



THE UNIVERSITY
of EDINBURGH



Biotechnology and
Biological Sciences
Research Council



THE ROYAL
SOCIETY

DNA variation, Phenotypes, & Lottery

Gregor Gorjanc

Athens, Greece
2025-01-30

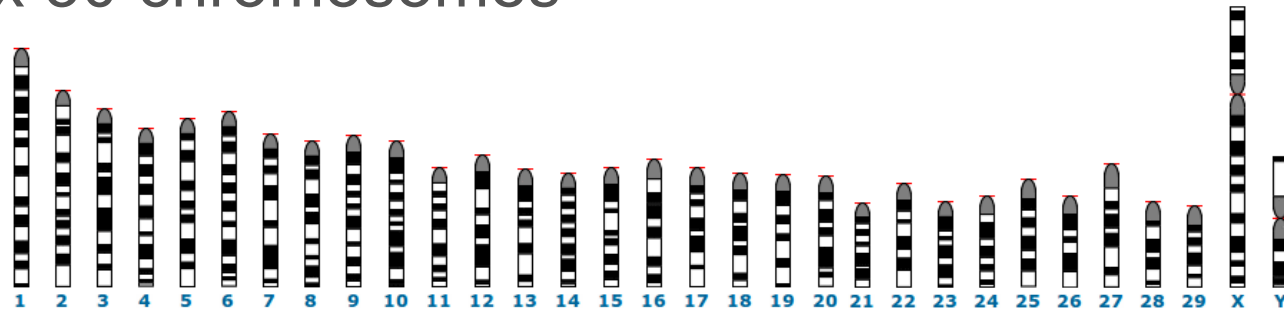


Learning objectives

- Encoding DNA variation
- Simulate DNA & phenotypes in AlphaSimR
- Simulate inheritance in AlphaSimR

Genome (cattle example)

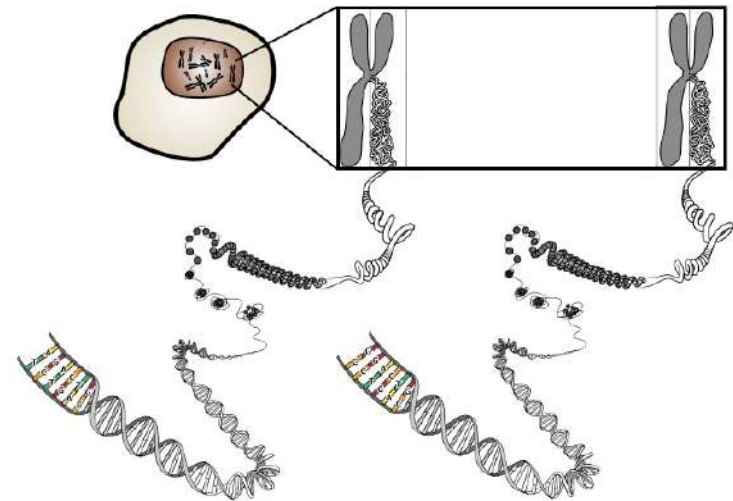
- 2 x 30 chromosomes



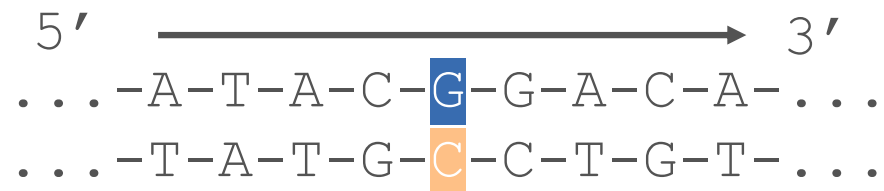
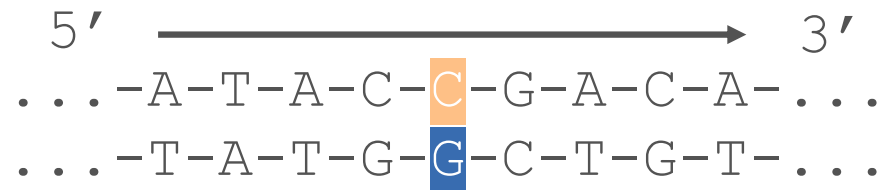
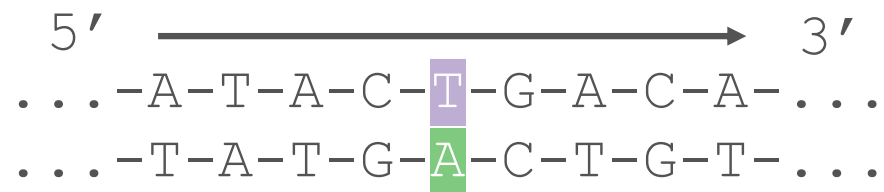
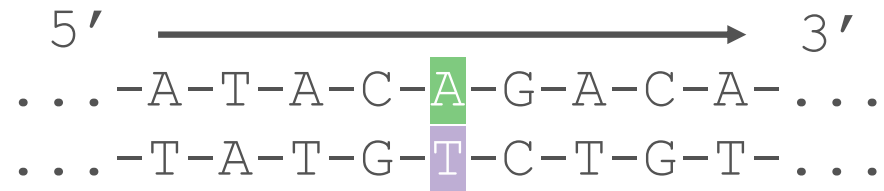
- DNA, 2 x 3 billion (10^9) base pairs

Adenine
Thymine
A
T

Cytosine
Guanine
C
G



Single Nucleotide Polymorphism



How many SNPs and other variants?

The sequences of 150,119 genomes in the UK biobank

www.biorxiv.org/content/10.1101/2021.11.16.468246v2

“... This constitutes a set of high quality variants, including 585,040,410 SNPs, representing 7.0% of all possible human SNPs, and 58,707,036 indels.”

→ ~600M SNPs

→ ~60M indels

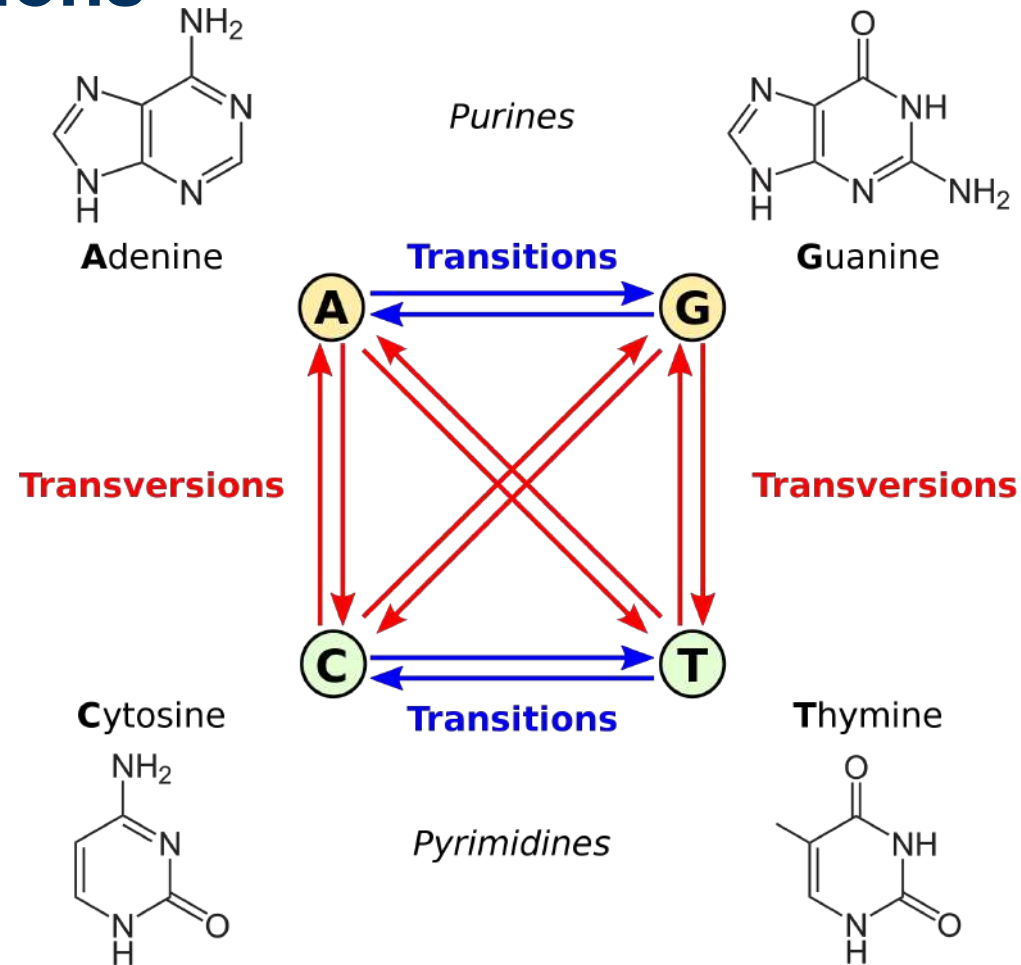
“... We identified 895,055 structural variants and 2,536,688 microsatellites, groups of variants typically excluded from large scale WGS studies.

→ ~1M structural variants!!!

→ ~3M microsatellites!!!

MEGA-SCALE DATA!!!

SNP mutations



<https://en.wikipedia.org/wiki/Transversion>

Bi-allelic SNP alleles, genotypes, & dosages

5' $\xrightarrow{\hspace{2cm}}$ 3'
...-A-T-A-C-**A**-G-A-C-A-...
...-T-A-T-G-**T**-C-T-G-T-... Ref. allele --> 0 --> 0

5' $\xrightarrow{\hspace{2cm}}$ 3' Ref. allele --> 0
...-A-T-A-C-**A**-G-A-C-A-...
...-T-A-T-G-**T**-C-T-G-T-...

Bi-allelic SNP alleles, genotypes, & dosages

5' → 3' ...-A-T-A-C-A-G-A-C-A-... ...-T-A-T-G-T-C-T-G-T-...	Ref. allele --> 0	
5' → 3' ...-A-T-A-C-A-G-A-C-A-... ...-T-A-T-G-T-C-T-G-T-...	Ref. allele --> 0	--> 0
5' → 3' ...-A-T-A-C-A-G-A-C-A-... ...-T-A-T-G-T-C-T-G-T-...	Ref. allele --> 0	
5' → 3' ...-A-T-A-C-G-G-A-C-A-... ...-T-A-T-G-C-C-T-G-T-...	Alt. allele --> 1	--> 1
5' → 3' ...-A-T-A-C-G-G-A-C-A-... ...-T-A-T-G-C-C-T-G-T-...	Alt. allele --> 1	
5' → 3' ...-A-T-A-C-G-G-A-C-A-... ...-T-A-T-G-C-C-T-G-T-...	Alt. allele --> 1	--> 2

Genome-wide haplotypes & genotype

Haplotype 1	0	1	1	0	0	1
Haplotype 2	1	1	1	1	0	0
Genotype	1	2	2	1	0	1

Take home message no. 1

Encoding haplotypes as a series of 0 & 1

Encoding genotypes as a series of 0, 1, & 2

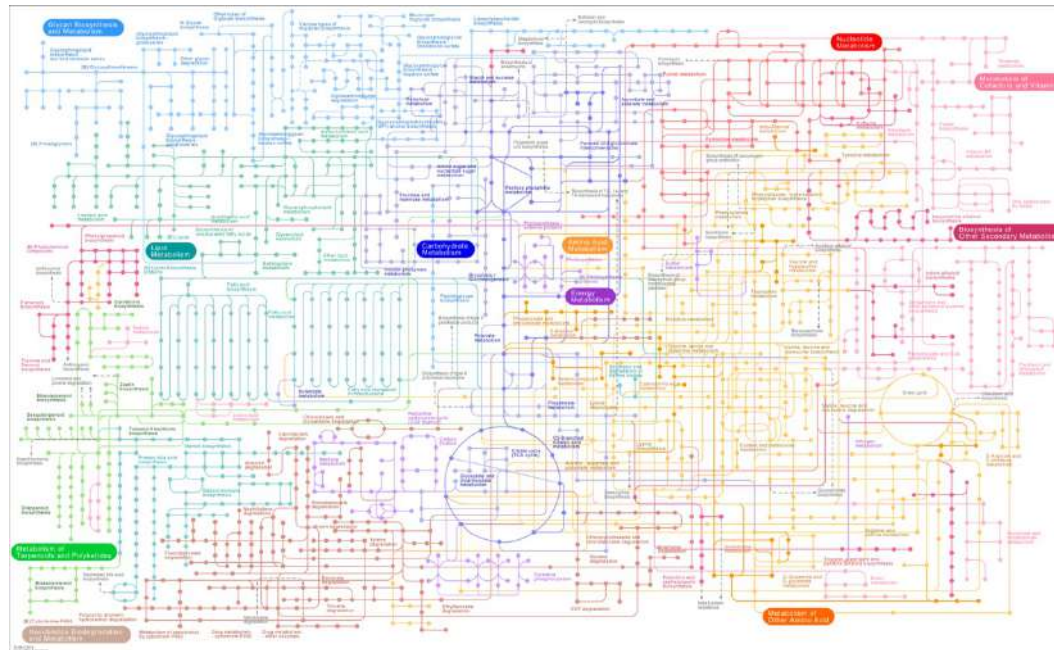
Practical

Work through `05_Simulating_DNA.Rmd`

From genomes to phenotypes?

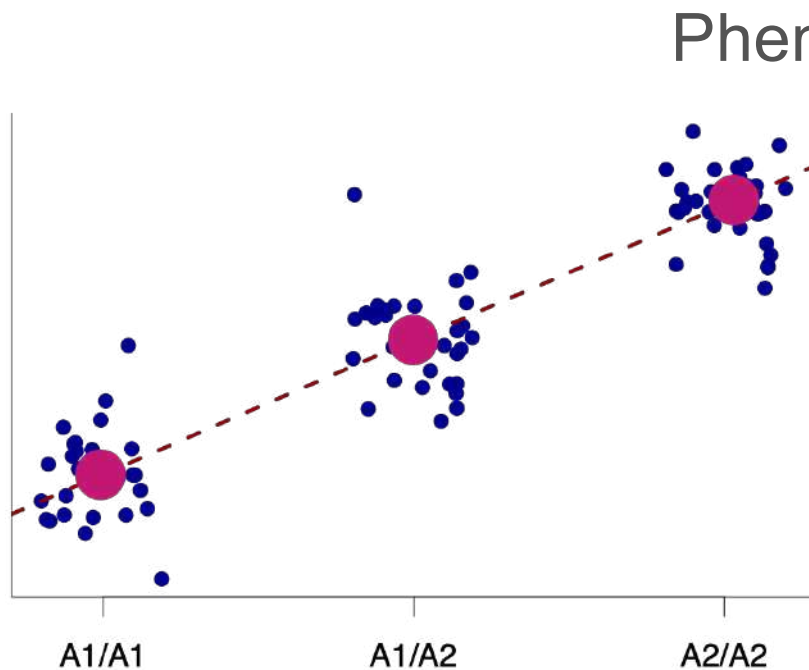
Phenotype = Function(Genomes, Environment)

But what is The Function?



Assume an additive function

- Simple 1st order approximation of The Function
- With additive effects and no environment interaction (more later!)

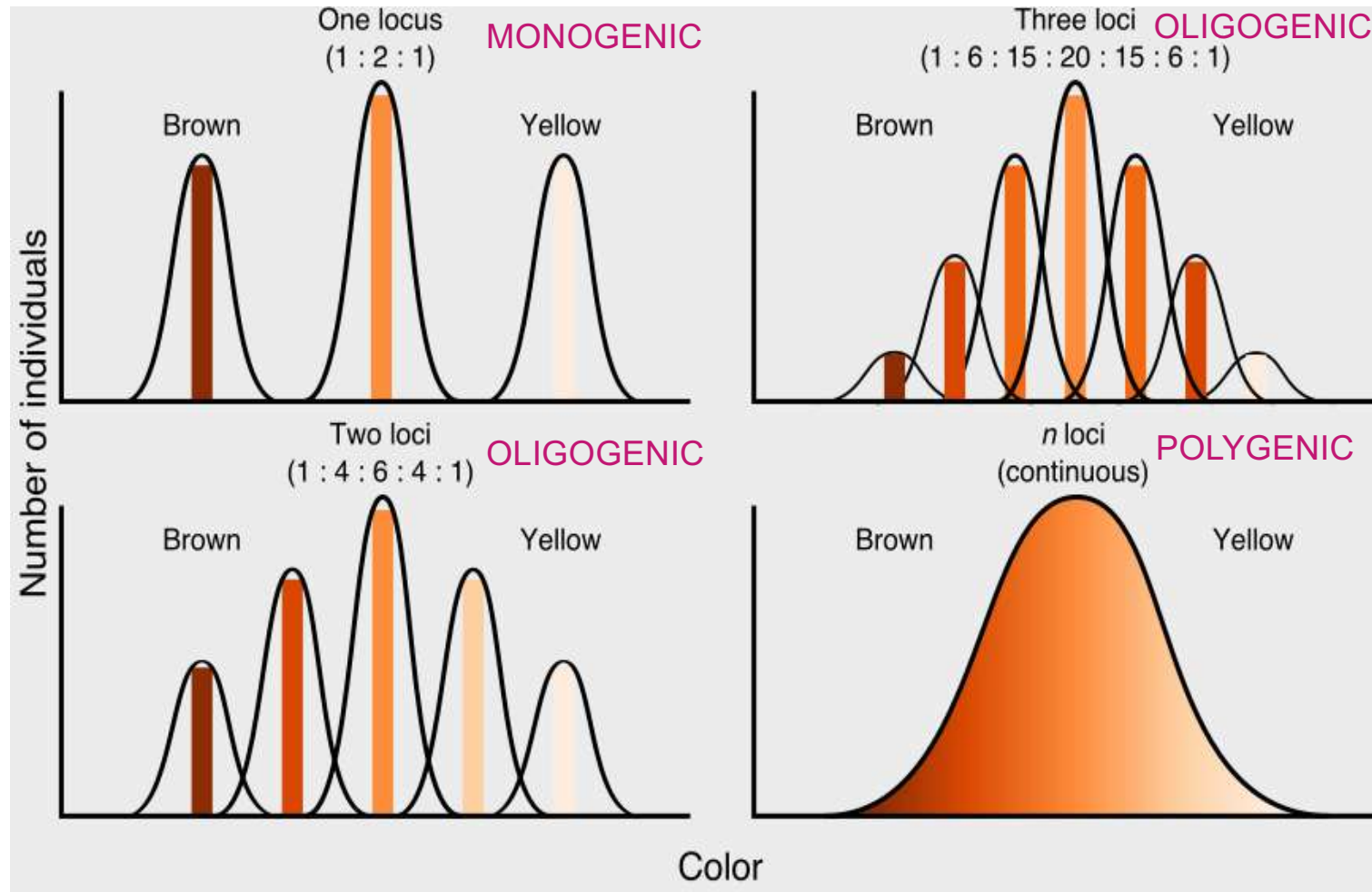


$$\begin{aligned} \text{Phenotype} = & \text{Dosage}_1 * \text{Effect}_1 + \\ & \text{Dosage}_2 * \text{Effect}_2 + \\ & \dots + \\ & \text{Dosage}_n * \text{Effect}_n + \\ & \text{Noise} \end{aligned}$$

Genome-wide haplotype & genotype values

Haplotype 1	0	1	1	0	0	1	
Haplotype 2	1	1	1	1	0	0	
Genotype	1	2	2	1	0	1	Allele dosages
	x						
	+1	+2	-1	+1	+1	-2	Effects
	↓						
Haplotype 1	0	+2	-1	0	0	-2	-1
Haplotype 2	+1	+2	-1	+1	0	0	+3
Genotype	+1	+4	-2	+1	0	-2	+2
							Values

Hypothetical architecture for cattle coat color



QTLs & SNPs

- Defined in simulation parameters
 - See ?SimParam_addTraitA
 - See ?SimParam_addSnpChip
- SNP chip overlap with QTL can be controlled
 - See ?SimParam_restrSegSites
- No genotyping error

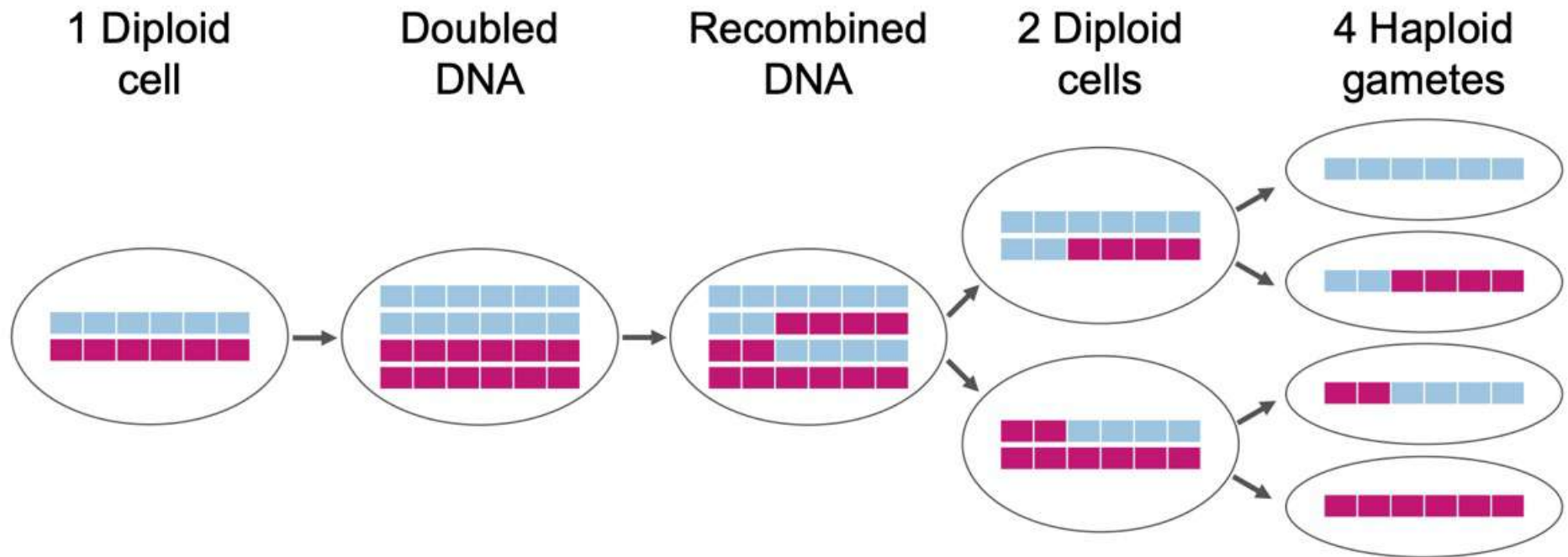
Take home message no. 2

**Simple DNA → Phenotype models give rise to
plenty of variation!**

Practical

Work through `06_Simulating_traits.Rmd`

Meiosis – Recombination & Segregation



Some numbers ...

- ~1 recombination per chromosome
 - recombination rate $\sim 1 \times 10^{-8}$
 - 1 Morgan (=100 cM) chromosome
 - $r \sim \text{Poisson}(l = 1)$ with Haldane mapping function
 - $r \sim \text{Gamma-sparkling}(l, v)$ general mapping function

Some numbers ...

- ~1 recombination per chromosome
 - recombination rate $\sim 1 \times 10^{-8}$
 - 1 Morgan (=100 cM) chromosome
 - $r \sim \text{Poisson}(l = 1)$ with Haldane mapping function
 - $r \sim \text{Gamma-sparkling}(l, v)$ general mapping function
- ~1 to 2 mutations per chromosome
 - mutation rate $\sim 1 \times 10^{-8}$ - $\sim 2 \times 10^{-8}$
 - In human $\sim 2.5 \times 10^{-8} \rightarrow 23 \times 2 \times 2 = \sim 100$ de-novo mutations
 - ~ 100 new + 2×50 old + 4×25 old-old + ...
 - ~1 new mutation has an effect?

Decoding germline *de novo* point mutations

Anne Goriely

NATURE GENETICS | VOLUME 48 | NUMBER 8 | AUGUST 2016

Analysis of a large whole-genome sequencing data set of 36,441 high-quality *de novo* mutations (DNMs) that arose in 816 family trios provides an unprecedented view into the landscape of DNMs in the germ line. This work both refines and challenges some of the views previously held on the nature and origin of DNMs.

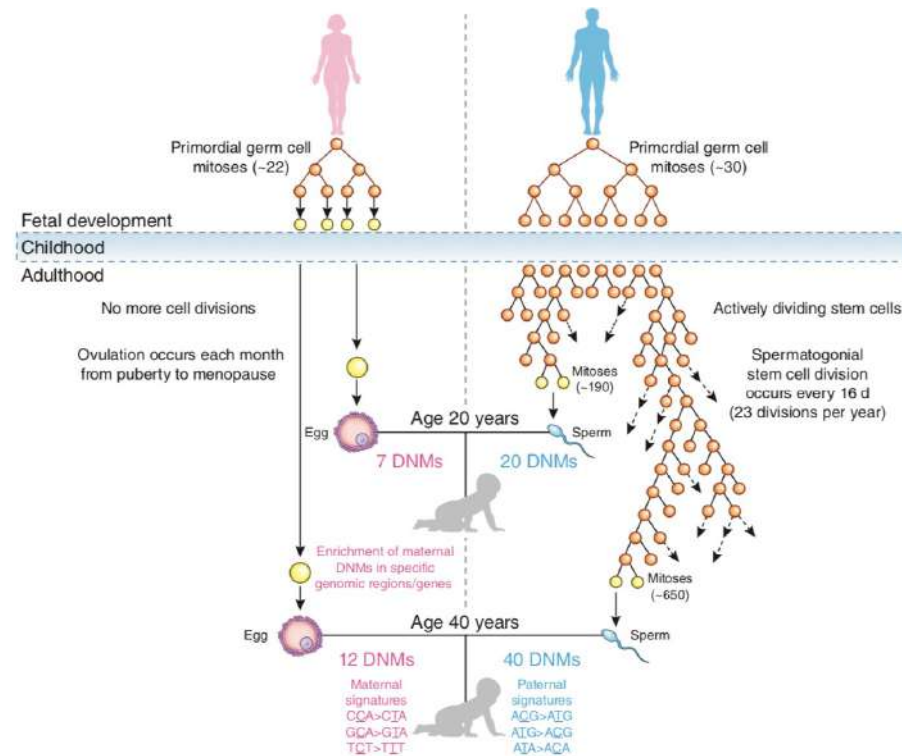


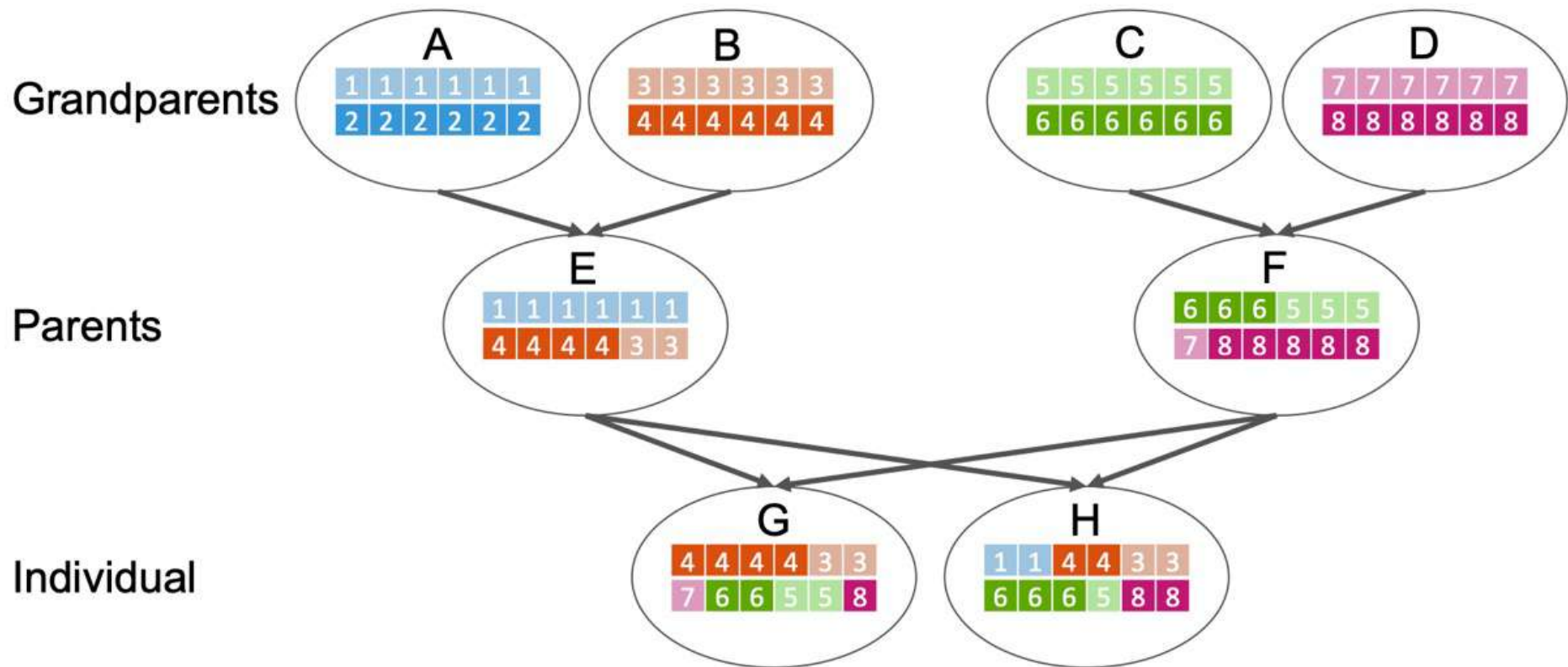
Figure 1 Gametogenesis differs in females and males. The sperm produced by a 20-year-old male has gone through ~190 cell divisions (mitoses), and this number increases to ~650 by the age of 40 years. In contrast, eggs do not replicate after birth. These sex-specific differences in germline biology are likely to explain the 3:1 excess of paternally derived DNMs observed in the progeny. However, maternal and paternal DNMs increase in number with parental age and show sex-specific mutational patterns. Orange cells, actively dividing stem cells; yellow cells, differentiating gametes undergoing meiosis.

Somatic mutations

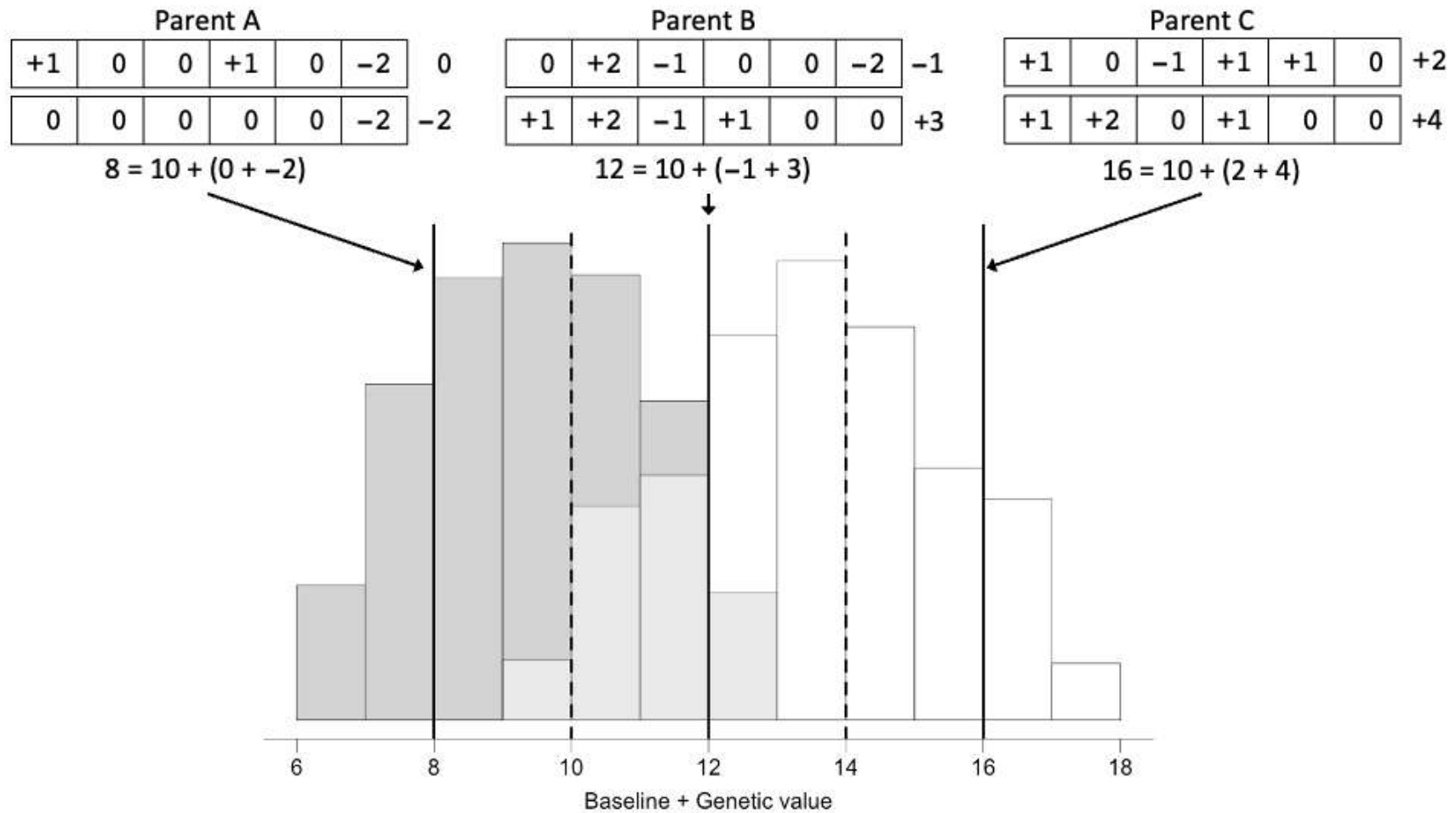
- $\sim 3 \times 10^{-7}$ --> ~ 10 x more common than germline!
- A somatic cell can then have 1000+ mutations!
100 from germline x 10+ = 1000+
- $\sim 10^{12 \text{ to } 13}$ cells in the body
- $\sim 10^{(12 \text{ to } 13)+3} = \sim 1 \times 10^{15-16}$ mutations in an adult with most nucleotides mutated in thousands of cells

Lynch (2016) <https://doi.org/10.1534/genetics.115.180471>

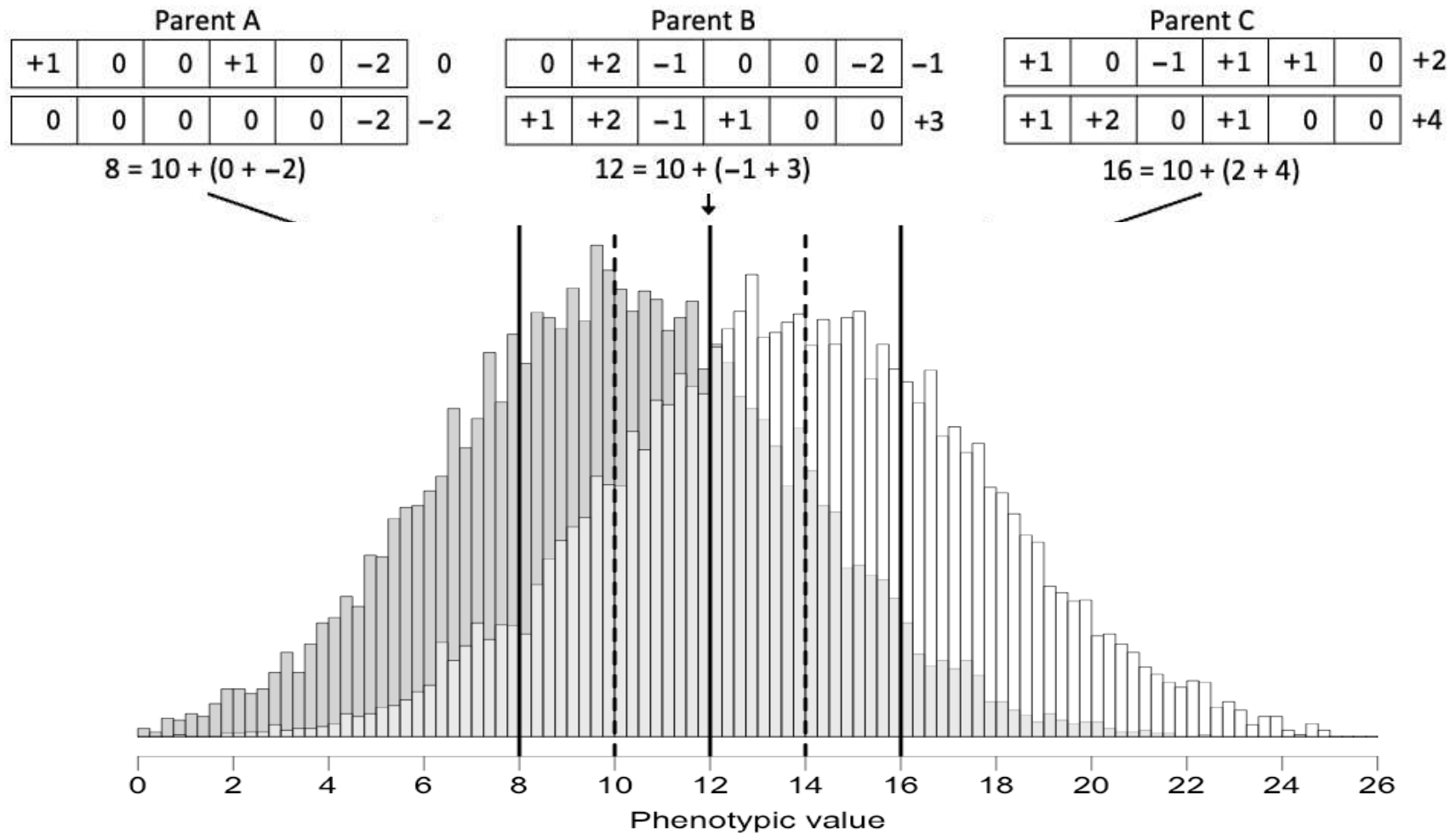
Meiosis in the context of a pedigree



Between and within family **genetic** variation



Between and within family phenotypic variation



Take home message no. 3

**Variation between & within families is substantial
and driven by meiosis!**

Takeaways

- Learning objectives
 - Encoding DNA variation
 - Simulate DNA & phenotypes in AlphaSimR
 - Simulate inheritance in AlphaSimR
- Take home messages
 - Encoding haplotypes (genotypes) as a series of 0 & 1 (0, 1, & 2)
 - Simple DNA → Phenotype models give rise to plenty of variation
 - Variation between & within families is substantial and driven by meiosis!



THE UNIVERSITY
of EDINBURGH



Biotechnology and
Biological Sciences
Research Council



THE ROYAL
SOCIETY

DNA variation, Phenotypes, & Lottery

Gregor Gorjanc

Athens, Greece
2025-01-30

