

Are Genes Everything?!



Alireza Habibzadeh
Biophysics Spring 2024 - Prof. Nader Reihani

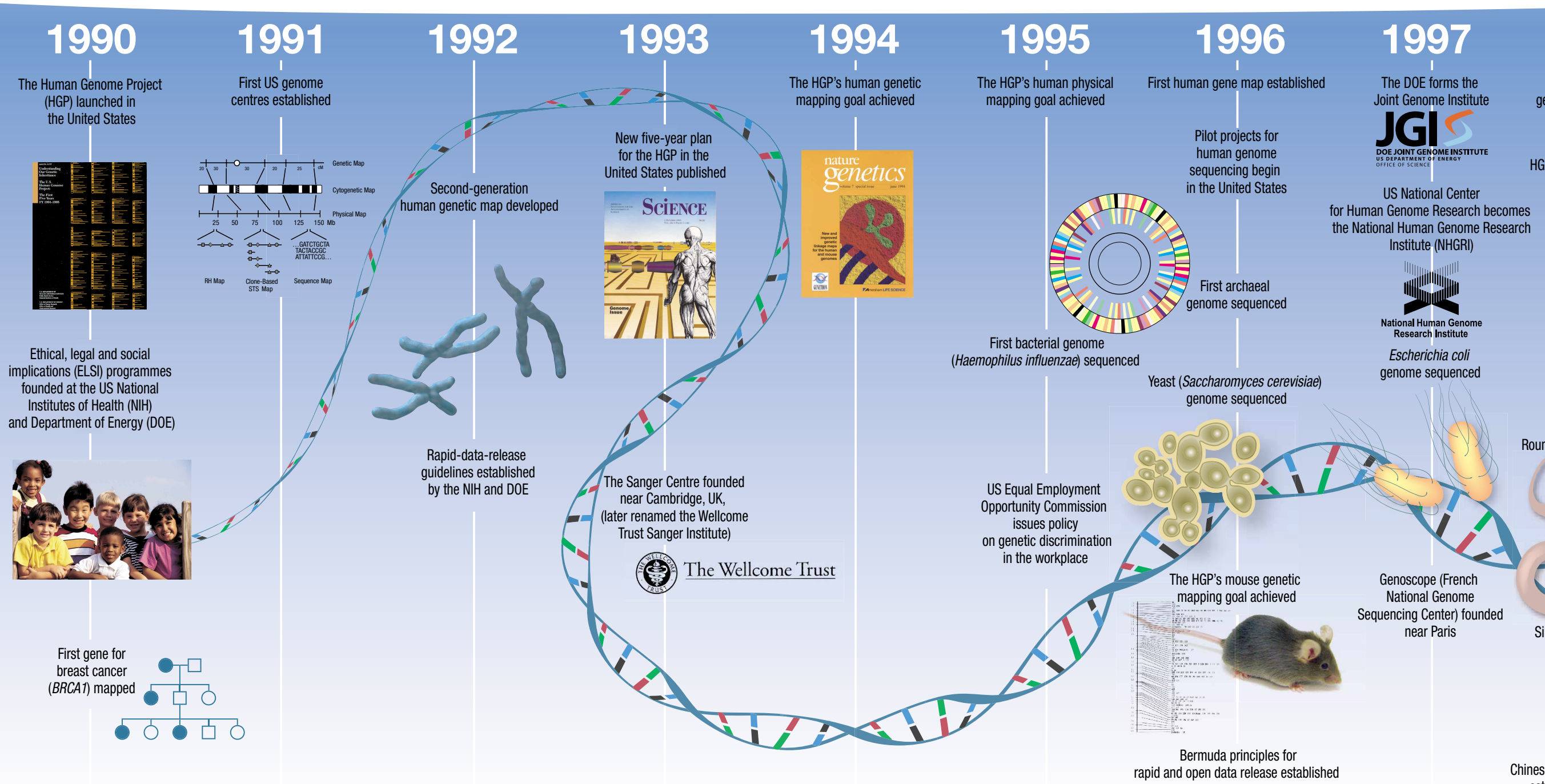
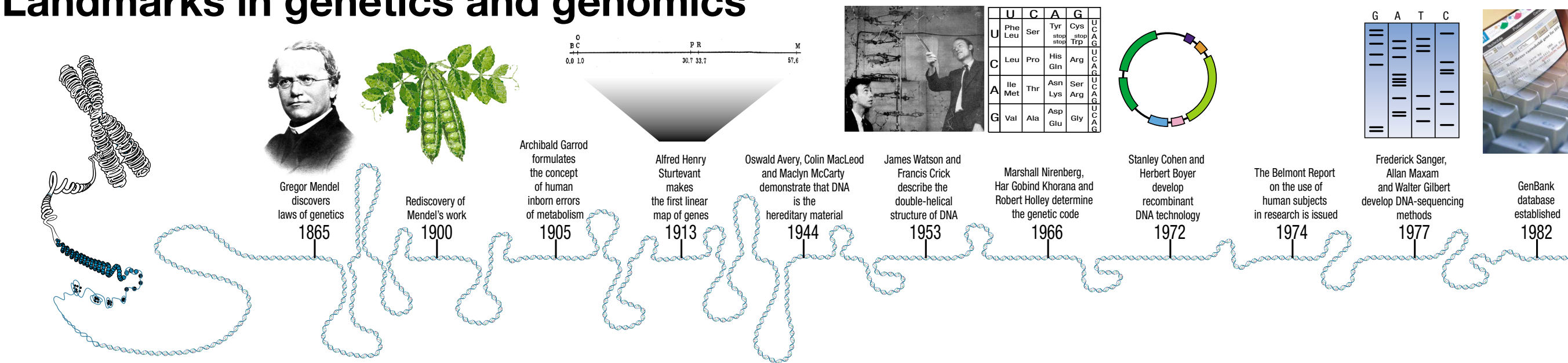
The Human Genome Project

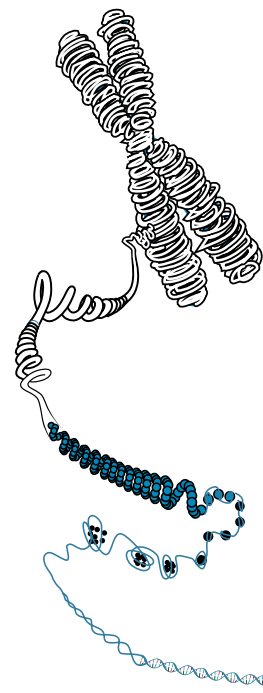
- 13 years from 1990 to 2003
The initial estimate was 15 years.
- \$3 billion in 1990 (overfund!)
In the end, approximately \$2.7 billion was used.
- The first (almost) complete human genome sequence (92%)
The remaining 8% took until the year ——. (guess the year)



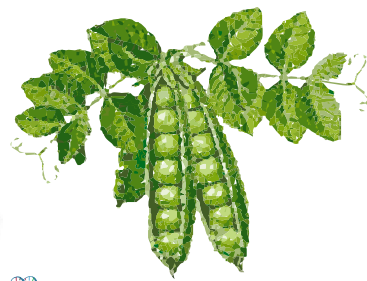
National Human Genome
Research Institute

Landmarks in genetics and genomics





Gregor Mendel
discovers
laws of genetics
1865



Rediscovery of
Mendel's work
1900

Archibald Garrod
formulates
the concept
of human
inborn errors
of metabolism
1905

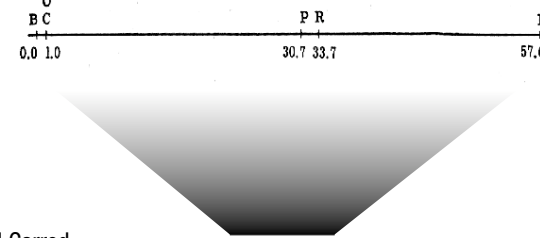
Alfred Henry
Sturtevant
makes
the first linear
map of genes
1913

Oswald Avery, Colin MacLeod
and Maclyn McCarty
demonstrate that DNA
is the
hereditary material
1944

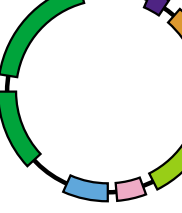
James Watson and
Francis Crick
describe the
double-helical
structure of DNA
1953

Marshall Nirenberg,
Har Gobind Khorana and
Robert Holley determine
the genetic code
1966

Stanley Cohen and
Herbert Boyer
develop
recombinant
DNA technology
1972



	Leu	Pro	His	Arg	stop
C			Gln		
A	Ile	Thr	Asn	Ser	
	Met		Lys	Arg	
G	Val	Ala	Asp	Gly	
			Glu		



1990

The Human Genome Project
(HGP) launched in
the United States

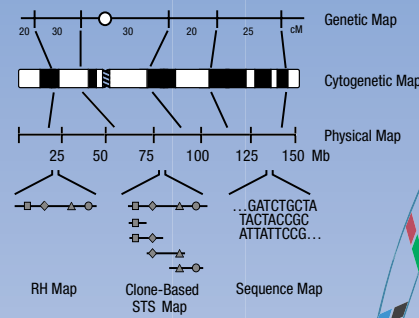


Ethical, legal and social
implications (ELSI) programmes
founded at the US National
Institutes of Health (NIH)
and Department of Energy (DOE)



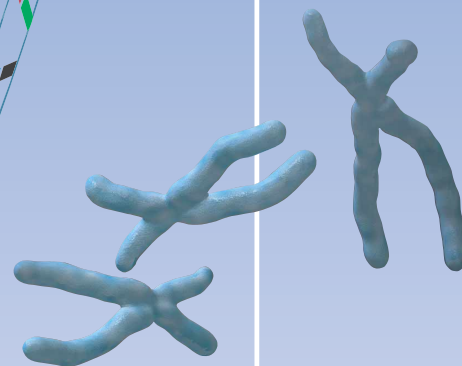
1991

First US genome
centres established



1992

Second-generation
human genetic map developed



Rapid-data-release
guidelines established
by the NIH and DOE

1993

New five-year plan
for the HGP in the
United States published



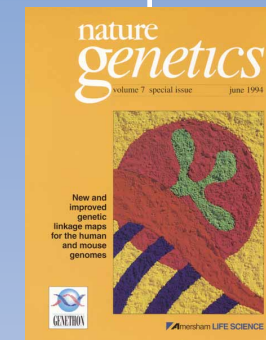
The Sanger Centre founded
near Cambridge, UK,
(later renamed the Wellcome
Trust Sanger Institute)



The Wellcome Trust

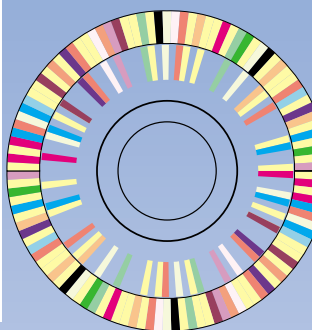
1994

The HGP's human genetic
mapping goal achieved



1995

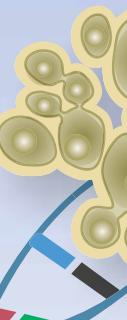
The HGP's human physical
mapping goal achieved

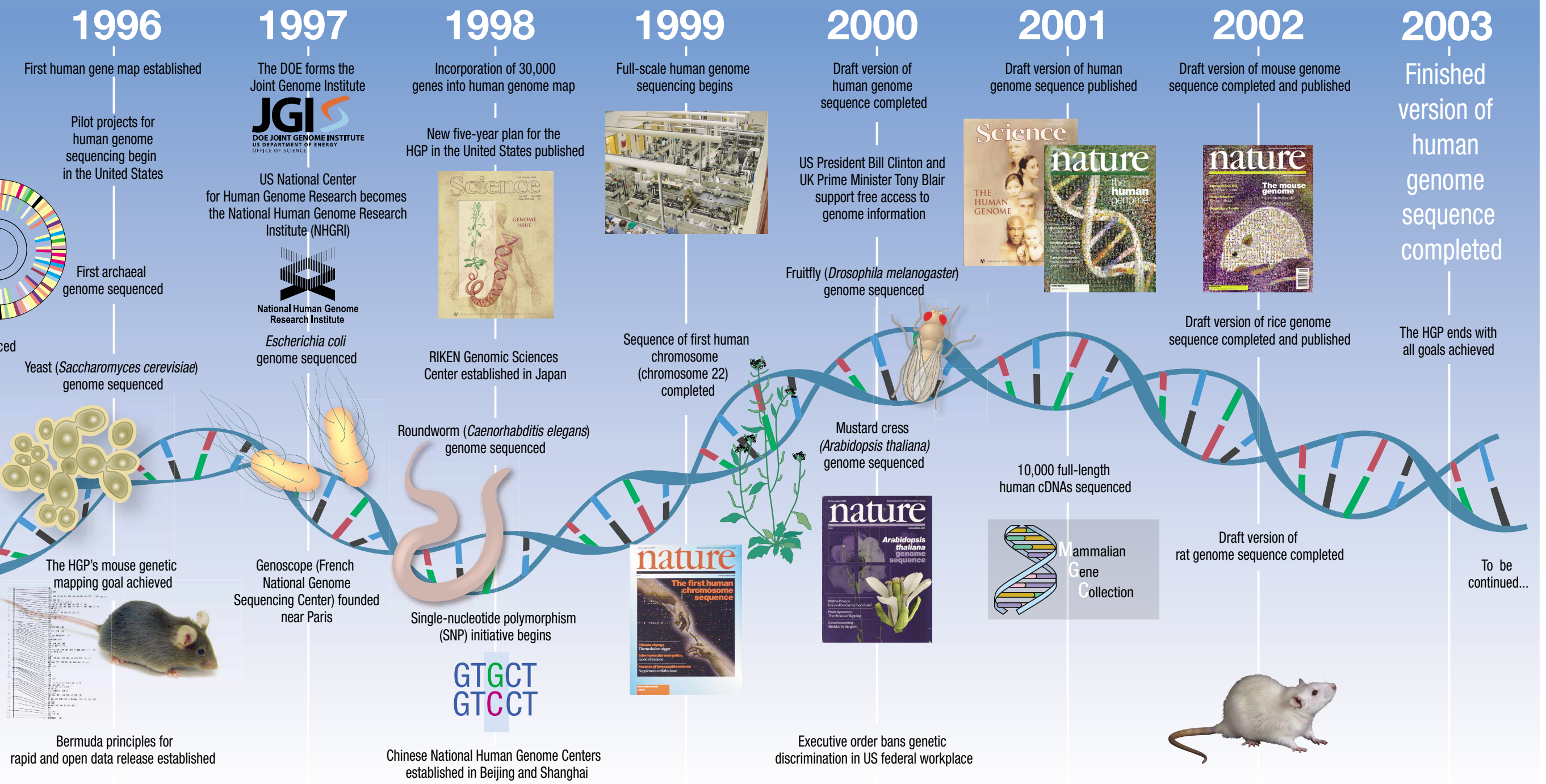
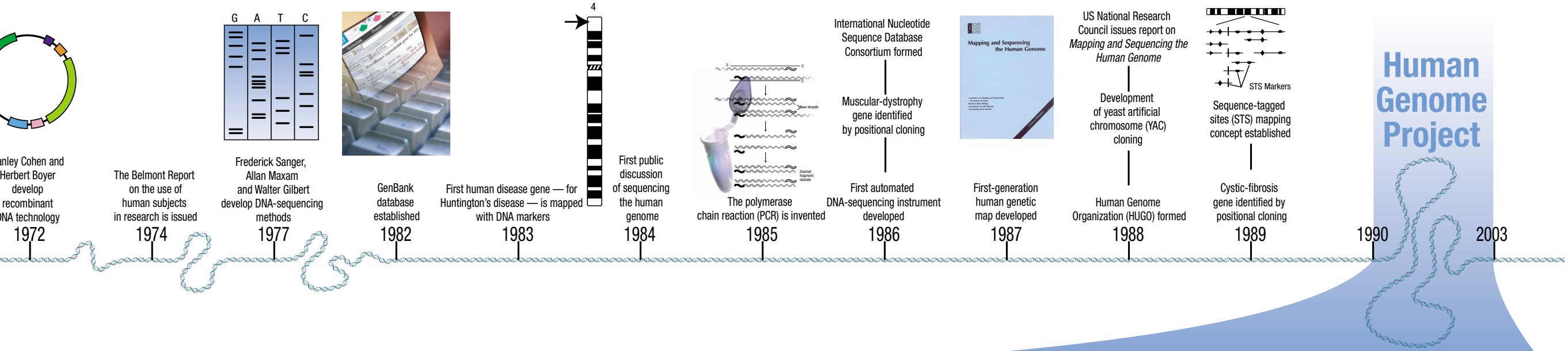


First bacterial genome
(*Haemophilus influenzae*) sequenced

Yeast (

US Equal Employment
Opportunity Commission
issues policy
on genetic discrimination
in the workplace





"Gene Therapy"

Lactose Intolerance

Mild Epilepsy

[illegible]

Whose gene is this?

- Some researchers suggested using a “normal” person. But who is really normal? :)
- 70% from one person of mixed race
- 30% from 19 people, mostly of European ancestry

WANTED
20 Volunteers
to participate in the
Human Genome Project
a very large international scientific research effort.

The goal is to decode the human hereditary information (*human blueprint*) that determines all individual traits inherited from parents. The outcome of the project will have tremendous impact on future progress of medical science and lead to improved diagnosis and treatment of hereditary diseases.

Volunteers will receive information about the project from the Clinical Genetics Service at Roswell Park, and sign a consent form before participating.

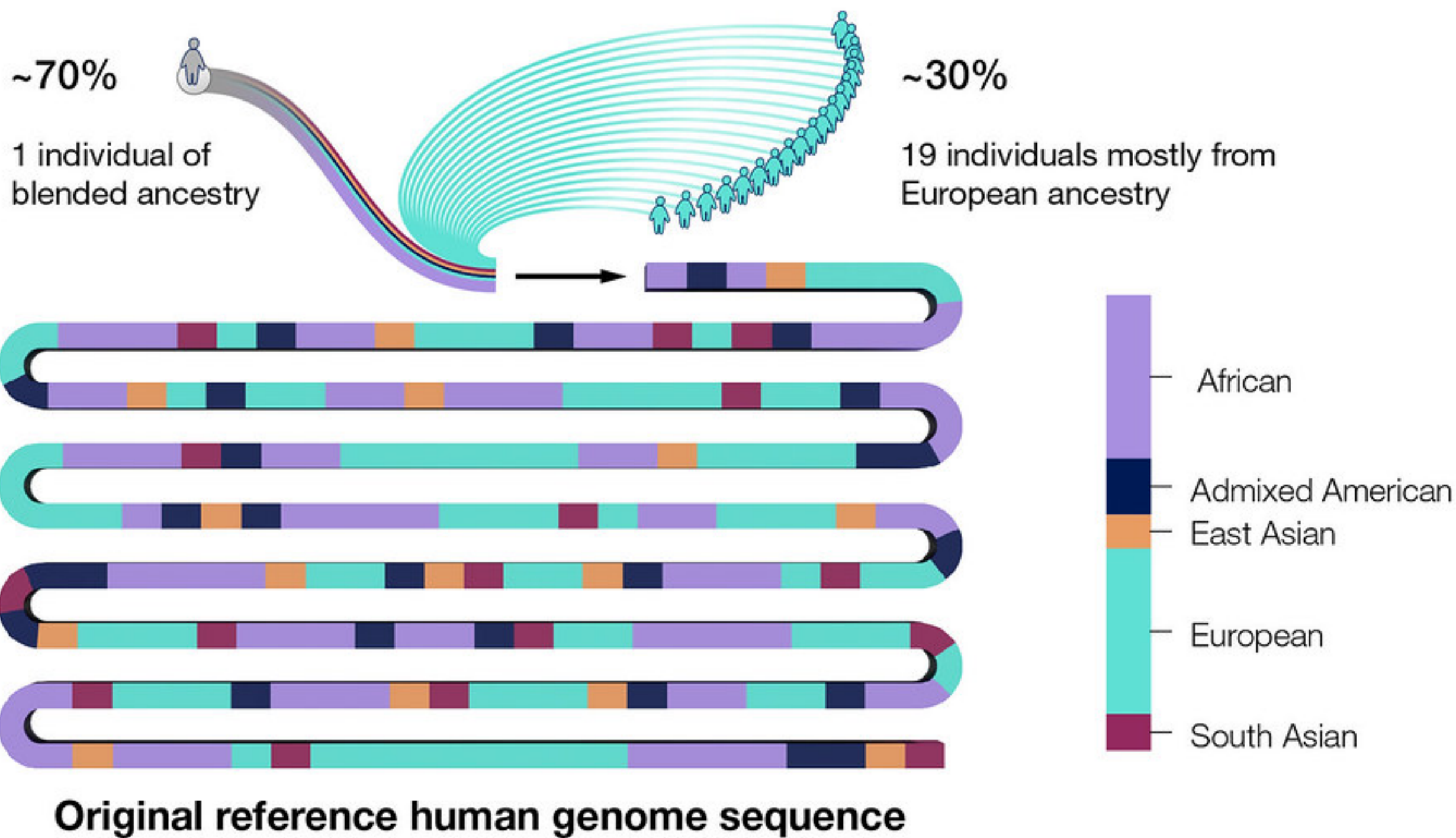
No personal information will be maintained or transferred.

Volunteers will provide a one-time donation of a small blood specimen. A small monetary reimbursement will be provided to the participants for their time and effort.

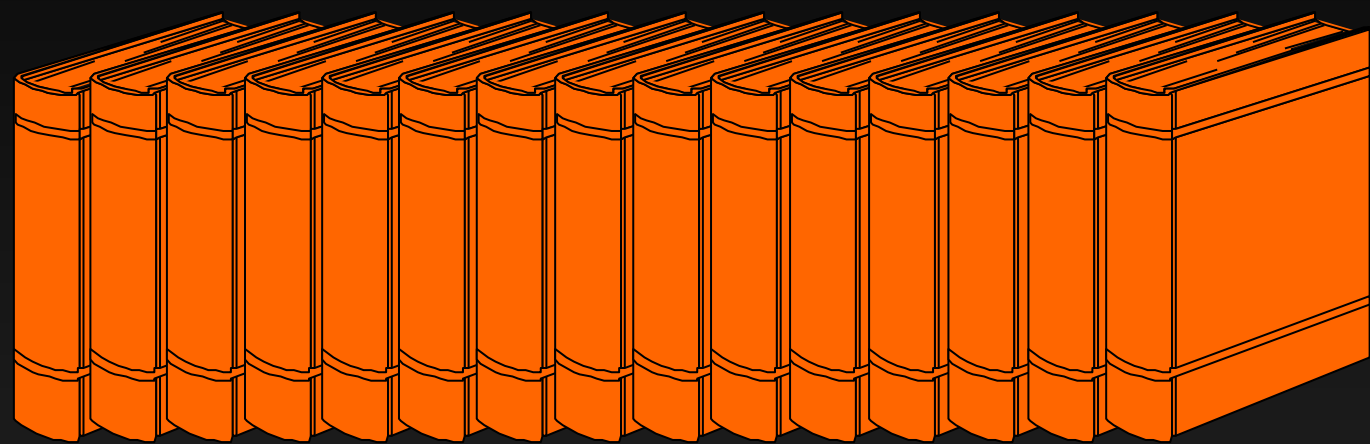
Individuals must be at least 18 years of age.
Persons who have undergone chemotherapy are not eligible.

**ROSWEILL
PARK**
CANCER INSTITUTE

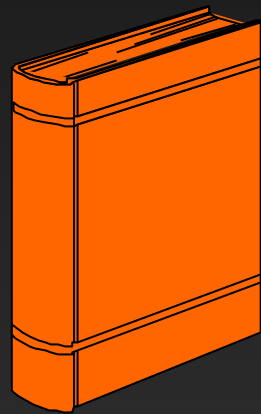
For more information please contact the
Clinical Genetics Service
845-5720 (9:00 am - 3:00 pm)
March 24 - 26, 1997



Scale of Genome, Chromosome, and Clone



Human Genome
(~3,000 Mb)



Human Chromosome
(~130 Mb)

Roughly size of
entire fruit fly or
nematode genome

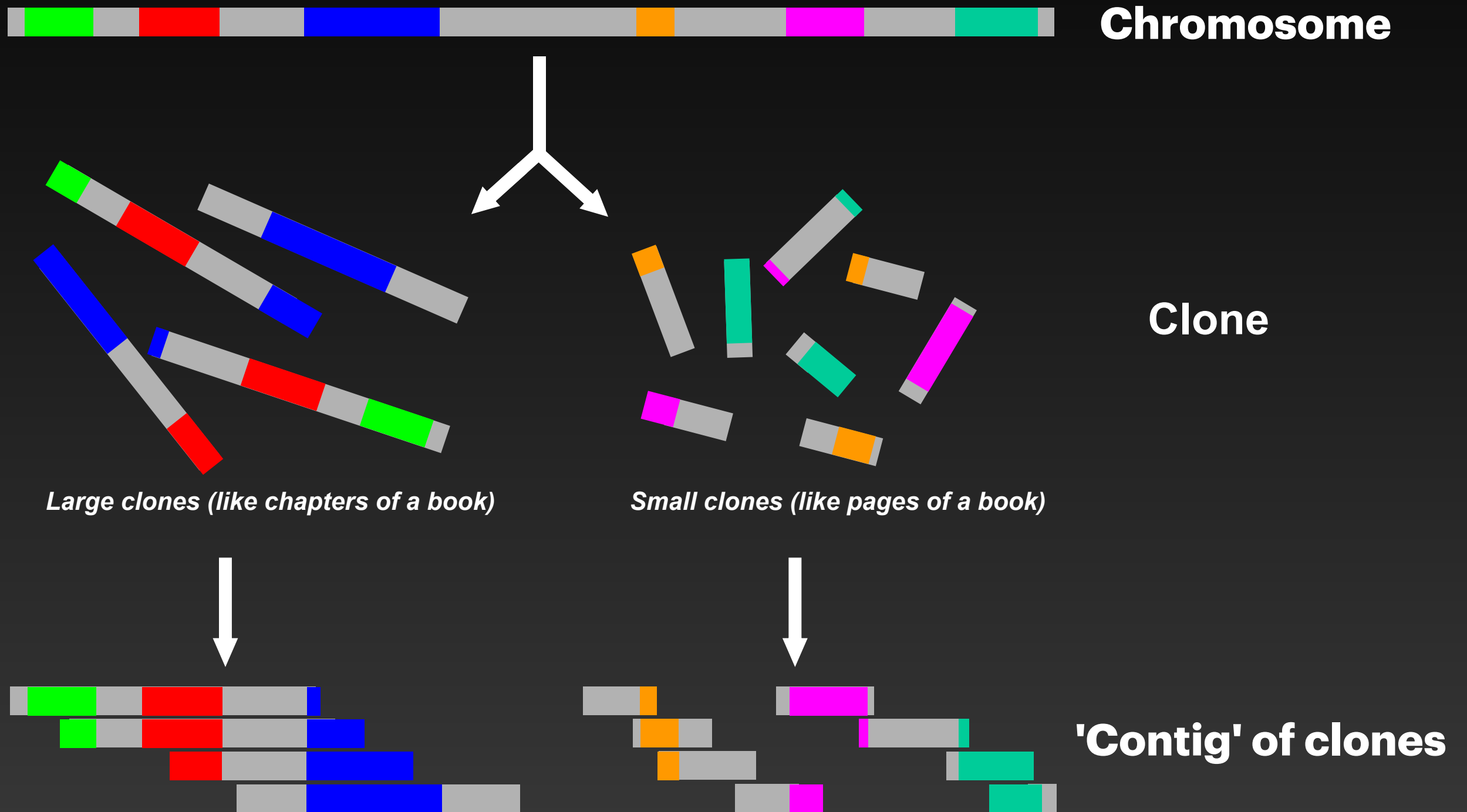
G	G	G	G	G	G	GATCGTCTAGAATCTC
G	G	G	G	G	G	GAGATCTCTGAGAGTC
G	G	G	G	G	G	GTGGGAACTGTGTGA
T	T	T	T	T	T	TGTGACTAGCCACAGT
T	T	T	T	T	T	TGTGACTAGCCACAGT
T	T	T	T	T	T	TACGTGTGAGAGATGT
A	A	A	A	A	A	ATGATGCACCTGACCC
G	G	G	G	G	G	GGTTTCACTCTCAAC
G	G	G	G	G	G	GA
C	C	C	C	C	C	CTCACTCCACCTCA
C	C	C	C	C	C	CCGGTTAGACATACAT
G	G	G	G	G	G	GAGGCCACCGCCGCT
G	G	G	G	G	G	GTGCACGTCCACCACC

Larger Clones
(~0.5-1.0 Mb)

G	G	G	G	G	G	GATCGTCTAGAATCTC
G	G	G	G	G	G	GAGATCTCTGAGAGTC
G	G	G	G	G	G	GTGGGAACTGTGTGA
T	T	T	T	T	T	TGTGACTAGCCACAGT
T	T	T	T	T	T	TAGGTATTGGGCAATT
T	T	T	T	T	T	TACGTGTGAGAGATGT
A	A	A	A	A	A	ATGATGCACCTGACCC
G	G	G	G	G	G	GGTTTCACTCTCAAC
G	G	G	G	G	G	GA
C	C	C	C	C	C	CTCACTCCACCTCA
C	C	C	C	C	C	CCGGTTAGACATACAT
G	G	G	G	G	G	GAGGCCACCGCCGCT
G	G	G	G	G	G	GTGCACGTCCACCACC

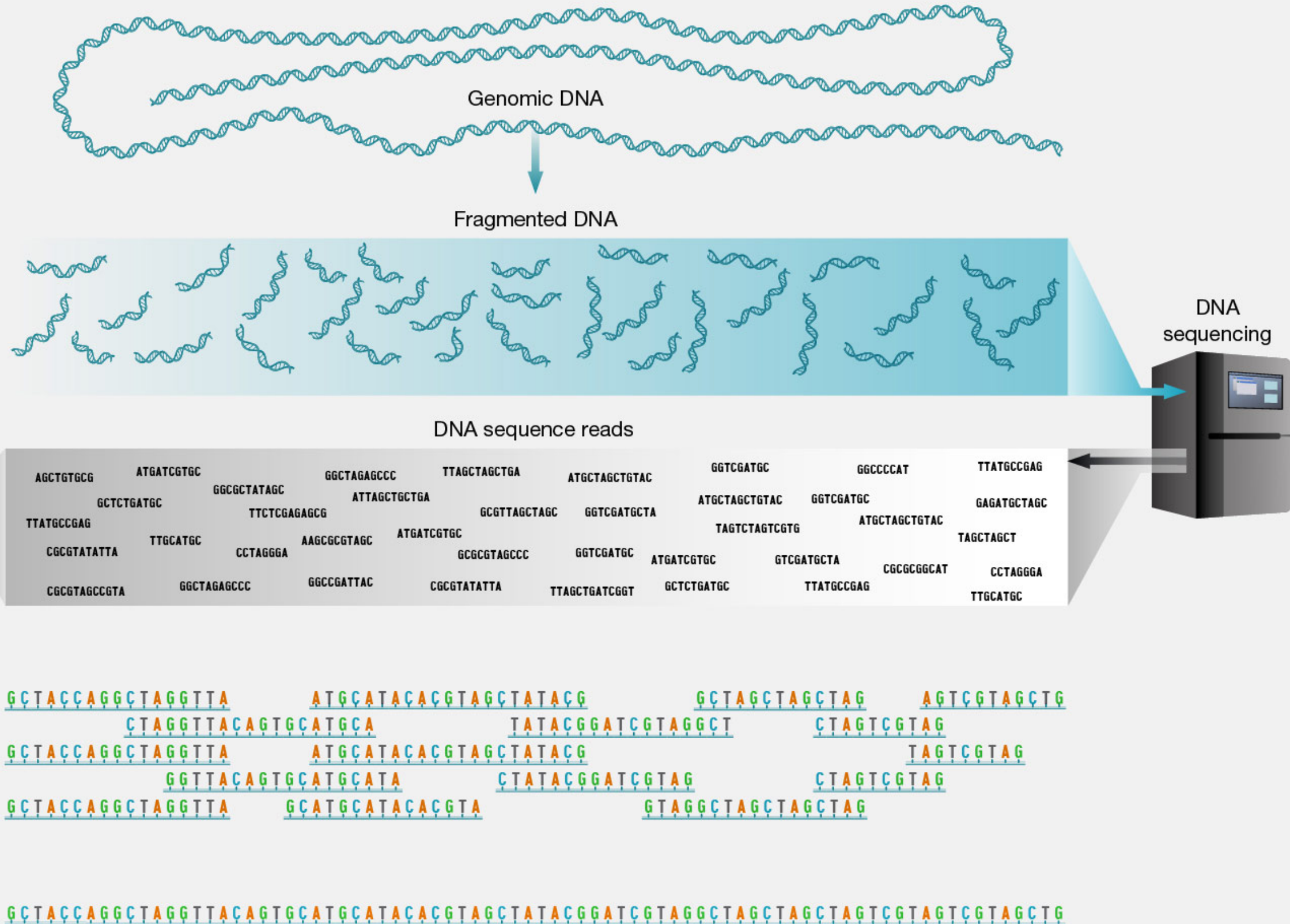
Smaller Clones
(~0.1-0.2 Mb)

Mapping DNA by Physical Cloning

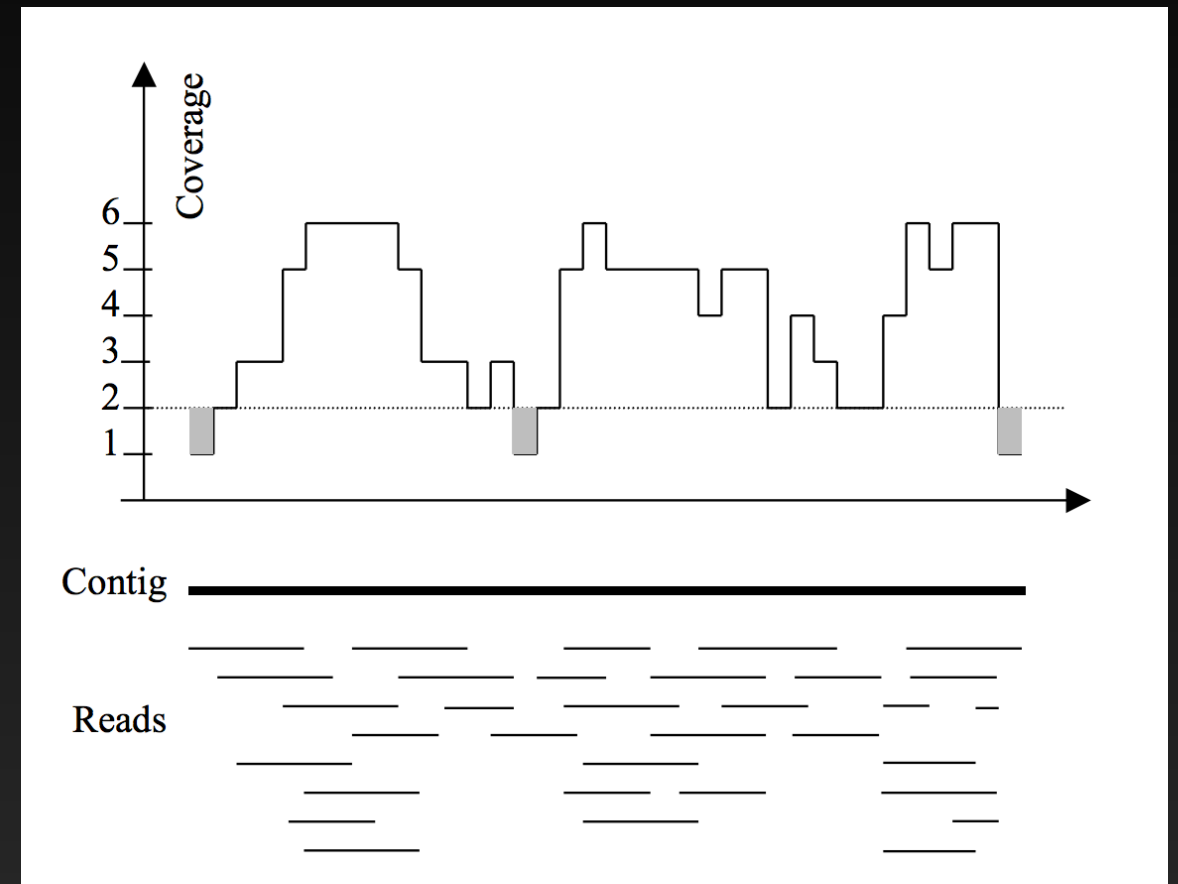
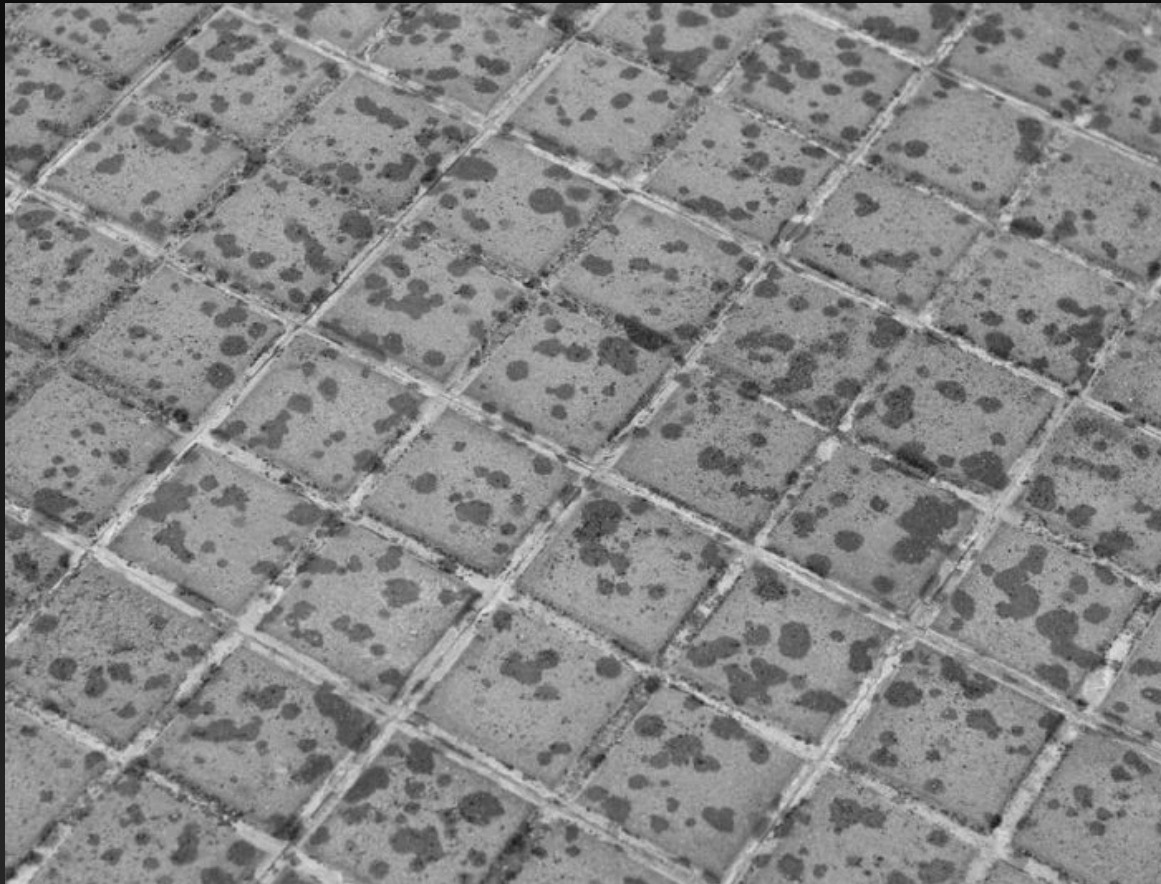


Shotgun sequencing

from genome.gov/genetics-glossary



Raindrops on a Sidewalk



- Because raindrops fall randomly in various places, some spots need to get extra drops for the entire sidewalk to get wet.
- Similarly, the final accuracy of sequencing depends on reading each part multiple times (for example, 30-50 times, which is called 'coverage').

Sequencing Saadi's Gulistan (Persian Literature)

دو برادر یکی خدمت سلطان کردی و دیگر به زور بازو نان خوردی.

باری، این توانگر گفت درویش را که: چرا خدمت نکنی تا از مشقّت کار کردن برهی؟

گفت: تو چرا کار نکنی تا از مَذَلَّتِ خدمت رهایی یابی؟ که خردمندان گفته‌اند: نان خود خوردن و نشستن به که کمرشمشیر زرّین به خدمت بستن.

دو برادر یکی خدمت

دو برادر یکی خدمت

یکی خدمت سلطان

یکی خدمت سلطان

یکی خدمت سلطان

خدمت سلطان مشغولم و به خیرش امیدوار و از عقوبتش

خدمت سلطان کردی و دیگر به زور بازو نان خوردی

خدمت سلطان کردی و دیگر به زور بازو نان خوردی

خدمت سلطان کردی و دیگر به زور بازو نان خوردی

خدمت سلطان کردی و دیگر به زور بازو نان خوردی

Coverage = 4

aligned reads

File Navigate Info Color Dim Misc

Help

yg.fasta.screen.ace.3

Contig32

Some Tags

Pos:

1

Search for String

Compl Cont

Compare Cont

Find Main Win

Exp Err/10kb:

12.17

8030

8040

8050

8060

8070

8080

8090

8100

8110

CONSENSUS

AGGAAAAGACTATCACAGCGTATTCCTGAAAGAGATGAACTATGAATTGAGTGTAGGCTTCTCTGCAGAGGCCAAA*GGTAGGATC

yg12h02.x1 cttgggggggggaaggaaagaacttttcccccggtttcctgaaggagaagacaccttgaaatggggggggggggttt*tttttttggg

yg03d09.y1 aggaAAAGACTatcAcagcgtatTcctgaaAGAGAtgaaCTATGAattGAgtaggcttctctgcagaggcaaaa*ggtaggato

yg09g04.x1 aggaaaagactatcacagcgtattcctgaaagagatgaactatgaattgagtgtaggcttcttgcagaggcaaaa*ggtaggato

yg13h04.x1 AGGAAAAGACTATCACAGCGTATTCCTGAAAGAGATGAACTATGAATTGAGTGTAGGCTTCTCTGCAGAGGCCAAA*GGTAGGATC

yg01e03.y1 AGGAAAAGACTATCACAGCGTATTCCTGAAAGAGATGAACTATGAATTGAGTGTAGGCTTCTCTGCAGAGGCCAAA*GGTAGGATC

yg08h10.x1 xxxxxxxxxxTatcAcagcgtattcctgaaagagatgaacTatgatGAgtaggctTCtctGcagaggcaaaa*ggtagGATC

yg04f11.y1 xxxxxxxxxxxxxxxxxxxxxxxxxxxxcagctcgccca

yg01g01.y1 acatcggttcaaagttgaacatccgtatxx*xxxxxxxx

yg01g07.y1 xxx*xxxxxxxx

yg02e04.y1 AGGAAAxx*xxxxxxxx

yg02f10.y1 AgGAAAAGACTatcgagCGTATTCCTGAAAGAGATGAACTATgaaTTGAGTGTAGGCTTCTCTGCAGAGGCCAAA*GGTAGGATC

yg02c10.y1 aggaAAAGACTATCACAGCGTAttcctGAAaGAGATGAACTATGAATTGAGTGTAGGCTTCTCTGCAGaggCAAA*GGTAGGATC

yg03g10.y1 AGGAAAAGACTATCACAGCGTATTCCTGAAAGAGATGAACTATGAATTGAGTGTAGGCTTCTCTGCAGAGGCCAAA*GGTAGGATC

yg18a10.y1 AGGAAAAGACTATCACAGCGTATTCCTGAAAGAGATGAACTxx*xxxxxxxx

yg08f02.y1 aggaAAAGACTATCACAGCGTAttcctGaAaGAGATGAACTATGAATTGAGTGTAGGCTTCTCTGCAGaggCAAA*GGTAGGATC

yg02h10.y1 agaaaaatcctatcgagcgtattcttaaaaaagatgaactatgaatcgagggtaggcttctctcaagag*cacagggtaggato

yg18e09.y1 aggaAaAGACTatcacagcgTAttcctGAAAGAGATGAACTATGAATTGAGTGTAGGCTTCTCTGCAGAGGCCAAA*GGTAGGATC

yg13d05.x1 ggtgccgcggtcactgtgccgcggtctgcgtcgccgca*tggcgcgcg

File

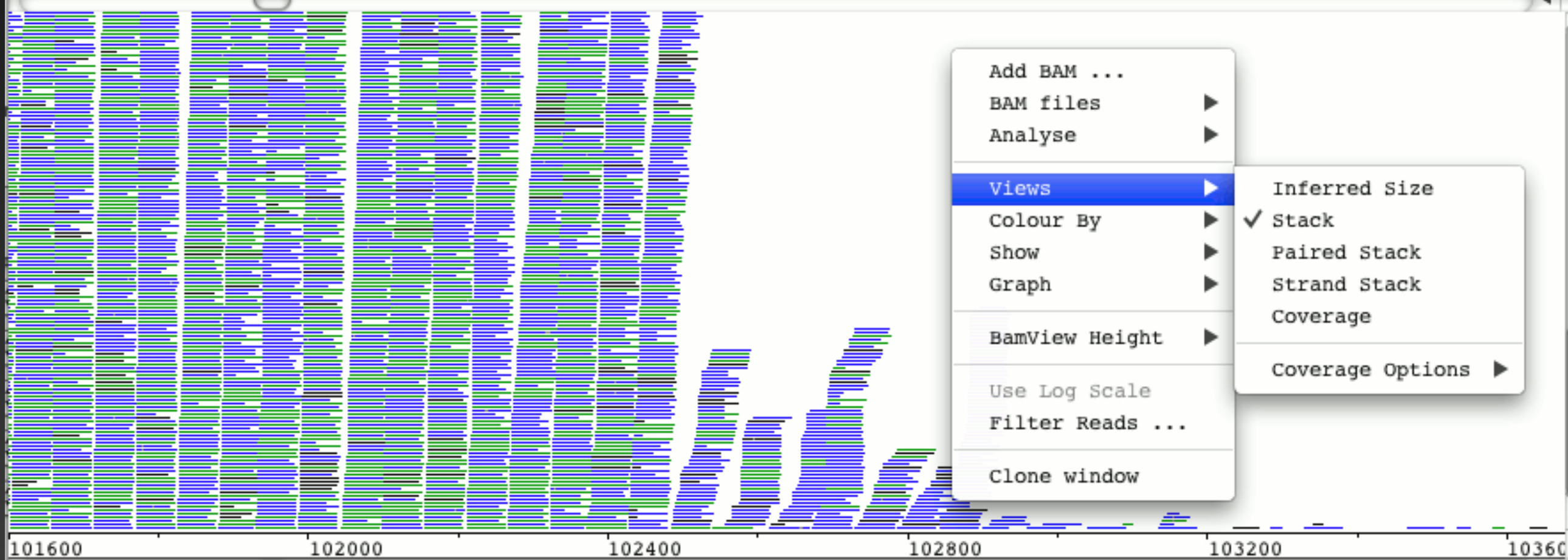
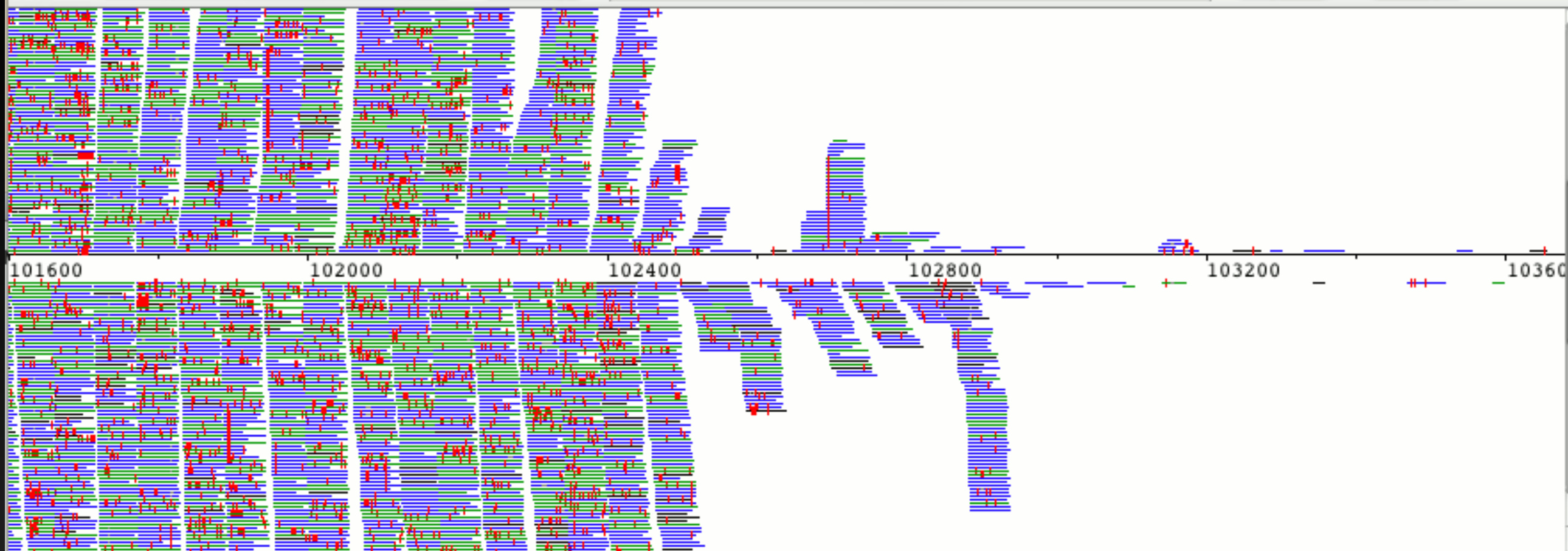
MAL1

GoTo:

101600

-

+

☐ Hide

Add BAM ...

BAM files ▶

Analyse ▶

Views ▶

Colour By ▶

Show ▶

Graph ▶

BamView Height ▶

Use Log Scale

Filter Reads ...

Clone window

Inferred Size

✓ Stack

Paired Stack

Strand Stack

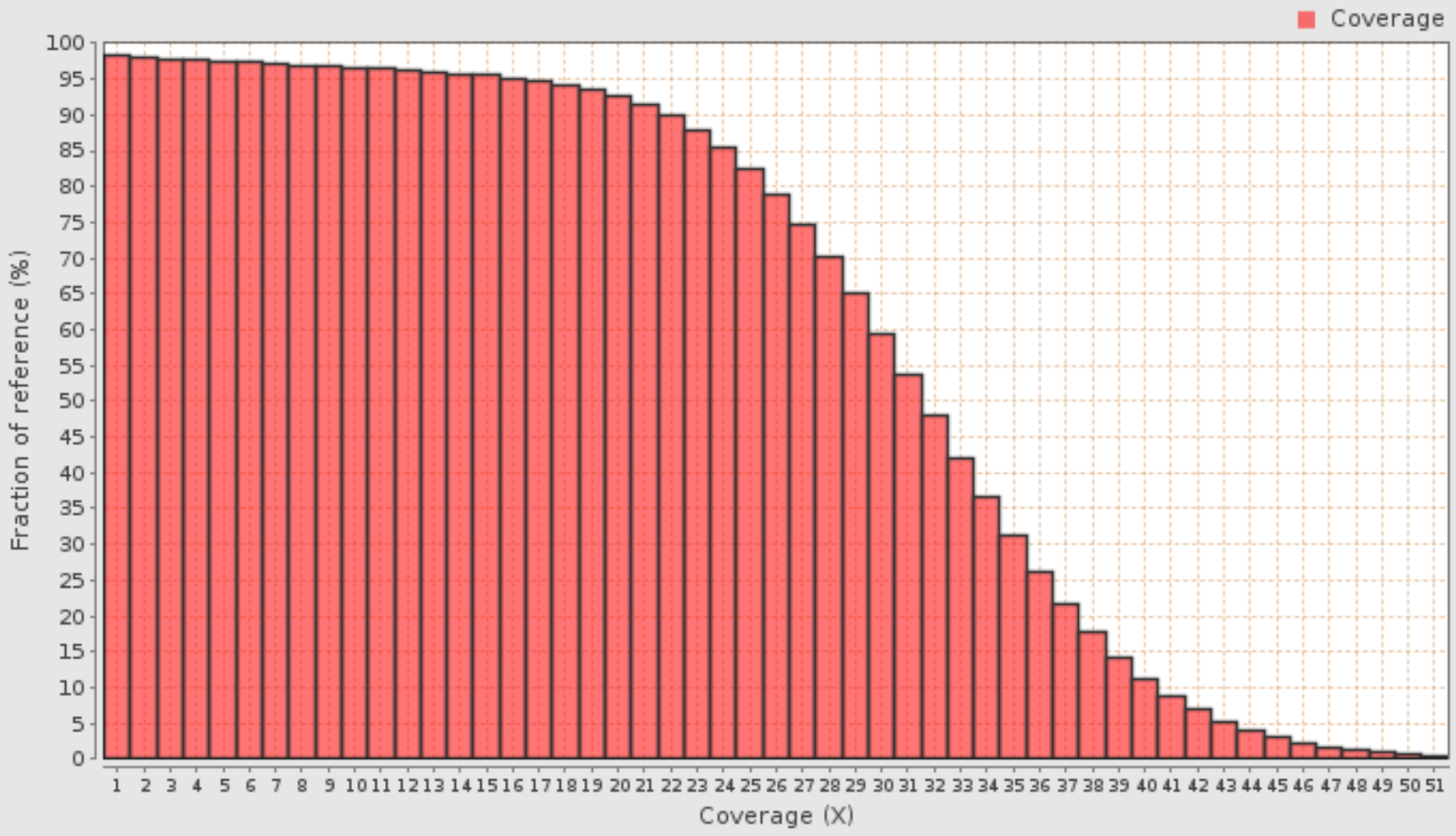
Coverage

Coverage Options ▶

- Results
- Input data & parameters
- Summary
- Coverage across reference
- Coverage Histogram
- Coverage Histogram (0-50X)
- Genome Fraction Coverage
- Duplication Rate Histogram
- Mapped Reads Nucleotide Content
- Mapped Reads GC-content Distributi
- Mapping Quality Across Reference
- Mapping Quality Histogram
- Insert Size Across Reference
- Insert Size Histogram

Genome Fraction Coverage

ERR089819.bam



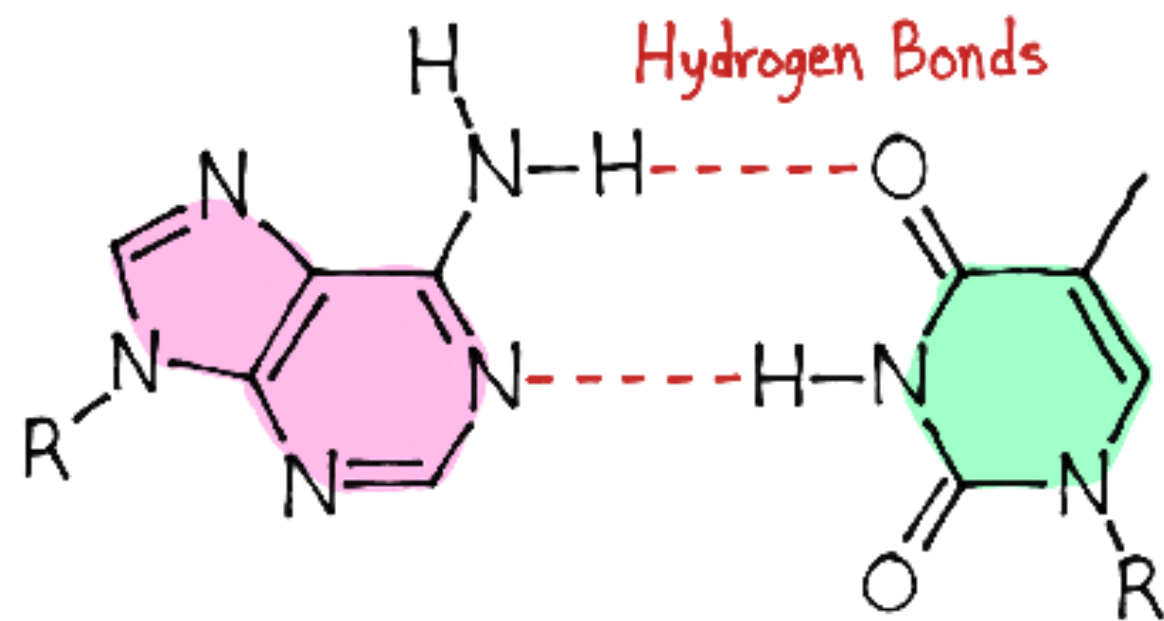
Qualimap v.2.0

File Tools Windows Help

RNA Seq QC: /home/kokonec

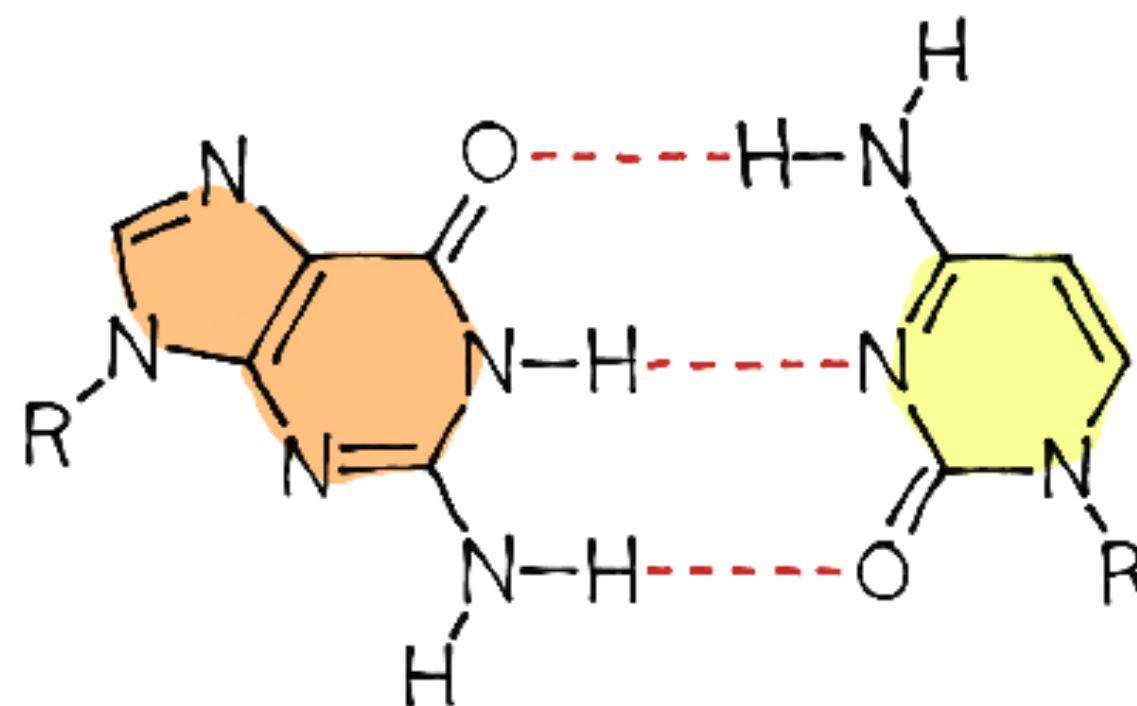
Results

- Input data & parameters
- Summary
- Coverage across reference
- Coverage Histogram
- Coverage Histogram (0-50X)
- Genome Fraction Coverage
- Duplication Rate Histogram
- Mapped Reads Nucleotide Content
- Mapped Reads GC-content Distribution
- Mapping Quality Across Reference
- Mapping Quality Histogram
- Insert Size Across Reference
- Insert Size Histogram



Adenine

Thymine



Guanine

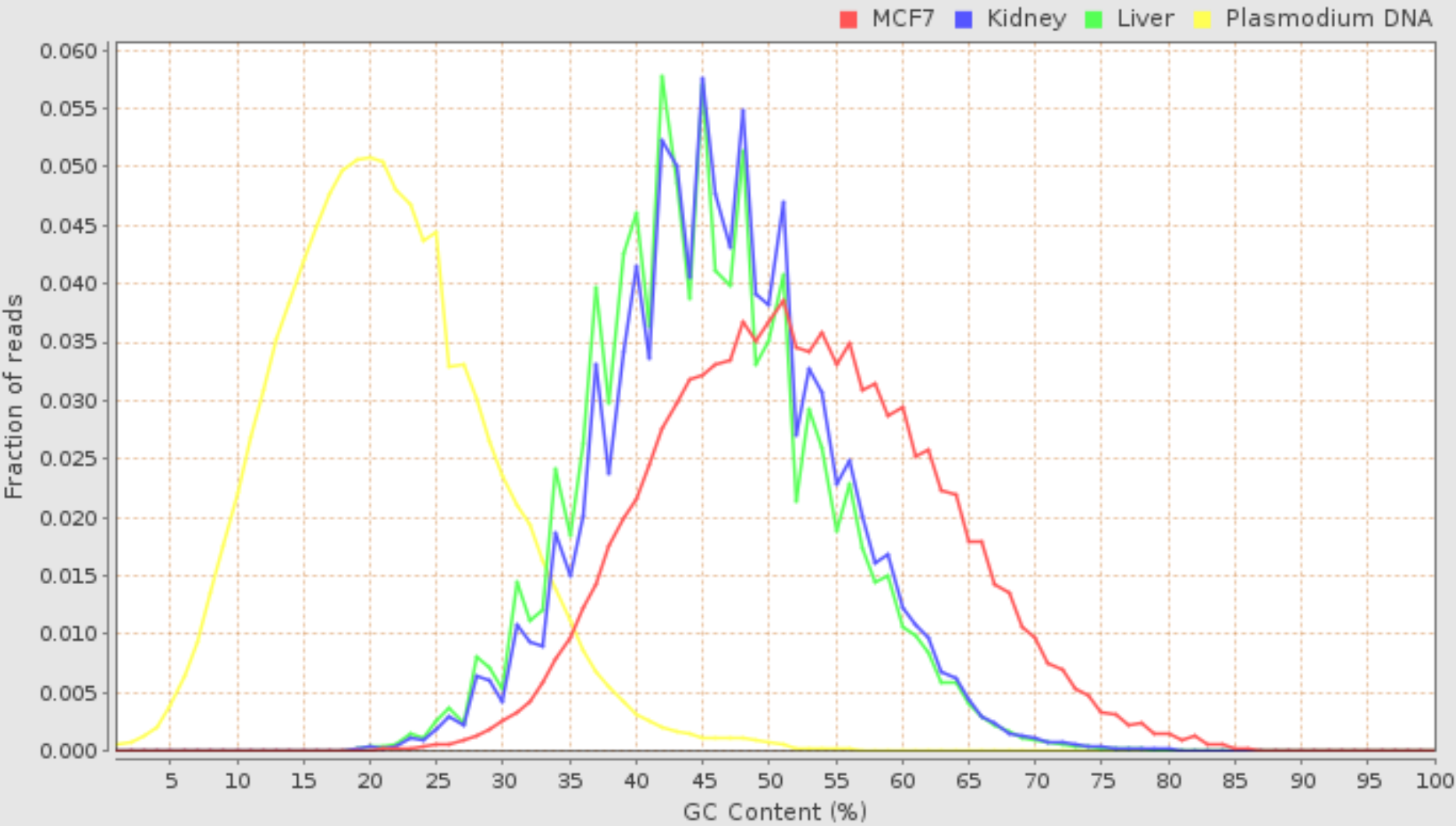
Cytosine

RNA Seq QC: ERR089819.bam

- Results
- Input data & parameters
- Summary
- PCA
- Coverage Across Reference
- Coverage Histogram (0-50X)
- Genome Fraction Coverage
- Duplication Rate Histogram
- Mapped reads GC-content
- Mapped Reads Clipping Profile
- Mapped Reads GC-content Distribution
- Mapping Quality Across Reference
- Mapping Quality Histogram
- Insert Size Across Reference
- Insert Size Histogram

Mapped Reads GC-content Distribution

Multi-sample BAM QC



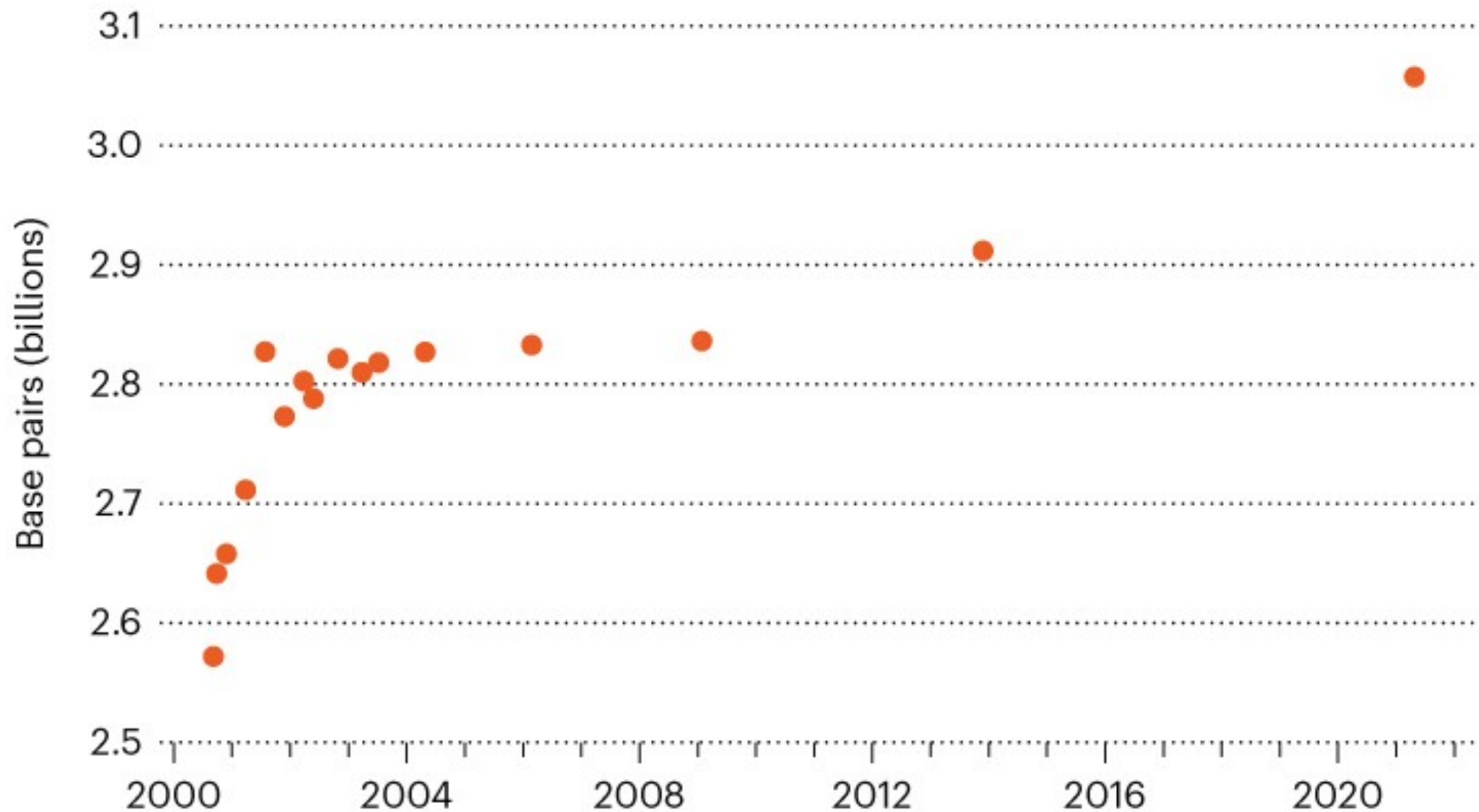
Reaching a Complete Genome

- The Human Genome Project (HGP) produced a high-quality sequence of the human genome, but it only covered 92% of the human genome.
- The remaining 8% was not sequenced due to the inability of the available DNA sequencing methods at the time, but these regions are important for structural (centromeres and telomeres) and medical reasons. (Previously, it was thought that these regions were useless.)
- In the past 20 years or so, several new and revolutionary methods for DNA sequencing have been developed.
- These new methods, along with better computational techniques, enabled a new group of researchers to finally produce a complete sequence of the human genome in 2022.



COMPLETING THE HUMAN GENOME

Researchers have been filling in incompletely sequenced parts of the human reference genome for 20 years, and have now almost finished it, with 3.05 billion DNA base pairs.



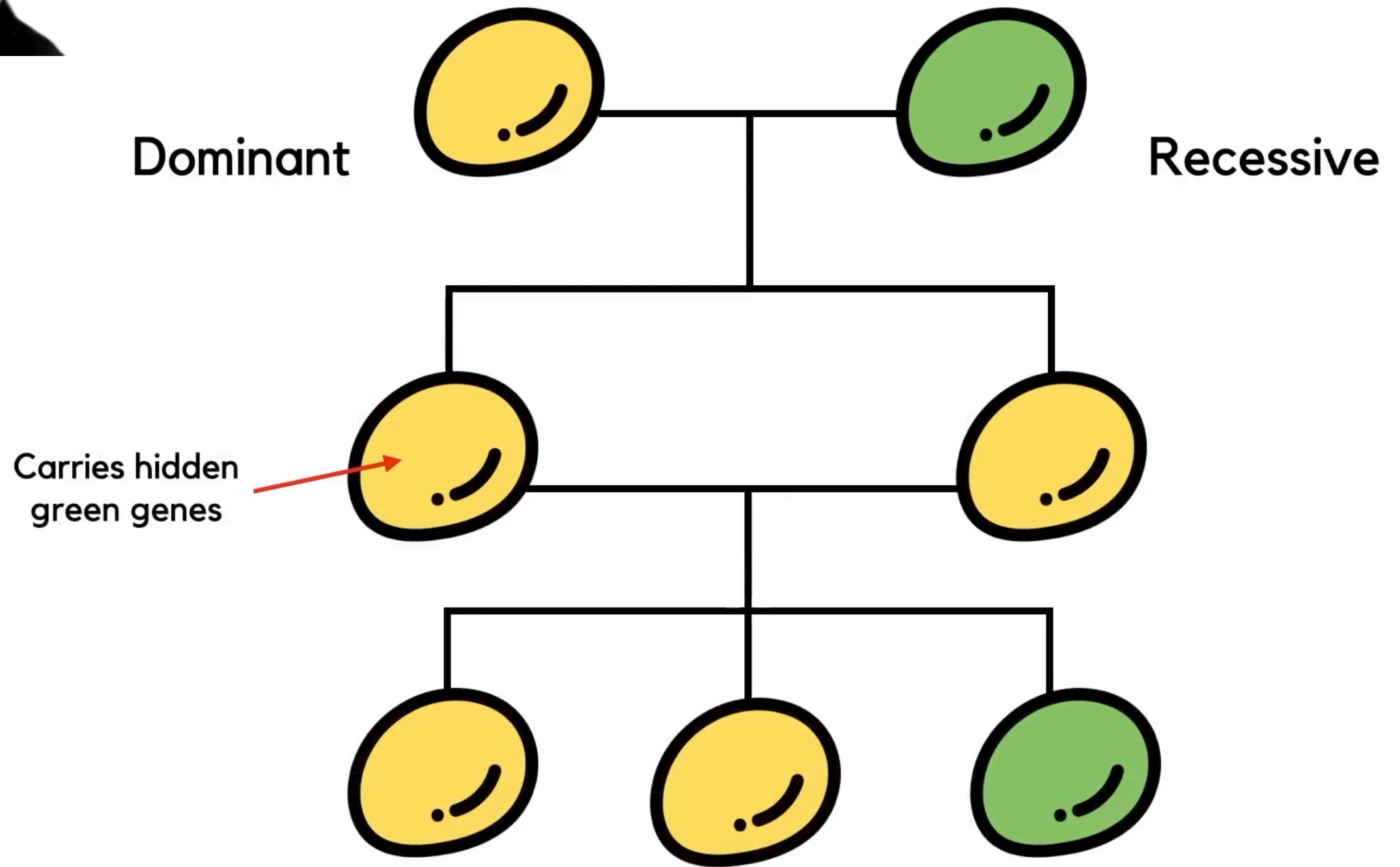
0.3% of sequence might still have errors. Includes X but not Y chromosome. Count excludes mitochondrial DNA.



Does DNA explain everything?


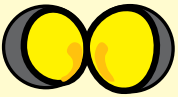
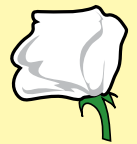






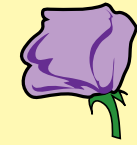


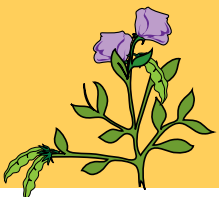



Mendel's model (1865)

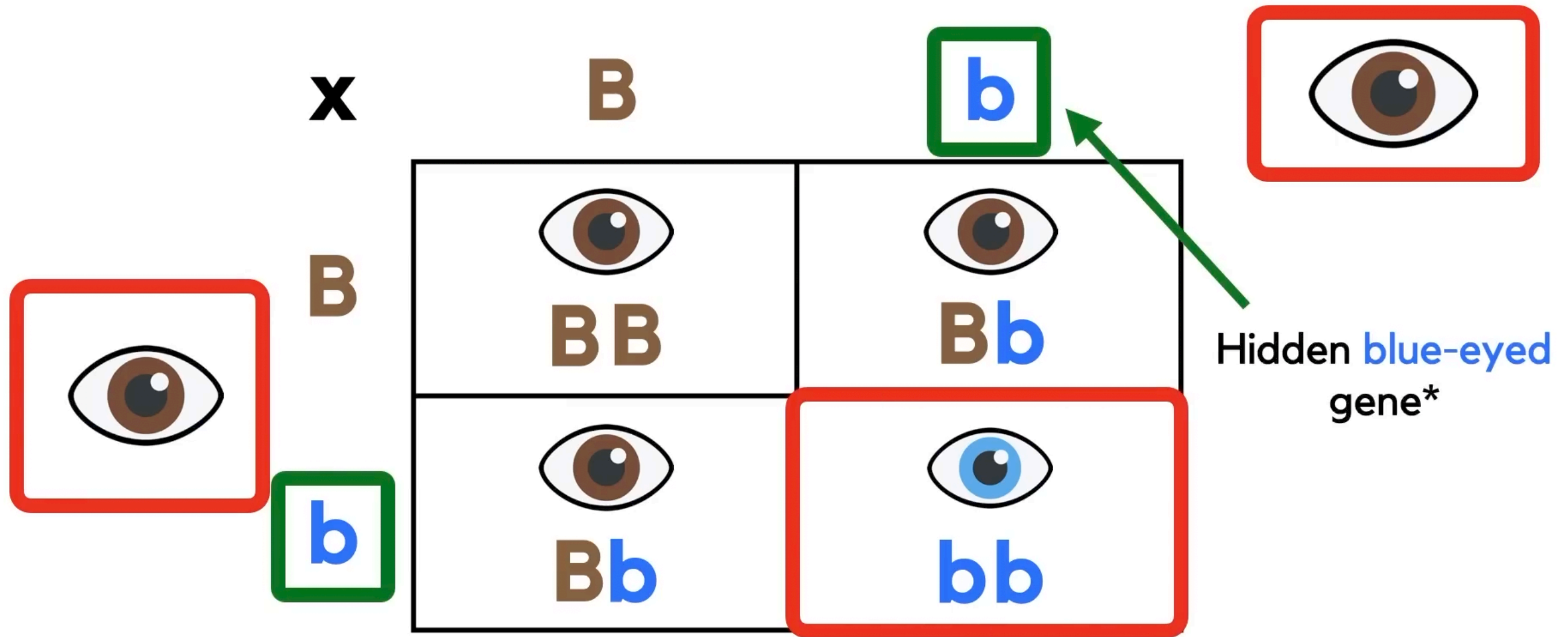


Mendel's model

- Every organism has two alleles for each trait, one received from the father and the other from the mother.
- Of the two alleles for a trait, one may completely express itself (dominant allele), while the other may show no observable effect (recessive allele).
- The two alleles for a trait may be identical or different.
- The two alleles for a trait separate from each other during gamete formation, and each gamete receives only one of them. During the formation of the zygote, one allele from the male gamete and one from the female gamete combine.

Seed		Flower	Pod		Stem	
Form	Cotyledons	Color	Form	Color	Place	Size
						
Grey & Round	Yellow	White	Full	Yellow	Axial pods, Flowers along	Long (6-7ft)
						
White & Wrinkled	Green	Violet	Constricted	Green	Terminal pods, Flowers top	Short ($\frac{3}{4}$ 1ft)
1	2	3	4	5	6	7

Punnett squares



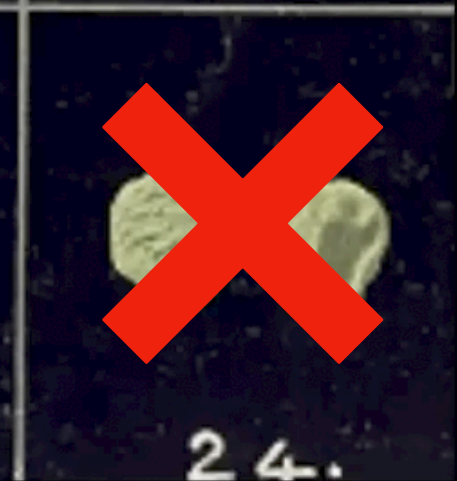
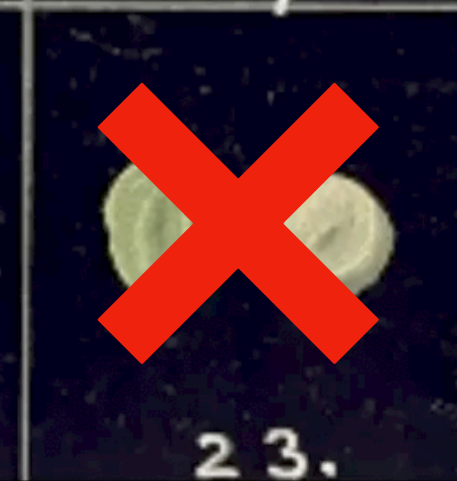
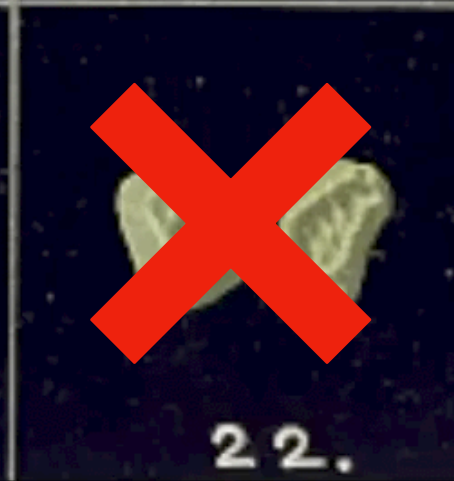
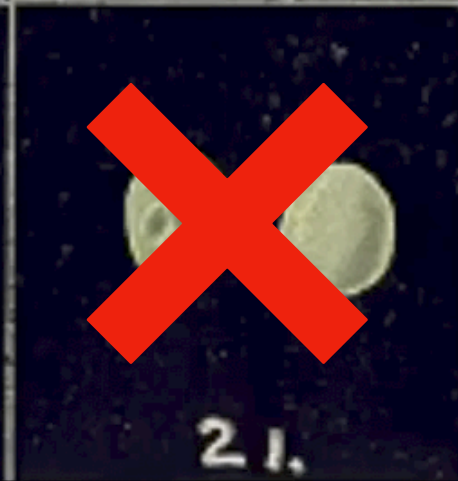
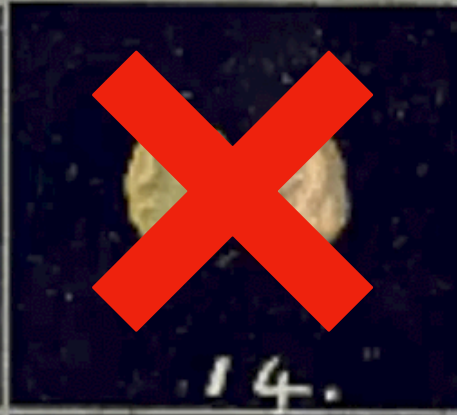
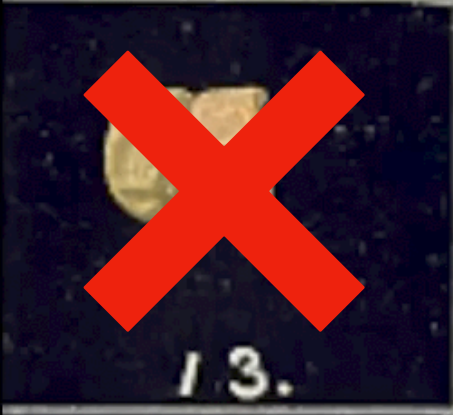
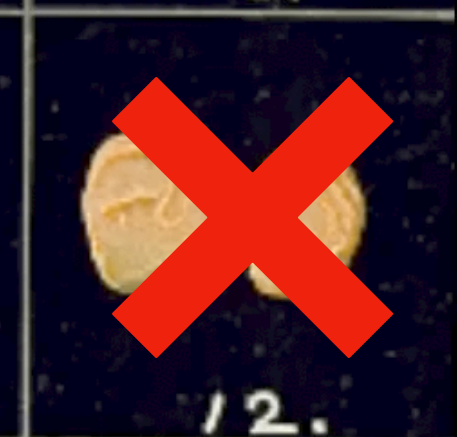
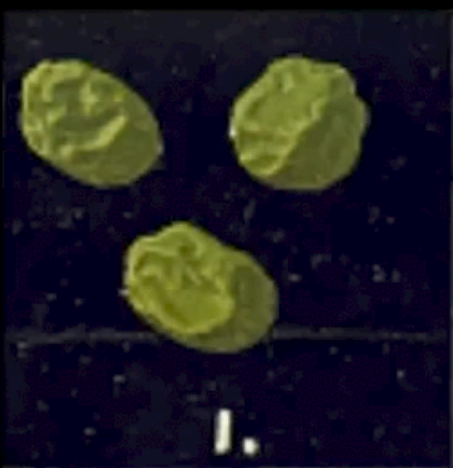
**Technically allele, not gene.*

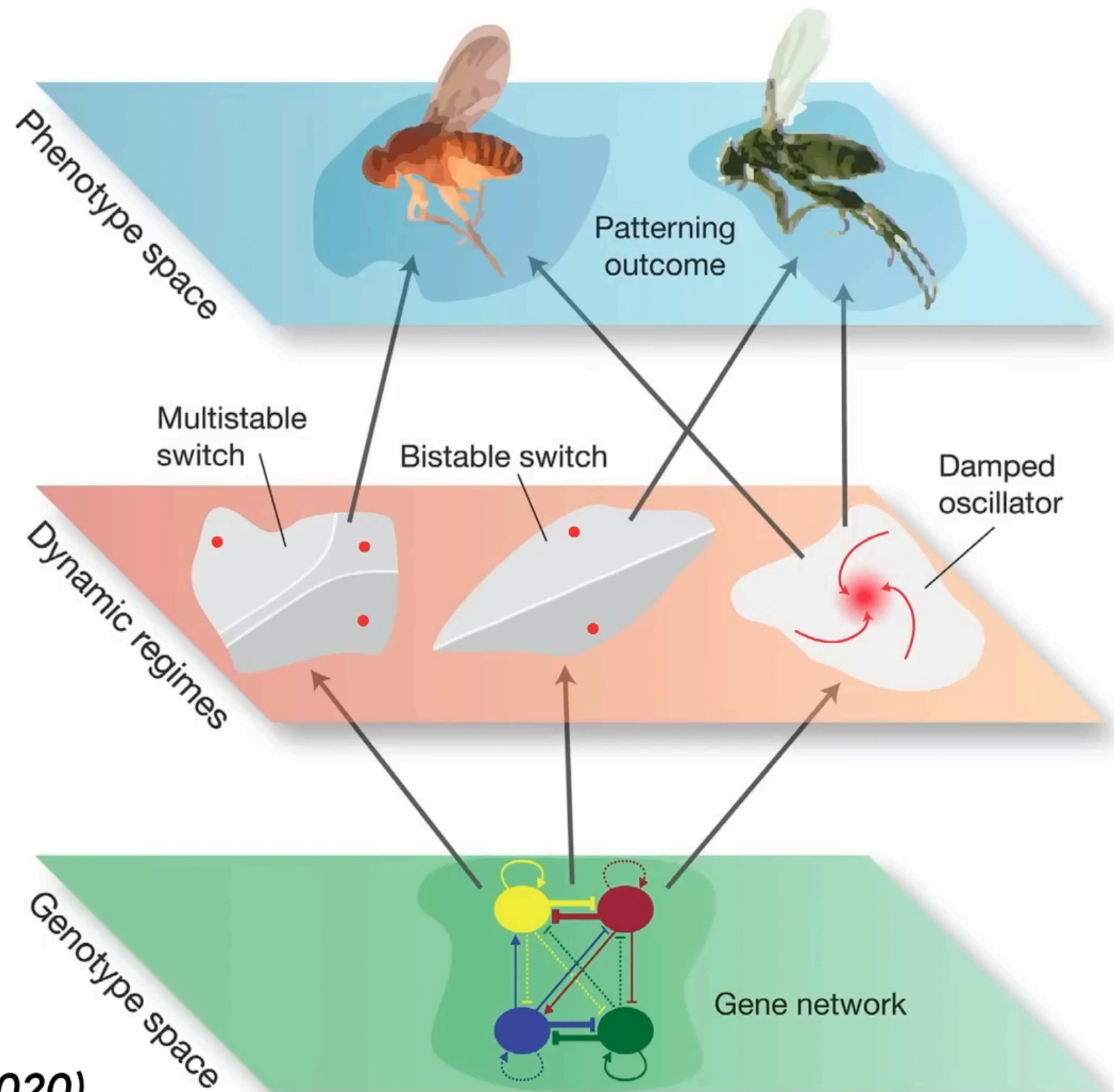
Criticisms of Mendel's theory

Weldon organized his findings into table form and sent them to Pearson in late November 1901. About Mendel, Weldon wrote: “He is either a black liar, or a wonderful man” – “wonderful” in the older, literal sense of “wonder-making.” For the most part, Weldon was inclined to think neither that Mendel was lying nor that he was miraculously lucky, but that he had reported truthfully on what he had observed in the particular varieties he worked with, under the conditions that he observed them in:

[I]f you take all Mendel's figures together, they are wonderfully good approximations to his hypothetically probable results. ~~Remembering his shaven crown, I can't help wondering if they are not too good?~~ I do not see that the results are so good as to be suspicious, so that I can see no alternative to the belief that Mendel's “laws” are absolutely true for his peas, and absolutely false for Laxton's, while those of Tschermak are intermediate But the fear of Mendel is before my eyes. Really one has never seen such perfectly devised observations, lasting over 8 years, give a result so absolutely untrustworthy. It seems to me to

...





DiFrisco & Jaeger (2020)

Nothing works all the time

Environmental conditions change things

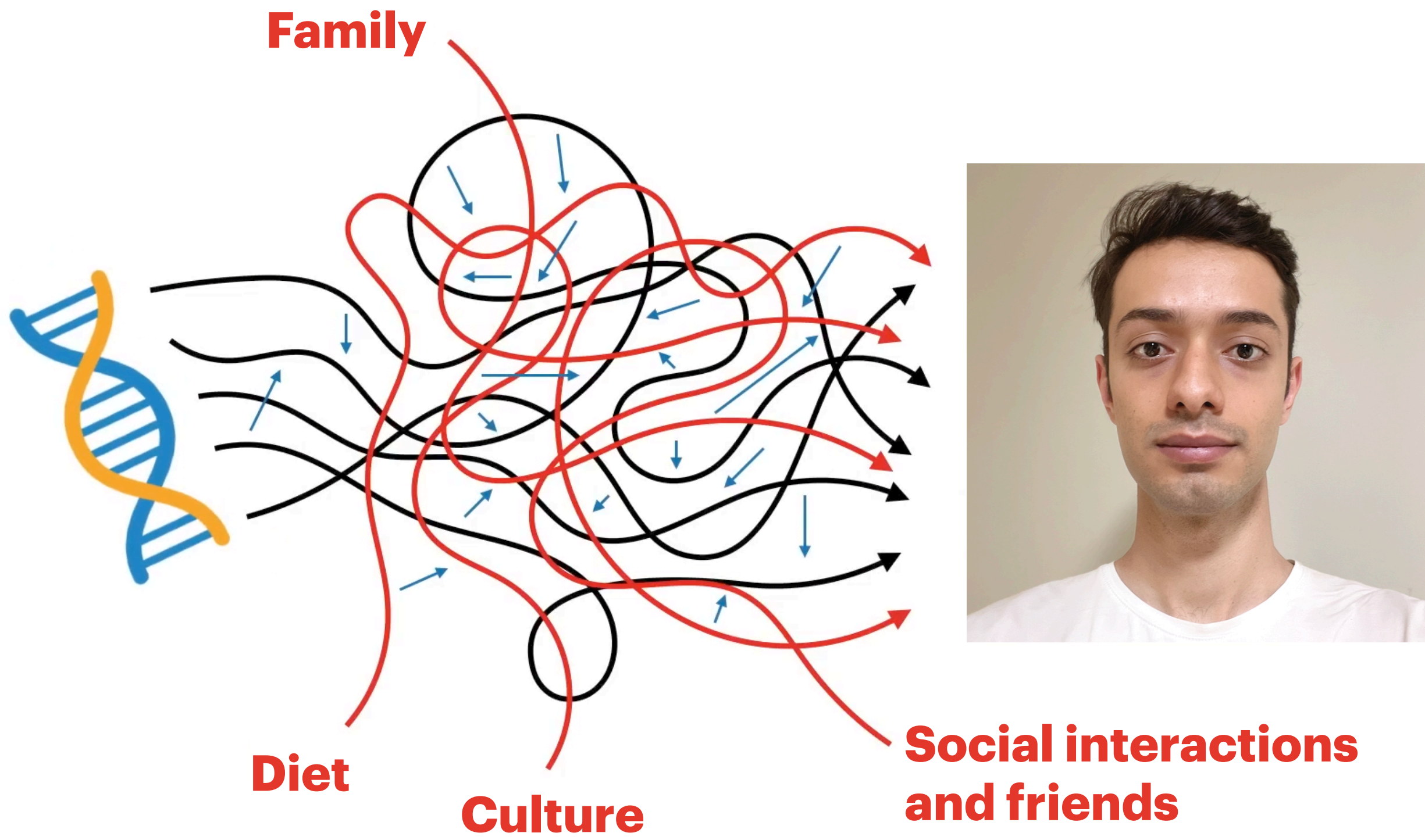
Table 5. Changes in Iris Color in Twin Pairs by Zygosity*

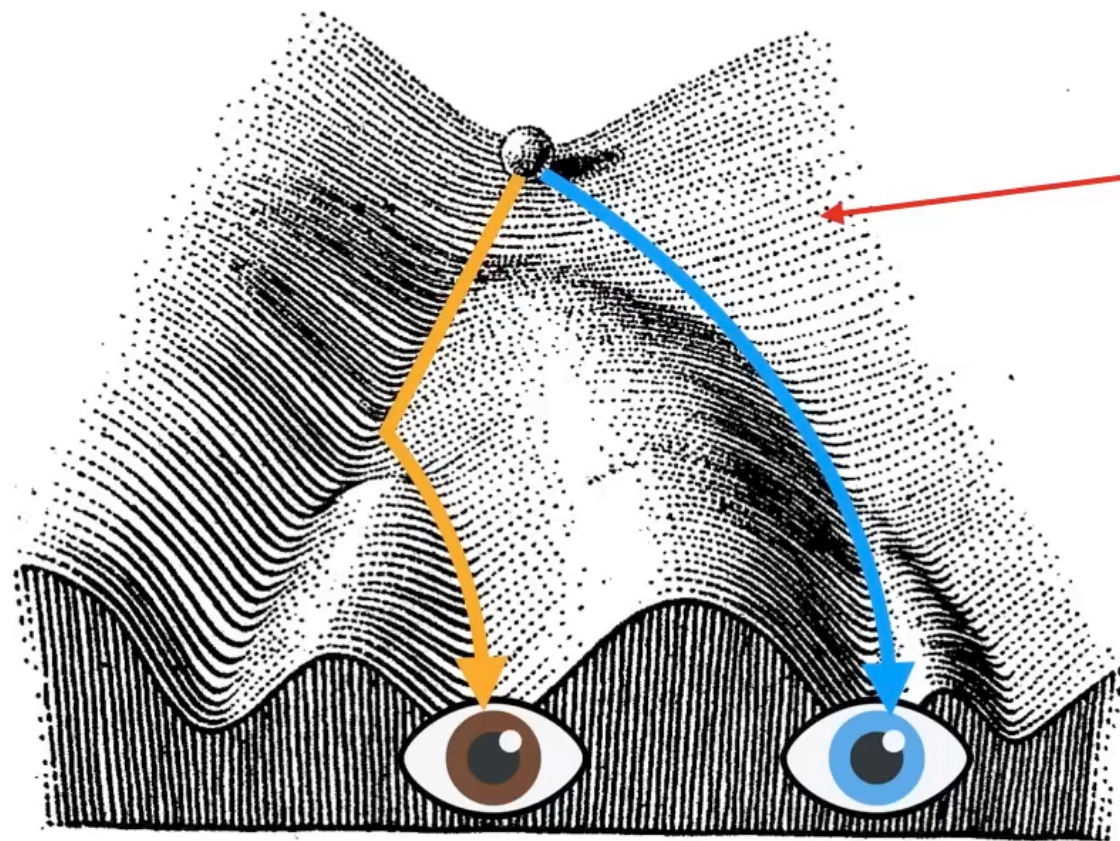
Age	Identical Twins			Fraternal Twins		
	No. of Subjects	Same,† %	Different, %	No. of Subjects	Same, %	Different, %
3 mo-6y‡	101	92.1	7.9	94	62.8	37.2
6-9 y	128	89.1	10.9	123	87.8	12.2
9-15 y‡	87	88.5	11.5	79	74.7	25.3
15 y-adult‡	40	85.0	15.0	27	59.3	40.7
6 y-adult	48	83.3	16.7	37	73.0	27.0

*White subjects only.

†Same indicates both twins changed in the same direction or both had no change; different, the pattern of change differed between twins.

‡ χ^2 Test ($P < .05$).

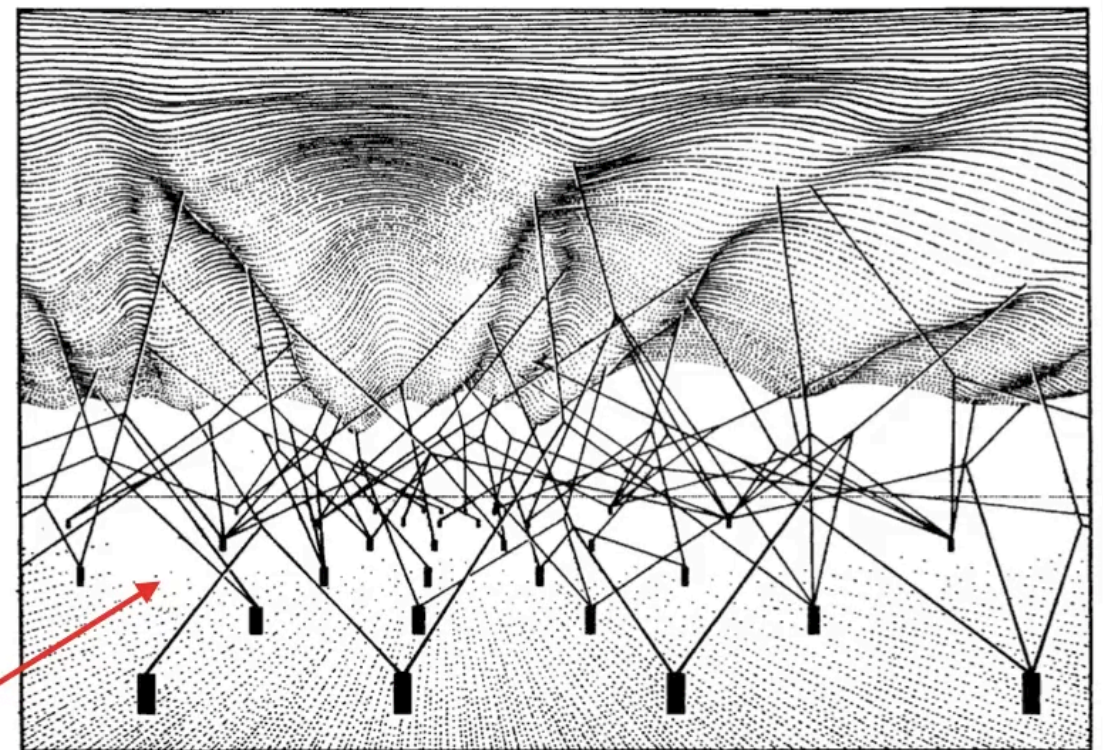




View from above

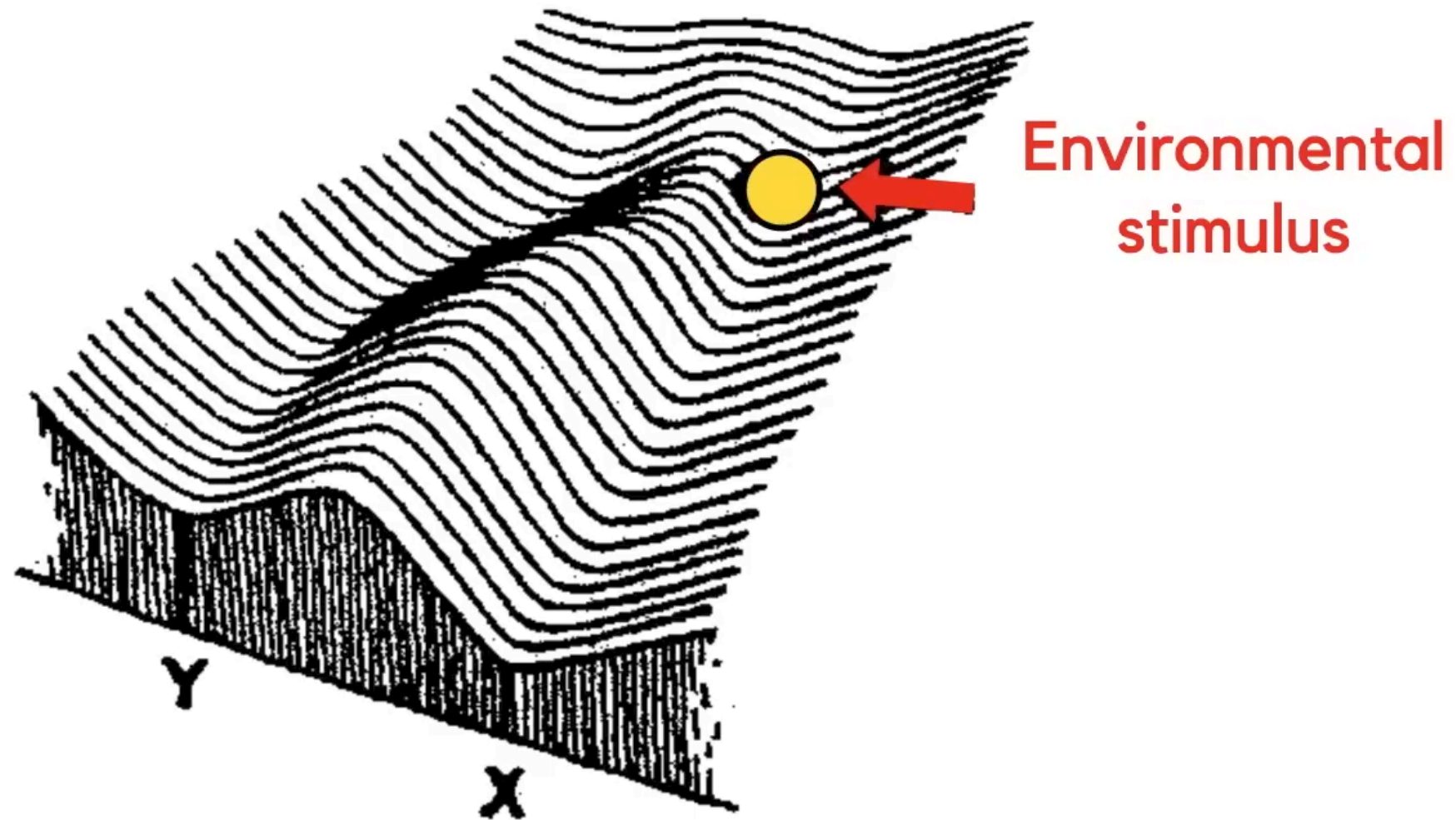
Guy-ropes underneath
the surface

(View from below the surface)



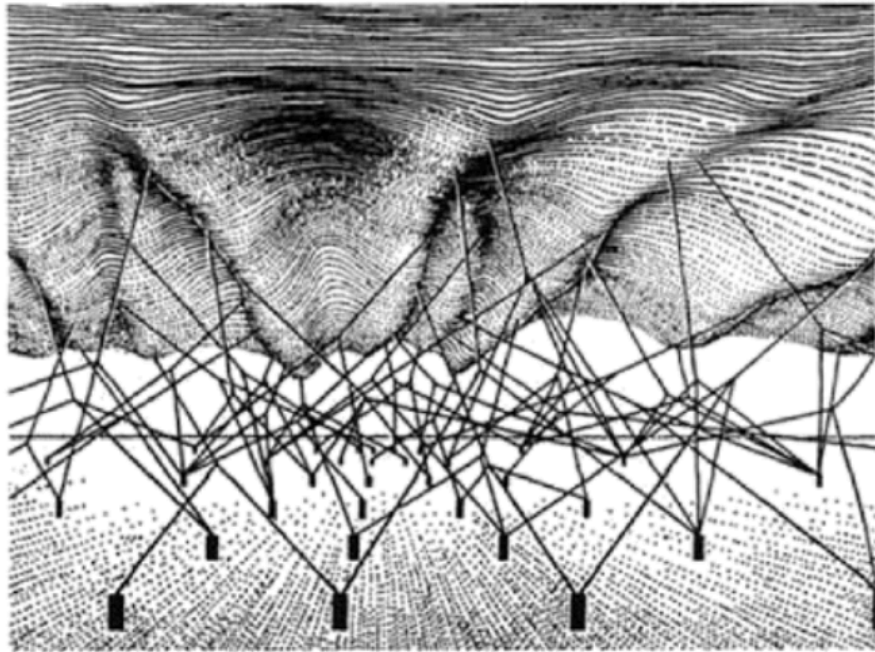
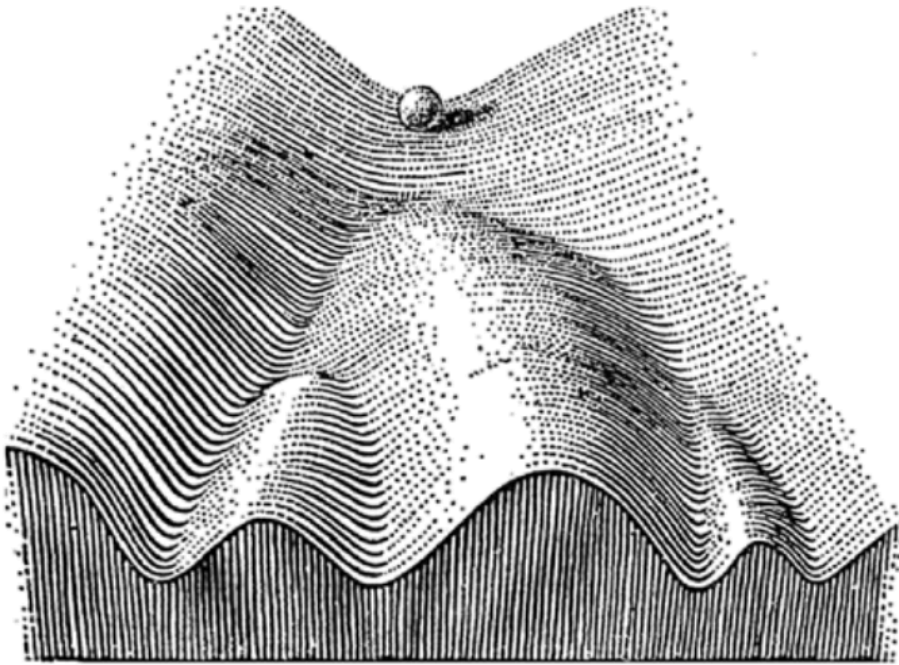
Waddington (1957)

From the book *The Strategy of the Genes*, written by Waddington in 1957

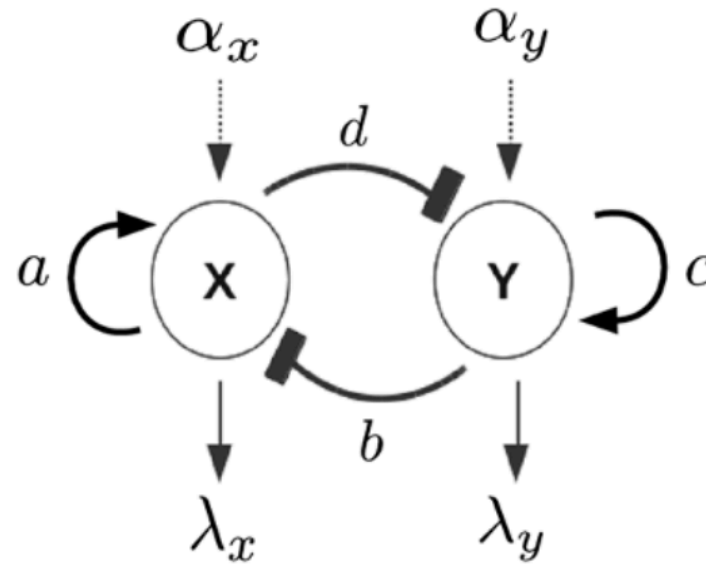


Waddington (1957)

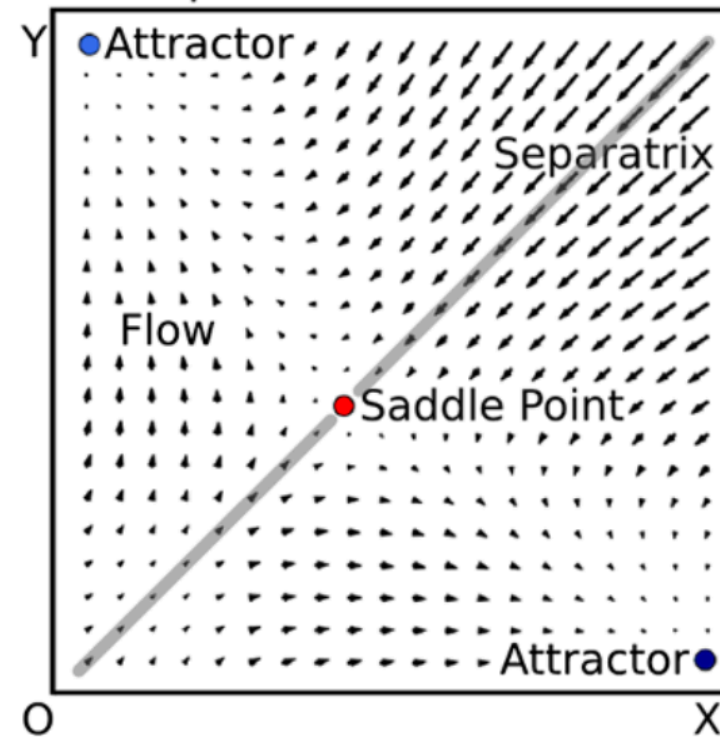
A Waddington's epigenetic landscape



B 1 Toggle switch model



3 Phase portrait

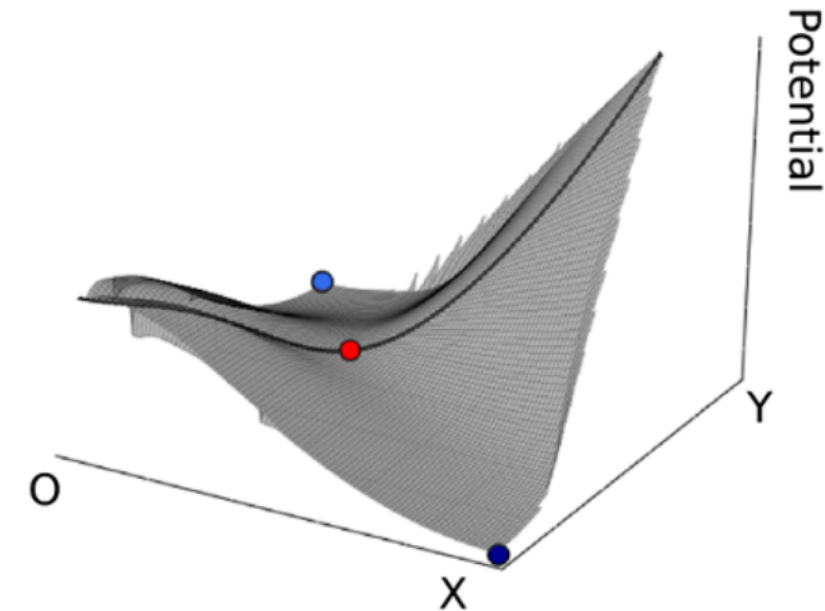


2 Equations

$$\frac{dx}{dt} = \left[\alpha_x + \frac{x^4}{a^4 + x^4} \right] \left[\frac{b^4}{b^4 + y^4} \right] - \lambda_x x$$

$$\frac{dy}{dt} = \left[\alpha_y + \frac{y^4}{c^4 + y^4} \right] \left[\frac{d^4}{d^4 + x^4} \right] - \lambda_y y$$

4 (Quasi) potential landscape




2 Equations

Ratio in the
final protein

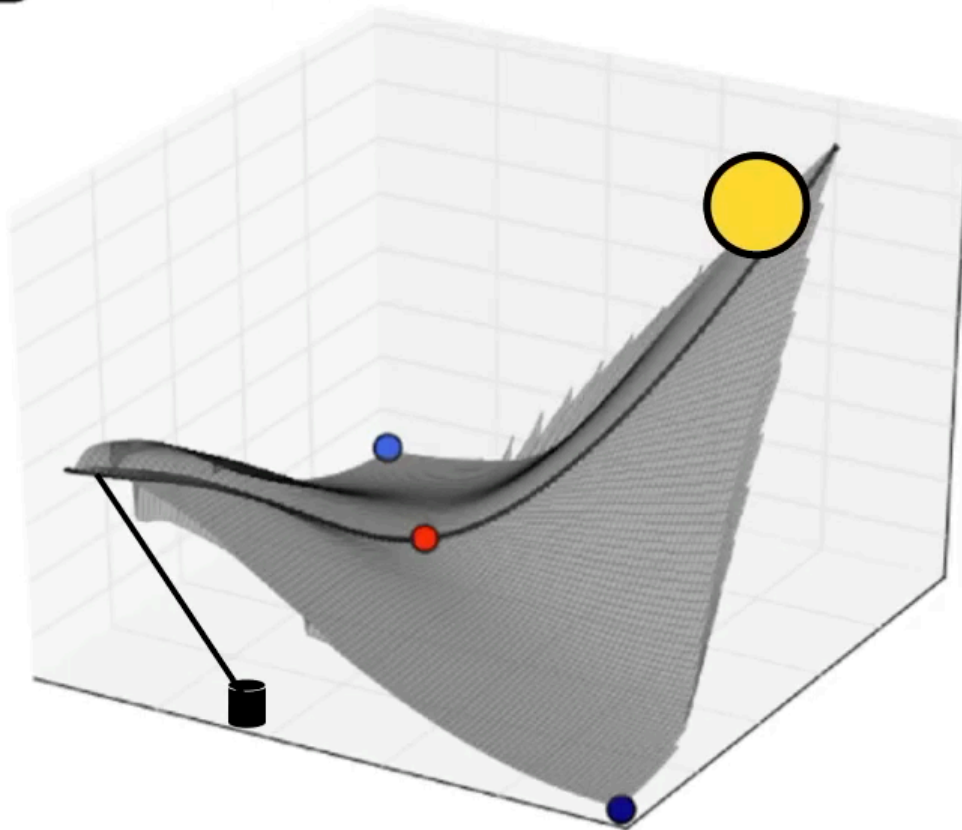
External gene
activation rate

Decay rate

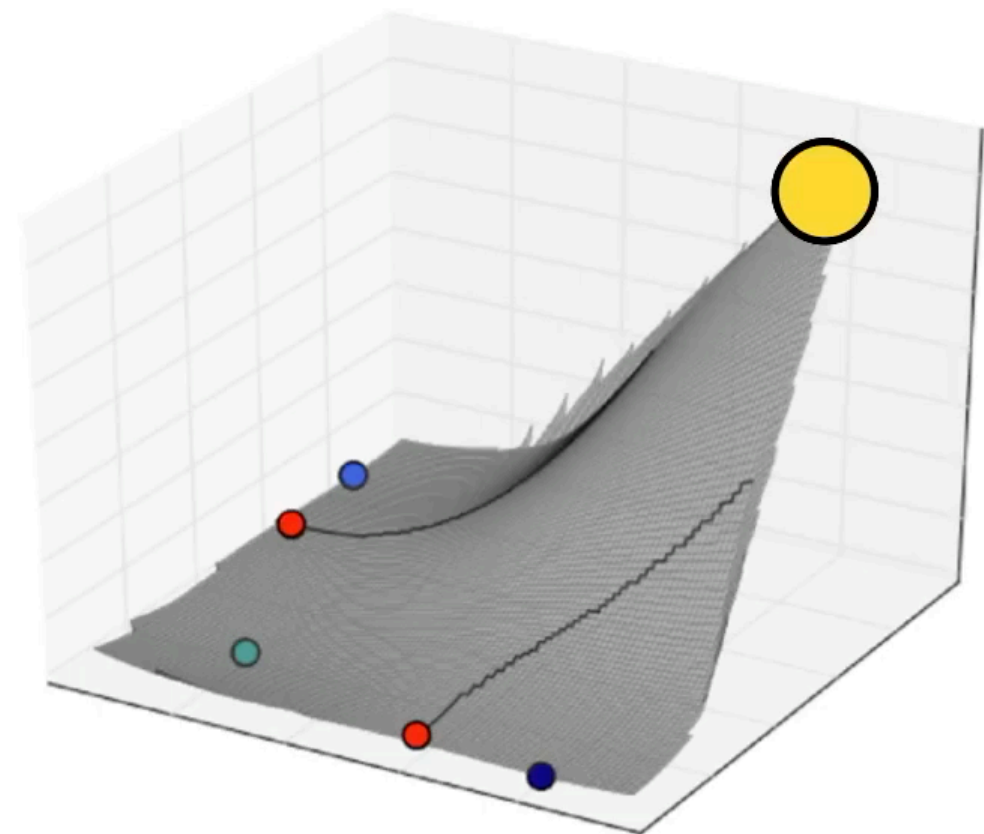

$$\begin{aligned}\frac{dx}{dt} &= \left[\alpha_x + \frac{x^4}{a^4 + x^4} \right] \left[\frac{b^4}{b^4 + y^4} \right] - \lambda_x x \\ \frac{dy}{dt} &= \left[\alpha_y + \frac{y^4}{c^4 + y^4} \right] \left[\frac{d^4}{d^4 + x^4} \right] - \lambda_y y\end{aligned}$$

4 (Quasi) potential landscape

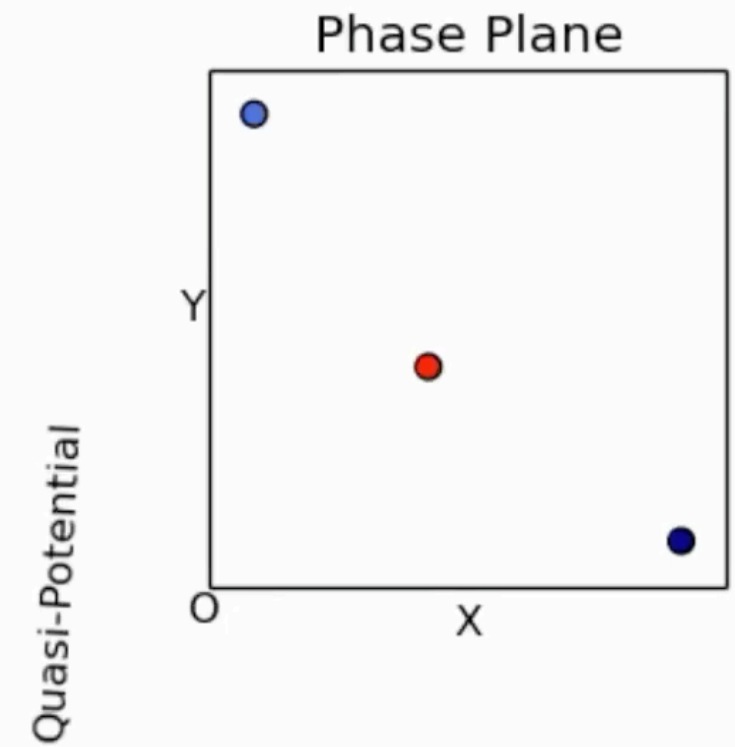
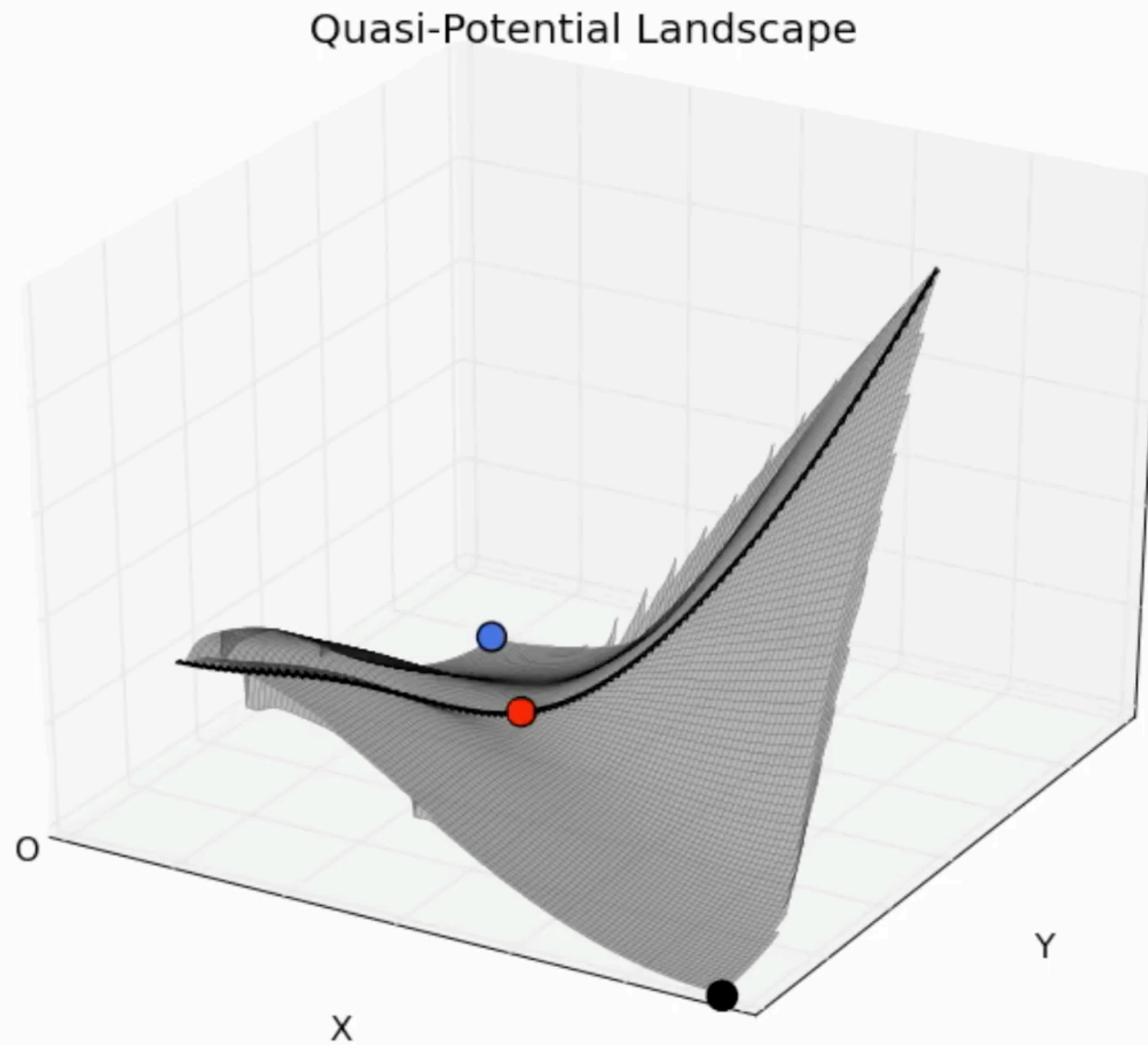
B Bistable regime



C Tristable regime



Verd et al. (2014)



Parameter Values

X autoactivation = 0.28
Y autoactivation = 0.28

X repression on *Y* = 0.4
Y repression on *X* = 0.4

X external activation = 0.25
Y external activation = 0.25

X degradation = 1.25
Y degradation = 1.25

References

- Radick, G. (2023). *Disputed Inheritance: **The Battle Over Mendel and the Future of Biology***. United Kingdom: University of Chicago Press.
- **The Human Genome Project**. Retrieved from <https://www.genome.gov/human-genome-project>.
- **You've Been Lied To About Genetics**. Retrieved from <https://www.subanima.org/mendel>.
- Kampourakis, K. (2021). **Should We Give Peas a Chance? An Argument for a Mendel-Free Biology Curriculum**. In: Haskel-Ittah, M., Yarden, A. (eds) *Genetics Education. Contributions from Biology Education Research*. Springer, Cham.
- Jamieson, A., Radick, G. (2013). **Putting Mendel in His Place: How Curriculum Reform in Genetics and Counterfactual History of Science Can Work Together**. In: Kampourakis, K. (eds) *The Philosophy of Biology. History, Philosophy and Theory of the Life Sciences*, vol 1. Springer, Dordrecht.
- Kampourakis, K. (2021). The Origin and Evolution of the Gene Concept. In: **Understanding Genes. Understanding Life**. Cambridge University Press; 2021:31-64.
- Verd, B., Crombach, A. & Jaeger, J. **Classification of transient behaviours in a time-dependent toggle switch model**. BMC Syst Biol 8, 43 (2014)
- Waddington, C. H. (1957). **The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology**. Kiribati: Allen & Unwin.

Thank you!

