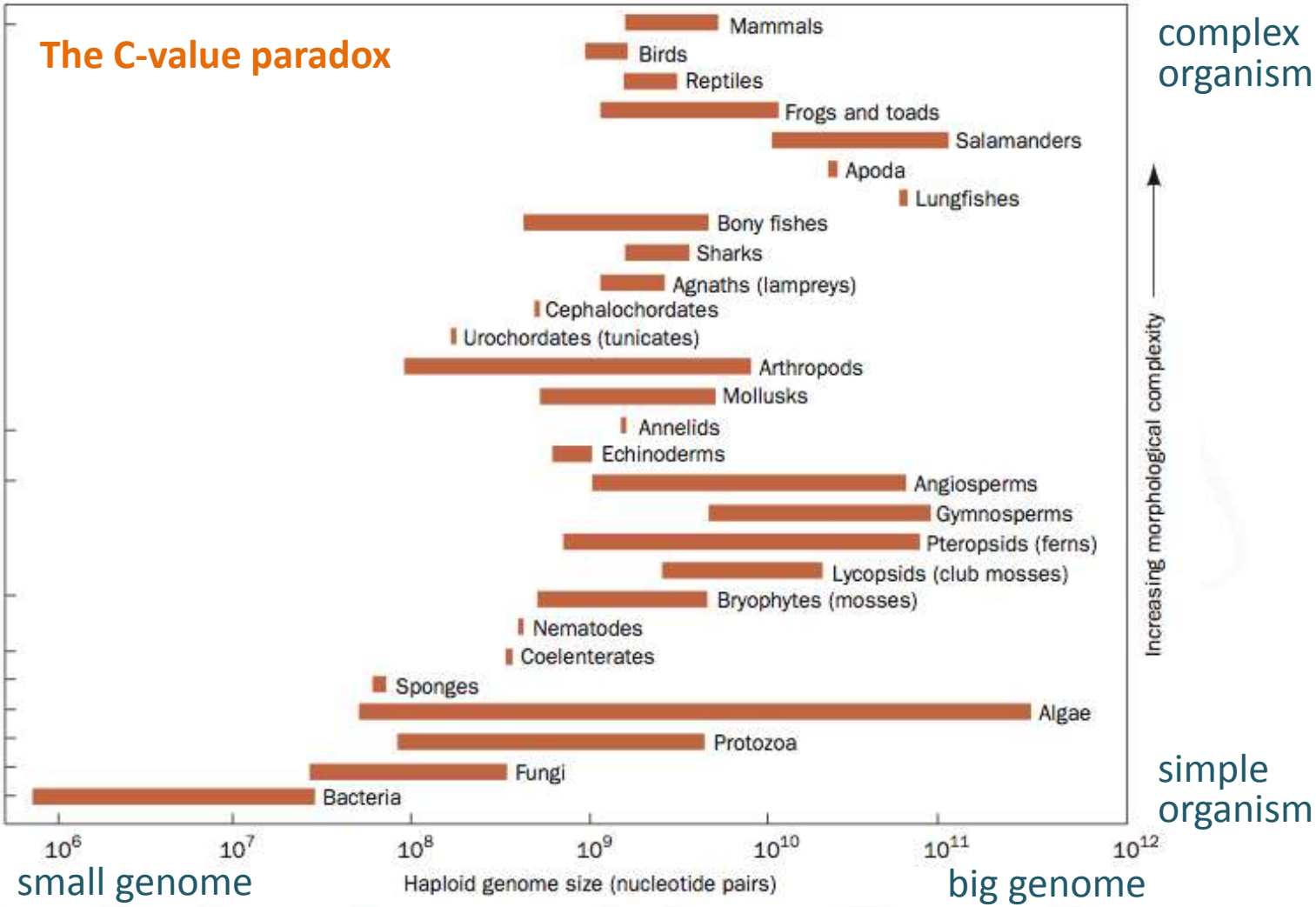
Homo sapiens
3 Gb

Protopterus aethiopicus
133 Gb

Why not all non-coding DNA matters



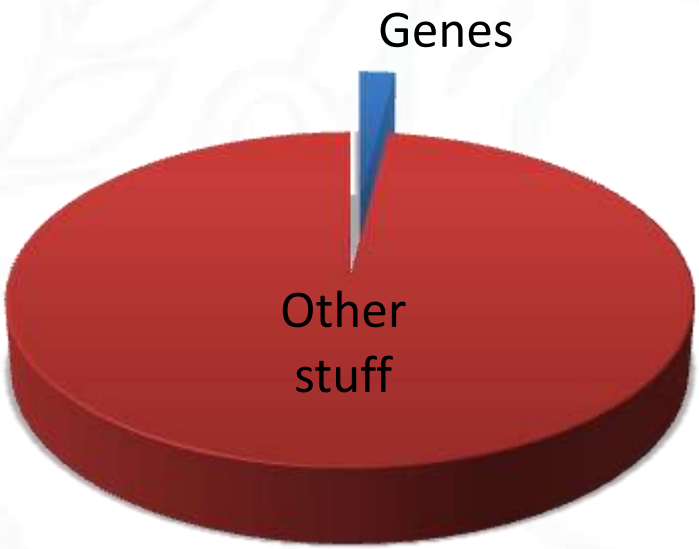
Why not all non-coding DNA matters



The complexity of an organism does not correlate with the size of its genome

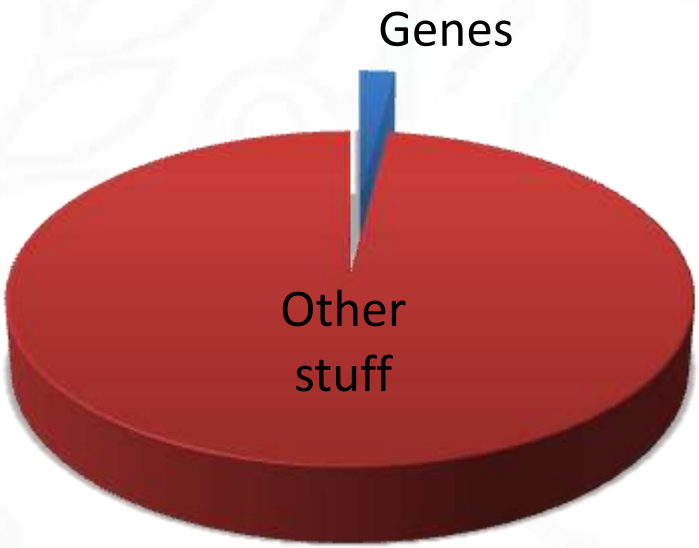
Why not all non-coding DNA matters

- **Nearly half** of the human genome is blocks of DNA sequence that are **repeated** in multiple locations across the genome.
- These sequences were once able to **copy and paste** themselves freely (and some still can).
- As a result, this **selfish** or **junk** DNA has accumulated over evolutionary time.



Why not all non-coding DNA matters

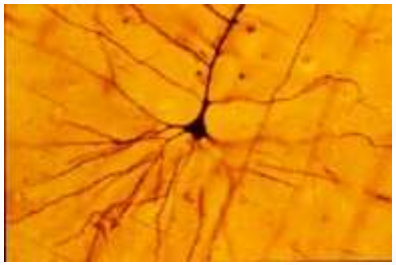
- The genome also contains many remnants of **broken** genes - **pseudogenes**.
- All this sequence may be **useless**, but it is not deleted because the **metabolic cost** of replicating it is **insignificant**.



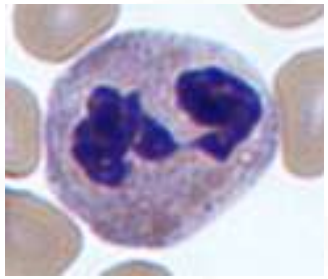
Why some non-coding DNA **does** matter

Same genome
Different-looking animals

Why some non-coding DNA **does** matter



Neuron
Storing and processing information



Neutrophil
Fighting bacteria

Same genome

Different cell types – **different** functions

Why some non-coding DNA **does** matter

DNA

↓ Transcription

RNA

↓ Translation

protein

↓ Folding

three-dimensional structure

↓

biological function

↓

integration of function into cell/organism

↓

observed phenotypes

Why some non-coding DNA **does** matter

DNA

↓ Transcription ← Key control point

RNA

↓ Translation

protein

↓ Folding

three-dimensional structure

↓

biological function

↓

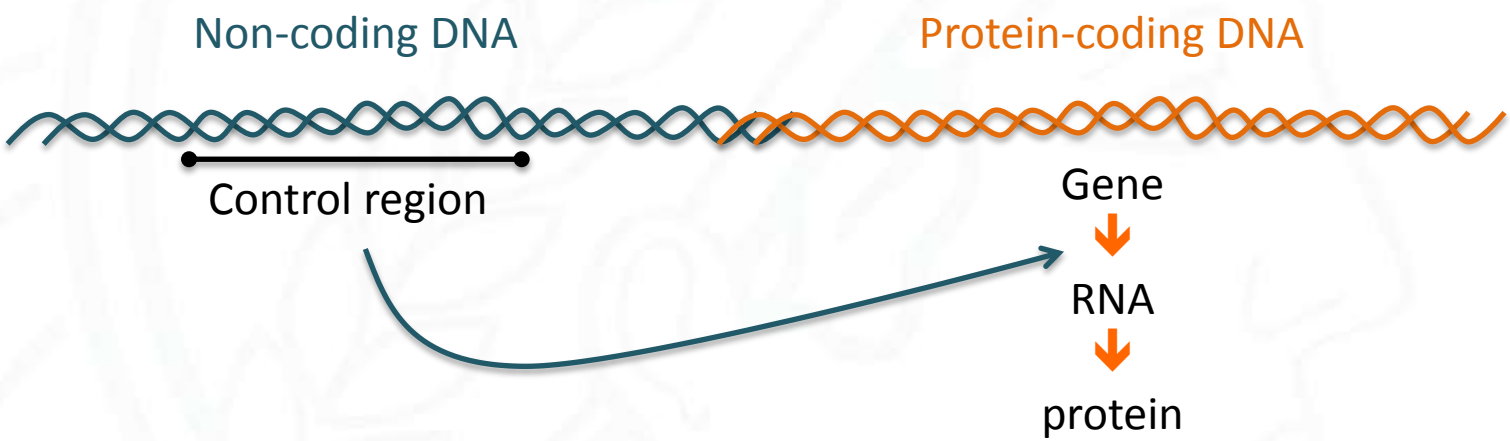
integration of function into cell/organism

↓

observed phenotypes

Why some non-coding DNA **does** matter

Regulation of transcription / **Regulation of gene expression**

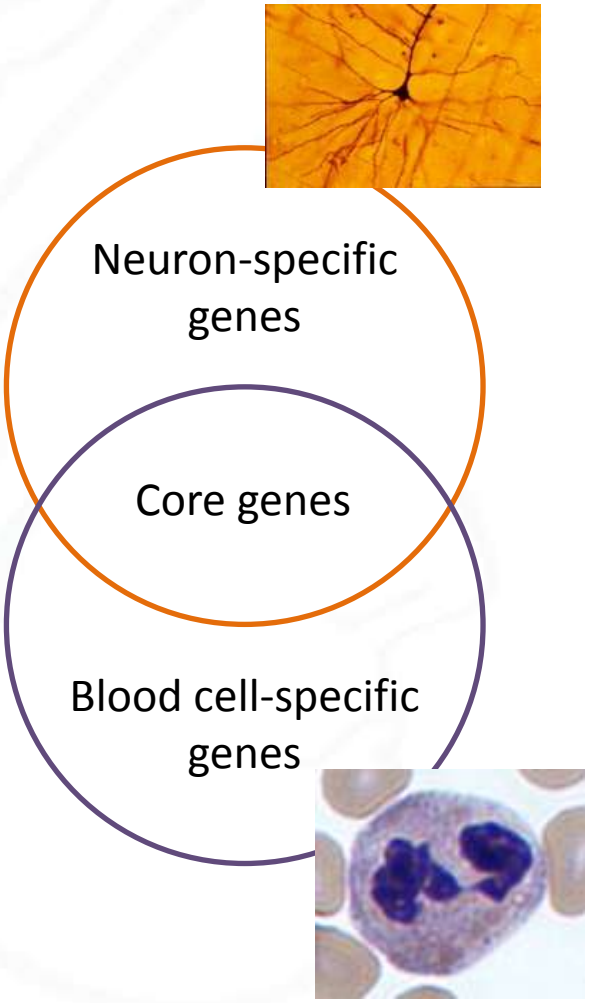


Regulation of gene expression changes according to

- cell type
- developmental stage
- environment

Why some non-coding DNA **does** matter

- Every cell contains the **same genome**.
- **Regulation of gene expression** enables cells to have specialized functions.
- **Dynamic** regulation of gene expression is what steers **development** of a fertilized egg into a human being.
- It is hard to say how much of the human genome is **crucial** for proper gene regulation – maybe **10%**.



Universality and Variability

LANGUAGE AND GENETICS DEPARTMENT

Part I: Universality

The core molecular principles underpinning living things.

Part II: Variability

How humans differ from other species and from each other.

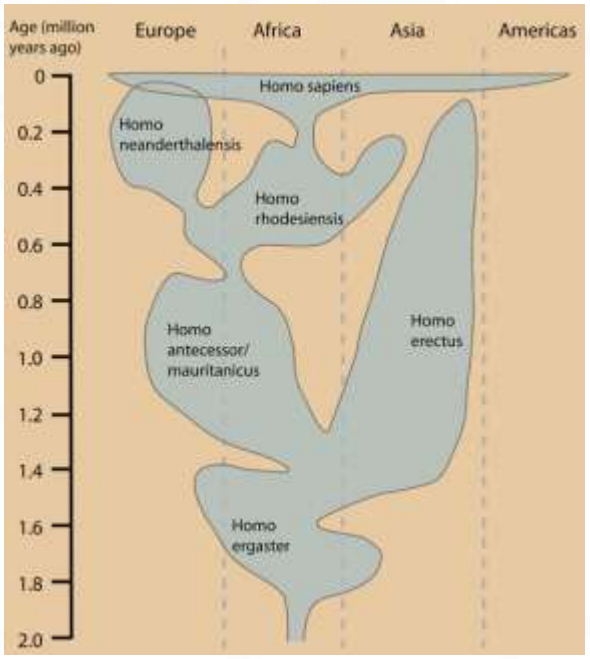
Variability

1. Differences between humans and other species

2. Differences between human individuals

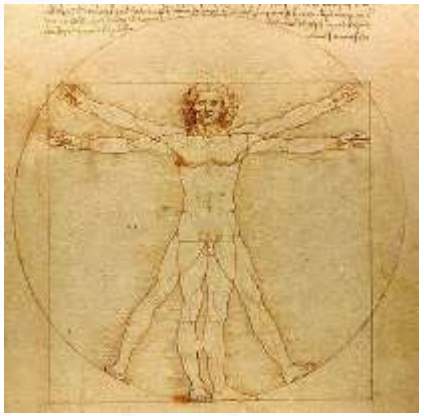
Between-species variability

- Why are humans the only extant species with language?
- How and when did language evolve in our ancestors?
- Do other species display some features of language?



Between-species variability

beta2-adrenergic receptor



```
ATGGGGCAACCCGGGAACGGCAGCGCCTTCTTGCTGGCACCCAATGGAAGCCATGCGCCGGACCACGACGTCACGCAGGAAAGGGACGAGGTGTGGGTGGTGGGCATGGGCATCGTCATG
.M..G..Q..P..G..N..G..S..A..F..L..L..A..P..N..G..S..H..A..P..D..H..D..V..T..Q..E..R..D..E..V..W..V..V..G..M..G..I..V..M.
TCTCTCATCGTCCTGGCCATCGTGTTTGGCAATGTGCTGGTCATCACAGCCATTGCCAAGTTCGAGCGTCTGCAGACGGTCACCAACTACTTCATCACTTCACTGGCCTGTGCTGATCTG
.S..L..I..V..L..A..I..V..F..G..N..V..L..V..I..T..A..I..A..K..F..E..R..L..Q..T..V..T..N..Y..F..I..T..S..L..A..C..A..D..L.
GTCATGGGCCTGGCAGTGGTGCCCTTTGGGGCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGAGTTTTGGACTTCCATTGATGTGCTGTGCGTCACGGCCAGC
.V..M..G..L..A..V..V..P..F..G..A..A..H..I..L..M..K..M..W..T..F..G..N..F..W..C..E..F..W..T..S..I..D..V..L..C..V..T..A..S.
ATTGAGACCCTGTGCGTGATCGCAGTGGATCGCTACTTTGCCATTACTTCACCTTTCAAGTACCAGAGCCTGCTGACCAAGAATAAGGCCCGGGTGATCATTCTGATGGTGTGGATTGTG
.I..E..T..L..C..V..I..A..V..D..R..Y..F..A..I..T..S..P..F..K..Y..Q..S..L..L..T..K..N..K..A..R..V..I..I..L..M..V..W..I..V.
TCAGGCCTTACCTCCTTCTTGCCCATTCAGATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGCTGTGACTTCTTCACGAACCAAGCCTATGCC
.S..G..L..T..S..F..L..P..I..Q..M..H..W..Y..R..A..T..H..Q..E..A..I..N..C..Y..A..N..E..T..C..C..D..F..F..T..N..Q..A..Y..A.
ATTGCCTCTTCCATCGTGTCCTTCTACGTTCCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCAGCTCCAGAAGATTGACAAATCTGAGGGCCGCTTC
.I..A..S..S..I..V..S..F..Y..V..P..L..V..I..M..V..F..V..Y..S..R..V..F..Q..E..A..K..R..Q..L..Q..K..I..D..K..S..E..G..R..F.
CATGTCCAGAACCTTAGCCAGGTGGAGCAGGATGGGCGGACGGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAAGGAGCACAAAGCCCTCAAGACGTTAGGCATCATCATGGGC
.H..V..Q..N..L..S..Q..V..E..Q..D..G..R..T..G..H..G..L..R..R..S..S..K..F..C..L..K..E..H..K..A..L..K..T..L..G..I..I..M..G.
ACTTTCACCCTCTGCTGGCTGCCCTTCTTCATCGTTAACATTGTGCATGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTCCTAAATTGGATAGGCTATGTCAATTCTGGT
.T..F..T..L..C..W..L..P..F..F..I..V..N..I..V..H..V..I..Q..D..N..L..I..R..K..E..V..Y..I..L..L..N..W..I..G..Y..V..N..S..G.
TTCAATCCCCTTATCTACTGCCGGAGCCCAGATTTCAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTATGGGAATGGCTACTCCAGCAACGGCAACACA
.F..N..P..L..I..Y..C..R..S..P..D..F..R..I..A..F..Q..E..L..L..C..L..R..R..S..S..L..K..A..Y..G..N..G..Y..S..S..N..G..N..T.
GGGGAGCAGAGTGGATATCACGTGGAACAGGAGAAAGAAAATAAACTGCTGTGTGAAGACCTCCCAGGCACGGAAGACTTTGTGGGCCATCAAGGTACTGTGCCTAGCGATAACATTGAT
.G..E..Q..S..G..Y..H..V..E..Q..E..K..E..N..K..L..L..C..E..D..L..P..G..T..E..D..F..V..G..H..Q..G..T..V..P..S..D..N..I..D.
TCACAAGGGAGGAATTGTAGTACAAATGACTCACTGCTG
.S..Q..G..R..N..C..S..T..N..D..S..L..L.
```

Between-species variability

LANGUAGE AND GENETICS DEPARTMENT

beta2-adrenergic receptor

99% identity to human

```
ATGGGGCAACCCGGGAACGGCAGCGCCTTCTTGCTGGCACCCAATGGAAGCCATGCGCCGGACCACGACGTCACGCAGGAAAGGGACGAGGTGTGGGTGGTGGGCATGGGCATCTTCATG
.M..G..Q..P..G..N..G..S..A..F..L..L..A..P..N..G..S..H..A..P..D..H..D..V..T..Q..E..R..D..E..V..W..V..V..G..M..G..I..F..M.
TCTCTCATCGTCCTGGCCATCGTGTTTGGCAATGTGCTGGTCATCACAGCCATTGCCAAGTTCGAGCGTCTGCAGACGGTCACCAACTACTTCATCACTTCACTGGCCTGTGCTGATCTG
.S..L..I..V..L..A..I..V..F..G..N..V..L..V..I..T..A..I..A..K..F..E..R..L..Q..T..V..T..N..Y..F..I..T..S..L..A..C..A..D..L.
GTCATGGGCCTGGCAGTGGTGCCCTTTGGGGCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGAGTTTTGGACTTCCATTGATGTGCTGTGCGTCACGGCCAGC
.V..M..G..L..A..V..V..P..F..G..A..A..H..I..L..M..K..M..W..T..F..G..N..F..W..C..E..F..W..T..S..I..D..V..L..C..V..T..A..S.
ATTGAGACCCTGTGCGTGATCGCAGTGGATCGCTACTTTGCCATTACTTCACCTTTCAAGTACCAGAGCCTGCTGACCAAGAATAAGGCCCGGGTGATCATTCTGATGGTGTGGATTGTG
.I..E..T..L..C..V..I..A..V..D..R..Y..F..A..I..T..S..P..F..K..Y..Q..S..L..L..T..K..N..K..A..R..V..I..I..L..M..V..W..I..V.
TCAGGCCTTACCTCCTTCTTGCCCATTCAGATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGCTGTGACTTCTTCACGAACCAAGCCTATGCC
.S..G..L..T..S..F..L..P..I..Q..M..H..W..Y..R..A..T..H..Q..E..A..I..N..C..Y..A..N..E..T..C..C..D..F..F..T..N..Q..A..Y..A.
ATTGCCTCTTCCATCGTGTCCTTCTACGTTCCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCAGCTCCAGAAGATTGACAAATCTGAGGGCCGCTTC
.I..A..S..S..I..V..S..F..Y..V..P..L..V..I..M..V..F..V..Y..S..R..V..F..Q..E..A..K..R..Q..L..Q..K..I..D..K..S..E..G..R..F.
CATGTCCAGAACCTTAGCCAGGTGGAGCAGGATGGGCGGACGGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAAGGAGCACAAAGCCCTCAAGACGTTAGGCATCATCATGGGC
.H..V..Q..N..L..S..Q..V..E..Q..D..G..R..T..G..H..G..L..R..R..S..S..K..F..C..L..K..E..H..K..A..L..K..T..L..G..I..I..M..G.
ACTTTCACCCTCTGCTGGCTGCCCTTCTTCATCGTTAACATTGTGCATGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTCCTAAATTGGATAGGCTATGTCAATTCTGGT
.T..F..T..L..C..W..L..P..F..F..I..V..N..I..V..H..V..I..Q..D..N..L..I..R..K..E..V..Y..I..L..L..N..W..I..G..Y..V..N..S..G.
TTCAATCCCCTTATCTACTGCCGGAGCCCAGATTTCAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTATGGGAATGGCTACTCCAGCAACGGCAACACA
.F..N..P..L..I..Y..C..R..S..P..D..F..R..I..A..F..Q..E..L..L..C..L..R..R..S..S..L..K..A..Y..G..N..G..Y..S..S..N..G..N..T.
GGGGAGCAGAGTGGATATCACGTGGAACAGGAGAAAGAAAATAAACTGCTGTGTGAAGACCTCCCAGGCACGGAAGACTTTGTGGGCCATGAAGGTACTGTGCCTAGCGATAACATTGAT
.G..E..Q..S..G..Y..H..V..E..Q..E..K..E..N..K..L..L..C..E..D..L..P..G..T..E..D..F..V..G..H..E..G..T..V..P..S..D..N..I..D.
TCACAAGGGAGGAATTGTAGTACAAATGACTCACTGCTG
.S..Q..G..R..N..C..S..T..N..D..S..L..L.
```

Between-species variability

LANGUAGE AND GENETICS DEPARTMENT

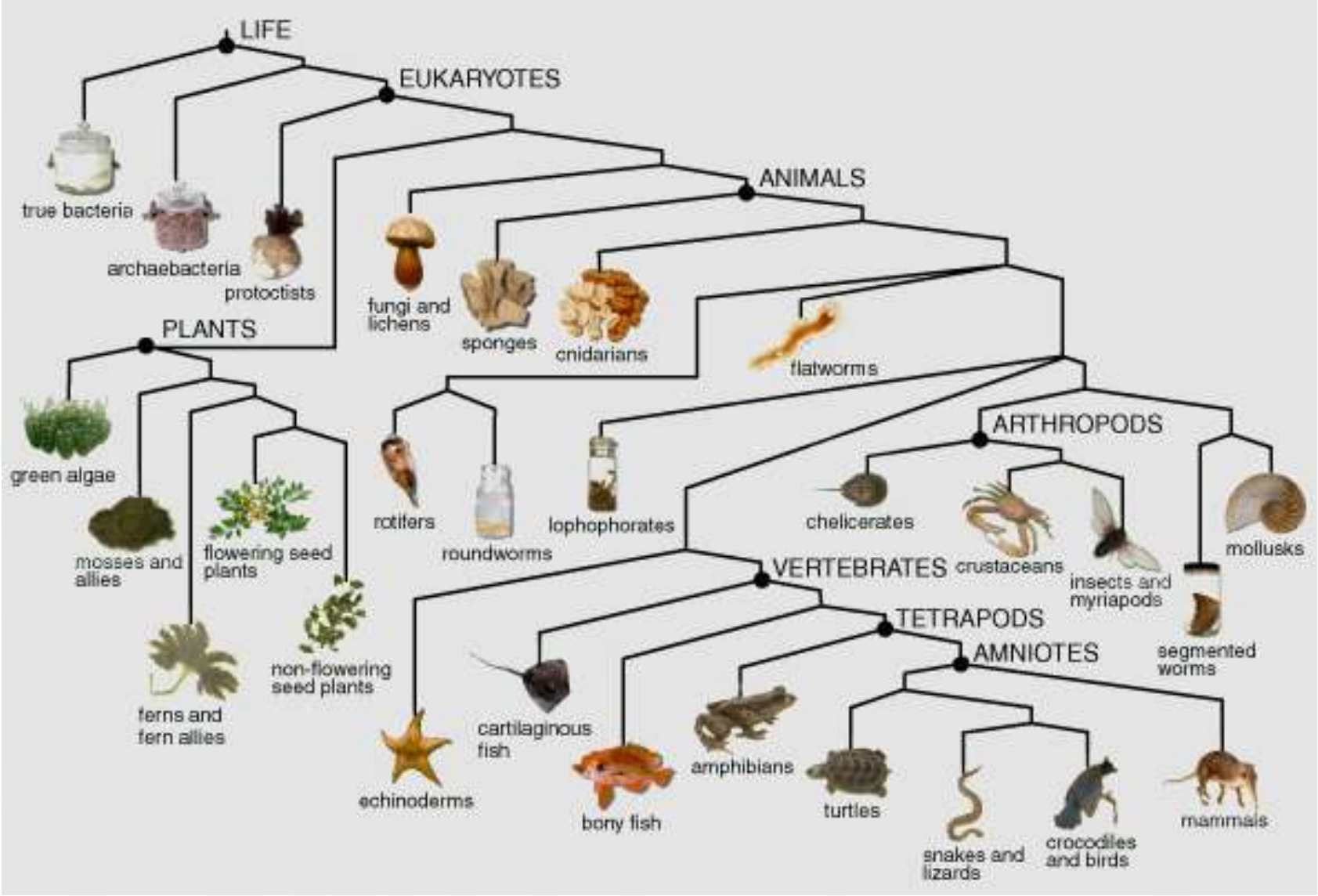
beta2-adrenergic receptor

90% identity to human

```
ATGGGGCAACCCGGGAACGGCAGCGCCTTCTTGCTGGCACCCAATGGAAGCCATGCGCCGGACCACGACGTCACGCAGGAAAGGGACGAGGTGTGGGTGGTGGGCATGGGCATCGTCATG
.M..G..Q..P..R..N..G..S..V..F..L..L..A..P..N..G..S..Q..A..P..D..H..D..A..T..Q..E..R..D..E..V..W..V..V..G..M..G..I..V..M.
TCTCTCATCGTCCTGGCCATCGTGTTTGGCAATGTGCTGGTCATCACAGCCATTGCCAAGTTCGAGCGTCTGCAGACGGTCACCAACTACTTCATCACTTCACTGGCCTGTGCTGATCTG
.S..L..I..V..L..A..I..V..F..G..N..V..L..V..I..T..A..I..A..K..F..E..R..L..Q..T..V..T..N..Y..F..I..T..S..L..A..C..A..D..L.
GTCATGGGCCTGGCAGTGGTGCCCTTTGGGGCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGAGTTTTGGACTTCCATTGATGTGCTGTGCGTCACGGCCAGC
.L..M..G..L..A..V..V..P..F..G..A..C..H..I..L..M..K..M..W..A..F..G..N..F..W..C..E..F..W..T..S..I..D..V..L..C..V..T..A..S.
ATTGAGACCCTGTGCGTGATCGCAGTGGATCGCTACTTTGCCATTACTTCACCTTTCAAGTACCAGAGCCTGCTGACCAAGAATAAGGCCCGGGTGATCATTCTGATGGTGTGGATTGTG
.I..E..T..L..C..V..I..A..V..D..R..Y..L..A..I..T..S..P..F..K..Y..Q..C..L..L..T..K..N..K..A..R..V..V..V..L..M..V..W..V..V.
TCAGGCCTTACCTCCTTCTTGCCCATTCAGATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGCTGTGACTTCTTCACGAACCAAGCCTATGCC
.S..G..L..T..S..F..L..P..I..Q..M..H..W..Y..R..A..T..H..Q..E..A..I..N..C..Y..A..K..E..T..C..C..D..F..F..T..N..Q..A..Y..A.
ATTGCCTCTTCCATCGTGTCCTTCTACGTTCCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCAGCTCCAGAAGATTGACAAATCTGAGGGCCGCTTC
.I..A..S..S..I..V..S..F..Y..L..P..L..V..V..M..V..F..V..Y..S..K..V..F..Q..V..A..K..R..Q..L..Q..K..I..D..K..S..E..G..R..F.
CATGTCCAGAACCTTAGCCAGGTGGAGCAGGATGGGCGGACGGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAAGGAGCACAAAGCCCTCAAGACGTTAGGCATCATCATGGGC
.H..T..Q..N..I..S..Q..M..E..Q..D..G..R..S..G..P..G..Q..R..R..A..S..K..F..C..L..K..E..H..K..A..L..K..T..L..G..I..I..M..G.
ACTTTCACCCTCTGCTGGCTGCCCTTCTTCATCGTTAACATTGTGCATGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTCCTAAATTGGATAGGCTATGTCAATTCTGGT
.T..F..T..L..C..W..L..P..F..F..I..V..N..I..V..H..V..I..Q..D..N..L..I..P..K..E..V..Y..I..L..L..N..W..V..G..Y..V..N..S..A.
TTCAATCCCCTTATCTACTGCCGGAGCCCAGATTTCAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTATGGGAATGGCTACTCCAGCAACGGCAACACA
.F..N..P..L..I..Y..C..R..S..P..D..F..R..I..A..F..Q..E..L..L..C..Q..R..K..S..S..L..K..A..S..G..D..G..Y..S..S..N..G..N..T.
GGGGAGCAGAGTGGATATCACGTGGAACAGGAGAAAGAAAATAAACTGCTGTGTGAAGACCTCCCAGGCACGGAAGACTTTGTGGGCCATCAAGGTACTGTGCCTAGCGATAACATTGAT
.G..E..Q..S..G..Y..H..V..E..Q..E..K..E..N..K..L..L..C..E..D..L..P..G..T..E..D..F..V..G..H..Q..G..T..V..P..S..D..N..I..D.
TCACAAGGGAGGAATTGTAGTACAAATGACTCACTGCTG
.S..Q..G..R..N..C..S..T..N..D..S..L..L.
```

sequence **conservation**

Between-species variability



The origin of genetic variation

- Occasionally a **change** (**mutation**) occurs in the DNA sequence of a cell's genome.

- Changes happen all the time:
  - through **damage** to the DNA
  - through copying **errors** when cells divide

- Where in the genome changes occur is **random**.

ATGCGTGTTTCACAGTG

ATGCGTGTATCACAGTG

The origin of genetic variation

- Changes in the DNA of eggs and sperm result in offspring with a new DNA variant.

- These DNA changes provide the raw material for **natural selection**.

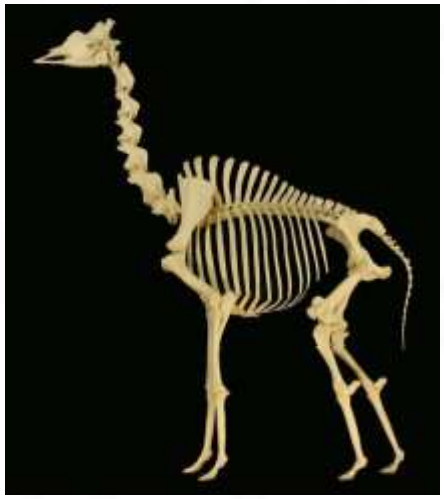
ATGCGTGTTTCACAGTG

↓

ATGCGTGTATCACAGTG

Evolution and the origin of species

- If a particular change helps an organism to **survive** and **proliferate**, that mutation will become **more common** in the next generation – over time, **all members** of the population may have the new variant.



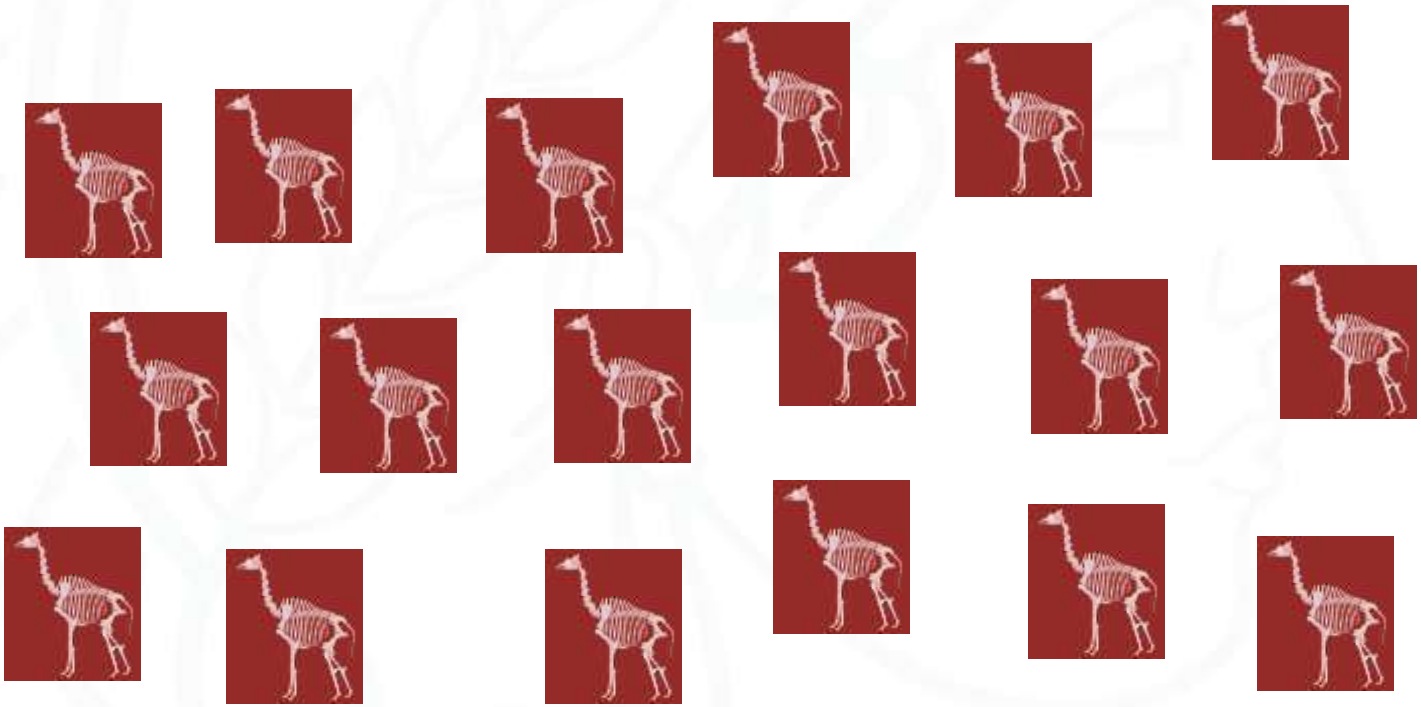
Parent
Short giraffe

Random
DNA change

Offspring
Tall giraffe

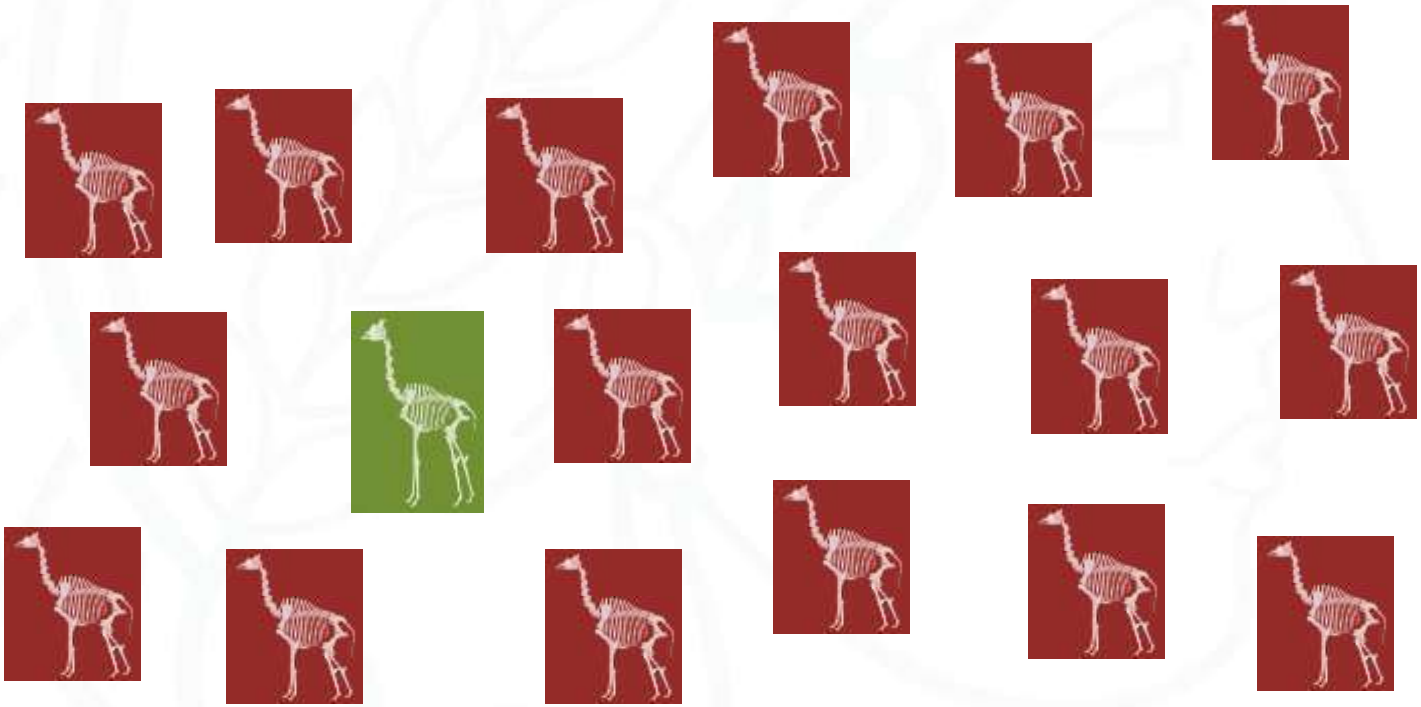
Evolution and the origin of species

- If a particular change helps an organism to **survive** and **proliferate**, that mutation will become **more common** in the next generation – over time, **all members** of the population may have the new variant.



Evolution and the origin of species

- If a particular change helps an organism to **survive** and **proliferate**, that mutation will become **more common** in the next generation – over time, **all members** of the population may have the new variant.

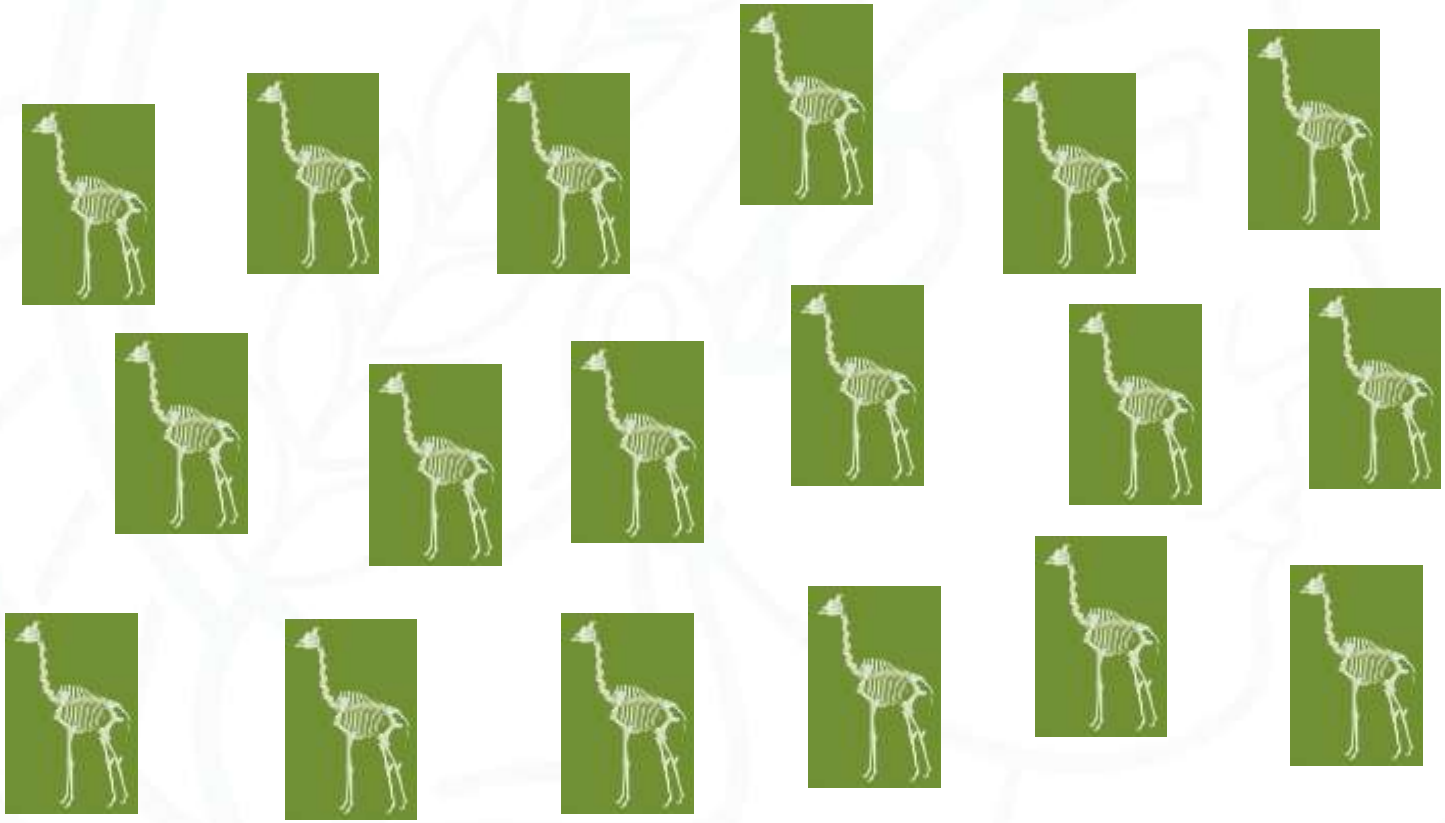


Evolution and the origin of species

- If a particular change helps an organism to **survive** and **proliferate**, that mutation will become **more common** in the next generation – over time, **all members** of the population may have the new variant.

Evolution and the origin of species

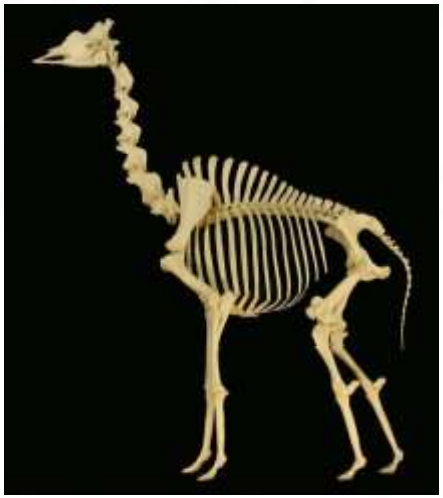
- If a particular change helps an organism to **survive** and **proliferate**, that mutation will become **more common** in the next generation – over time, **all members** of the population may have the new variant.



Evolution and the origin of species

- If a particular change is **detrimental** to an organism's ability to survive and proliferate, that mutation will not be passed on to the next generation and will **disappear**.

Random
DNA change

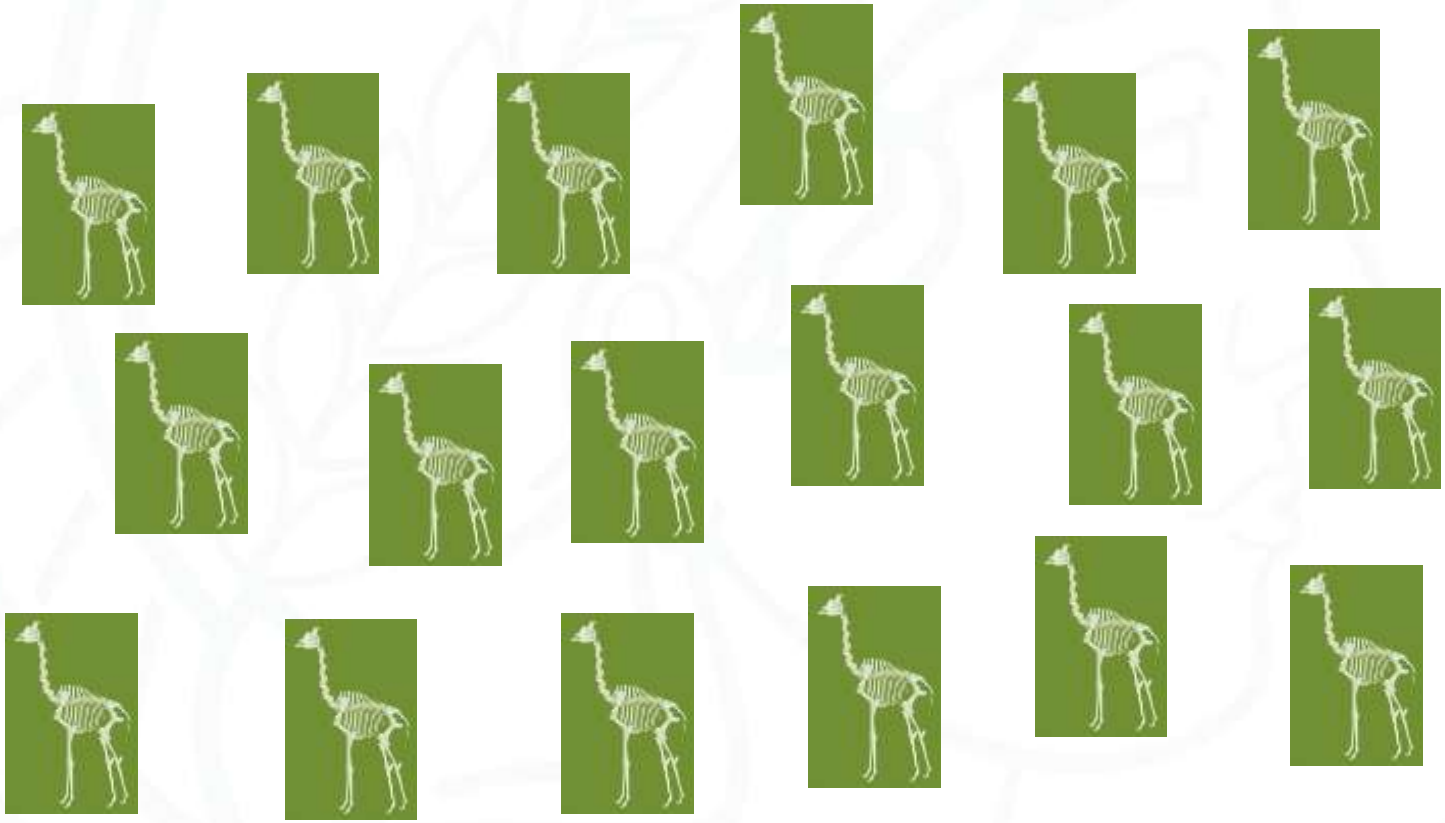


Parent
Tall giraffe

Offspring
Short giraffe

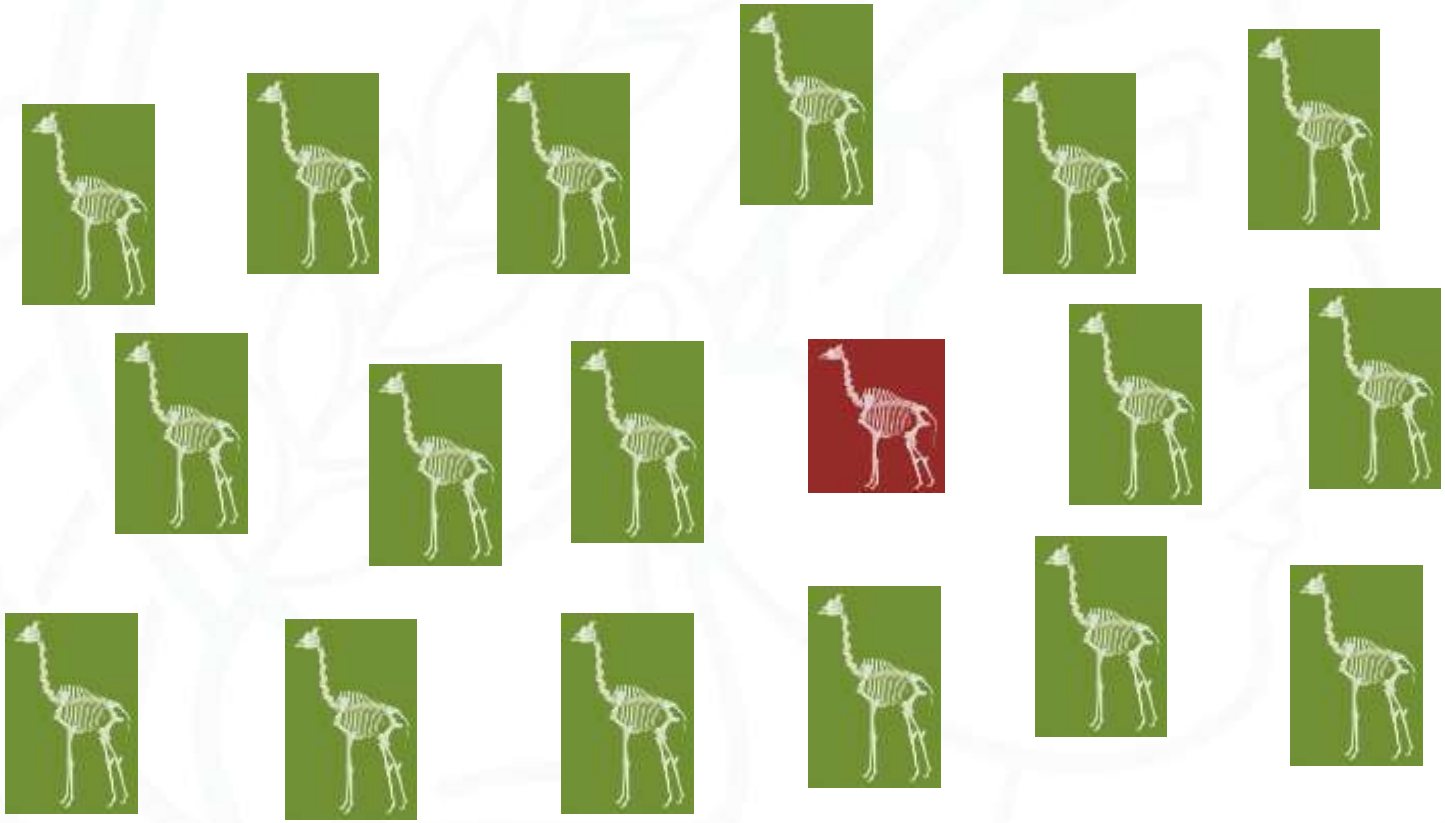
Evolution and the origin of species

- If a particular change is **detrimental** to an organism's ability to survive and proliferate, that mutation will not be passed on to the next generation and will **disappear**.



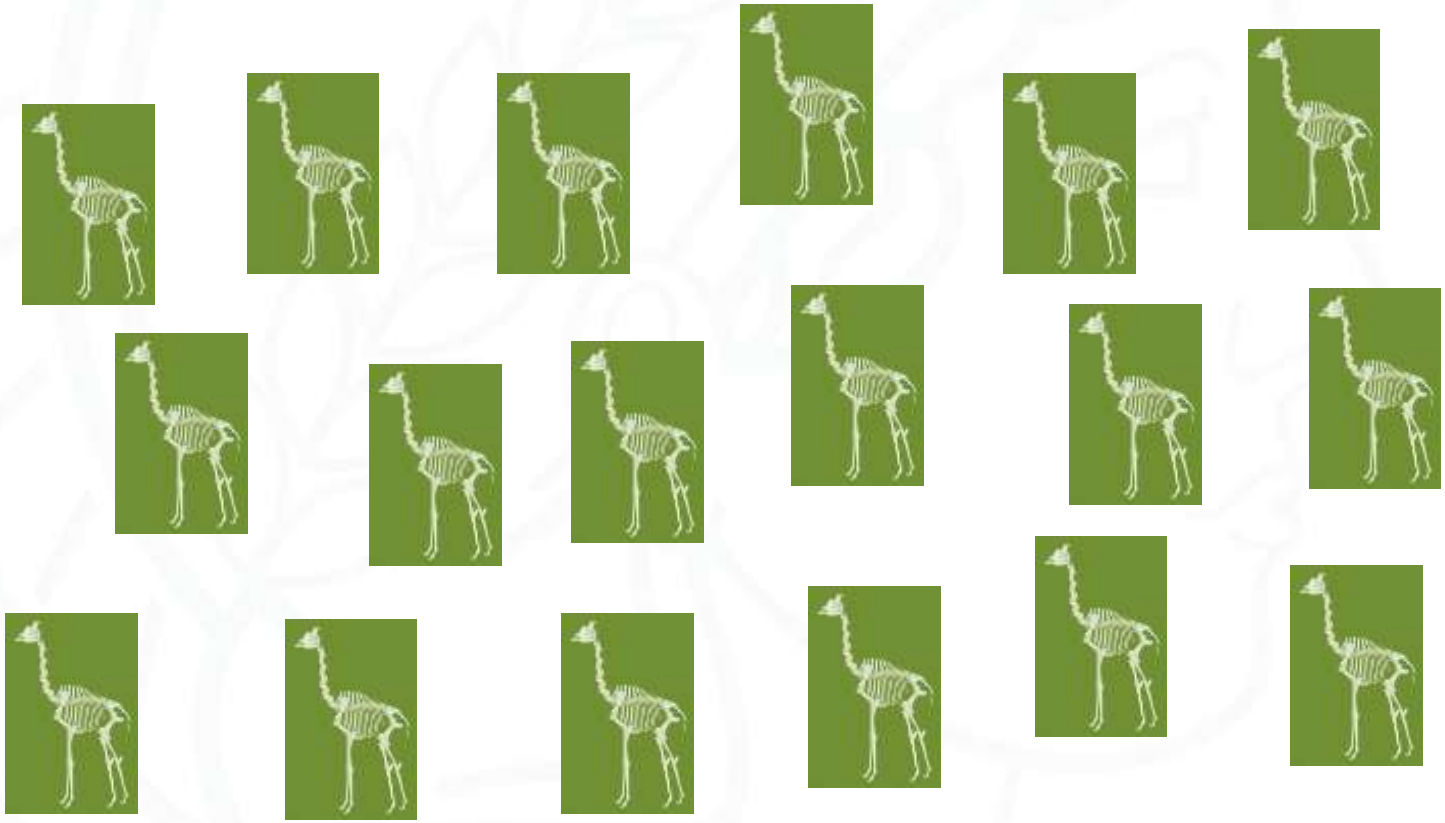
Evolution and the origin of species

- If a particular change is **detrimental** to an organism's ability to survive and proliferate, that mutation will not be passed on to the next generation and will **disappear**.



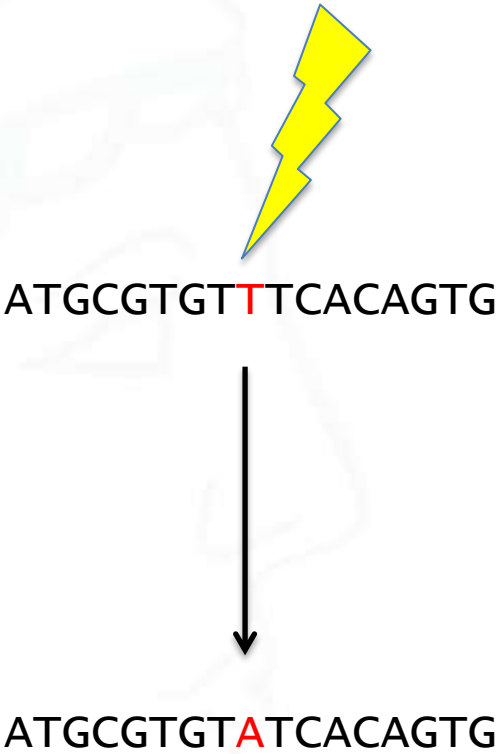
Evolution and the origin of species

- If a particular change is **detrimental** to an organism's ability to survive and proliferate, that mutation will not be passed on to the next generation and will **disappear**.



Evolution and the origin of species

- Over generations, DNA changes with **strong beneficial** effects become **fixed**, while changes with **strong detrimental** effects disappear.
- Most changes **do not** have such strong effects.
- They may be beneficial sometimes, detrimental at other times.
- Or they may have **no significant effect** over an organism's ability to survive and proliferate.



Evolution and the origin of species

- DNA changes with weak or neutral effects **persist** in the population over generations.
- The frequency of neutral changes is influenced by **probabilistic** factors more than by natural selection.
  - Genetic **drift**
  - Population bottlenecks
  - Migration
- These effects are prominent in the genetic differences between human populations.
- **Culture** also plays a role.

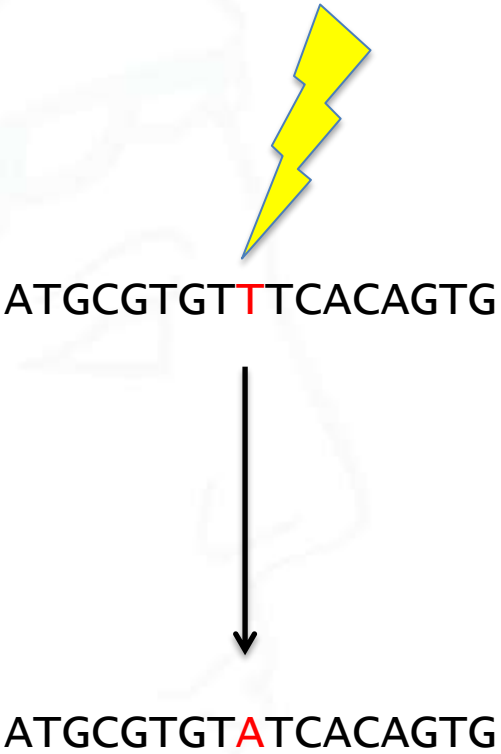
ATGCGTGTTTCACAGTG

↓

ATGCGTGTATCACAGTG

Evolution and the origin of species

- DNA changes with weak or neutral effects that persist over generations constitute a species' **genetic diversity**.
- Diversity increases a species' **resilience** to environmental changes (disease, climate).



Variability

1. Differences between humans and other species

2. Differences between human individuals

Within-species variability

LANGUAGE AND GENETICS DEPARTMENT

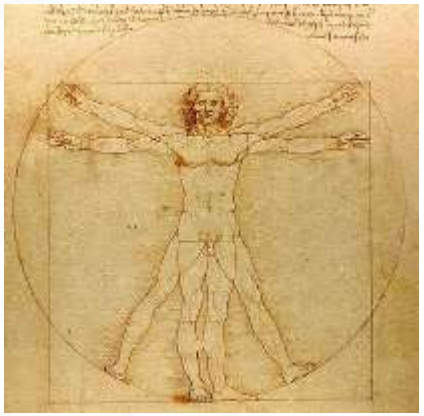
- Why do children acquire native language at different rates?
- Why do some people have difficulty with specific aspects of speech, language or reading?
- Why do people differ in their aptitude for second language learning?

Within-species variability

LANGUAGE AND GENETICS DEPARTMENT

Beta2-adrenergic receptor

Reference sequence



```
ATGGGGCAACCCGGGAACGGCAGCGCCTTCTTGCTGGCACCCAATGGAAGCCATGCGCCGGACCACGACGTCACGCAGGAAAGGGACGAGGTGTGGGTGGTGGGCATGGGCATCGTCATG
.M..G..Q..P..G..N..G..S..A..F..L..L..A..P..N..G..S..H..A..P..D..H..D..V..T..Q..E..R..D..E..V..W..V..V..G..M..G..I..V..M.
TCTCTCATCGTCCTGGCCATCGTGTTTGGCAATGTGCTGGTCATCACAGCCATTGCCAAGTTCGAGCGTCTGCAGACGGTCACCAACTACTTCATCACTTCACTGGCCTGTGCTGATCTG
.S..L..I..V..L..A..I..V..F..G..N..V..L..V..I..T..A..I..A..K..F..E..R..L..Q..T..V..T..N..Y..F..I..T..S..L..A..C..A..D..L.
GTCATGGGCCTGGCAGTGGTGCCCTTTGGGGCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGAGTTTTGGACTTCCATTGATGTGCTGTGCGTCACGGCCAGC
.V..M..G..L..A..V..V..P..F..G..A..A..H..I..L..M..K..M..W..T..F..G..N..F..W..C..E..F..W..T..S..I..D..V..L..C..V..T..A..S.
ATTGAGACCCTGTGCGTGATCGCAGTGGATCGCTACTTTGCCATTACTTCACCTTTCAAGTACCAGAGCCTGCTGACCAAGAATAAGGCCCGGGTGATCATTCTGATGGTGTGGATTGTG
.I..E..T..L..C..V..I..A..V..D..R..Y..F..A..I..T..S..P..F..K..Y..Q..S..L..L..T..K..N..K..A..R..V..I..I..L..M..V..W..I..V.
TCAGGCCTTACCTCCTTCTTGCCCATTCAGATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGCTGTGACTTCTTCACGAACCAAGCCTATGCC
.S..G..L..T..S..F..L..P..I..Q..M..H..W..Y..R..A..T..H..Q..E..A..I..N..C..Y..A..N..E..T..C..C..D..F..F..T..N..Q..A..Y..A.
ATTGCCTCTTCCATCGTGTCCTTCTACGTTCCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCAGCTCCAGAAGATTGACAAATCTGAGGGCCGCTTC
.I..A..S..S..I..V..S..F..Y..V..P..L..V..I..M..V..F..V..Y..S..R..V..F..Q..E..A..K..R..Q..L..Q..K..I..D..K..S..E..G..R..F.
CATGTCCAGAACCTTAGCCAGGTGGAGCAGGATGGGCGGACGGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAAGGAGCACAAAGCCCTCAAGACGTTAGGCATCATCATGGGC
.H..V..Q..N..L..S..Q..V..E..Q..D..G..R..T..G..H..G..L..R..R..S..S..K..F..C..L..K..E..H..K..A..L..K..T..L..G..I..I..M..G.
ACTTTCACCCTCTGCTGGCTGCCCTTCTTCATCGTTAACATTGTGCATGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTCCTAAATTGGATAGGCTATGTCAATTCTGGT
.T..F..T..L..C..W..L..P..F..F..I..V..N..I..V..H..V..I..Q..D..N..L..I..R..K..E..V..Y..I..L..L..N..W..I..G..Y..V..N..S..G.
TTCAATCCCCTTATCTACTGCCGGAGCCCAGATTTCAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTATGGGAATGGCTACTCCAGCAACGGCAACACA
.F..N..P..L..I..Y..C..R..S..P..D..F..R..I..A..F..Q..E..L..L..C..L..R..R..S..S..L..K..A..Y..G..N..G..Y..S..S..N..G..N..T.
GGGGAGCAGAGTGGATATCACGTGGAACAGGAGAAAGAAAATAAACTGCTGTGTGAAGACCTCCCAGGCACGGAAGACTTTGTGGGCCATCAAGGTACTGTGCCTAGCGATAACATTGAT
.G..E..Q..S..G..Y..H..V..E..Q..E..K..E..N..K..L..L..C..E..D..L..P..G..T..E..D..F..V..G..H..Q..G..T..V..P..S..D..N..I..D.
TCACAAGGGAGGAATTGTAGTACAAATGACTCACTGCTG
.S..Q..G..R..N..C..S..T..N..D..S..L..L.
```

Within-species variability

Beta2-adrenergic receptor

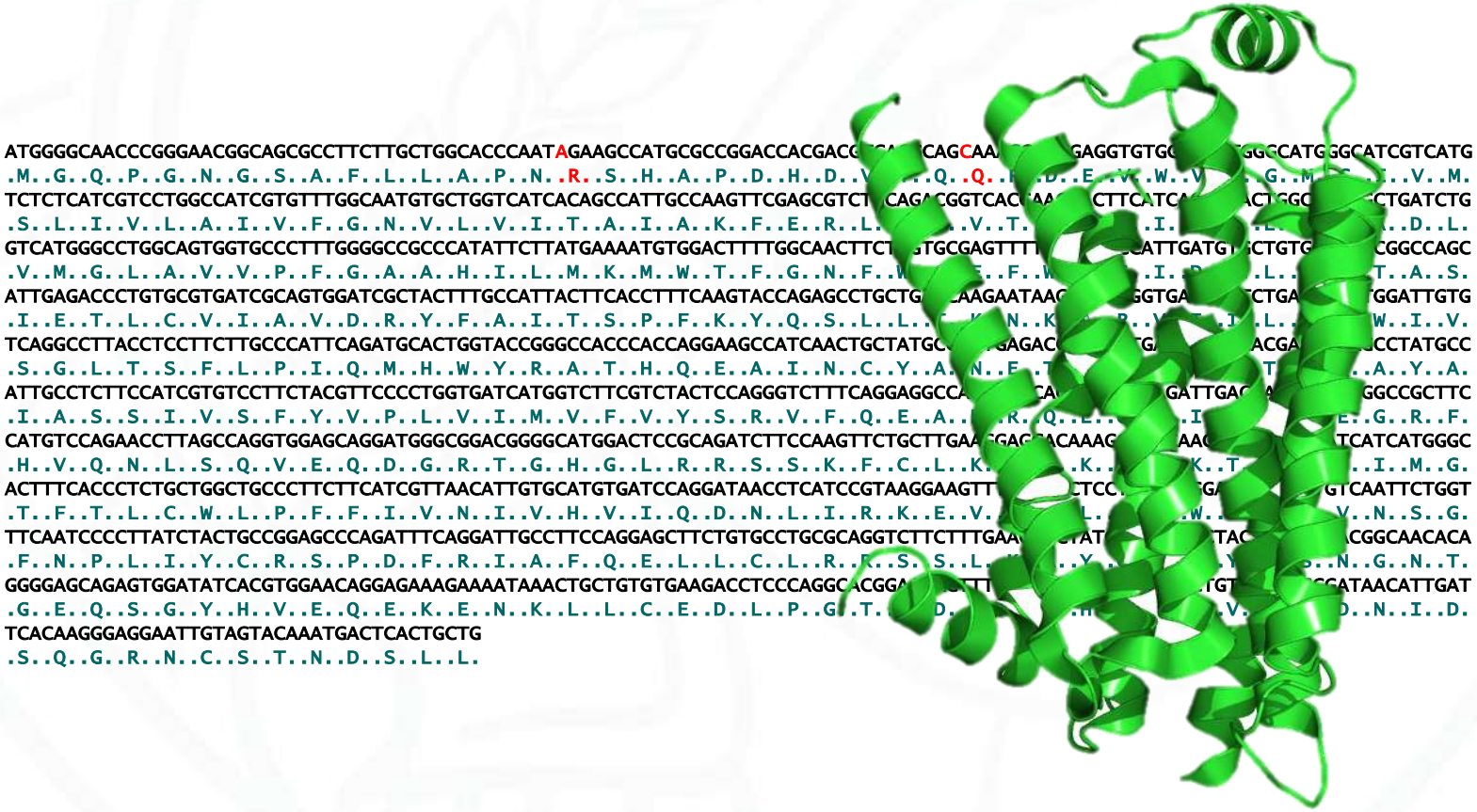
Changes compared to the **reference** are called **polymorphisms**

```
ATGGGGCAACCCGGGAACGGCAGCGCCTTCTTGCTGGCACCCAATAGAAGCCATGCGCCGGACCACGACGTCACGCAGCAAAGGGACGAGGTGTGGGTGGTGGGCATGGGCATCGTCATG
.M..G..Q..P..G..N..G..S..A..F..L..L..A..P..N..R..S..H..A..P..D..H..D..V..T..Q..Q..R..D..E..V..W..V..V..G..M..G..I..V..M.
TCTCTCATCGTCCTGGCCATCGTGTTTGGCAATGTGCTGGTCATCACAGCCATTGCCAAGTTCGAGCGTCTGCAGACGGTCACCAACTACTTCATCACTTCACTGGCCTGTGCTGATCTG
.S..L..I..V..L..A..I..V..F..G..N..V..L..V..I..T..A..I..A..K..F..E..R..L..Q..T..V..T..N..Y..F..I..T..S..L..A..C..A..D..L.
GTCATGGGCCTGGCAGTGGTGCCCTTTGGGGCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGAGTTTTGGACTTCCATTGATGTGCTGTGCGTCACGGCCAGC
.V..M..G..L..A..V..V..P..F..G..A..A..H..I..L..M..K..M..W..T..F..G..N..F..W..C..E..F..W..T..S..I..D..V..L..C..V..T..A..S.
ATTGAGACCCTGTGCGTGATCGCAGTGGATCGCTACTTTGCCATTACTTCACCTTTCAAGTACCAGAGCCTGCTGACCAAGAATAAGGCCCGGGTGATCATTCTGATGGTGTGGATTGTG
.I..E..T..L..C..V..I..A..V..D..R..Y..F..A..I..T..S..P..F..K..Y..Q..S..L..L..T..K..N..K..A..R..V..I..I..L..M..V..W..I..V.
TCAGGCCTTACCTCCTTCTTGCCCATTCAGATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGCTGTGACTTCTTCACGAACCAAGCCTATGCC
.S..G..L..T..S..F..L..P..I..Q..M..H..W..Y..R..A..T..H..Q..E..A..I..N..C..Y..A..N..E..T..C..C..D..F..F..T..N..Q..A..Y..A.
ATTGCCTCTTCCATCGTGTCCTTCTACGTTCCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCAGCTCCAGAAGATTGACAAATCTGAGGGCCGCTTC
.I..A..S..S..I..V..S..F..Y..V..P..L..V..I..M..V..F..V..Y..S..R..V..F..Q..E..A..K..R..Q..L..Q..K..I..D..K..S..E..G..R..F.
CATGTCCAGAACCTTAGCCAGGTGGAGCAGGATGGGCGGACGGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAAGGAGCACAAAGCCCTCAAGACGTTAGGCATCATCATGGGC
.H..V..Q..N..L..S..Q..V..E..Q..D..G..R..T..G..H..G..L..R..R..S..S..K..F..C..L..K..E..H..K..A..L..K..T..L..G..I..I..M..G.
ACTTTCACCCTCTGCTGGCTGCCCTTCTTCATCGTTAACATTGTGCATGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTCCTAAATTGGATAGGCTATGTCAATTCTGGT
.T..F..T..L..C..W..L..P..F..F..I..V..N..I..V..H..V..I..Q..D..N..L..I..R..K..E..V..Y..I..L..L..N..W..I..G..Y..V..N..S..G.
TTCAATCCCCTTATCTACTGCCGGAGCCCAGATTTCAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTATGGGAATGGCTACTCCAGCAACGGCAACACA
.F..N..P..L..I..Y..C..R..S..P..D..F..R..I..A..F..Q..E..L..L..C..L..R..R..S..S..L..K..A..Y..G..N..G..Y..S..S..N..G..N..T.
GGGGAGCAGAGTGGATATCACGTGGAACAGGAGAAAGAAAATAAACTGCTGTGTGAAGACCTCCCAGGCACGGAAGACTTTGTGGGCCATCAAGGTACTGTGCCTAGCGATAACATTGAT
.G..E..Q..S..G..Y..H..V..E..Q..E..K..E..N..K..L..L..C..E..D..L..P..G..T..E..D..F..V..G..H..Q..G..T..V..P..S..D..N..I..D.
TCACAAGGGAGGAATTGTAGTACAAATGACTCACTGCTG
.S..Q..G..R..N..C..S..T..N..D..S..L..L.
```

Within-species variability

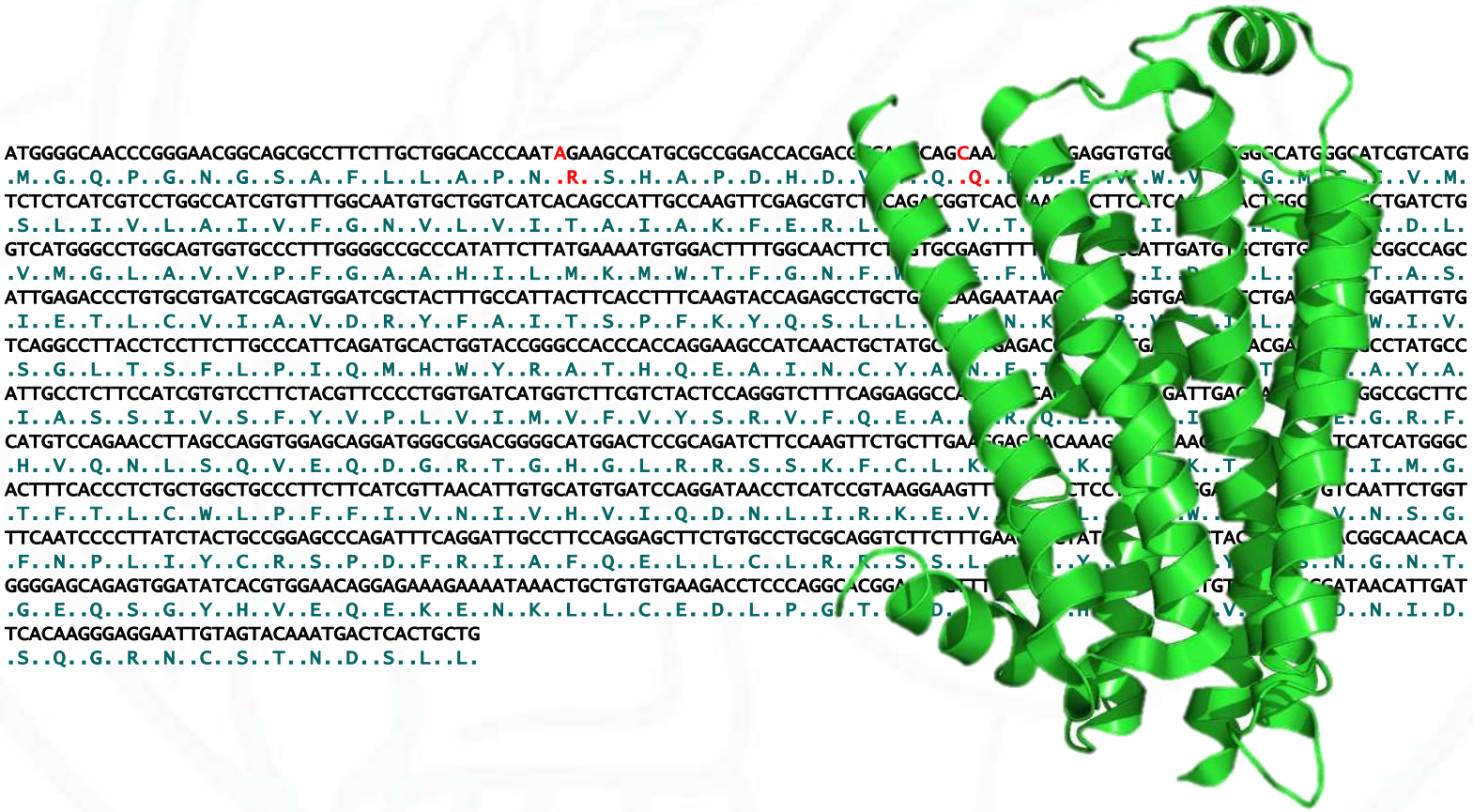
Beta2-adrenergic receptor

Changes compared to the **reference** are called **polymorphisms**



Within-species variability

Effects – possible **very small** differences in susceptibility to asthma, obesity.



Within-species variability

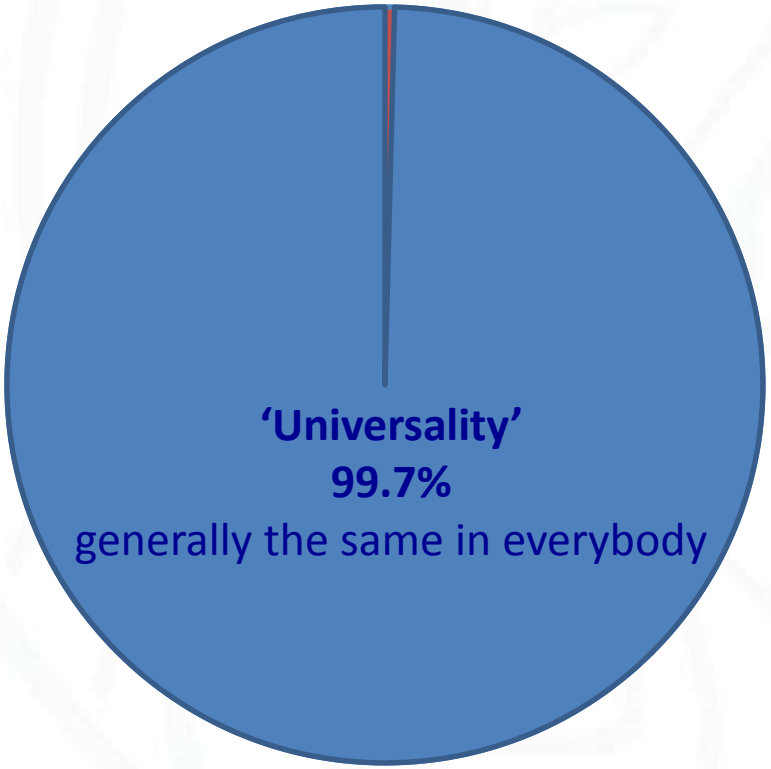
LANGUAGE AND GENETICS DEPARTMENT

Single nucleotide polymorphisms – SNPs

```
ATGGGGCAACCCGGGAACGGCAGCGCCTTCTTGCTGGCACCCAATAGAAGCCATGCGCCGGACCACGACGTCACGCAGCAAAGGGACGAGGTGTGGGTGGTGGGCATGGGCATCGTCATG
.M..G..Q..P..G..N..G..S..A..F..L..L..A..P..N..R..S..H..A..P..D..H..D..V..T..Q..Q..R..D..E..V..W..V..V..G..M..G..I..V..M.
TCTCTCATCGTCCTGGCCATCGTGTTTGGCAATGTGCTGGTCATCACAGCCATTGCCAAGTTCGAGCGTCTGCAGACGGTCACCAACTACTTCATCACTTCACTGGCCTGTGCTGATCTG
.S..L..I..V..L..A..I..V..F..G..N..V..L..V..I..T..A..I..A..K..F..E..R..L..Q..T..V..T..N..Y..F..I..T..S..L..A..C..A..D..L.
GTCATGGGCCTGGCAGTGGTGCCCTTTGGGGCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGAGTTTTGGACTTCCATTGATGTGCTGTGCGTCACGGCCAGC
.V..M..G..L..A..V..V..P..F..G..A..A..H..I..L..M..K..M..W..T..F..G..N..F..W..C..E..F..W..T..S..I..D..V..L..C..V..T..A..S.
ATTGAGACCCTGTGCGTGATCGCAGTGGATCGCTACTTTGCCATTACTTCACCTTTCAAGTACCAGAGCCTGCTGACCAAGAATAAGGCCCGGGTGATCATTCTGATGGTGTGGATTGTG
.I..E..T..L..C..V..I..A..V..D..R..Y..F..A..I..T..S..P..F..K..Y..Q..S..L..L..T..K..N..K..A..R..V..I..I..L..M..V..W..I..V.
TCAGGCCTTACCTCCTTCTTGCCCATTCAGATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGCTGTGACTTCTTCACGAACCAAGCCTATGCC
.S..G..L..T..S..F..L..P..I..Q..M..H..W..Y..R..A..T..H..Q..E..A..I..N..C..Y..A..N..E..T..C..C..D..F..F..T..N..Q..A..Y..A.
ATTGCCTCTTCCATCGTGTCCTTCTACGTTCCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCAGCTCCAGAAGATTGACAAATCTGAGGGCCGCTTC
.I..A..S..S..I..V..S..F..Y..V..P..L..V..I..M..V..F..V..Y..S..R..V..F..Q..E..A..K..R..Q..L..Q..K..I..D..K..S..E..G..R..F.
CATGTCCAGAACCTTAGCCAGGTGGAGCAGGATGGGCGGACGGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAAGGAGCACAAAGCCCTCAAGACGTTAGGCATCATCATGGGC
.H..V..Q..N..L..S..Q..V..E..Q..D..G..R..T..G..H..G..L..R..R..S..S..K..F..C..L..K..E..H..K..A..L..K..T..L..G..I..I..M..G.
ACTTTCACCCTCTGCTGGCTGCCCTTCTTCATCGTTAACATTGTGCATGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTCCTAAATTGGATAGGCTATGTCAATTCTGGT
.T..F..T..L..C..W..L..P..F..F..I..V..N..I..V..H..V..I..Q..D..N..L..I..R..K..E..V..Y..I..L..L..N..W..I..G..Y..V..N..S..G.
TTCAATCCCCTTATCTACTGCCGGAGCCCAGATTTCAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTATGGGAATGGCTACTCCAGCAACGGCAACACA
.F..N..P..L..I..Y..C..R..S..P..D..F..R..I..A..F..Q..E..L..L..C..L..R..R..S..S..L..K..A..Y..G..N..G..Y..S..S..N..G..N..T.
GGGGAGCAGAGTGGATATCACGTGGAACAGGAGAAAGAAAATAAACTGCTGTGTGAAGACCTCCCAGGCACGGAAGACTTTGTGGGCCATCAAGGTACTGTGCCTAGCGATAACATTGAT
.G..E..Q..S..G..Y..H..V..E..Q..E..K..E..N..K..L..L..C..E..D..L..P..G..T..E..D..F..V..G..H..Q..G..T..V..P..S..D..N..I..D.
TCACAAGGGAGGAATTGTAGTACAAATGACTCACTGCTG
.S..Q..G..R..N..C..S..T..N..D..S..L..L.
```

SNPs

- The human species has accumulated a lot of genetic variation.



'Variability'
0.3%
= 10 million nucleotides
different in >1% of the population
↓
the **SNPs**

SNPs

Effects at the molecular level

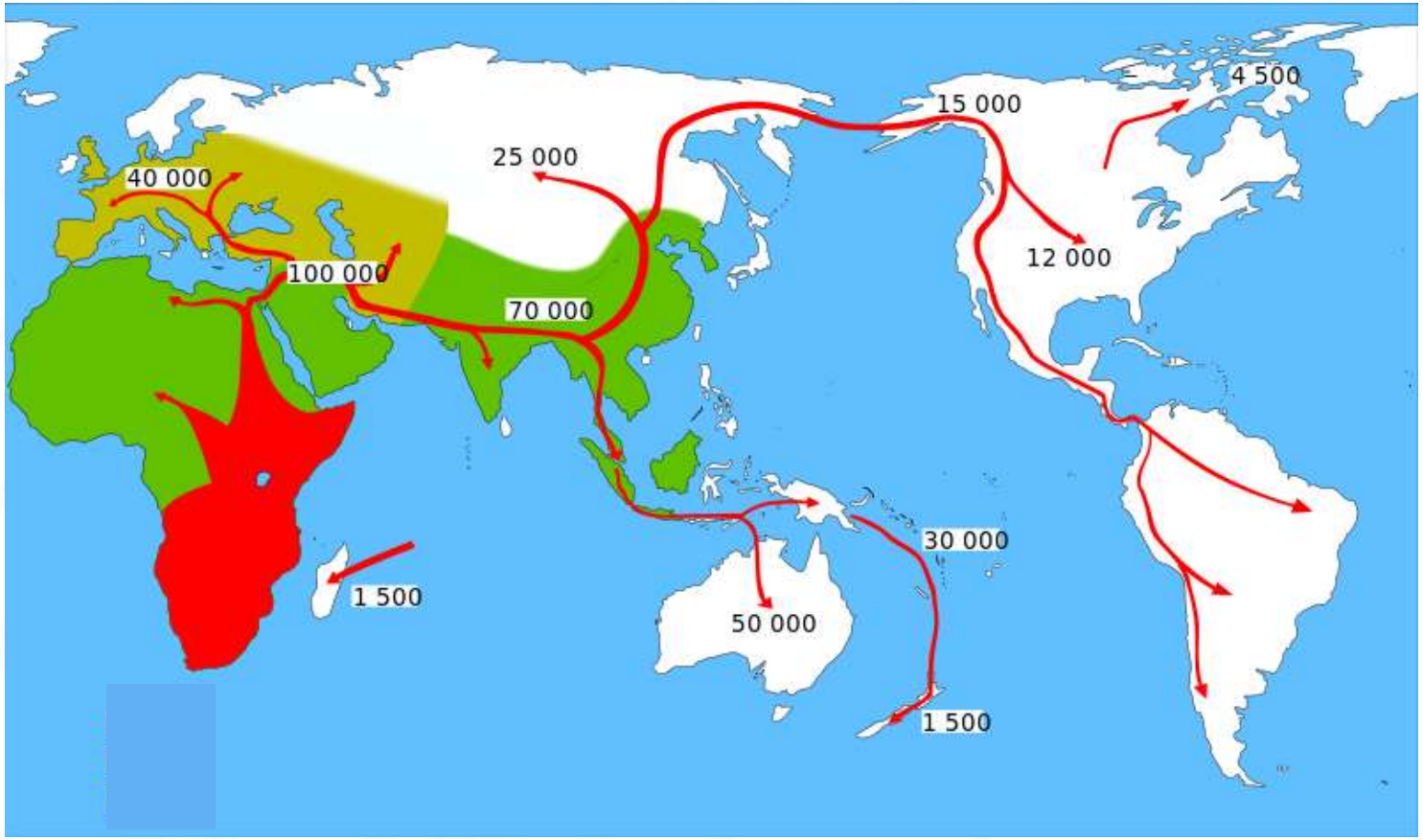
- A minority (<1%) **change protein sequences**.
- Some affect **regulation of gene expression**.
- Many may have **no effect** at all.

Effects at the phenotype level

- Probably responsible for much of the **genetic** variation in:
  - how we **look**
  - how we **think**
  - what **medical conditions** we get and how we respond to treatment

SNPs

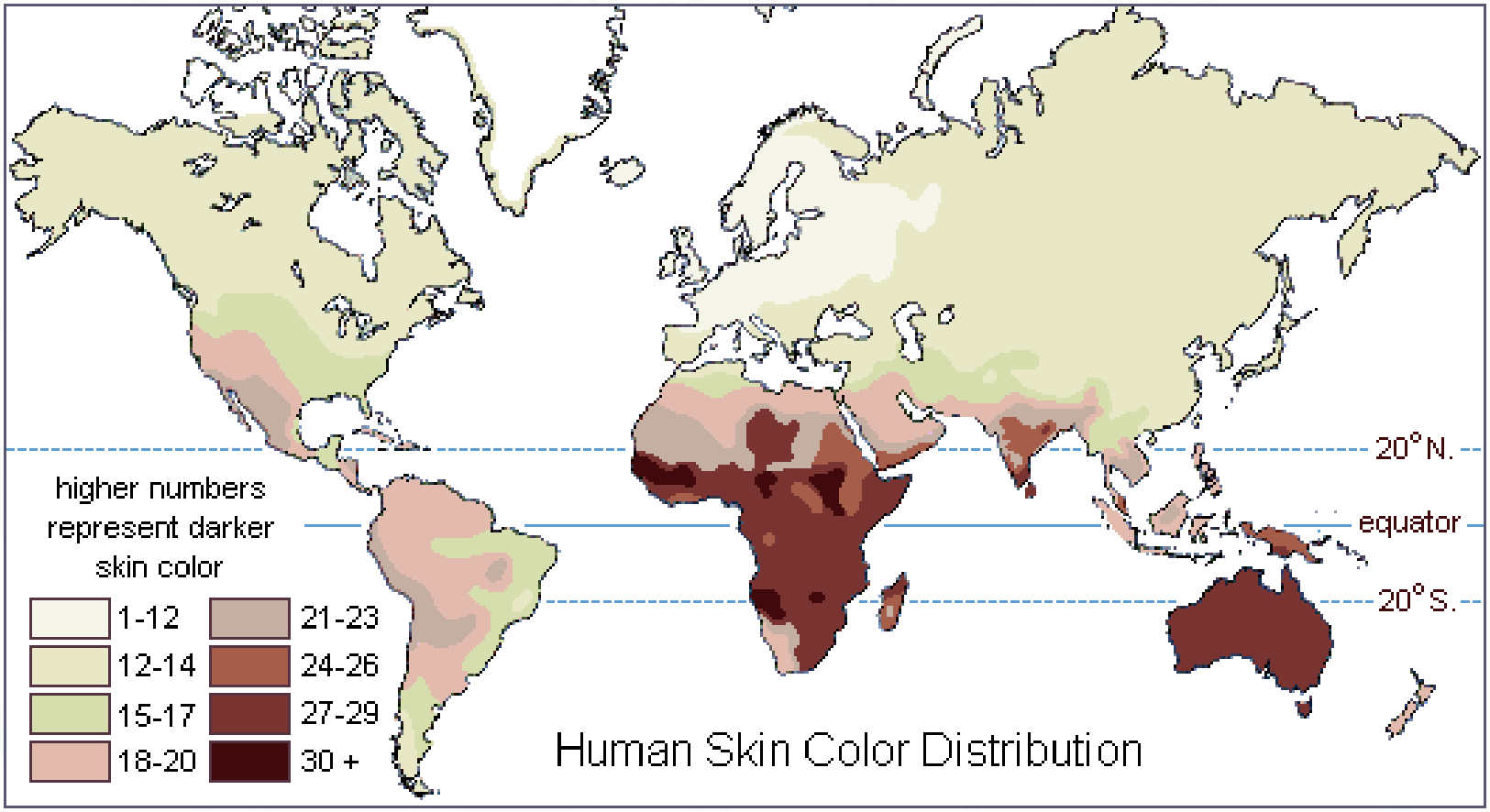
- Most of the SNPs arose during our **early evolution** in Africa and are now found in populations all over the world.



LANGUAGE AND GENETICS DEPARTMENT

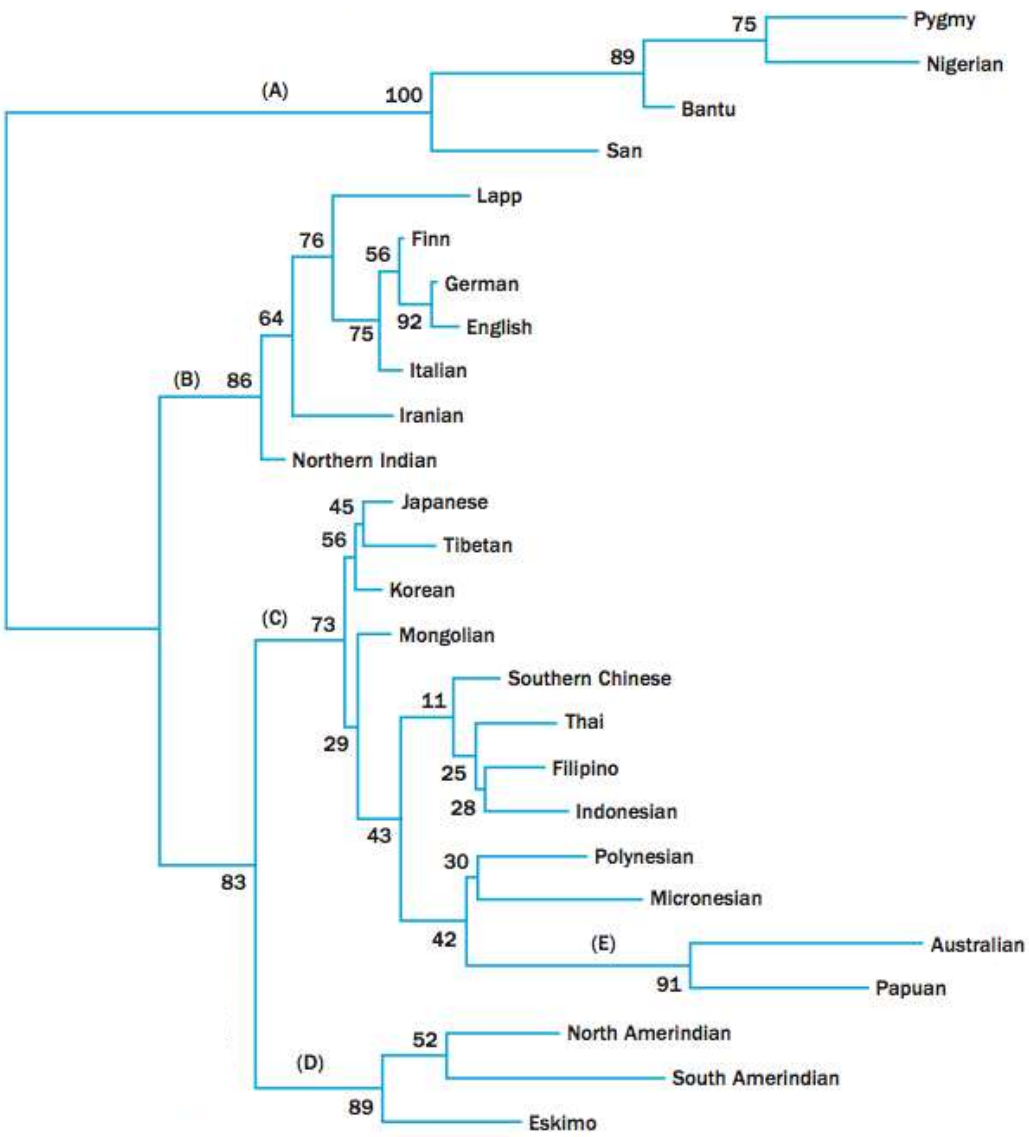
SNPs

- Some SNPs, such as those associated with skin pigmentation, have arisen **since** the migration out of Africa. Some of these may have conferred an **advantage in new environments**.



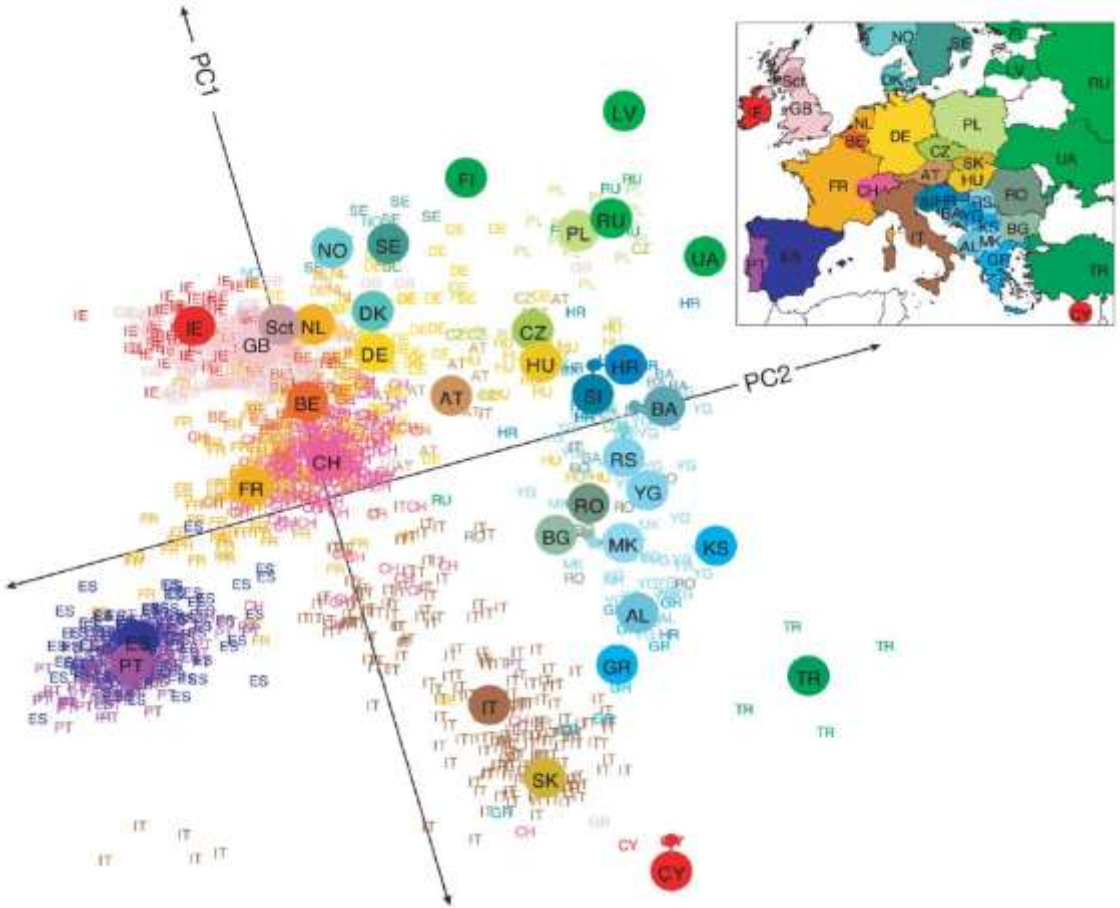
LANGUAGE AND GENETICS DEPARTMENT

SNPs


Pygmy
Nigerian
Bantu
San
(A)
100
89
75
Lapp
Finn
German
English
Italian
Iranian
Northern Indian
76
56
92
75
64
(B)
86
Japanese
Tibetan
Korean
Mongolian
45
56
(C)
73
Southern Chinese
Thai
Filipino
Indonesian
11
25
28
29
43
Polynesian
Micronesian
30
42
83
(E)
Australian
Papuan
91
North Amerindian
South Amerindian
Eskimo
52
(D)
89

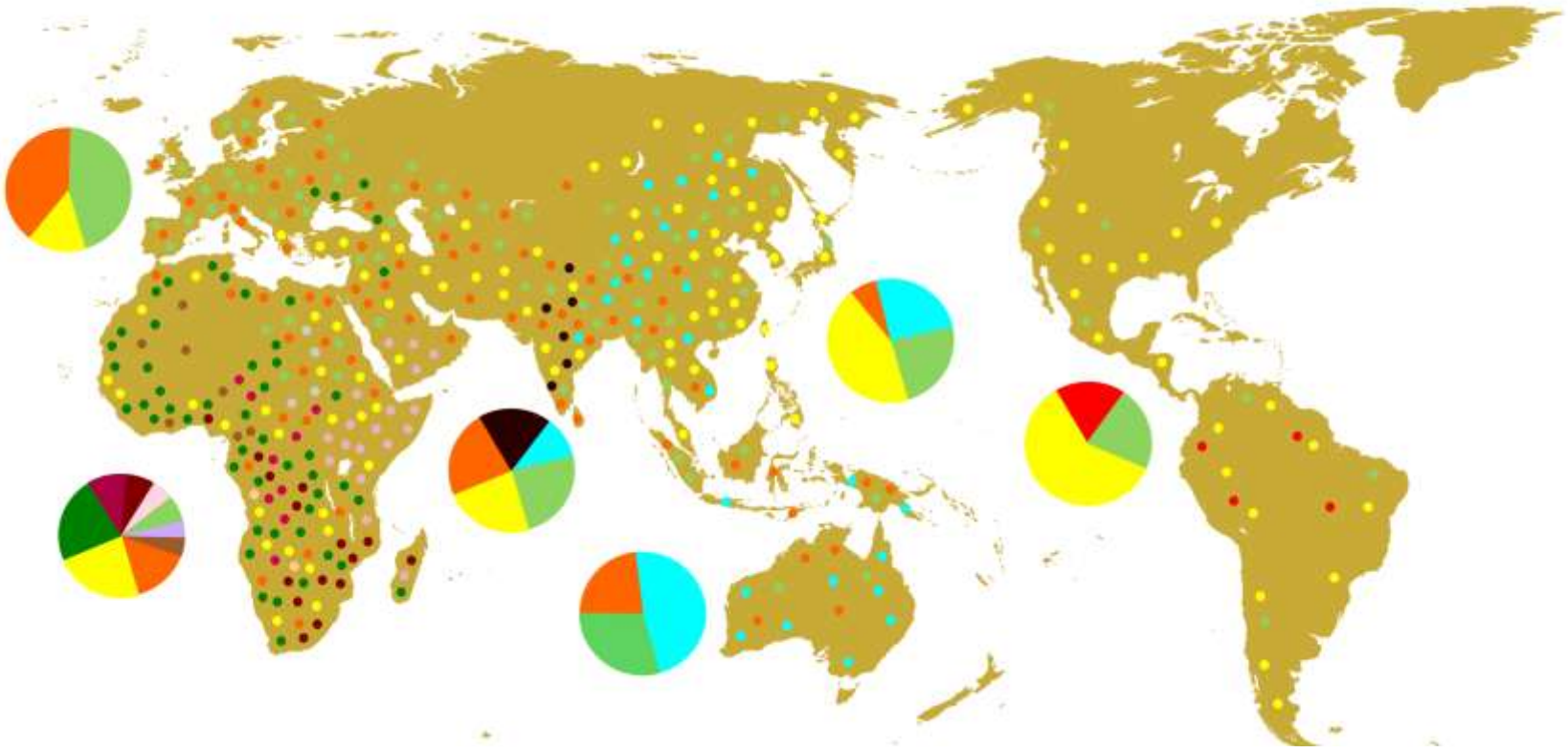
SNPs

- We are genetically most similar to people from the same place as us – but differences **within** populations (ancient SNPs) are much bigger than differences **between** populations (newer SNPs).



SNPs

- **Greatest genetic diversity is in Africa** – after migration out of Africa, diversity was **lost** through serial **founder effects** faster than it was generated by new changes.



Variation in action

An extract from the **ABCC11** gene on chromosome 16

…ACCAAGTCTGCCACTTACTGGCCTGAGTACACTGGCAATGCAGAAG…
…ACCAAGTCTGCCACTTACTGGCCCGAGTACACTGGCAATGCAGAAG…

SNP
rs17922931

Two different **alleles** of ABCC11

Variation in action

LANGUAGE AND GENETICS DEPARTMENT

Heterozygous

...ACCAAGTCTGCCACTTACTGGCCTGAGTACACTGGCAATGCAGAAG...
...ACCAAGTCTGCCACTTACTGGCCCGAGTACACTGGCAATGCAGAAG...

Homozygous

...ACCAAGTCTGCCACTTACTGGCCTGAGTACACTGGCAATGCAGAAG...
...ACCAAGTCTGCCACTTACTGGCCTGAGTACACTGGCAATGCAGAAG...

...ACCAAGTCTGCCACTTACTGGCCCGAGTACACTGGCAATGCAGAAG...
...ACCAAGTCTGCCACTTACTGGCCCGAGTACACTGGCAATGCAGAAG...

Genotype

Variation in action

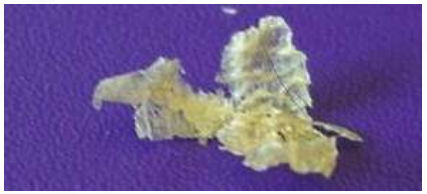
Heterozygous

T
C

wet earwax

Homozygous

T
T



dry earwax

C
C

wet earwax

Genotype

Phenotype ← just one of many phenotypes that may be associated with this genotype

From genotype to phenotype

DNA

↓ Transcription

RNA

↓ Translation

protein

↓ Folding

three-dimensional structure

↓

biological function

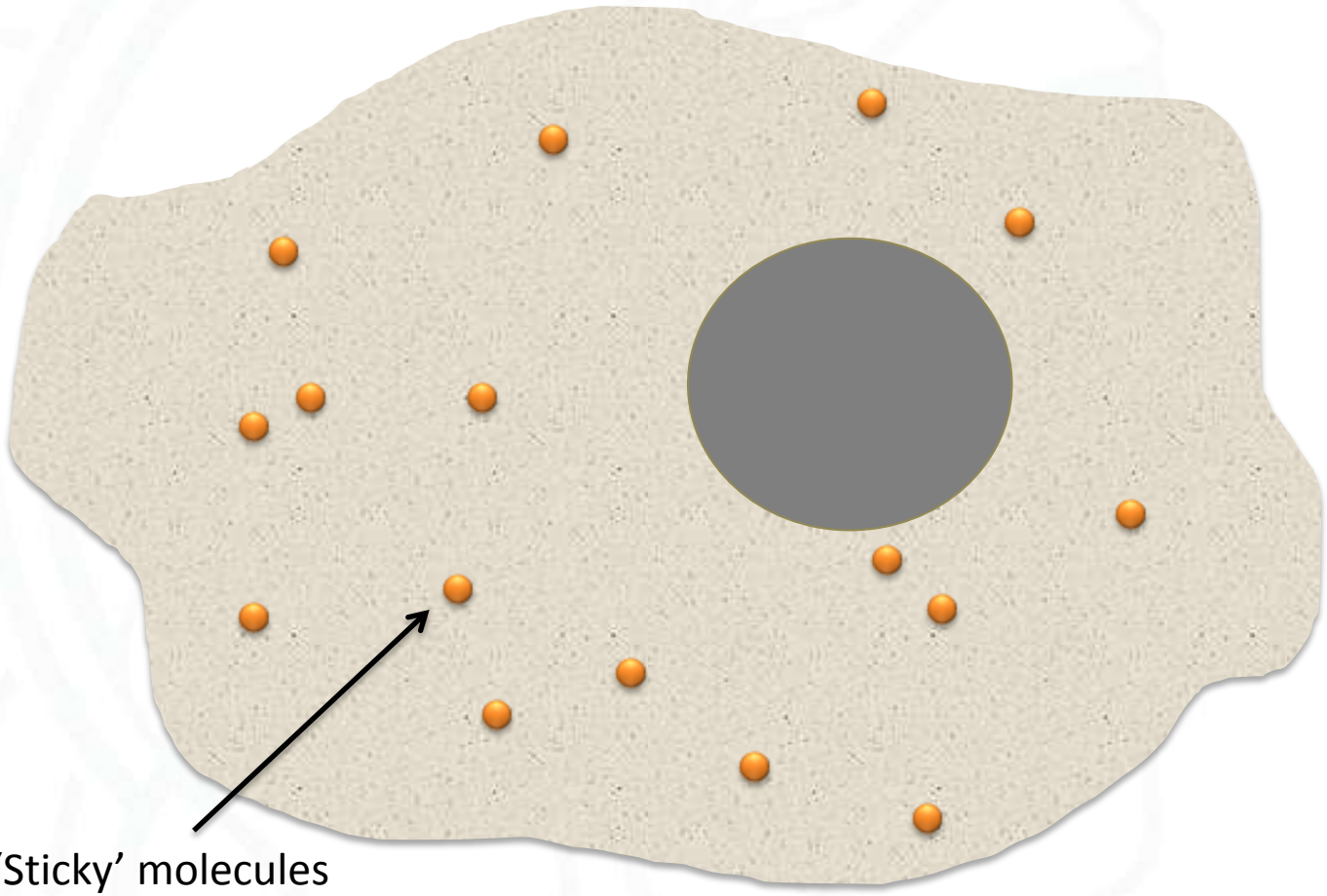
↓

integration of function into cell/organism

↓

observed phenotypes

From genotype to phenotype

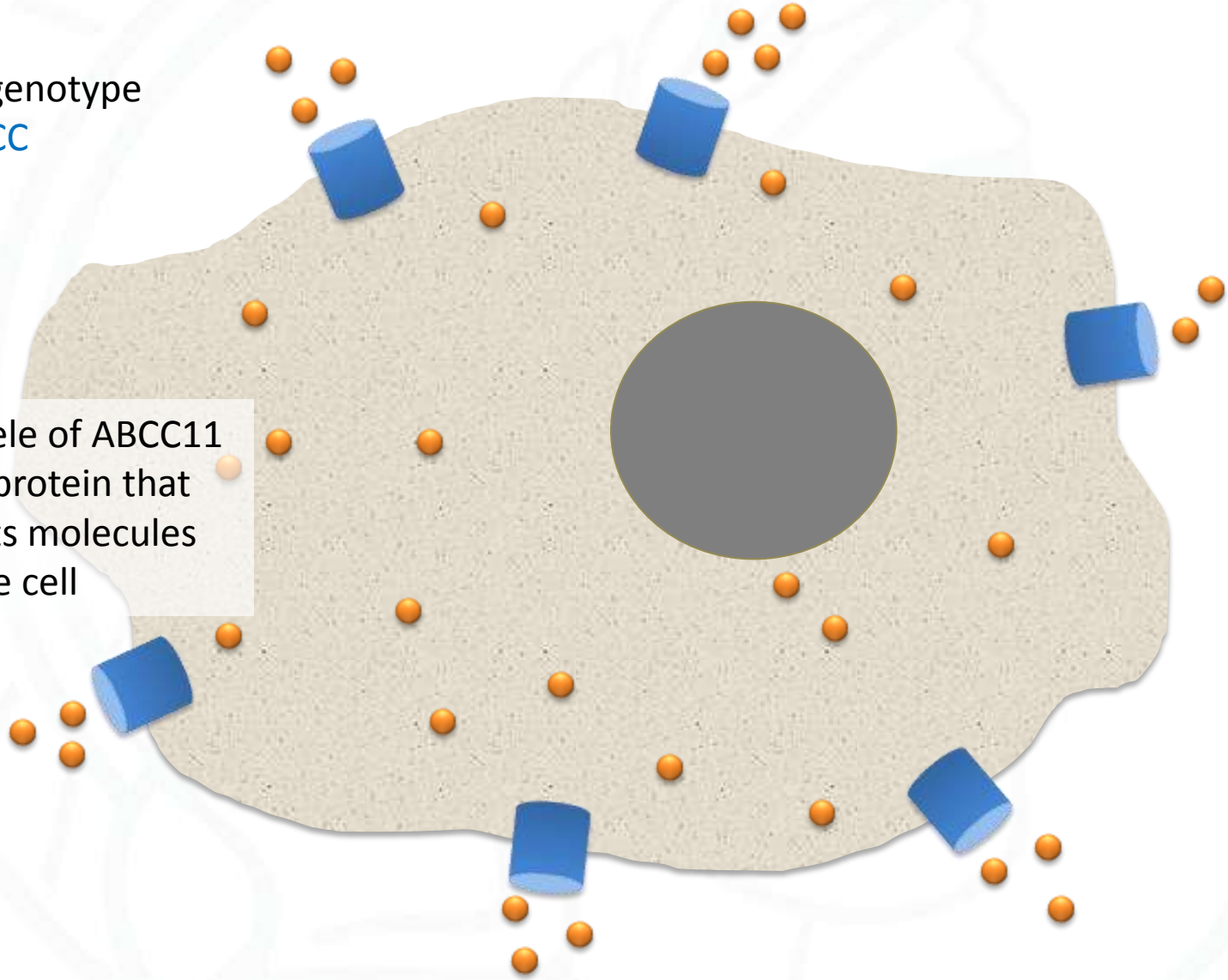


'Sticky' molecules are produced inside the cell

From genotype to phenotype

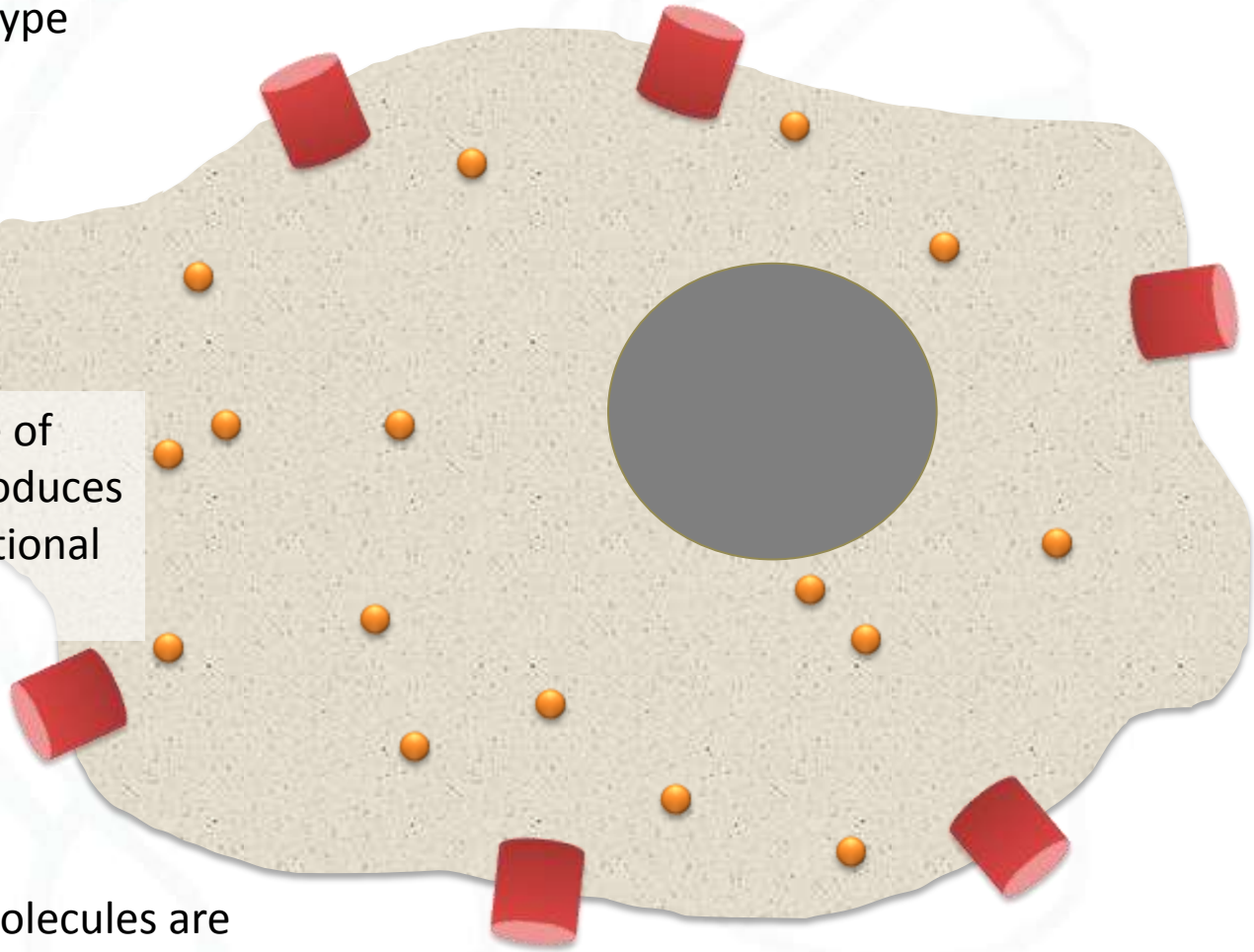
genotype
CC

The C allele of ABCC11 makes a protein that transports molecules out of the cell



From genotype to phenotype

genotype
TT

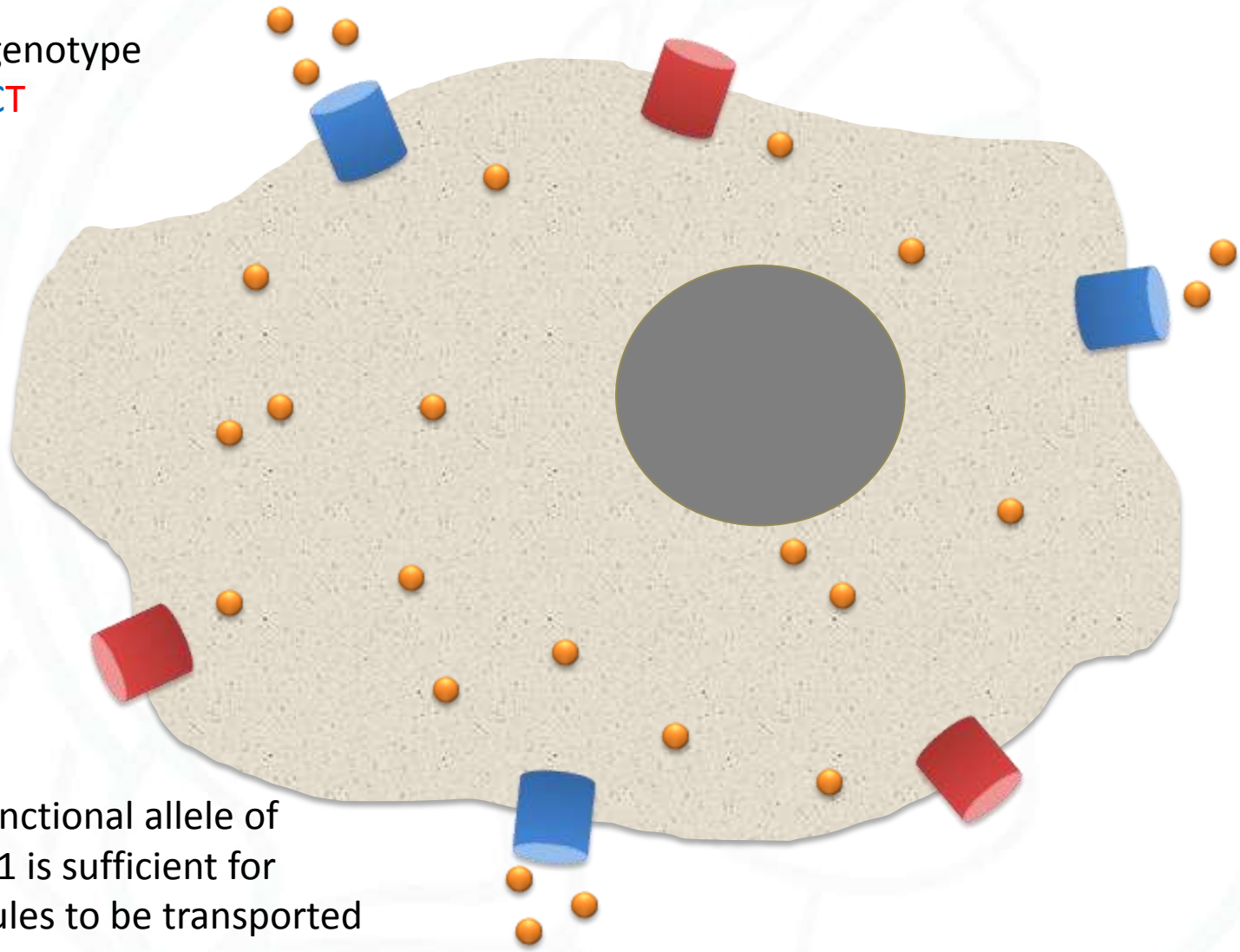


The T allele of ABCC11 produces a non-functional protein

No ‘sticky’ molecules are transported out of the cell

From genotype to phenotype

genotype
CT



One functional allele of ABCC11 is sufficient for molecules to be transported

From genotype to phenotype

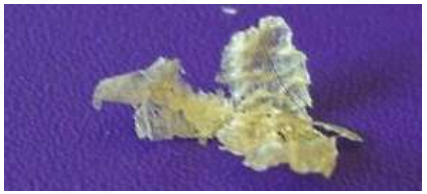
Heterozygous

T
C

wet earwax

Homozygous

T
T



dry earwax

C
C

wet earwax

Genotype

Phenotype

Dominant and recessive

T
C

- 1 copy of the C allele is sufficient to produce the 'wet' phenotype.
- This is a **dominant** allele/phenotype.

T
T

- 2 copies of the T allele are required to give the 'dry' phenotype.
- This is a **recessive** allele/phenotype.

Dominant and recessive

- Most genetic disorders are **recessive**:
  - one functional copy of a gene provides sufficient protein.
  - loss-of-function mutation
- But disorders are **dominant** when:
  - half is not enough.
  - loss-of-function mutation
  - the alternative allele makes a toxic version of the protein.
  - gain-of-function mutation

Variation in action

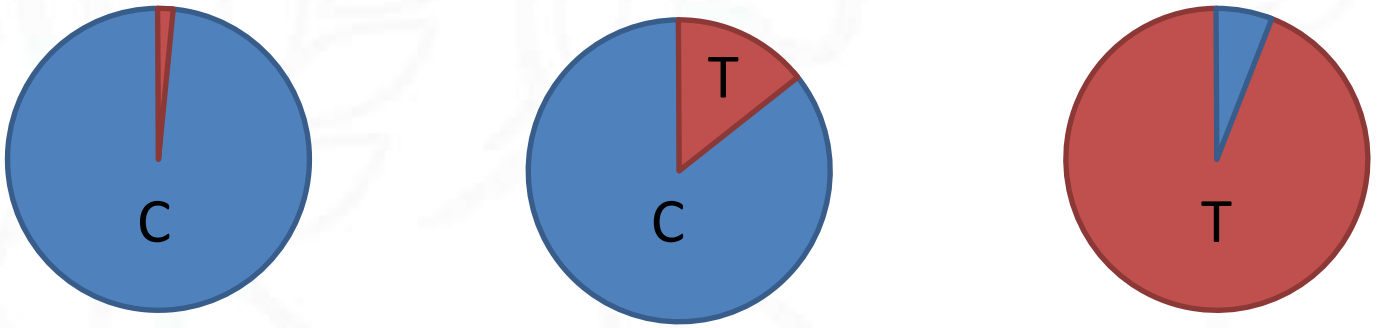
In ABCC11:

The C allele (functional protein) is the **ancestral** allele.

The T allele (non-functional protein) first appeared about 2000 generations ago, as a random new mutation in one person.

Allele frequencies in modern populations

SNP rs17922931



African | European | Chinese, Japanese

Monogenic traits

- Earwax type is a **monogenic trait** – a characteristic determined by variation in only a **single gene**.
- Monogenic traits are **categorical** – eg. ‘wet’ or ‘dry’.

Monogenic traits

- **Monogenic traits**
  - have a straightforward **relationship between genotype and phenotype**.

CC

CT

TT

Monogenic traits

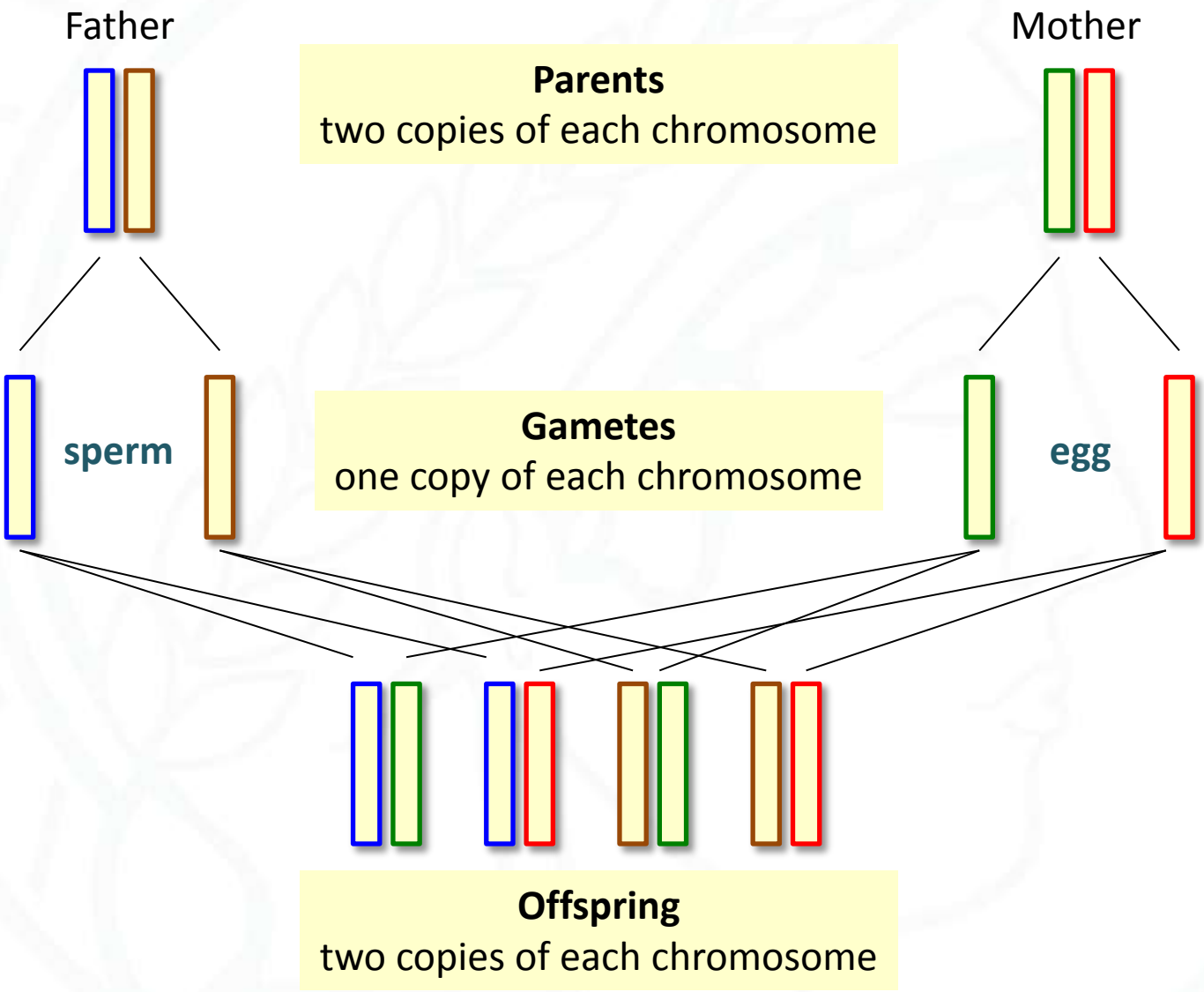
- **Monogenic traits**
  - have a straightforward **relationship between genotype and phenotype**.
  - have straighforward **patterns of inheritance**.

CC

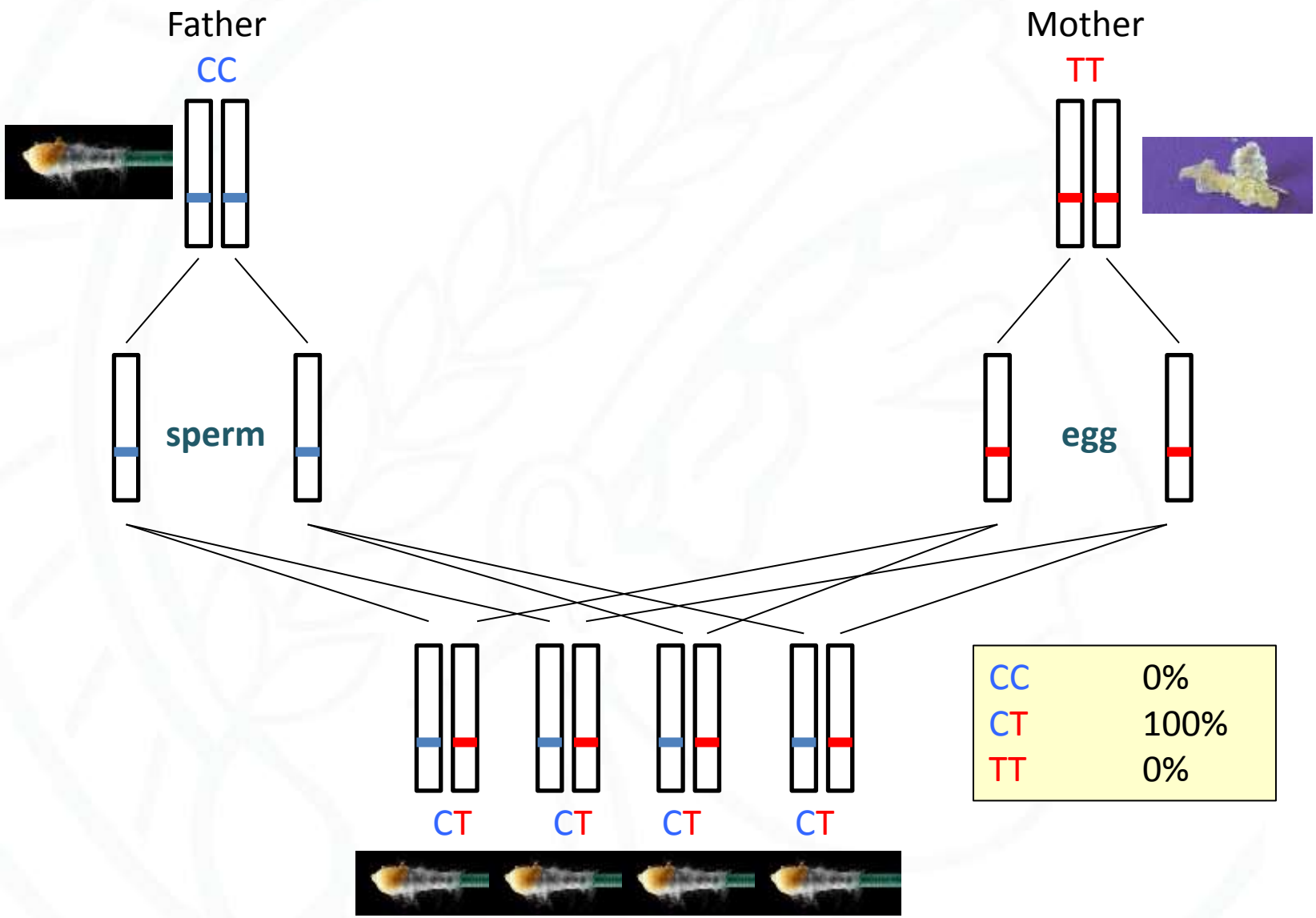
CT

TT

Inheritance

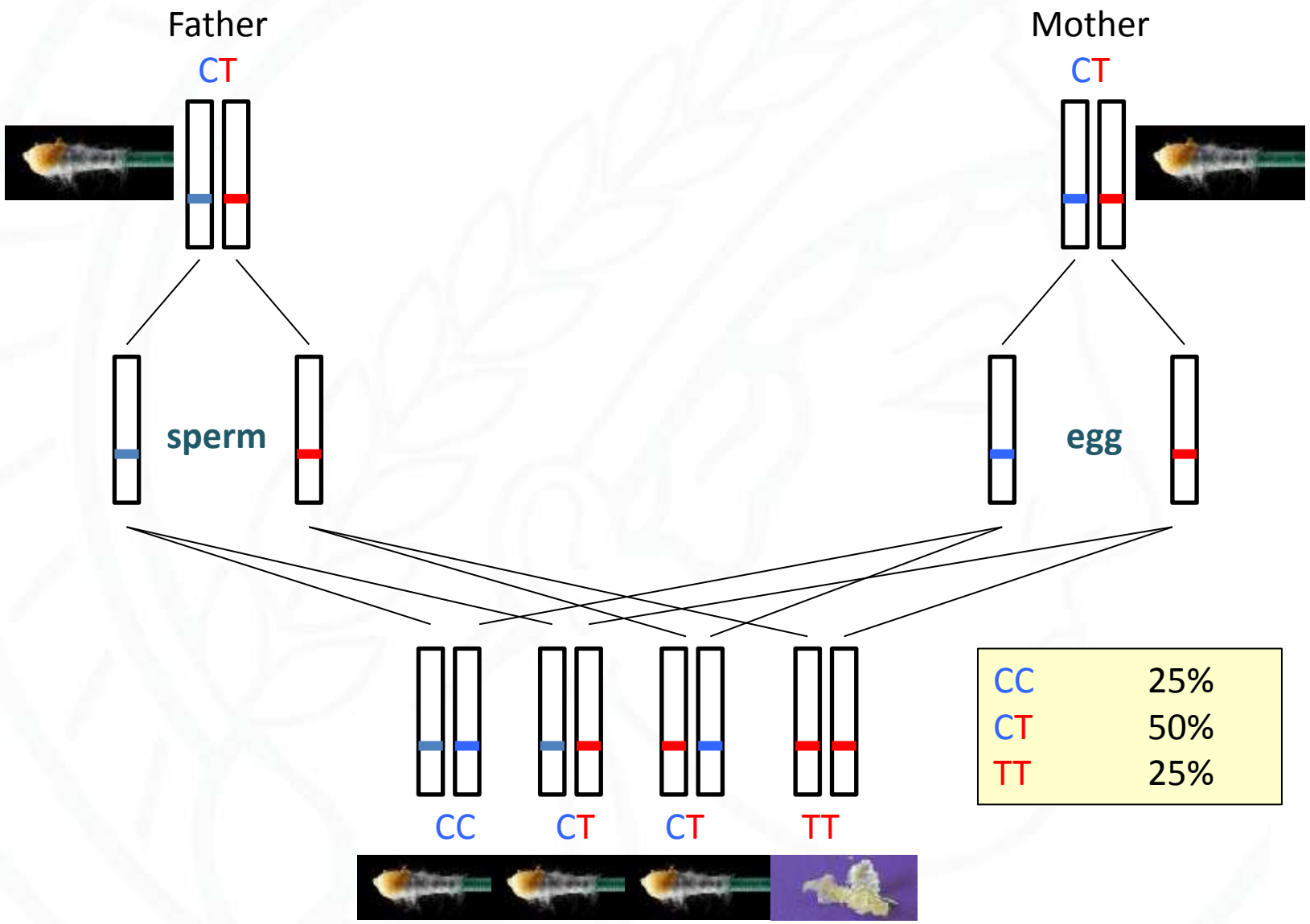


Inheritance of monogenic traits



CC	0%
CT	100%
TT	0%

Inheritance of monogenic traits



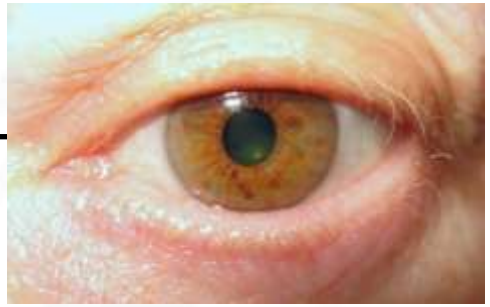
CC	25%
CT	50%
TT	25%

Monogenic vs polygenic traits

Trait type	Example	Number of genes	Phenotype variation	Genotype-to-phenotype	Inheritance pattern
Monogenic	Earwax	1	Categorical	Simple	Simple
Polygenic	Height	2 - many	Continuous	Complex	Complex

Polygenic traits

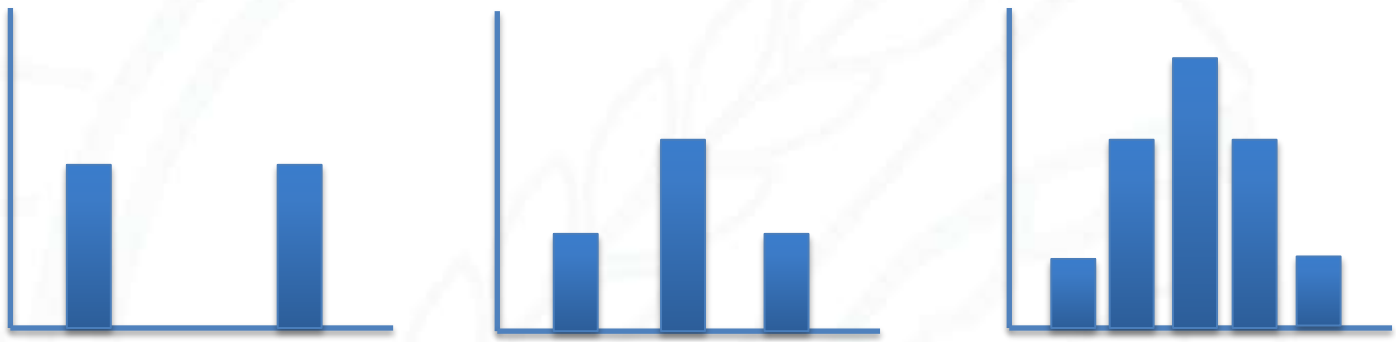
Father



Mother

Child

Polygenic traits



Categorical
1 gene

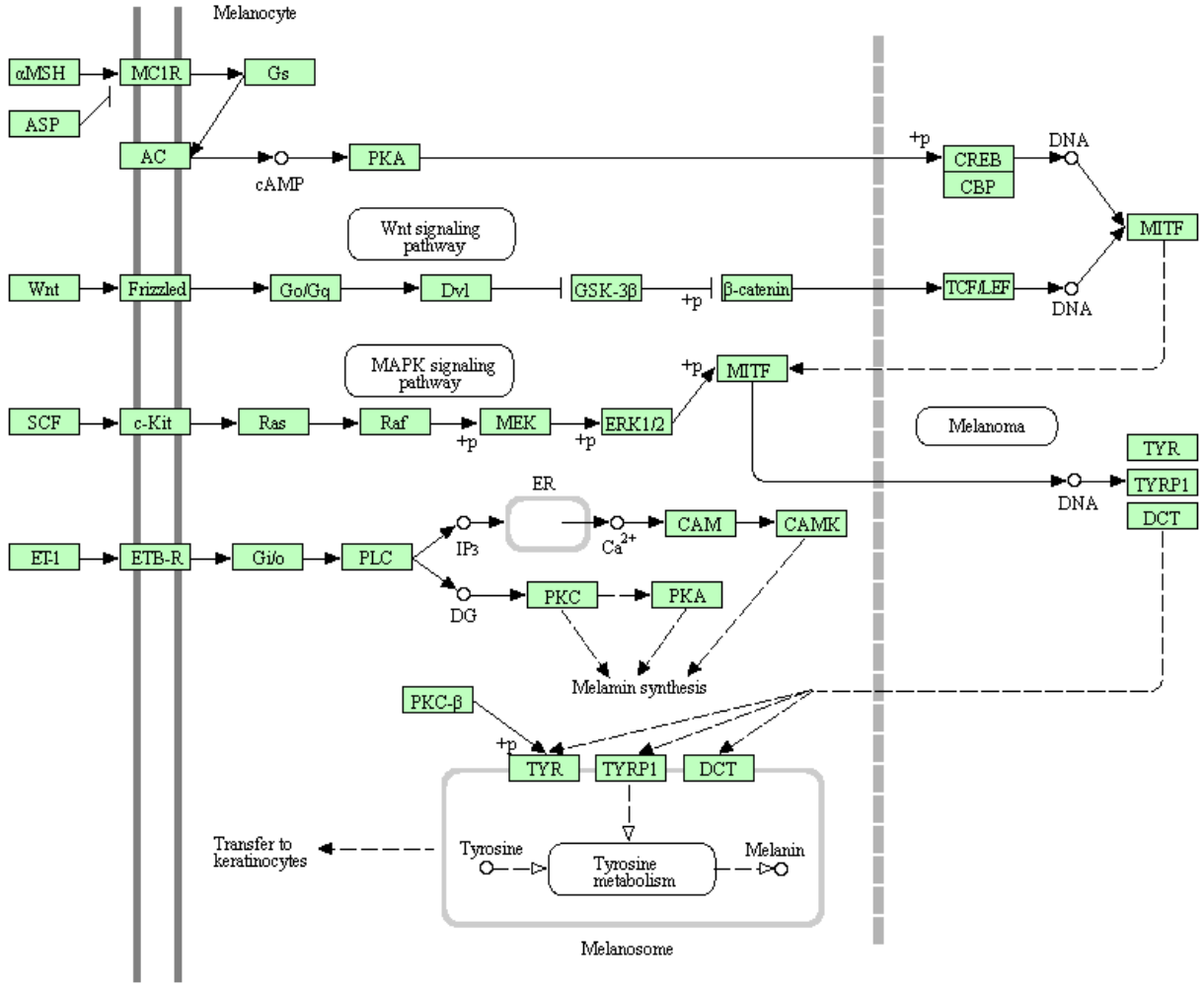
Continuous
many genes

- Only takes a few genes to produce a **normal distribution**, eg height, IQ.
- Noise from **random effects** smoothes the distribution.
- **Environmental factors** can shift the whole distribution up or down.

Polygenic traits

Proteins in hair/eye/skin pigmentation

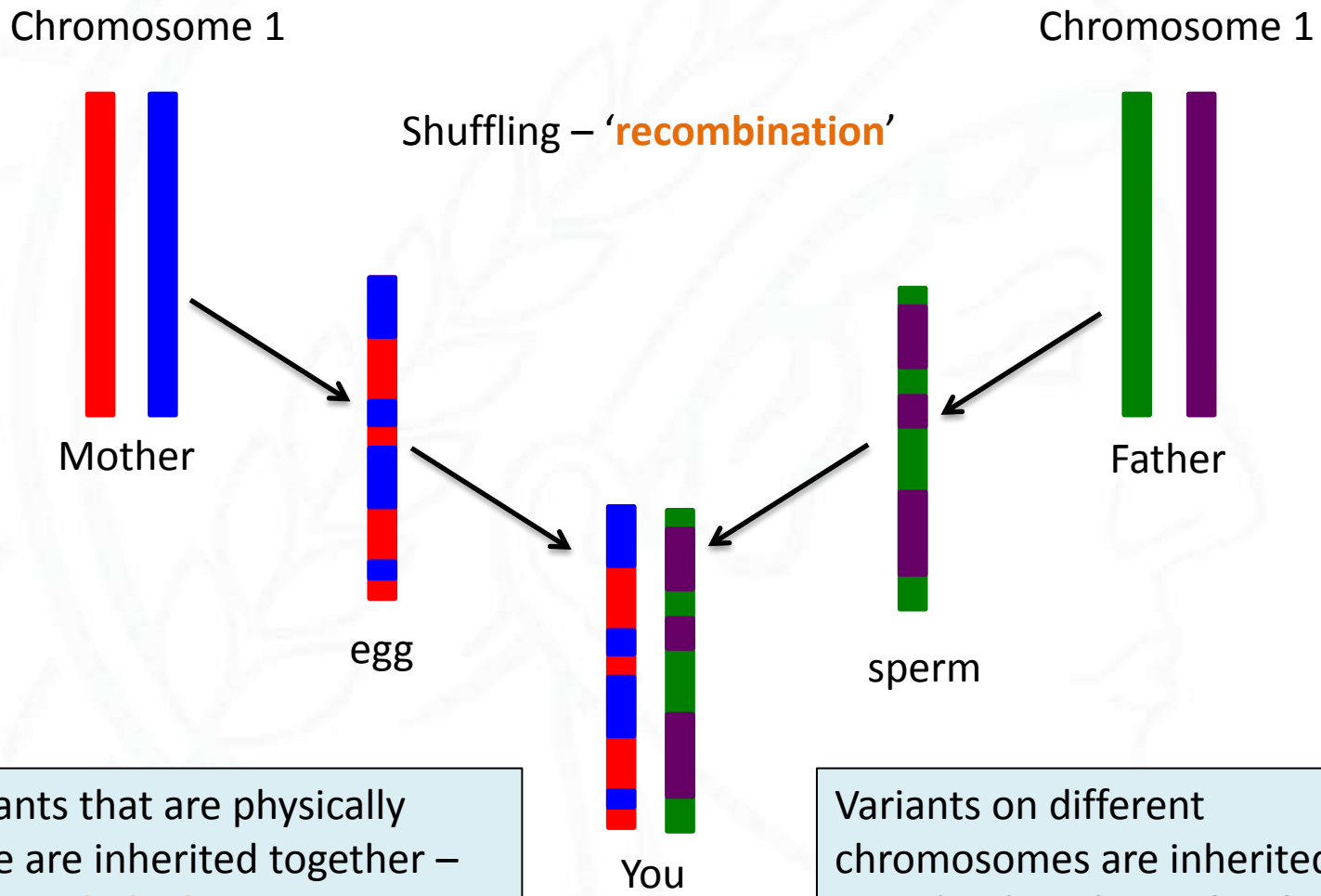
red hair



albinism

LANGUAGE AND GENETICS DEPARTMENT

The inheritance of complex traits

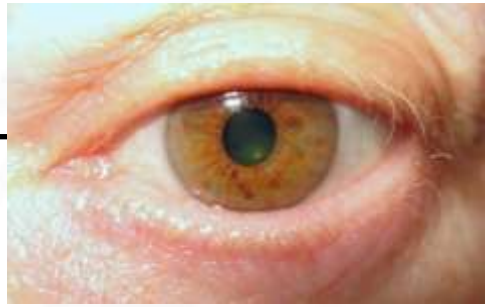


Variants that are physically close are inherited together – they are **linked**.

Variants on different chromosomes are inherited completely **independently**.

The inheritance of complex traits

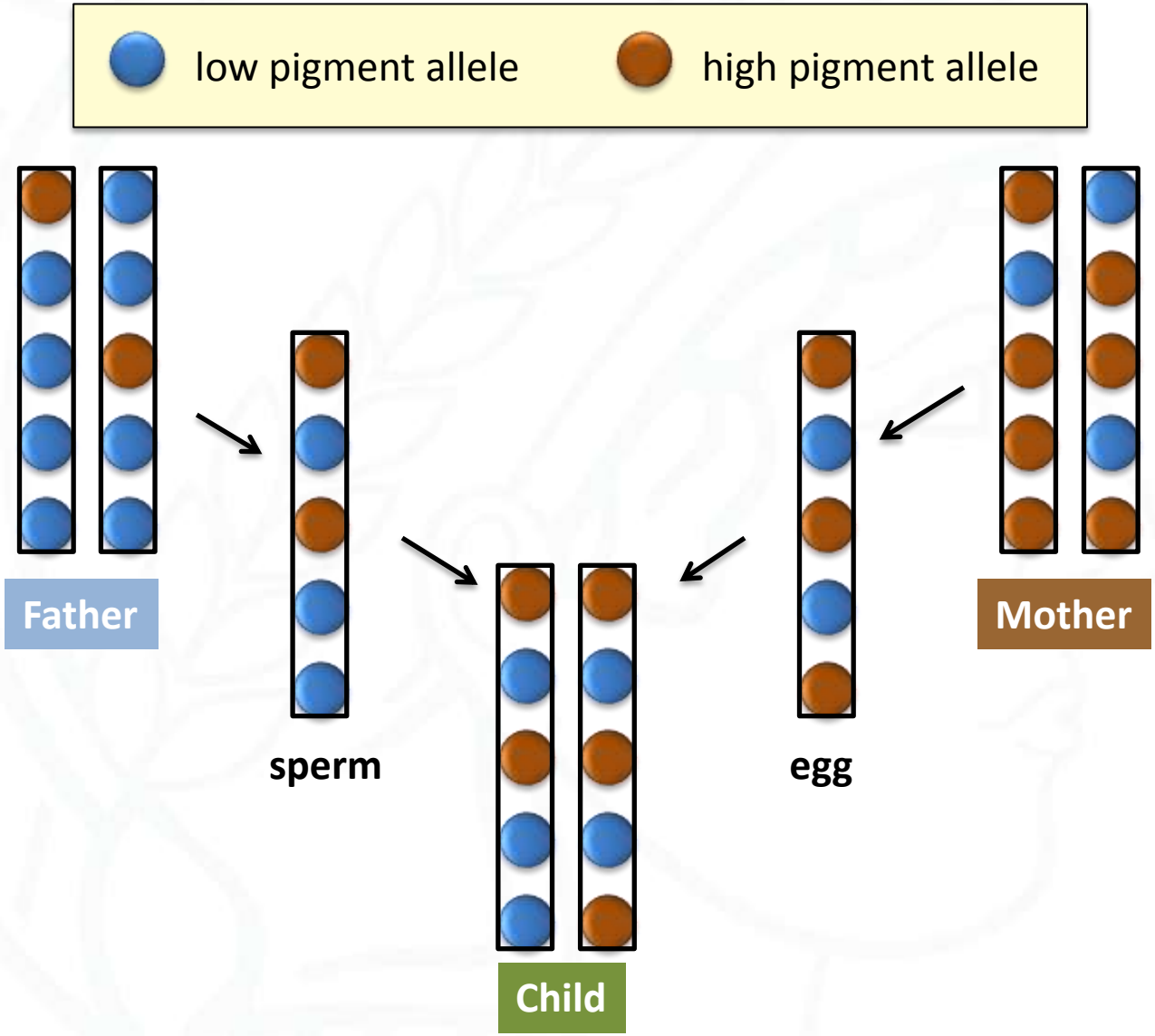
Father



Mother

Child

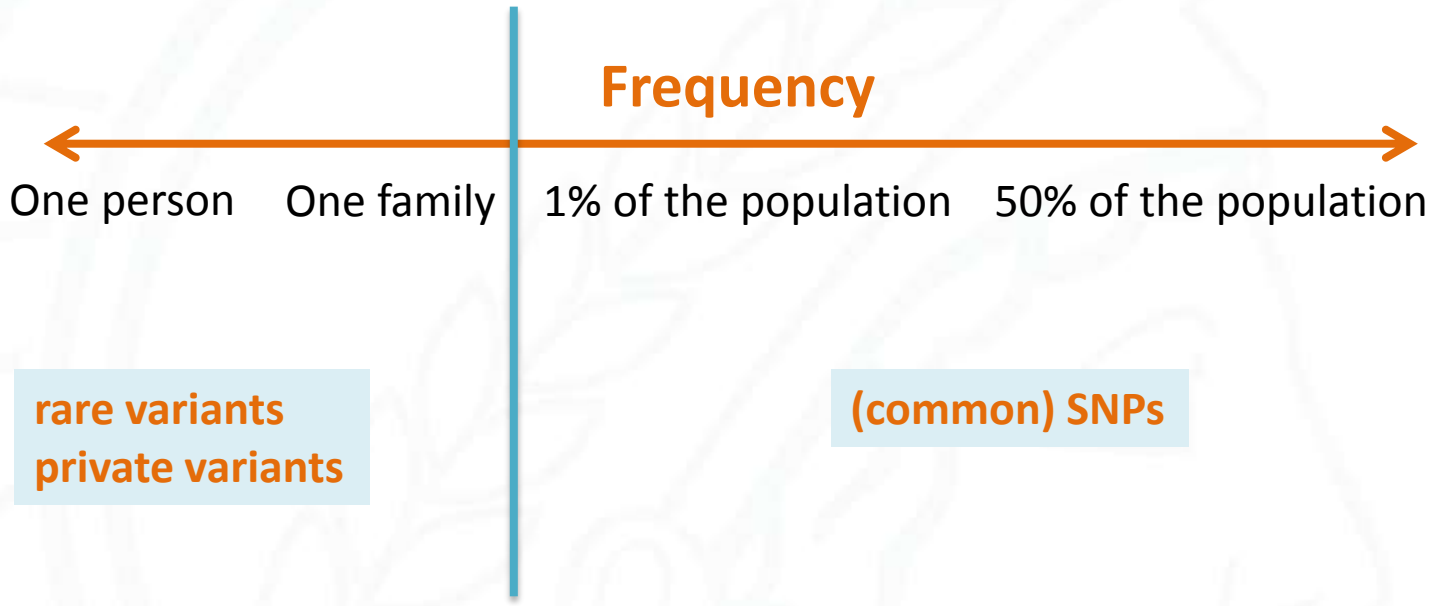
The inheritance of complex traits



SNP summary

- Single Nucleotide Polymorphisms.
- Common variation **- minor allele frequency** >1%.
- Explain **normal variation**.
- Can affect **protein structure** or **regulation of expression**.

Variation… varies



- **De novo** changes occur in every individual.
- They can alter **protein sequence**, alter **gene regulation**, or have **no effect**.
- On average, every individual has 1 new protein-inactivating change.
- De novo variation means that traits can be genetic **without** being inherited.

Variation… varies

LANGUAGE AND GENETICS DEPARTMENT

Frequency

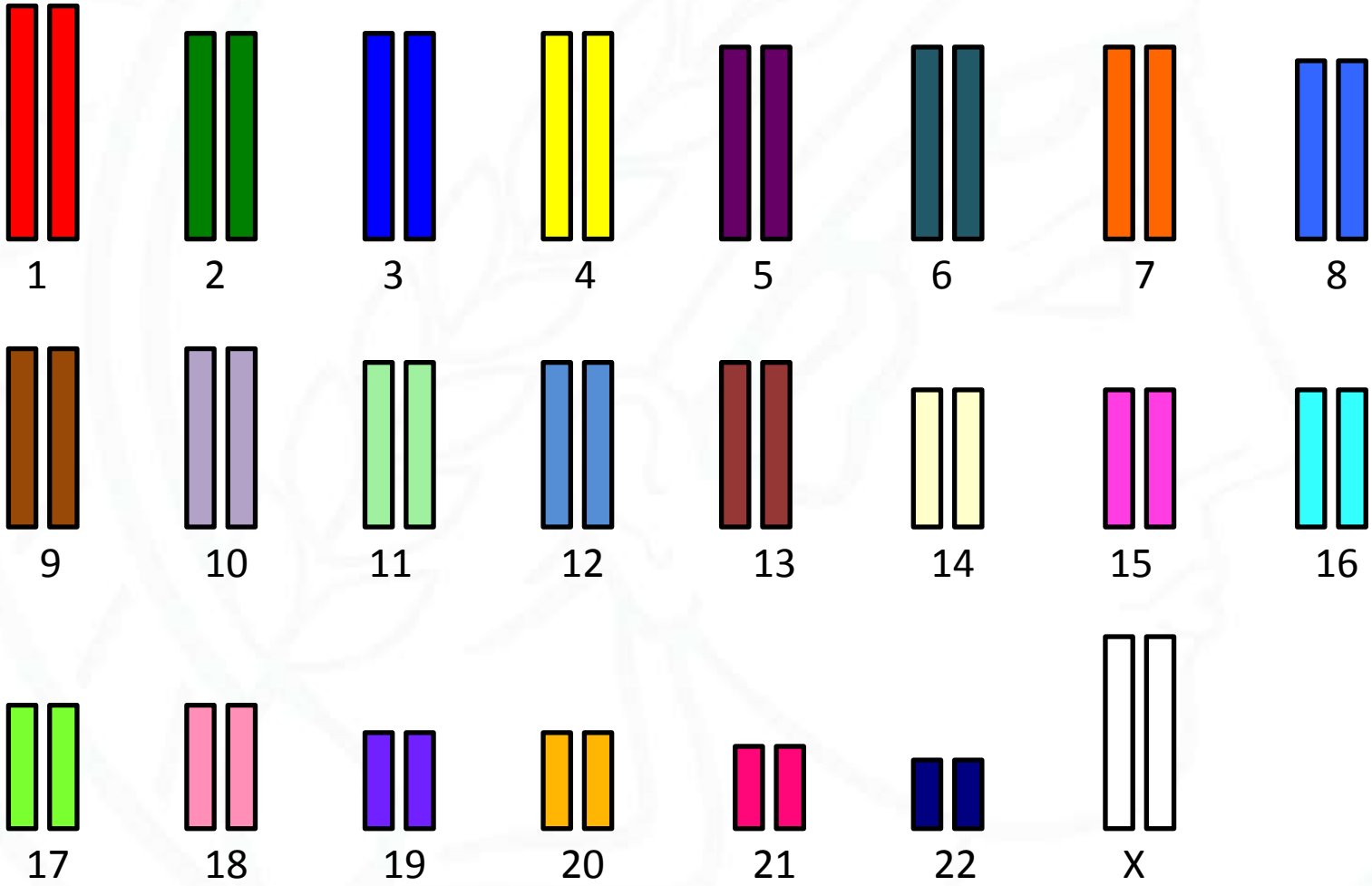
One person | One family | 1% of the population | 50% of the population

Size

1 nucleotide | A few nucleotides | 100s of nucleotides | 1000s of nucleotides | Whole chromosome

Variation… varies

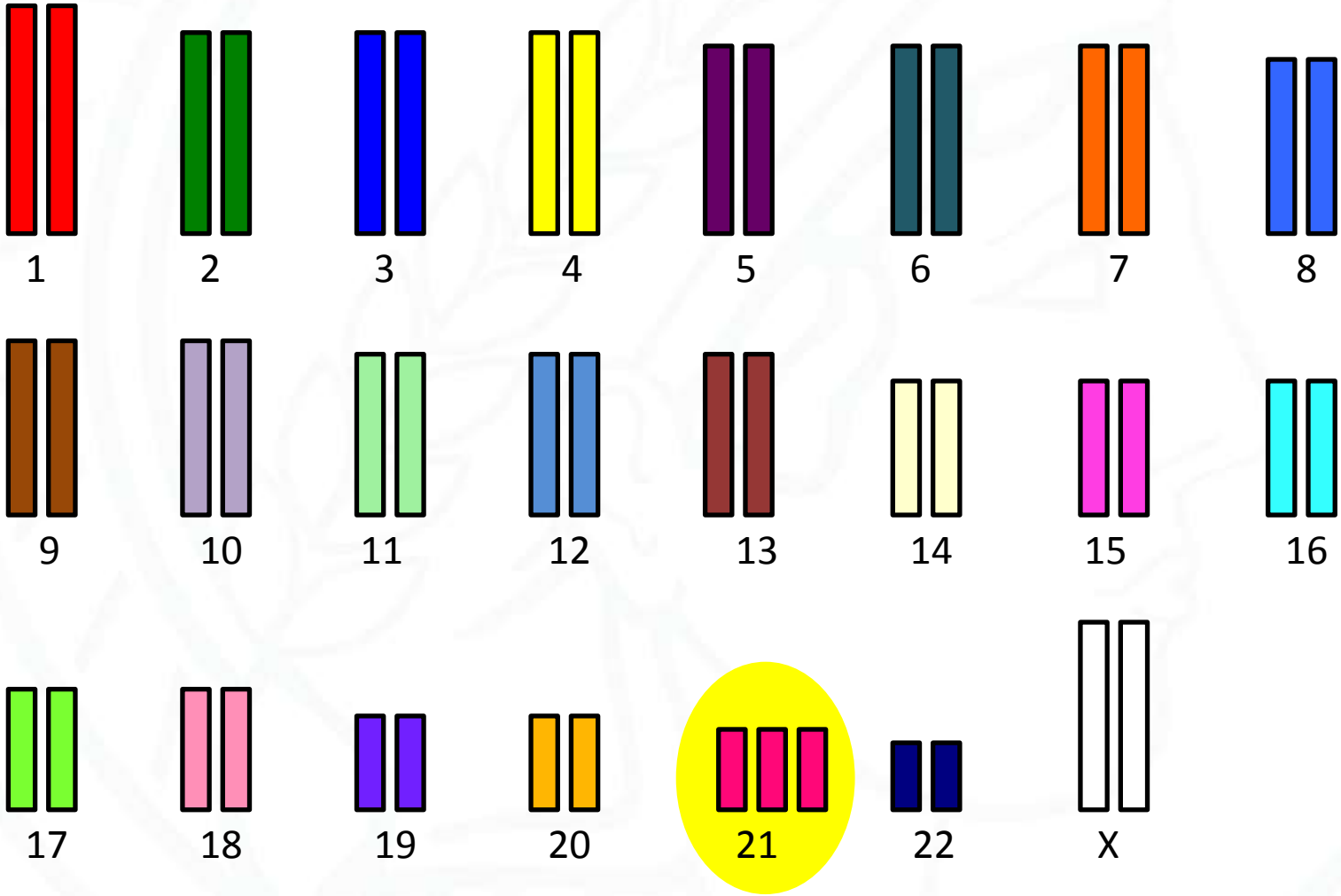
Normal human **karyotype**



Variation… varies

LANGUAGE AND GENETICS DEPARTMENT

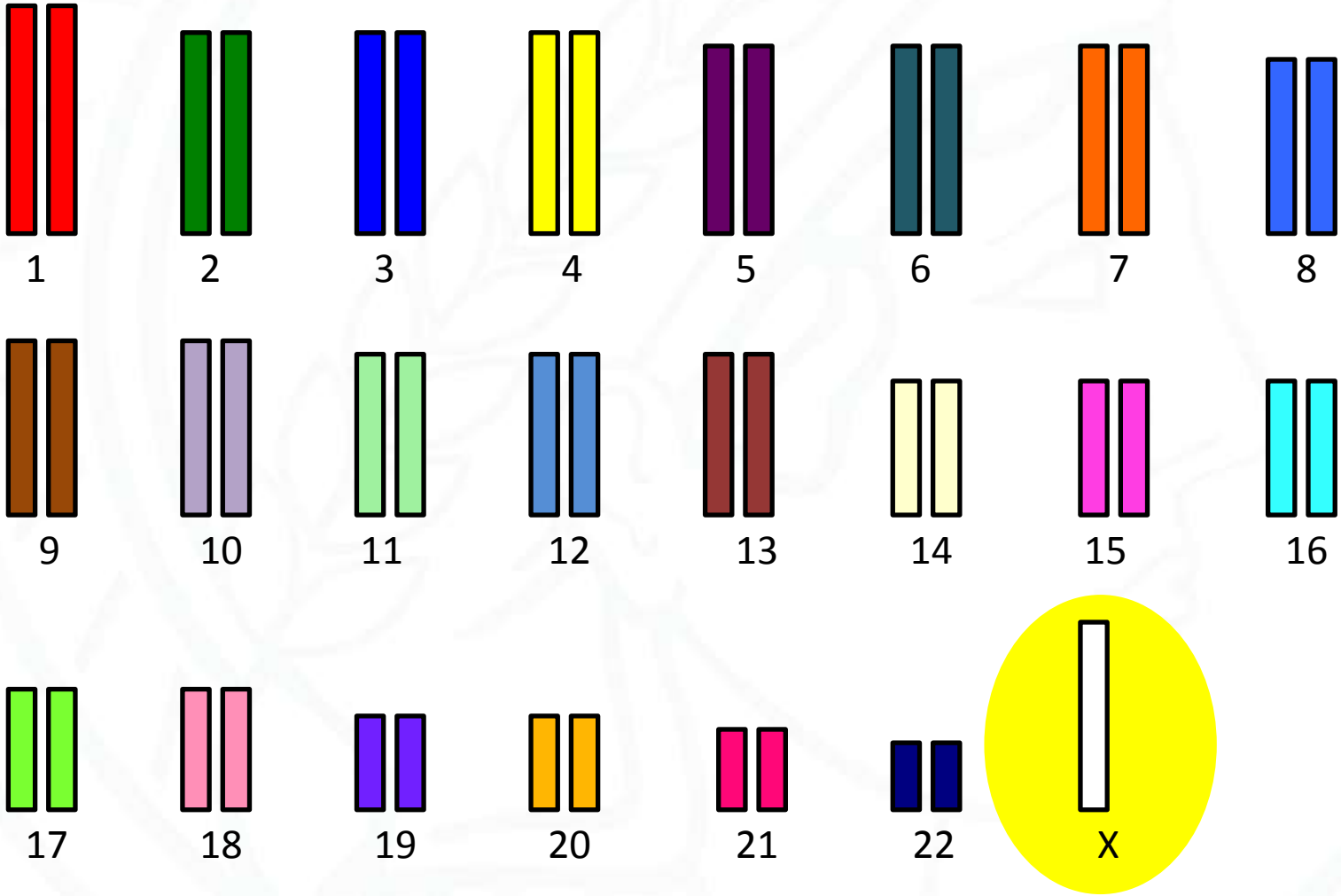
Down syndrome



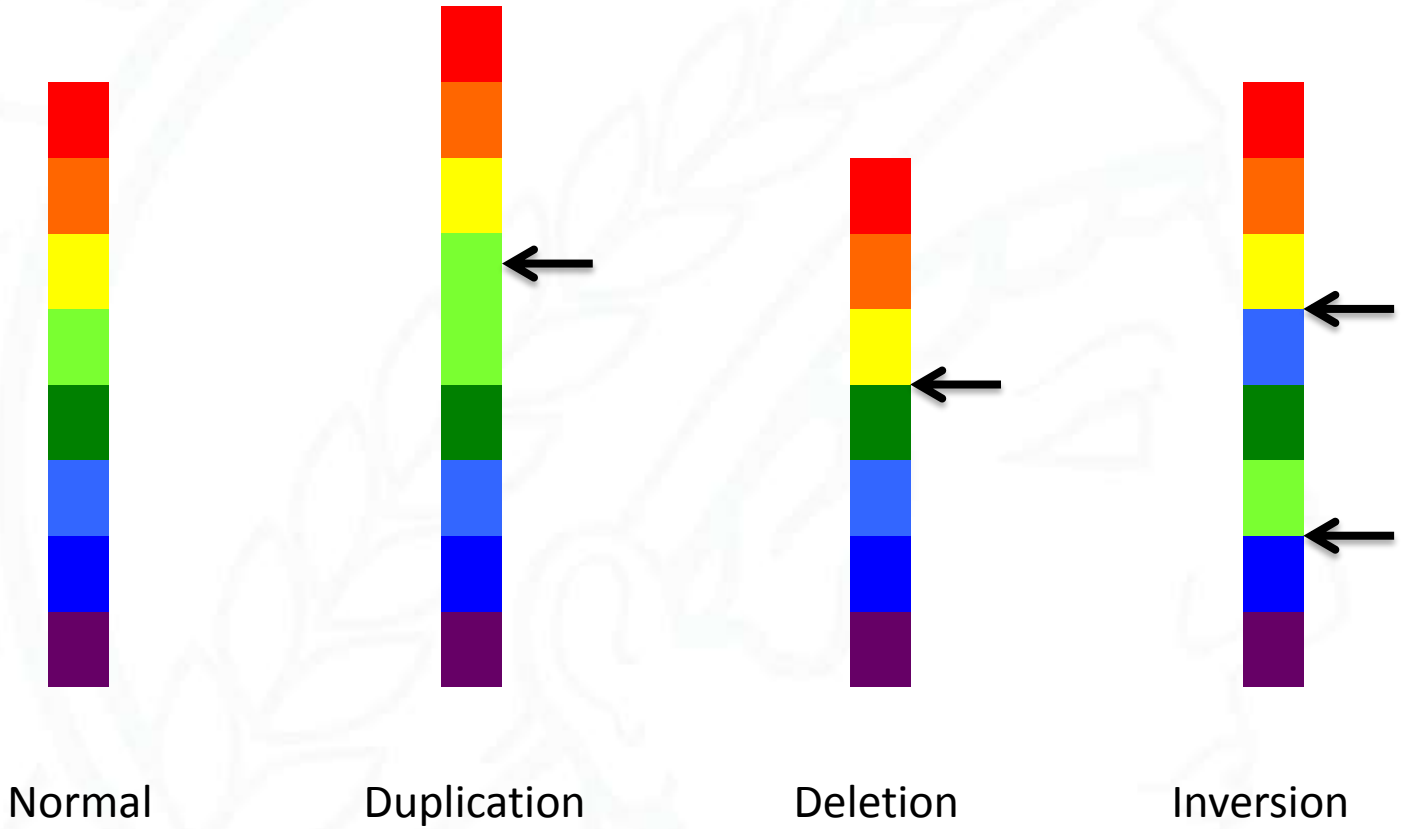
Variation… varies

LANGUAGE AND GENETICS DEPARTMENT

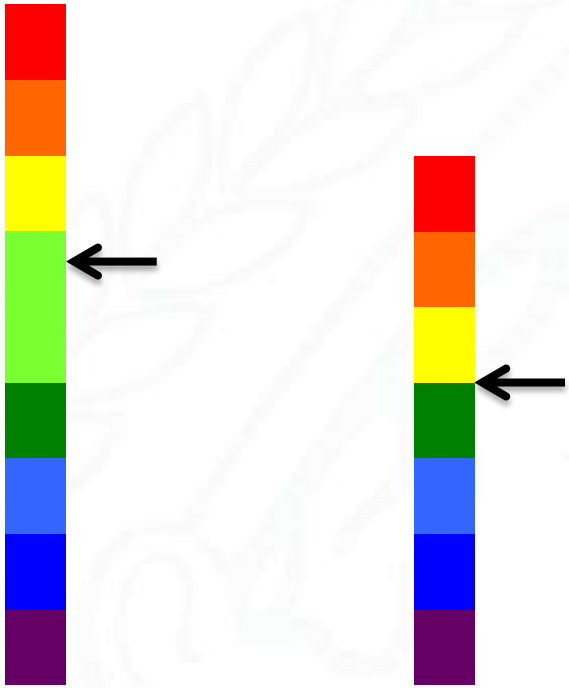
Turner syndrome



Variation… varies



Variation… varies



Duplication Deletion

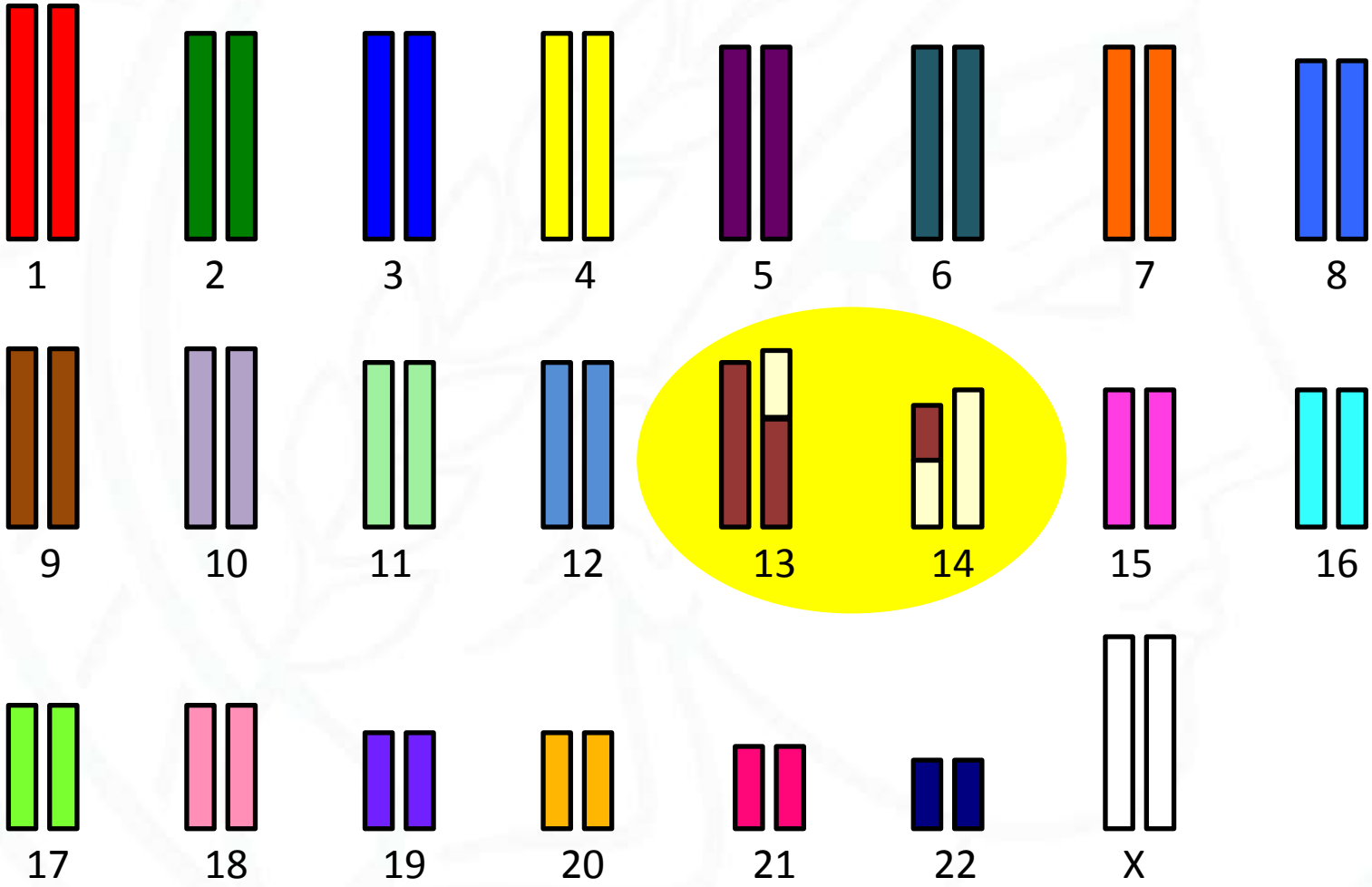
Copy number variations – CNVs

1000s of nucleotides duplicated or deleted, can include 1 or more genes.
Affects **>5% of the genome** of normal healthy individuals.

Variation… varies

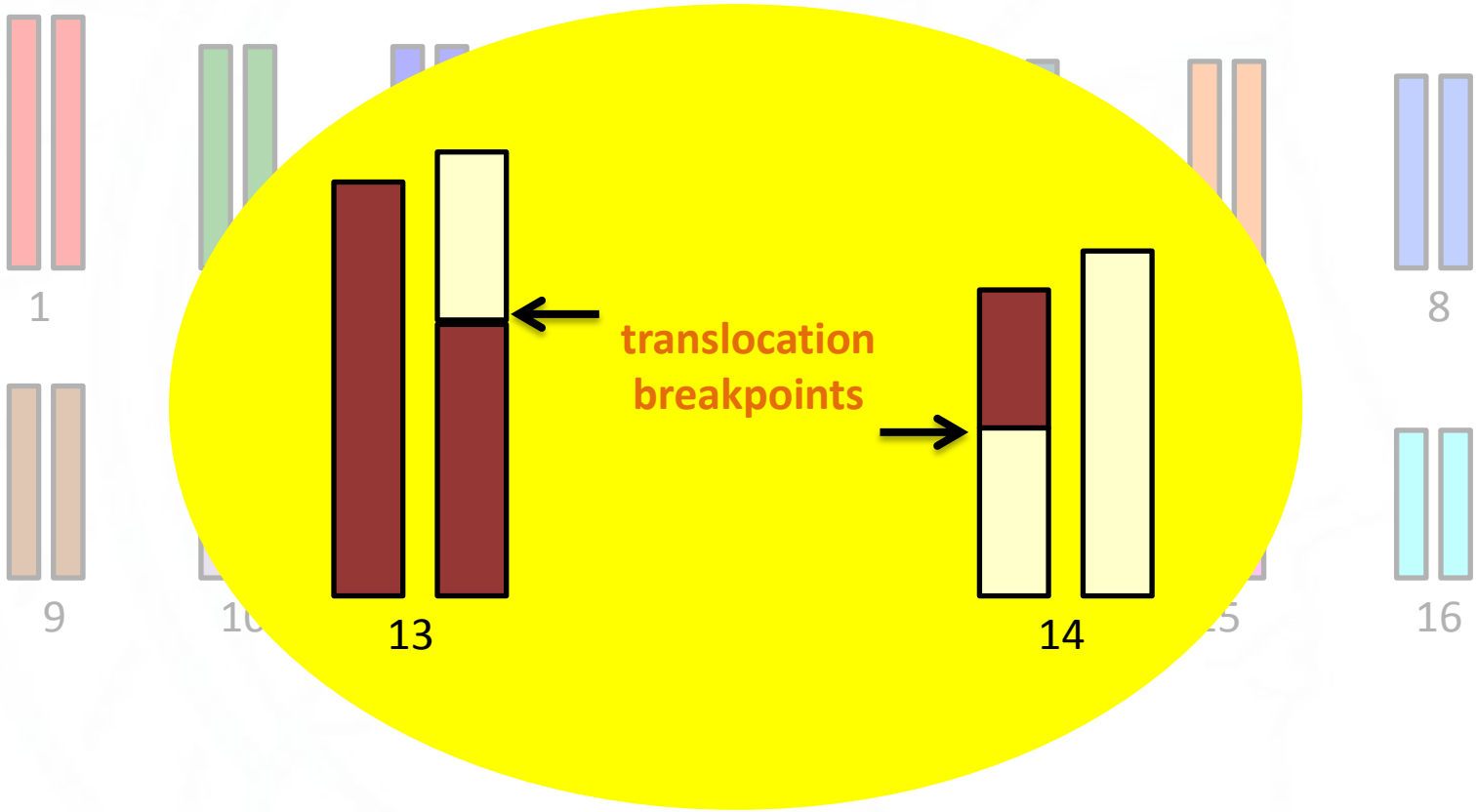
LANGUAGE AND GENETICS DEPARTMENT

A **balanced translocation**



Variation… varies

A **balanced translocation**

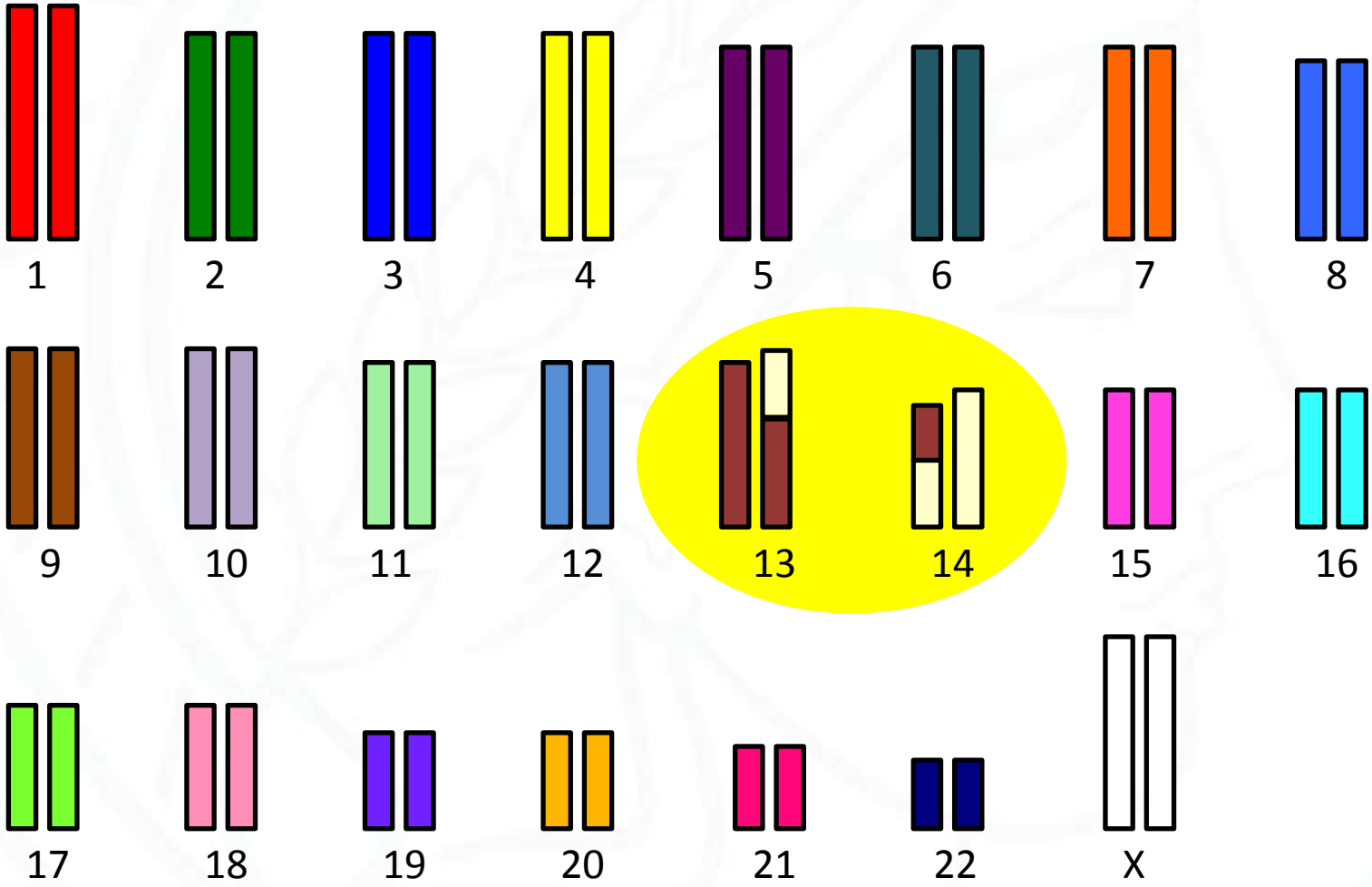


- Usually **harmless** – breakpoints in non-essential regions of the genome.
- But damaging if the breakpoint occurs in a gene.

Variation… varies

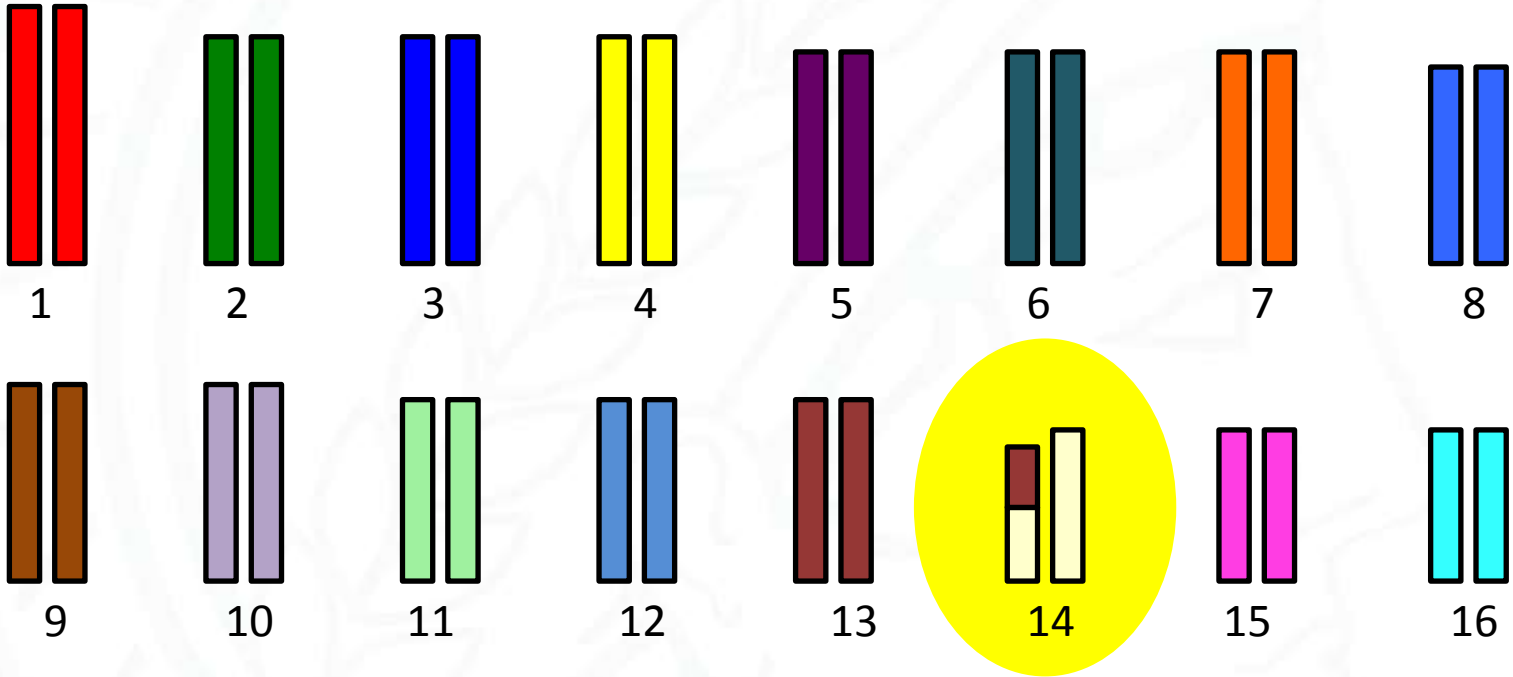
LANGUAGE AND GENETICS DEPARTMENT

Parent with a balanced translocation



Variation… varies

Child with inherited translocation



- **Probably damaging** – too many copies of chromosome 13 and too few of chromosome 14, many genes affected.

Variation… varies

The phenotype resulting from a genetic variant depends on:

- other genetic variants present
- environmental influences
- random processes

Complex phenotypes are the products of complex interactions.

Gene-gene interactions

Gene-environment interactions

Gene-randomness interactions

Summary

DNA

↓ Transcription ← **Key checkpoint**

RNA

DNA variation

↓ Translation

protein

↓ Folding

DNA variation three-dimensional structure

↓

biological function

↓

integration of function into cell/organism

↓

observed phenotypes