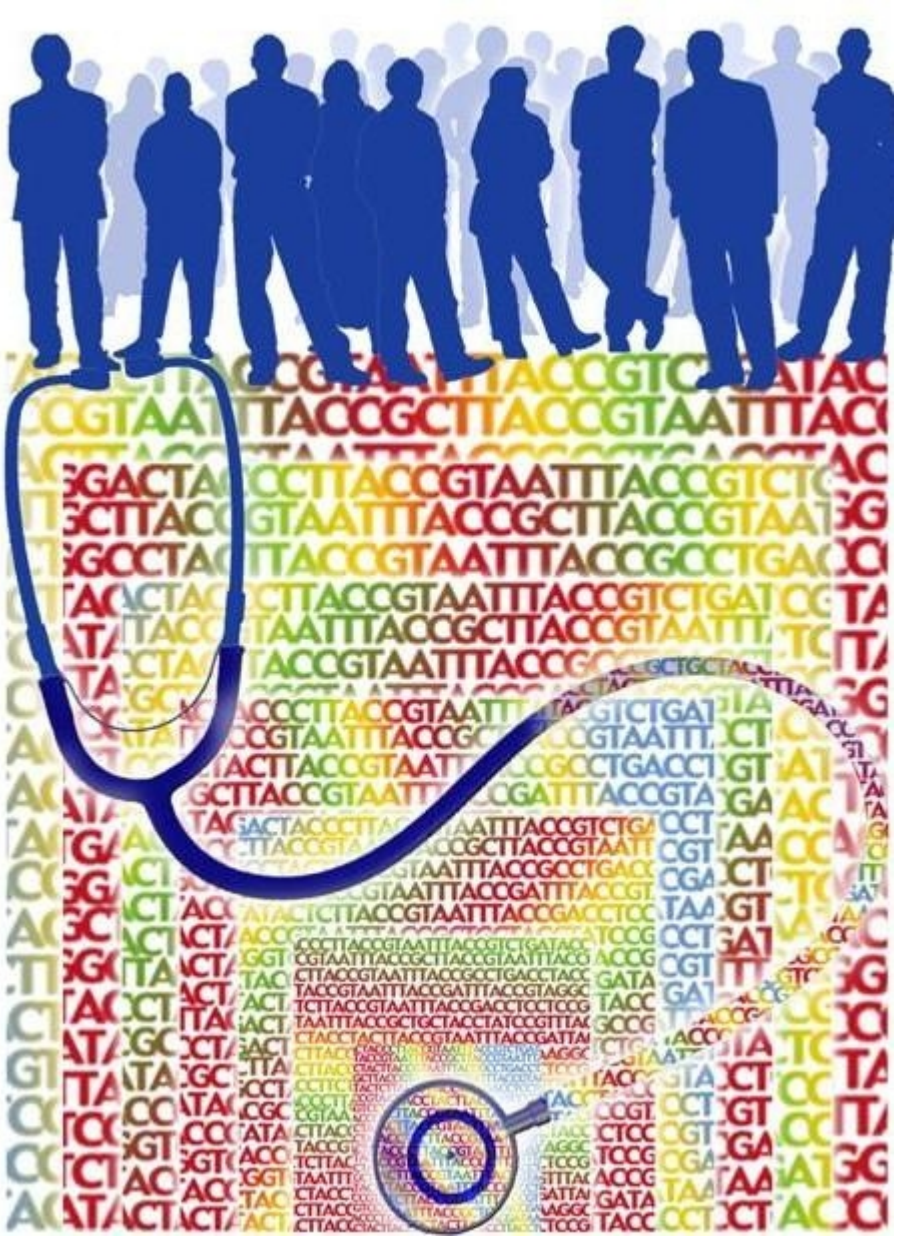
Noise, Cryptic Variation, Robustness and Quantitative Genetics

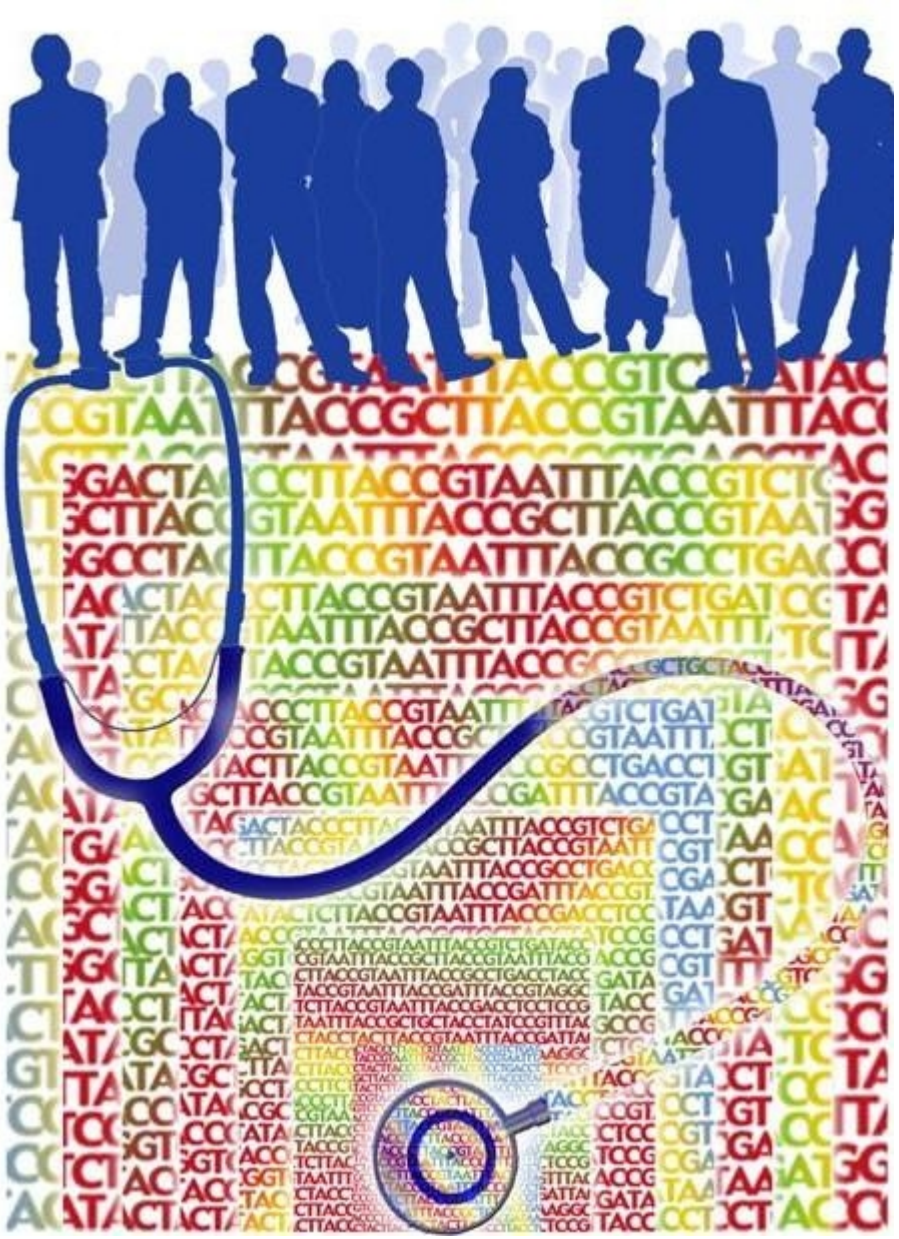
Virginie Courtier-Orgogozo
Institut Jacques Monod, Paris

Human genetic diversity



- Genome size: 2.9 Gb
- Gene number: 25 000
- (1% of coding sequences)
-
- In one individual:
  ~70 new mutations compared to his parents
- ~20 lethal mutations (heterozygous)
-
-
-
-
- *Genetic difference between two humans?*
-
- *Genetic differences between humans and chimps?*
-
-

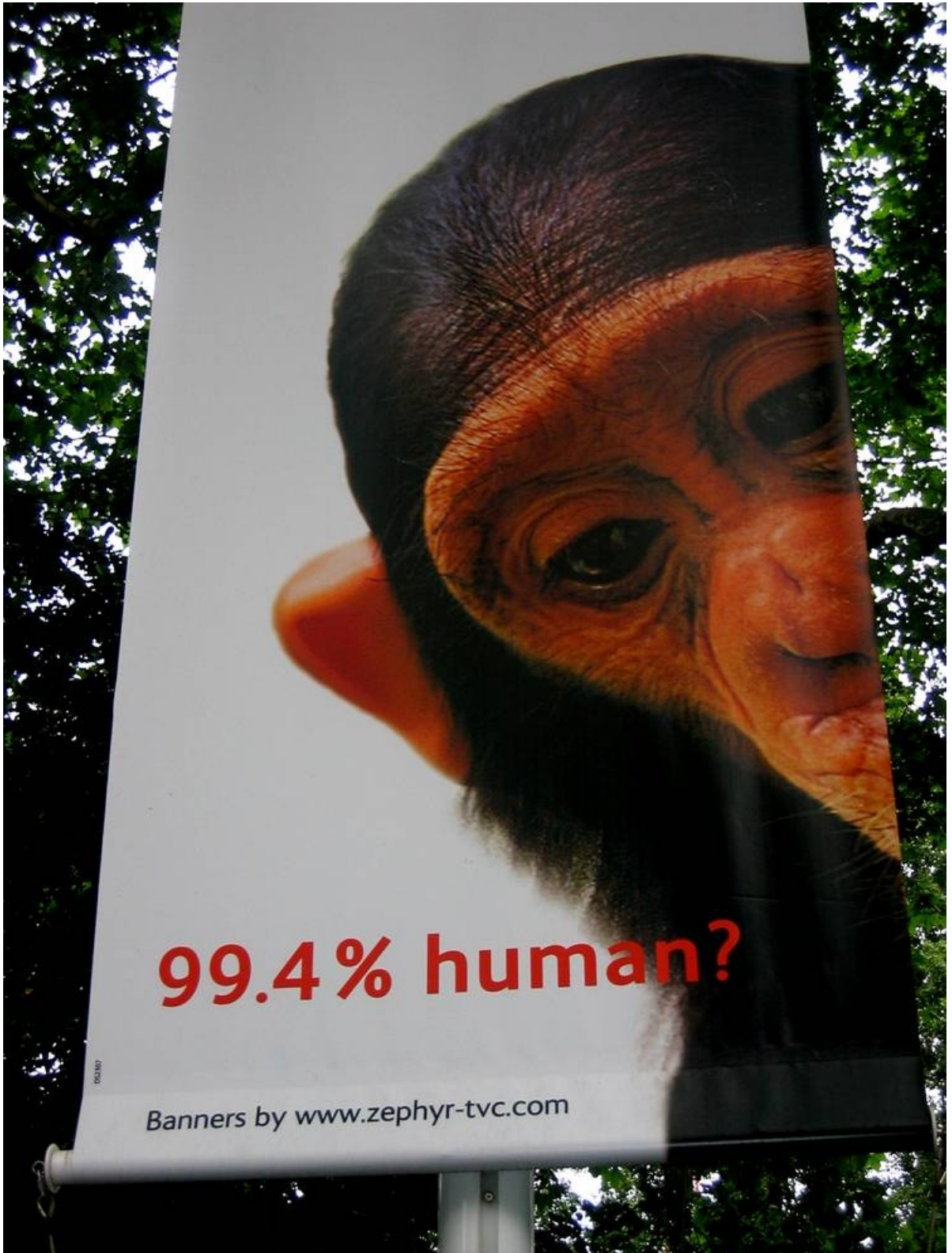
Human genetic diversity



- Genome size: 2.9 Gb
- Gene number: 25 000
- (1% of coding sequences)
-
- In one individual:
  ~70 new mutations compared to his parents
- ~20 lethal mutations (heterozygous)
-
-
-
-
- *Genetic difference between two humans?*

  ~0.1%
-
- *Genetic differences between humans and chimps?*
-

  ~4% (<1% for coding sequences)
-


99.4% human?
Banners by www.zephyr-tvc.com

From laboratory to “real-life” data

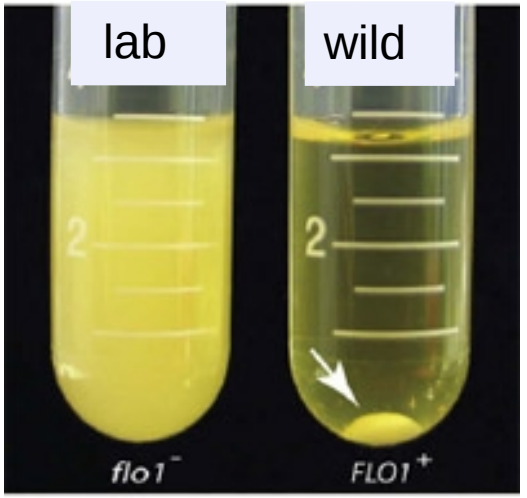
Knock out

Natural variation

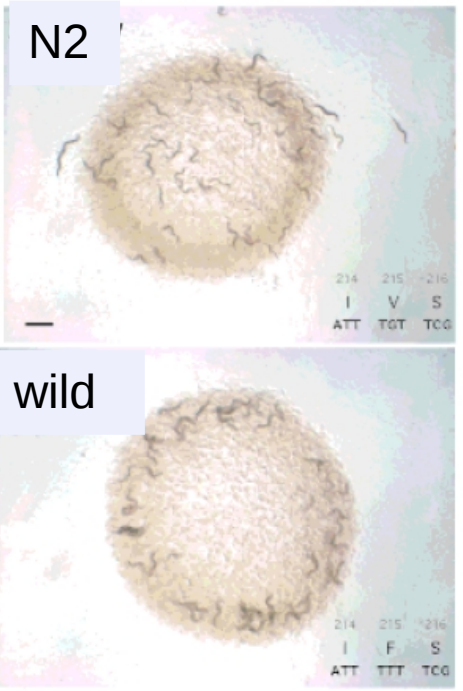
Domestication of laboratory strains

Arabidopsis thaliana

Saccharomyces cerevisiae



Caenorhabditis elegans



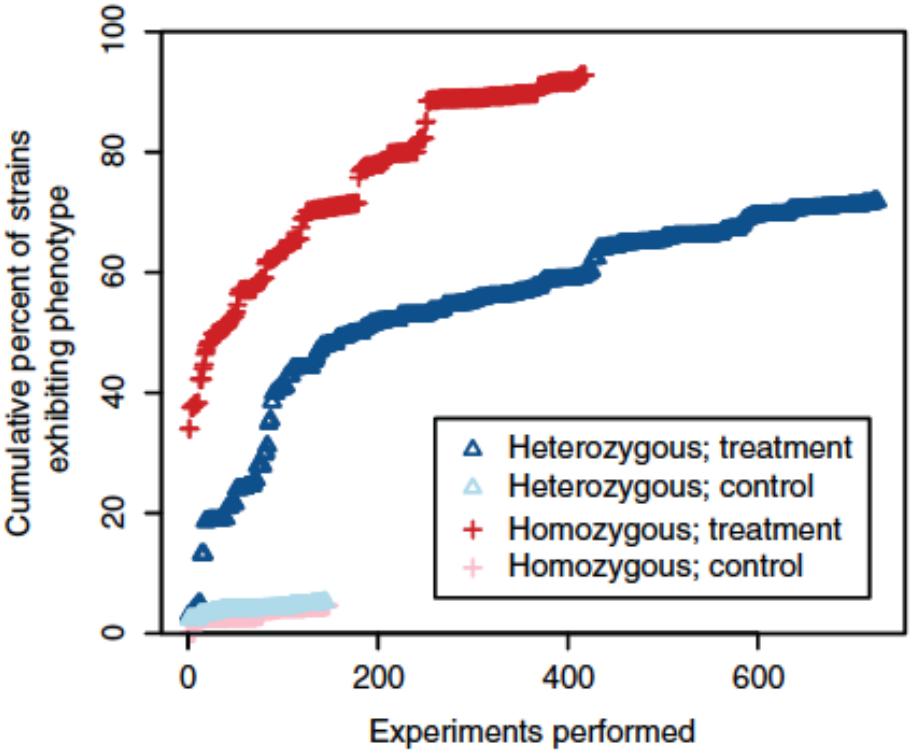
Domestication of laboratory strains
results in extreme phenotypic values
for many traits:
artificial selection and pleiotropy

Choice of laboratory environment

ca. 10-20 years ago: surprise at not finding phenotypes in gene knockouts

The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes
Maureen E. Hillenmeyer, *et al.*
Science **320**, 362 (2008);

1144 growth environments
for *S. cerevisiae*



Laboratory mutations

- Not in nature
- Extreme effects
- Would likely be lost under selection
- Must be induced

- Interrogates (nearly) all regions
- Readily cloned
- Strong effects

QTL

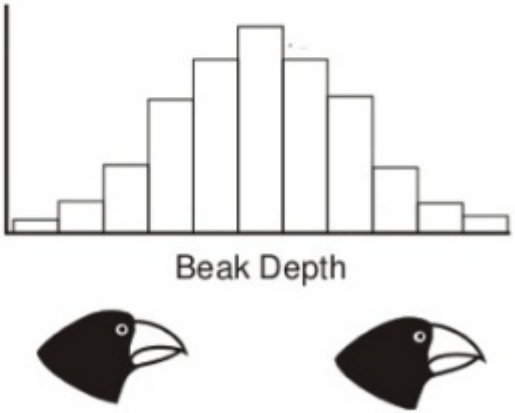
- Representative of nature
- Variants with small effects
- Sustained under selection
- Readily available

- Interrogates only variable regions
- Difficult to map
- Small effects

Is natural variation discrete or continuous? Biometricians against Mendelians

Karl Pearson
Walter Weldon

- Continuous variation
- Pre-existing variation
- Gradual change



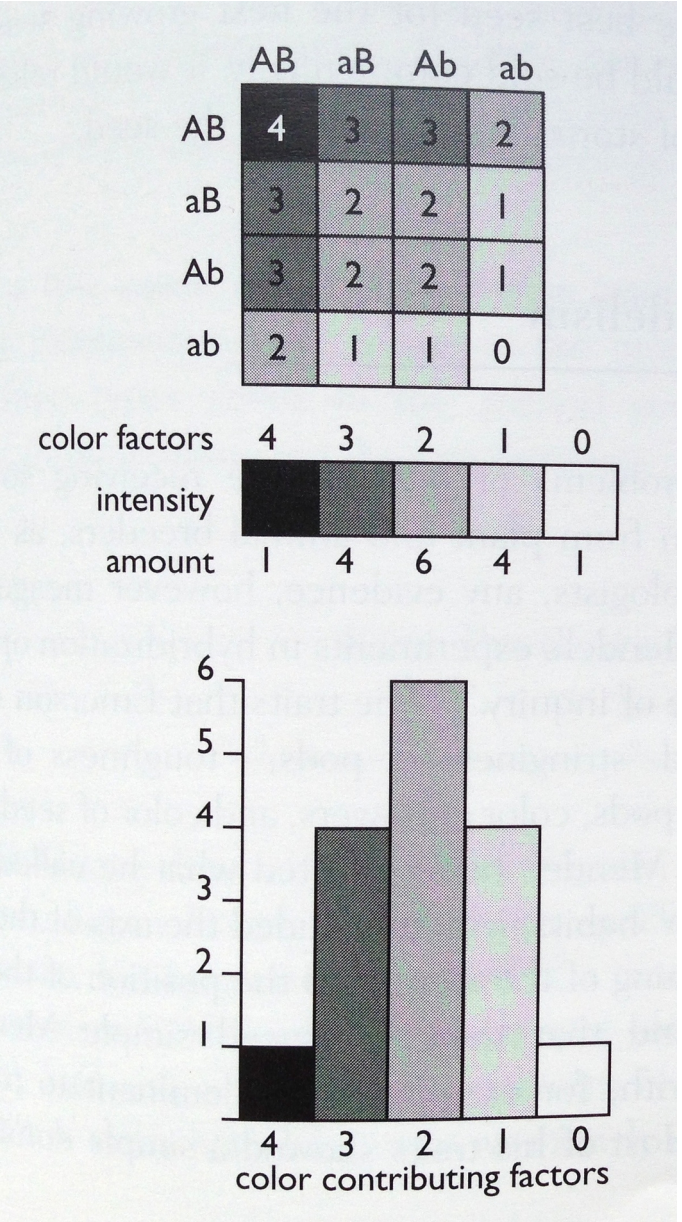
William Bateson
Hugo de Vries

- Discontinuous variation with discrete heredity factors
- Mutation
- Evolutionary jumps

Reconciliation of Mendelian genetics and heredity of quantitative characters

Nilsson-Ehle (cereals)
East (corn)

example with only two factors with additive action:

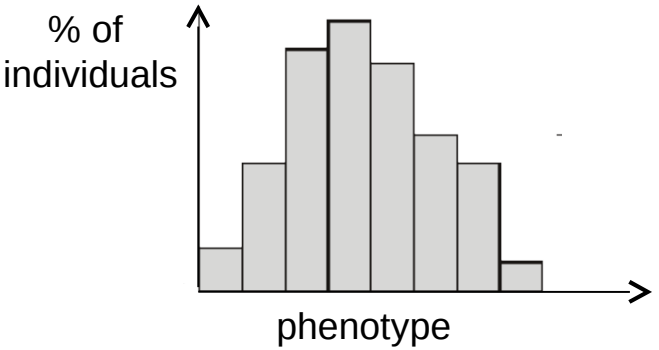


Carlson 2004 book *Mendel's Legacy*

Quantitative genetics

Quantitative genetics

- If to each genotype corresponds a distribution of phenotypes
  = variable expressivity
  the character itself is quantitative

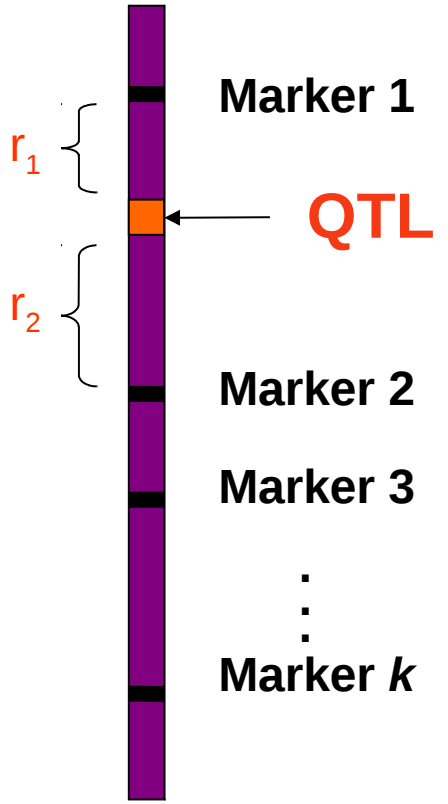


and/or

- If the variation of many genes is involved in the phenotypic difference between two strains/individuals
  the segregation of the character is quantitative

Quantitative Trait Loci (QTL) mapping

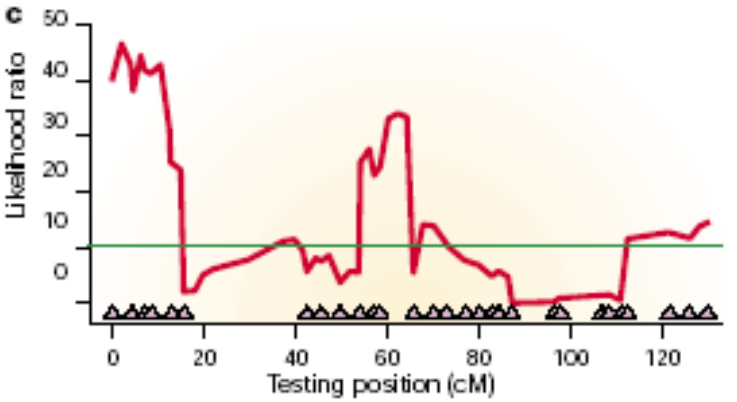
- QTL are specific genetic loci that affect quantitative traits.
- QTL can be detected by markers that are linked with it.



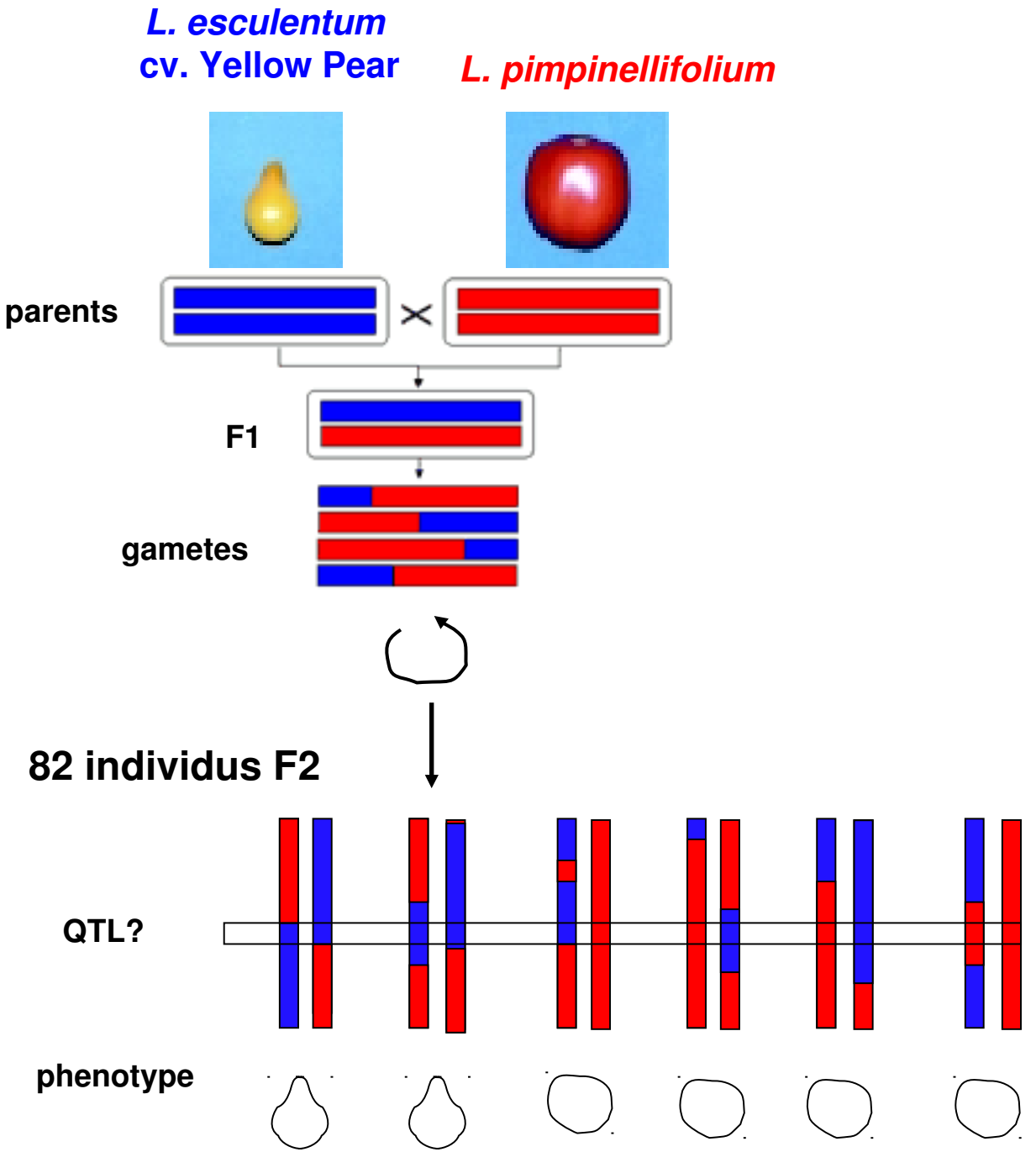
Two goals:

Identify the location of the QTL

Estimate the genetic effects of the QTL

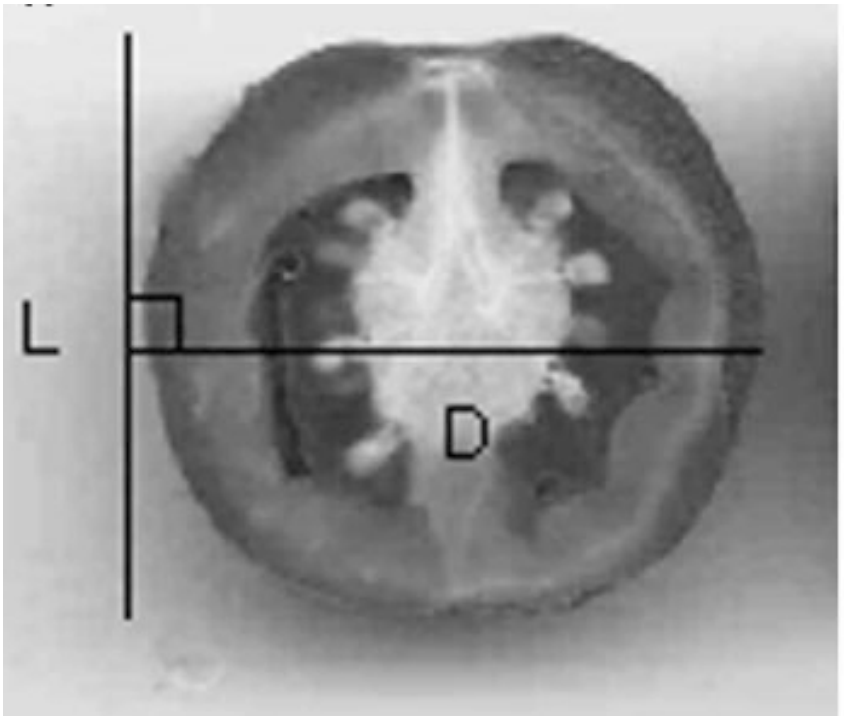


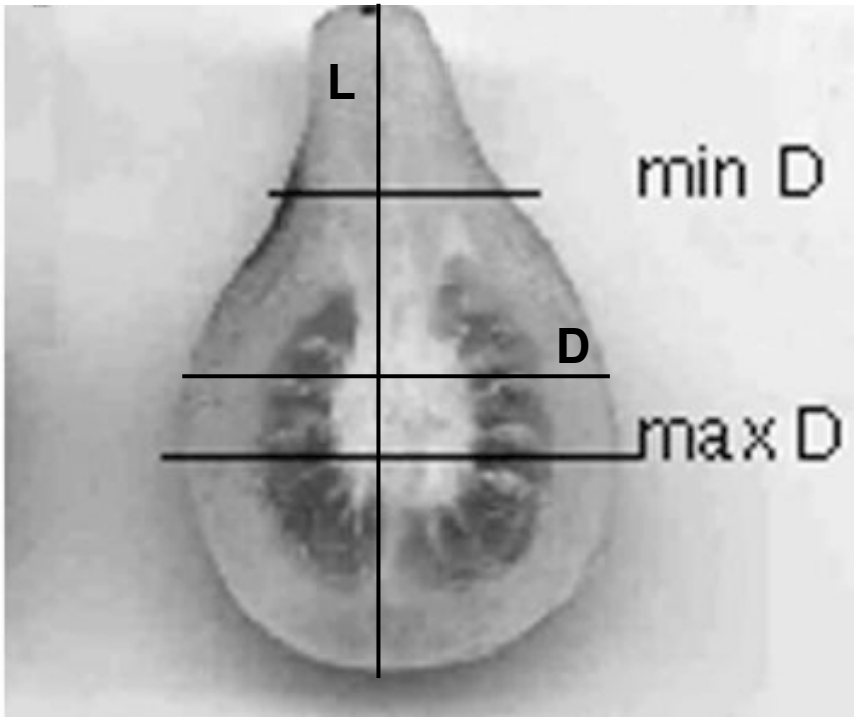
QTL mapping


L. esculentum
cv. Yellow Pear
L. pimpinellifolium
parents
F1
gametes
82 individus F2
QTL?
phenotype

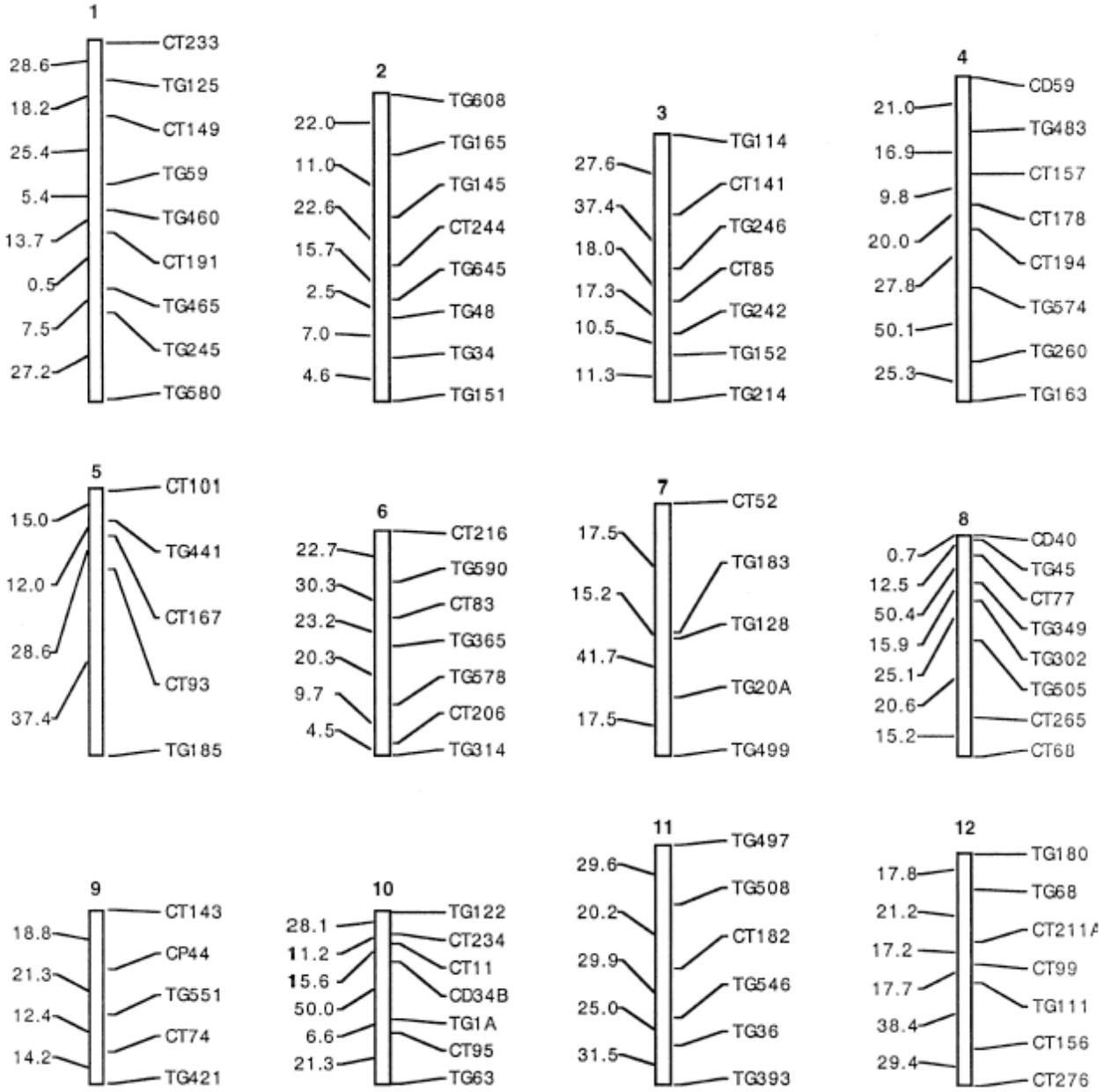
Quantitative measure of the phenotype

Measure of 2 indexes L/D and Dmin/Dmax for 10 fruits per plant
L/D : L= length, D = diameter at equator
Dmin/Dmax





82 molecular markers on the 12 tomato chromosomes



Two main files

Markers file

```
-start
-Chromosome 1
CF5475        0.4
CF5573        24.7
CT7895        41.0
CT8903        59.0
CF5613        67.7
CT7892        76.0
CT890         89.0
CT233         39.0
Telomere      50.0
-Chromosome 2
CF5671        0
CF5675        10.4
CF5673        34.7
CT789         41.0
CT890         89.0
CT567         115.0
Telomere      130.0
…
```

Genotypes and phenotype(s) file

```
-start individuals markers
Ind_1 0 0 1 1 0 0 0 0 0 1 2 2 2 2
Ind_2 0 0 0 1 0 1 0 0 1 1 1 1 0 0
Ind_3 2 2 2 2 2 1 0 1 1 1 1 0 0 0
Ind_4 0 1 0 0 0 0 1 1 1 2 2 1 1 1
Ind_5 0 1 0 0 0 0 1 1 1 1 2 2 2 2
Ind_6 1 1 1 1 1 1 1 1 1 0 0 0 0 0
Ind_7 1 1 1 1 1 1 1 0 1 n n 1 1 1
Ind_8 2 2 2 1 1 1 1 0 1 1 1 1 1 0
Ind_9 1 1 1 1 1 1 1 0 0 1 1 1 1 1
Ind_1 0 2 2 1 1 1 1 1 0 0 0 1 1 2
-stop individuals markers

-start individuals traits 1 LoverD named
Ind_1    5.5
Ind_2    3.0
Ind_3    4.0
Ind_4    7.0
Ind_5    6.5
Ind_6    5.0
Ind_7    3.5
Ind_8    6.0
```

Simple linear regression for each marker

L/D of individual i = a + b.xi + ε

xi = 0 if Le/Le, = 1 if Le/Lp, = 2 if Lp/Lp

a,b = best fit parameters (least square regression)

ε assumed to have a normal distribution

Test Ho: b = 0 *versus* H1: b = estimated b

Likelihood ratio test statistic

$$D = -2(\ln(\text{likelihood for null model}) - \ln(\text{likelihood for alternative model}))$$
$$= -2\ln\left(\frac{\text{likelihood for null model}}{\text{likelihood for alternative model}}\right).$$

The probability distribution of the test statistic can be approximated by a chi-square distribution with (df1 − df2) degrees of freedom, where df1 and df2 are the degrees of freedom of models 1 and 2 respectively

Interval mapping

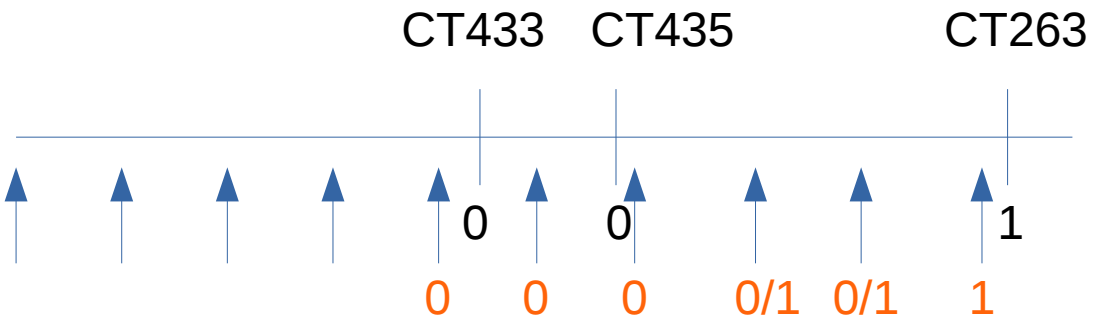
L/D of individual i = a + b.xi + e

xi = indicator variable specifying the probabilities of an individual being in different genotypes for the tested position, constructed by flanking makers

xi = 0 if Le/Le, = 1 if Le/Lp, = 2 if Lp/Lp

a,b = best fit parameters (maximum likelihood)

Test Ho: b=0 versus H1: b=estimated b



Interval mapping

L/D of individual i = a + b.xi + e

xi = indicator variable specifying the probabilities of an individual being in different genotypes for the tested position, constructed by flanking makers

xi = 0 if Le/Le, = 1 if Le/Lp, = 2 if Lp/Lp

a,b = best fit parameters (maximum likelihood)

Test Ho: b=0 versus H1: b=estimated b

Composite Interval mapping

L/D of individual i = a + b.xi + c.xi + e

xi = indicator variable specifying the probabilities of an individual being in different genotypes for the tested position, constructed by flanking makers

xi = 0 if Le/Le, = 1 if Le/Lp, = 2 if Lp/Lp

yi = 0 if Le/Le, = 1 if Le/Lp, = 2 if Lp/Lp at marker y

LOD score

L/D of individual i = a + b.xi + e

Test Ho: b = 0 *versus* H1: b = estimated b

Lo = pr (data | no QTL) – phenotypes assumed to follow a normal distribution
L1 = pr (data | QTL at tested position)

$$LOD = -\log \frac{L_0}{L_1}$$

The likelihood ratio test statistic (LR) is

$$LR = -2 \ln \frac{L_0}{L_1} = -2 \ln 10^{-LOD} = 2(\ln 10) LOD = 4.605 LOD$$

and thus

$$LOD = -\log \exp(-\frac{LR}{2}) = \frac{1}{2}(\log e) LR = 0.217 LR$$

Interval mapping

L/D of individual i = a + b.xi + e

xi = indicator variable specifying the probabilities of an individual being in different genotypes for the tested position, constructed by flanking makers

xi = 0 if Le/Le, = 1 if Le/Lp, = 2 if Lp/Lp

a,b = best fit parameters (maximum likelihood)

Test Ho: b=0 versus H1: b=estimated b

Composite Interval mapping

L/D of individual i = a + b.xi + c.xi + e

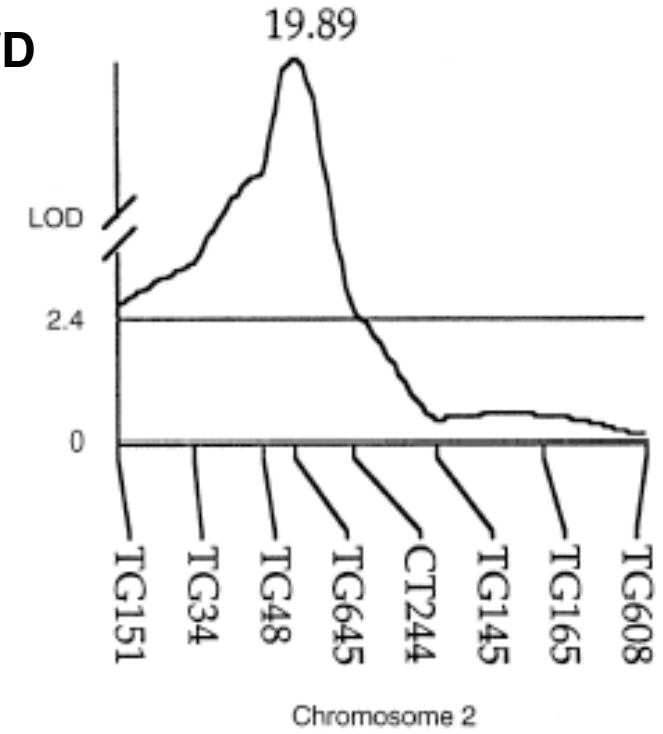
xi = indicator variable specifying the probabilities of an individual being in different genotypes for the tested position, constructed by flanking makers

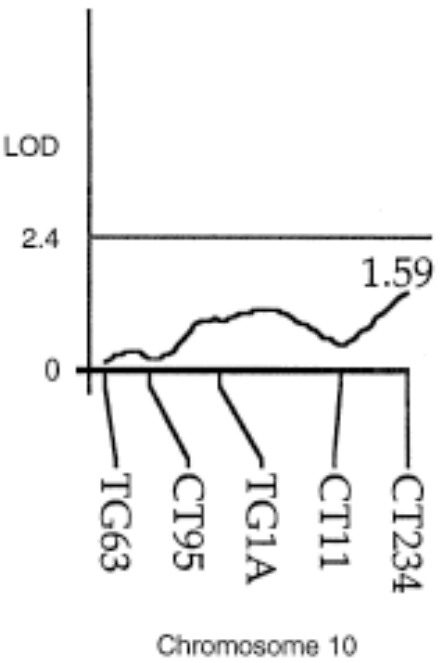
xi = 0 if Le/Le, = 1 if Le/Lp, = 2 if Lp/Lp

yi = 0 if Le/Le, = 1 if Le/Lp, = 2 if Lp/Lp at marker y

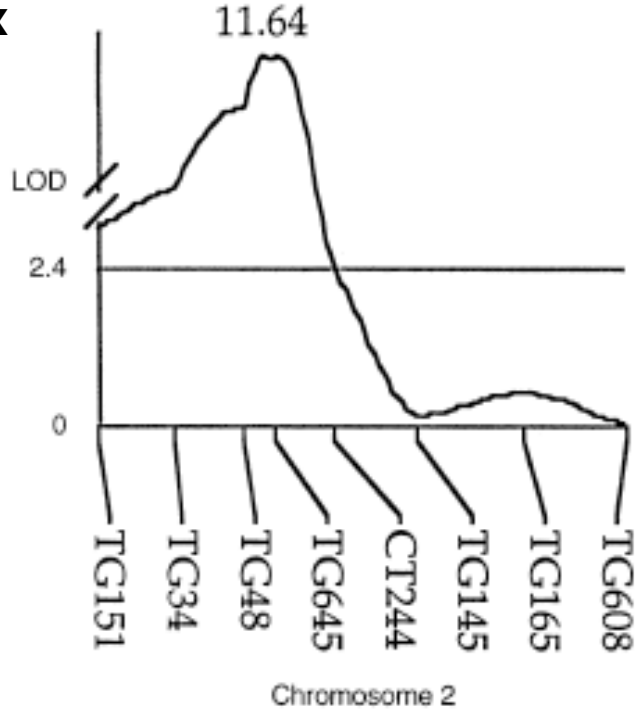
One major locus near marker TG645

L/D





Dmin/Dmax



responsible for 67% of L/D variance

allele YP = recessive

Corrections for multiple testing

- Correction of *p* value: Bonferroni correction

$p_{genome\text{-}wide} = p_{nominal} / n_{tests}$

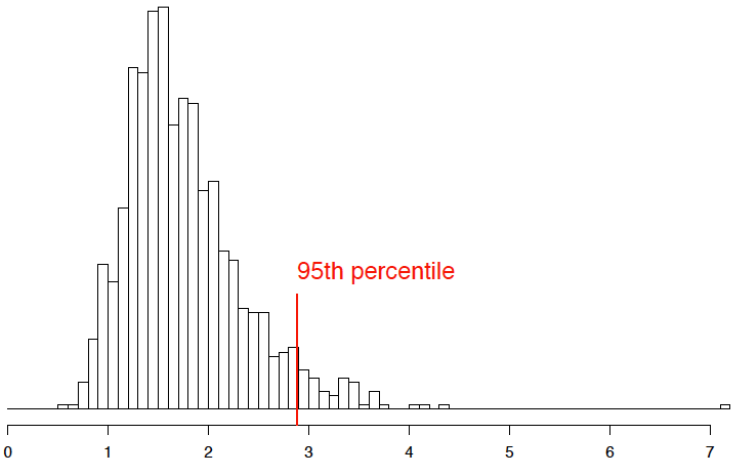
very “conservative” correction
some less conservative variants such as sequential Bonferroni

- Empirical permutation test:

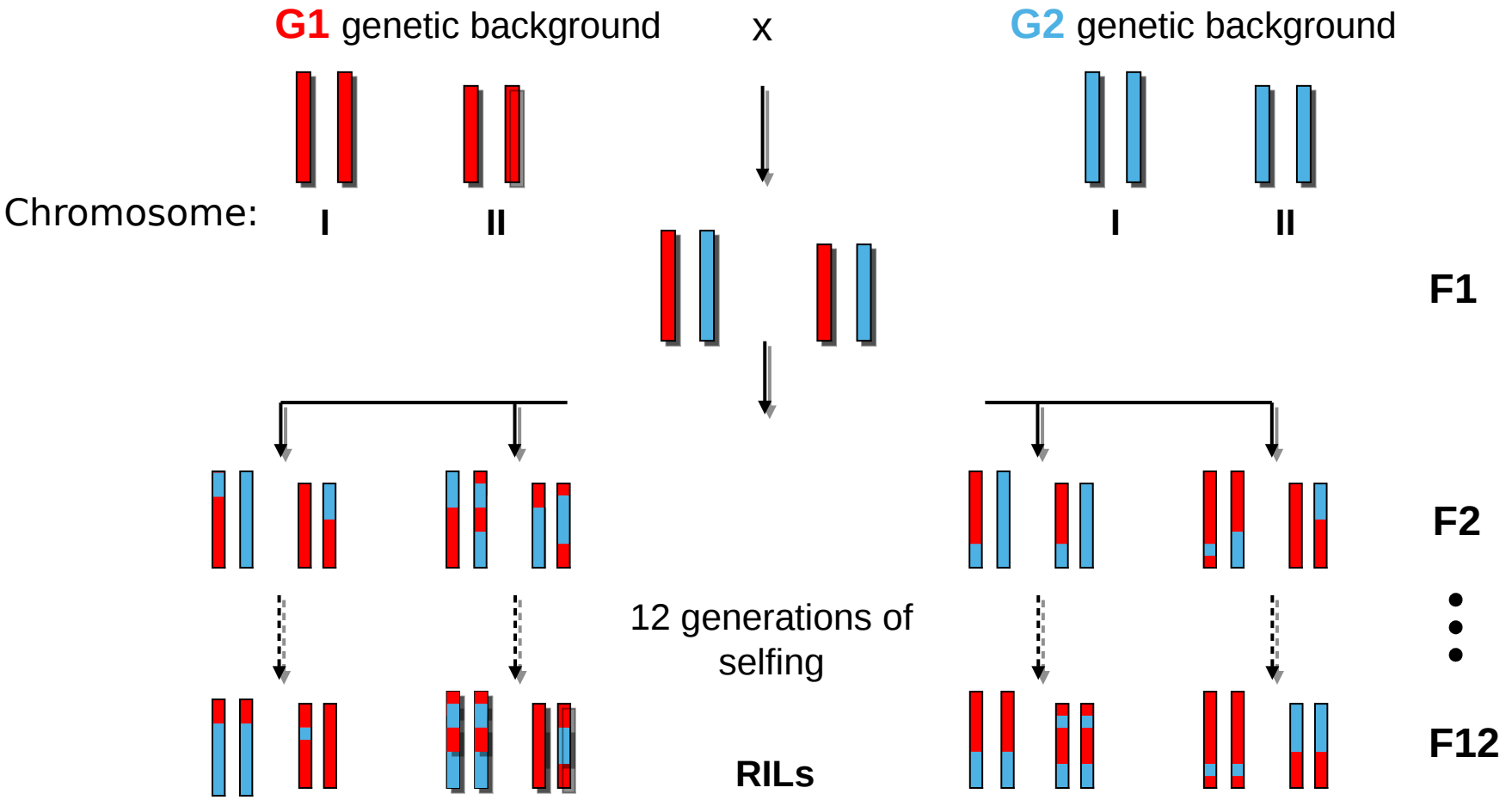
takes into account the structure of the data

Permutation of the data (here shuffling genotype and phenotype), many times (example: 10,000)
At each permutation i: Pi = min(*p* over all markers)

$p_{5\%}$ = threshold of *p* value where only 5% permutations pass the test
can then be used in the true dataset



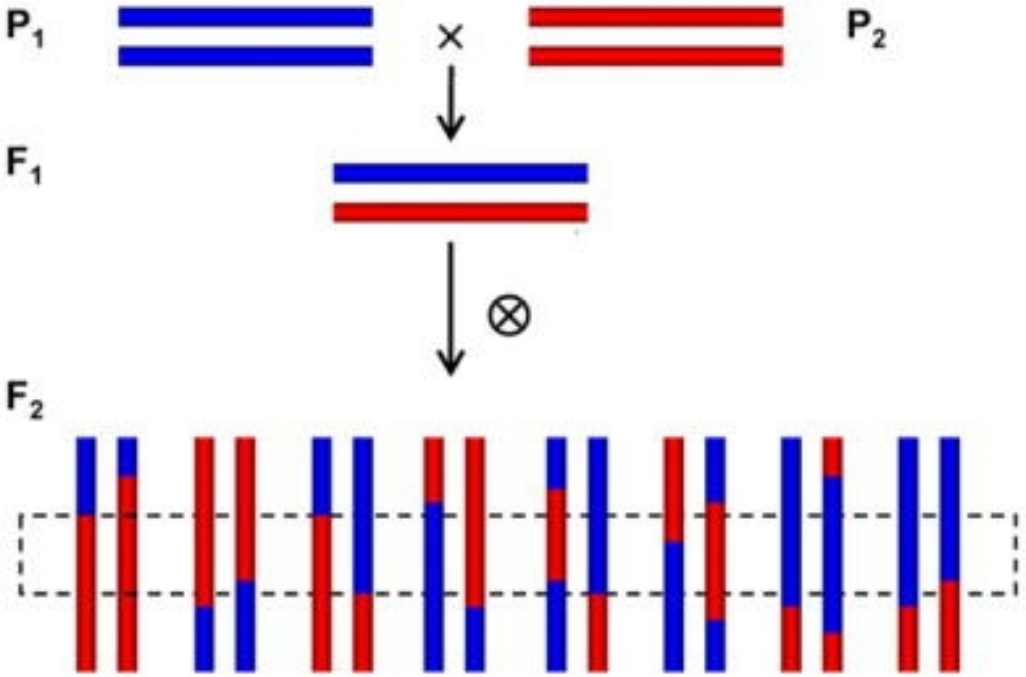
Recombinant Inbred Lines (RIL)

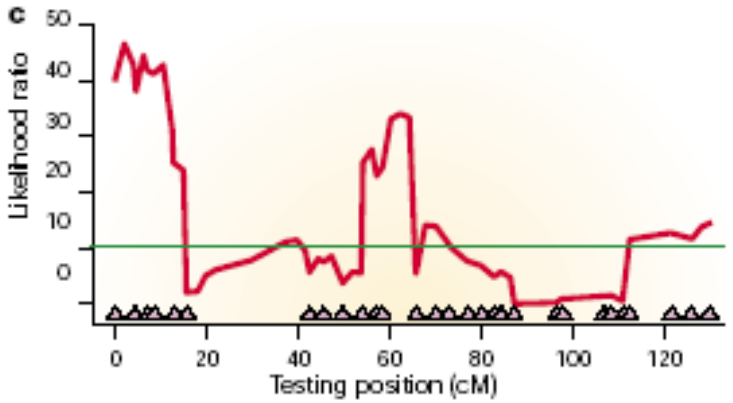


Analysis of multiple individuals of the same genotype

Linkage Mapping

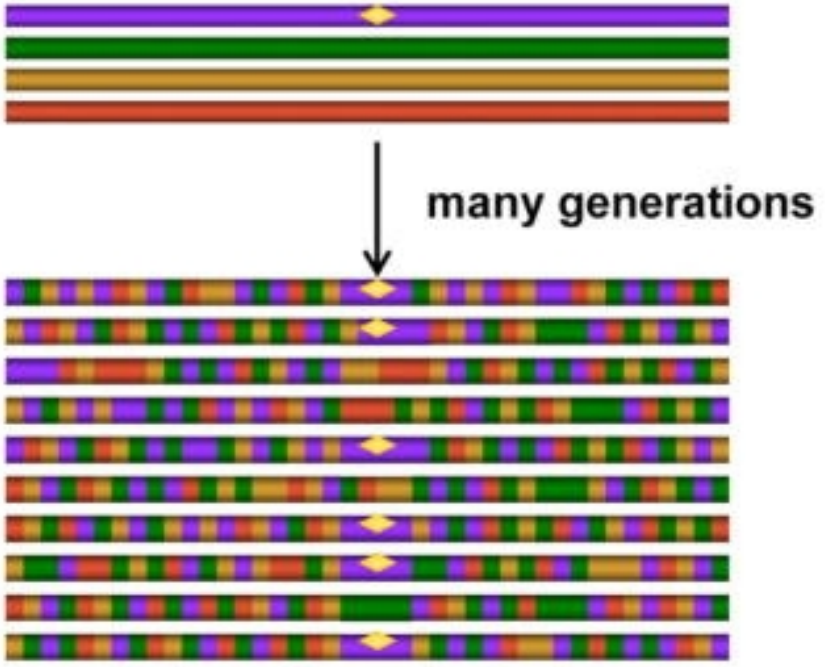
Crosses in the lab

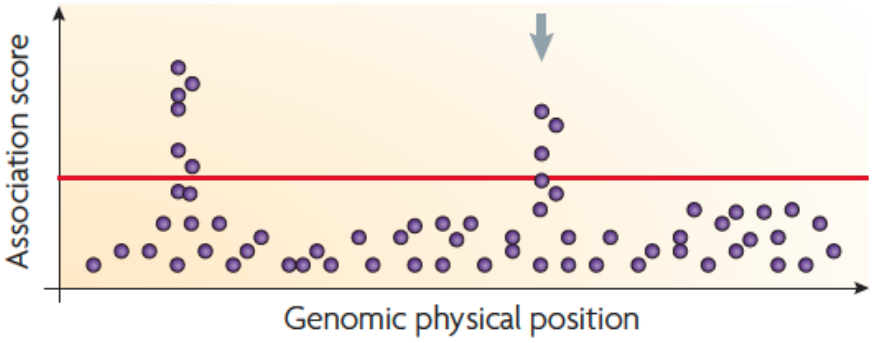


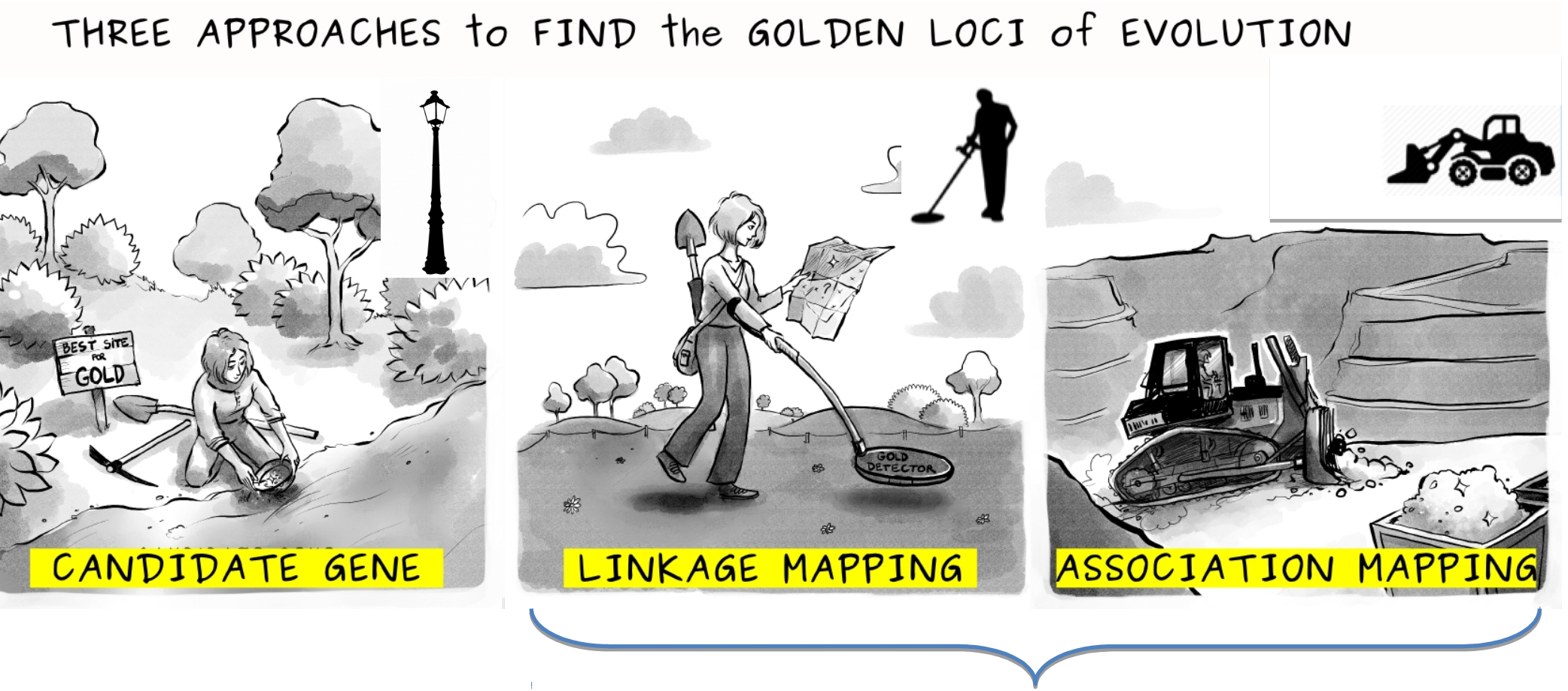


Association Mapping

Past crosses in natural populations







REVERSE GENETICS
From genes to traits

FORWARD GENETICS
From traits to genes

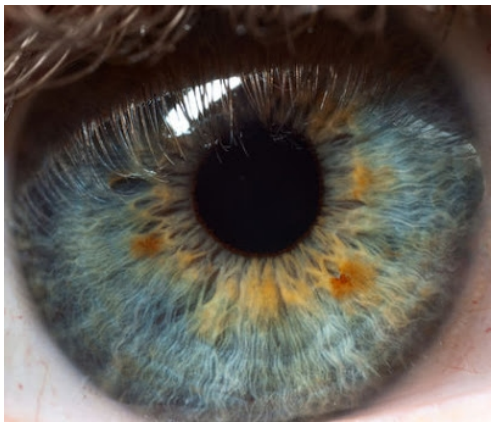
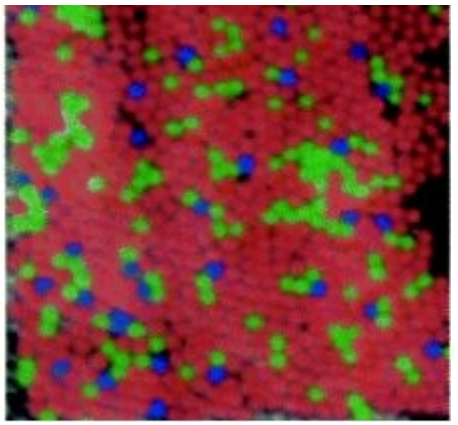
Little Ascertainment Bias, but

Requires the intermixing of two gene pools or lineages

Noise

Developmental noise

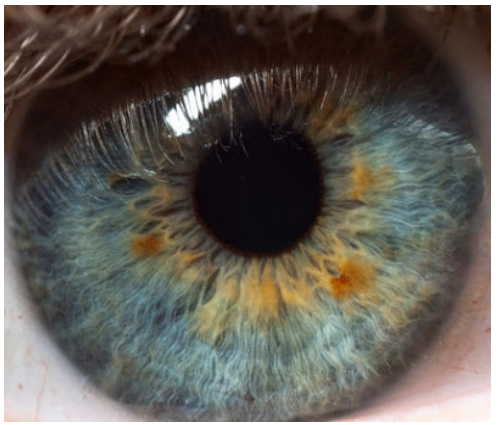
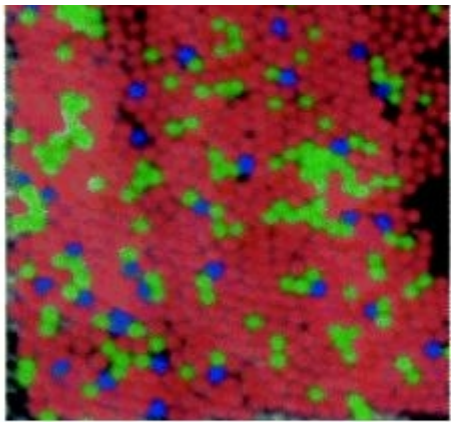
Differences between left and right sides of the body


ear shape, neuron connectivity, olfactory receptor gene expression, X inactivation pattern, organ cell number and size...

Developmental noise

Differences between left and right sides of the body


ear shape, neuron connectivity, olfactory receptor gene expression, X inactivation pattern, organ cell number and size...

Differences between twins

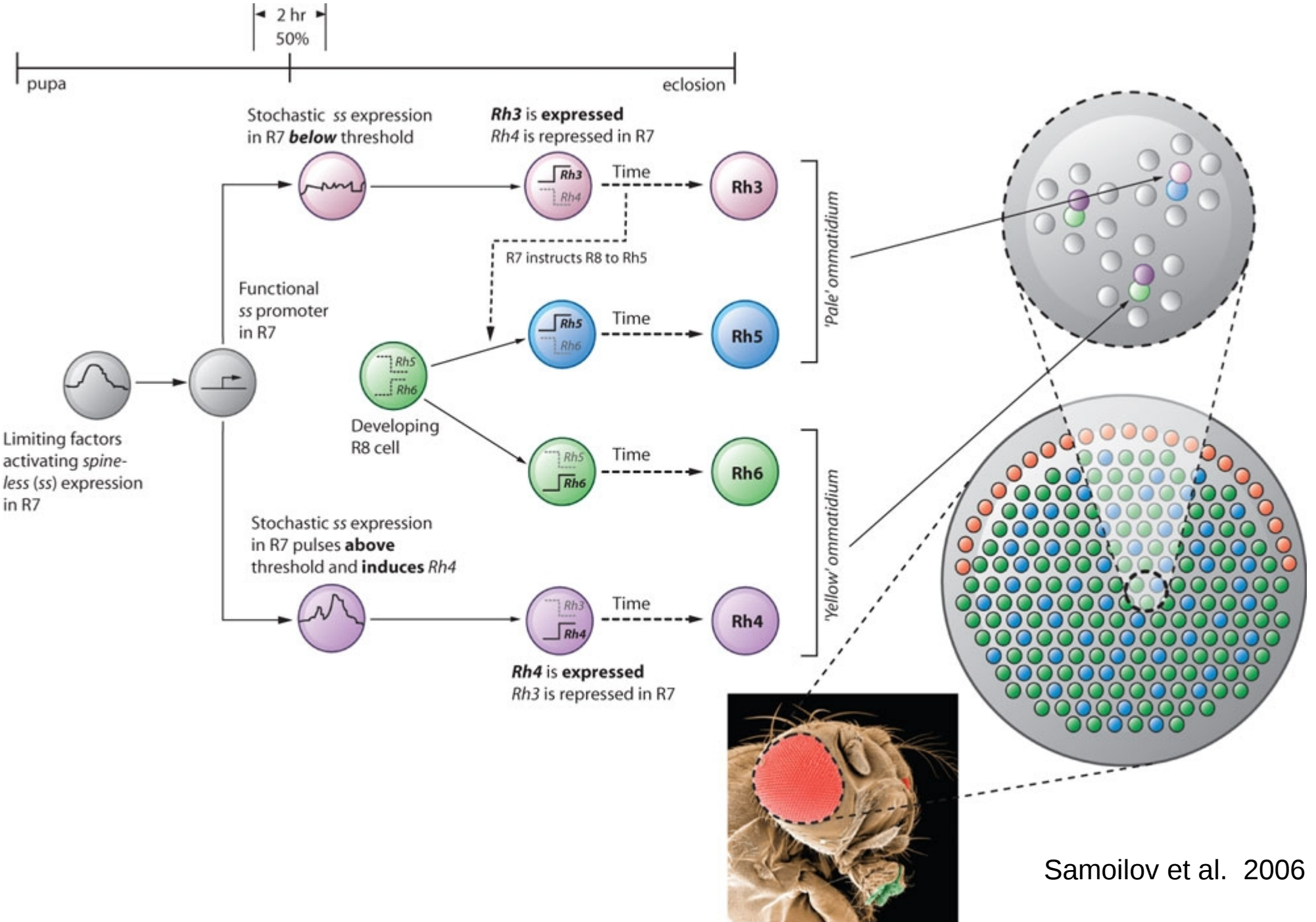
immune system cells, gait, arms crossing, voice, heart beat, brain waves...

Some can be attributed to variation in the number of determinant molecules

During terminal differentiation of mouse 3T3-L1 pre-adipocytes, individual TF abundance differs dramatically (from ~250 to >300,000 copies per nucleus) and the dynamic range can vary up to fivefold during differentiation.

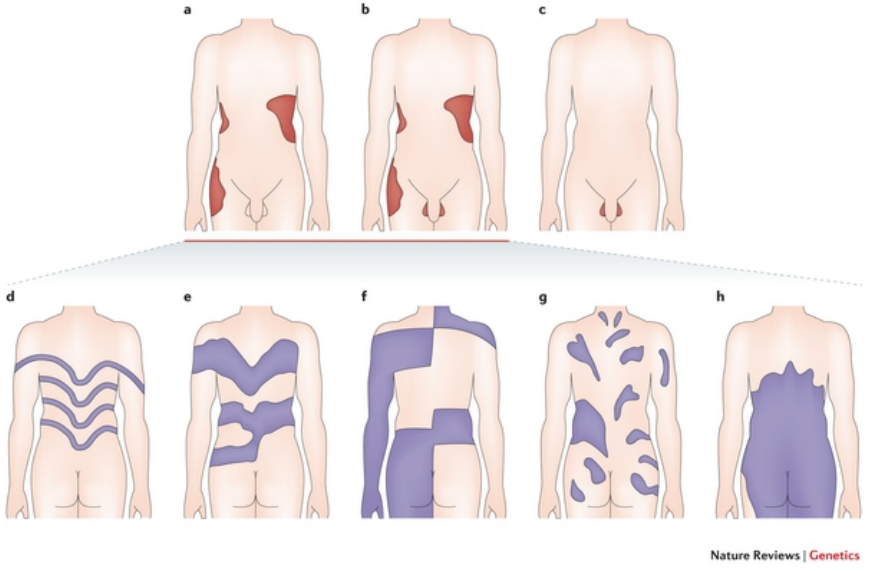
Simicevic 2013 Nature

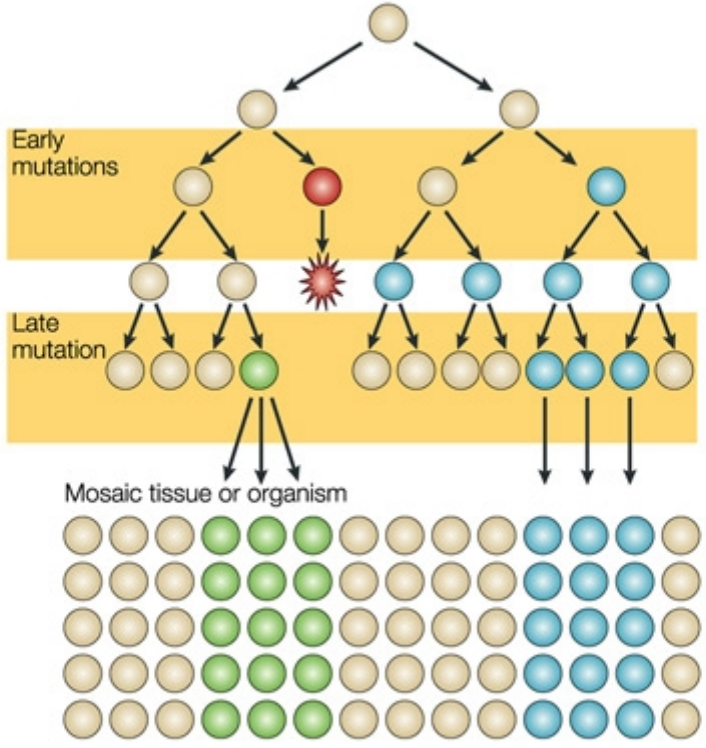
Developmental noise can be “good”



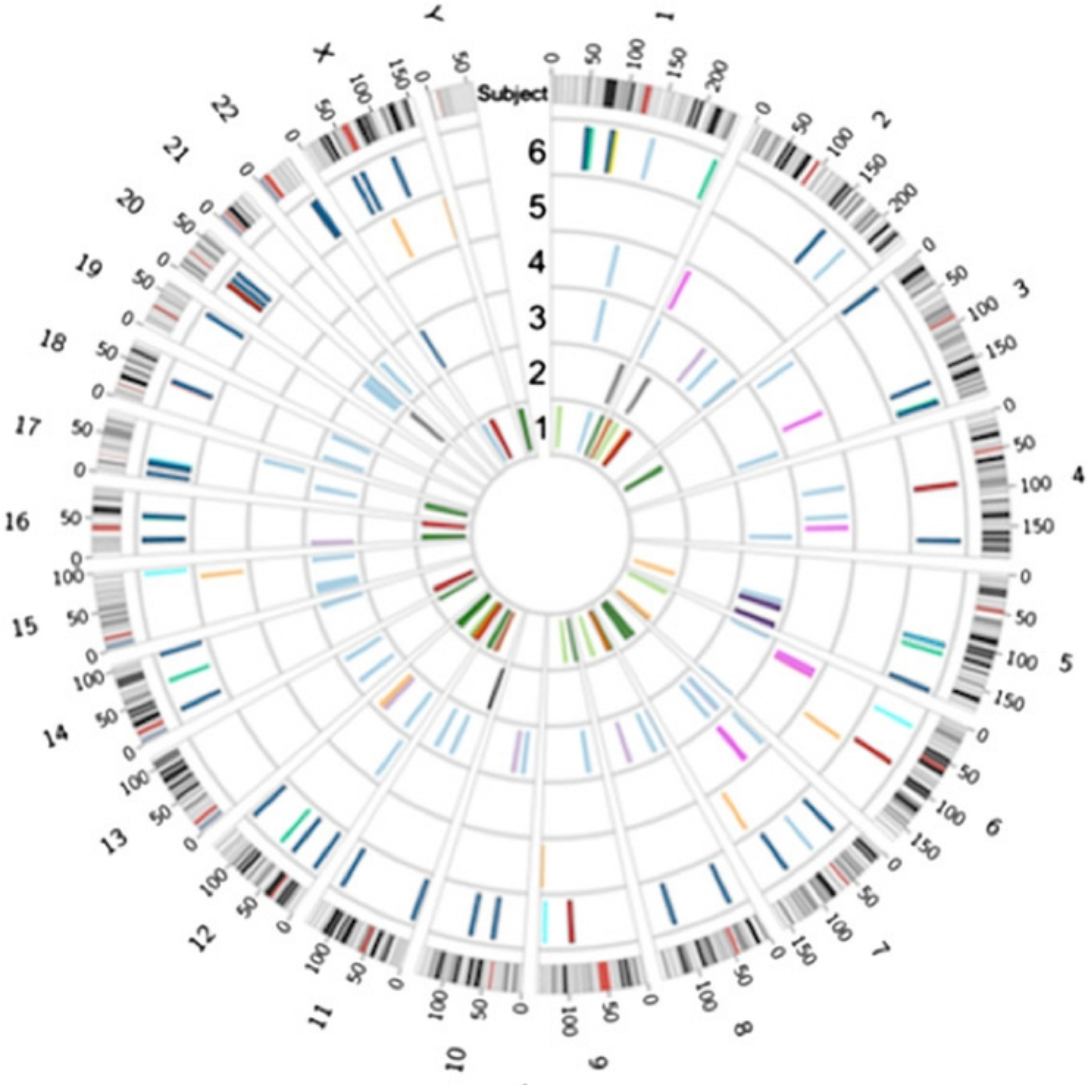
Samoilov et al. 2006

Somatic mosaicism



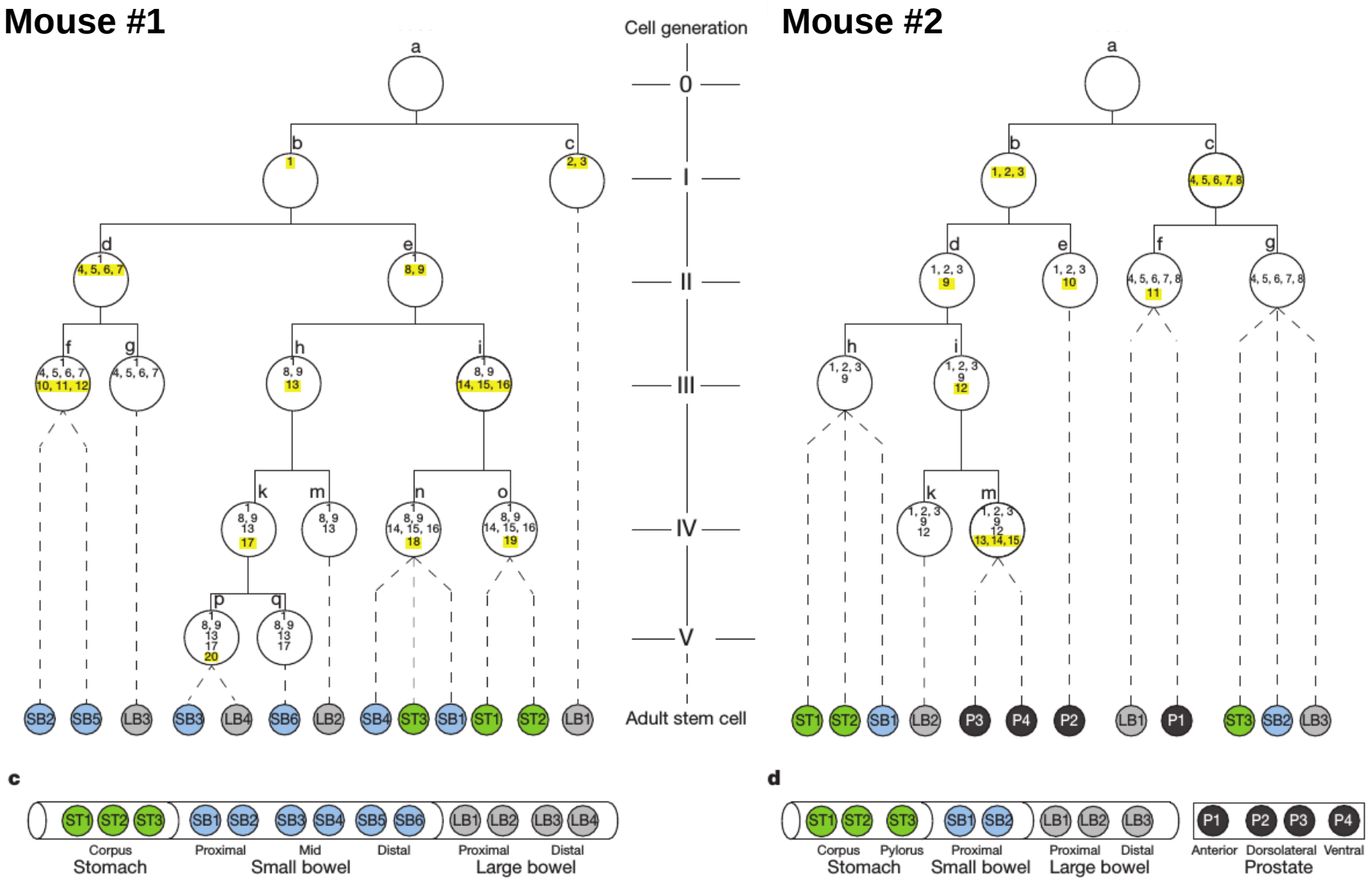


73 somatic CNVs in 11 tissues of six persons



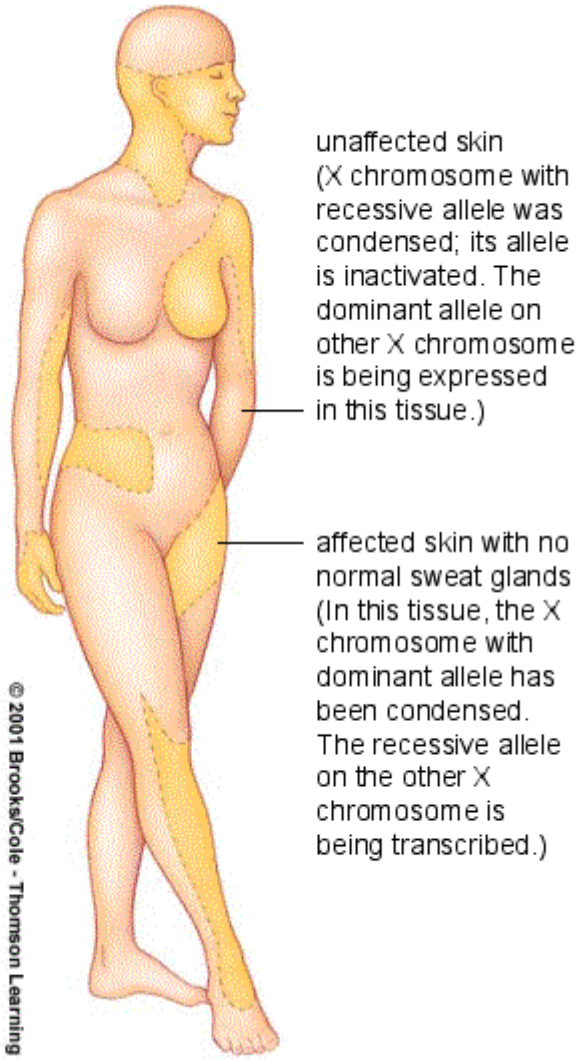
- O'Huallachain 2012 PNAS

Somatic mosaicism used to reconstruct cell lineages

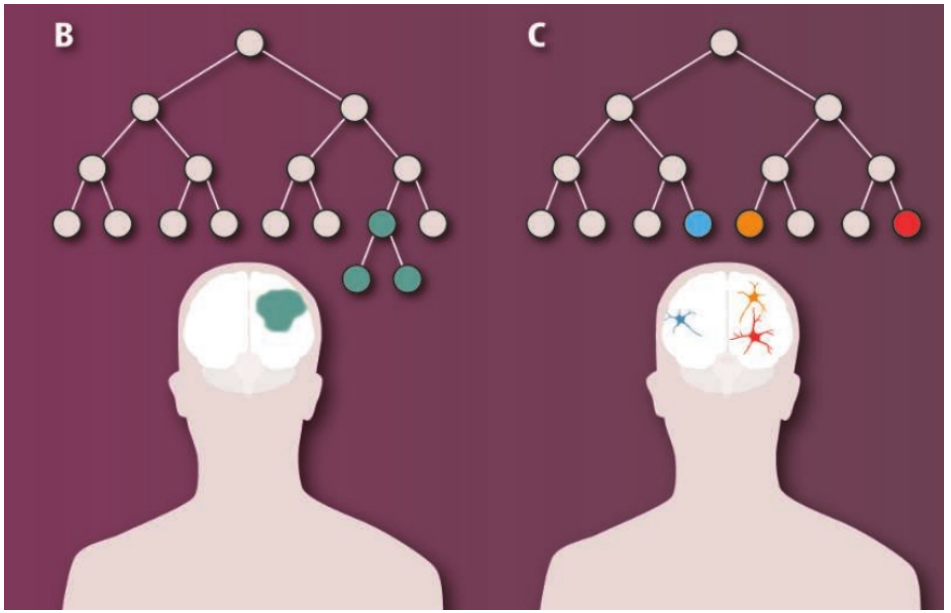


Behjati 2014 Nature

Female mosaicism : X inactivation pattern



Somatic transposition in human brain

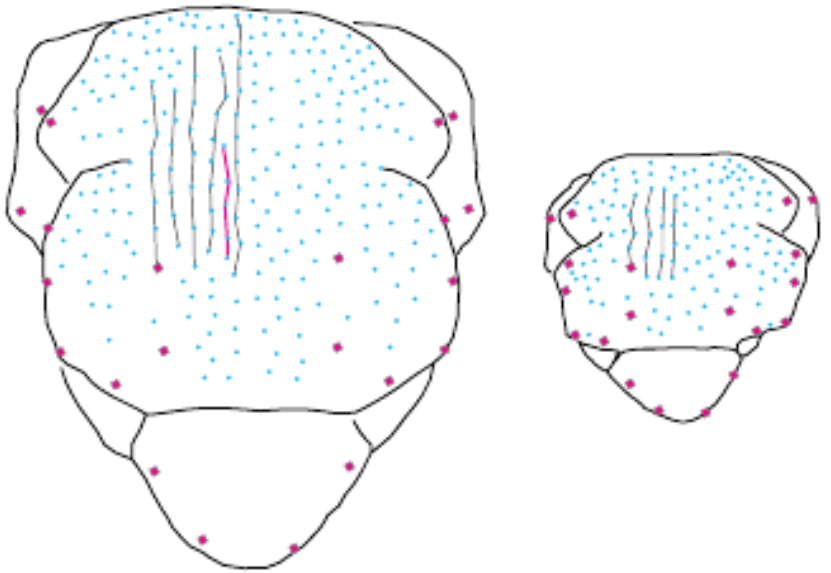


In three individuals:

in the hippocampus and caudate nucleus
7,743 somatic L1 insertions, 13,692 somatic Alu insertions and 1,350 SVA insertions

Baillie 2011 Nature

Robustness



Well-fed Semi-starved

Robustness

Absence or low variation of a phenotype when faced with an incoming variation

1) Of what?
2) To what?

To either:
- stochastic variation
- environmental variation: specify
- genetic variation: specify

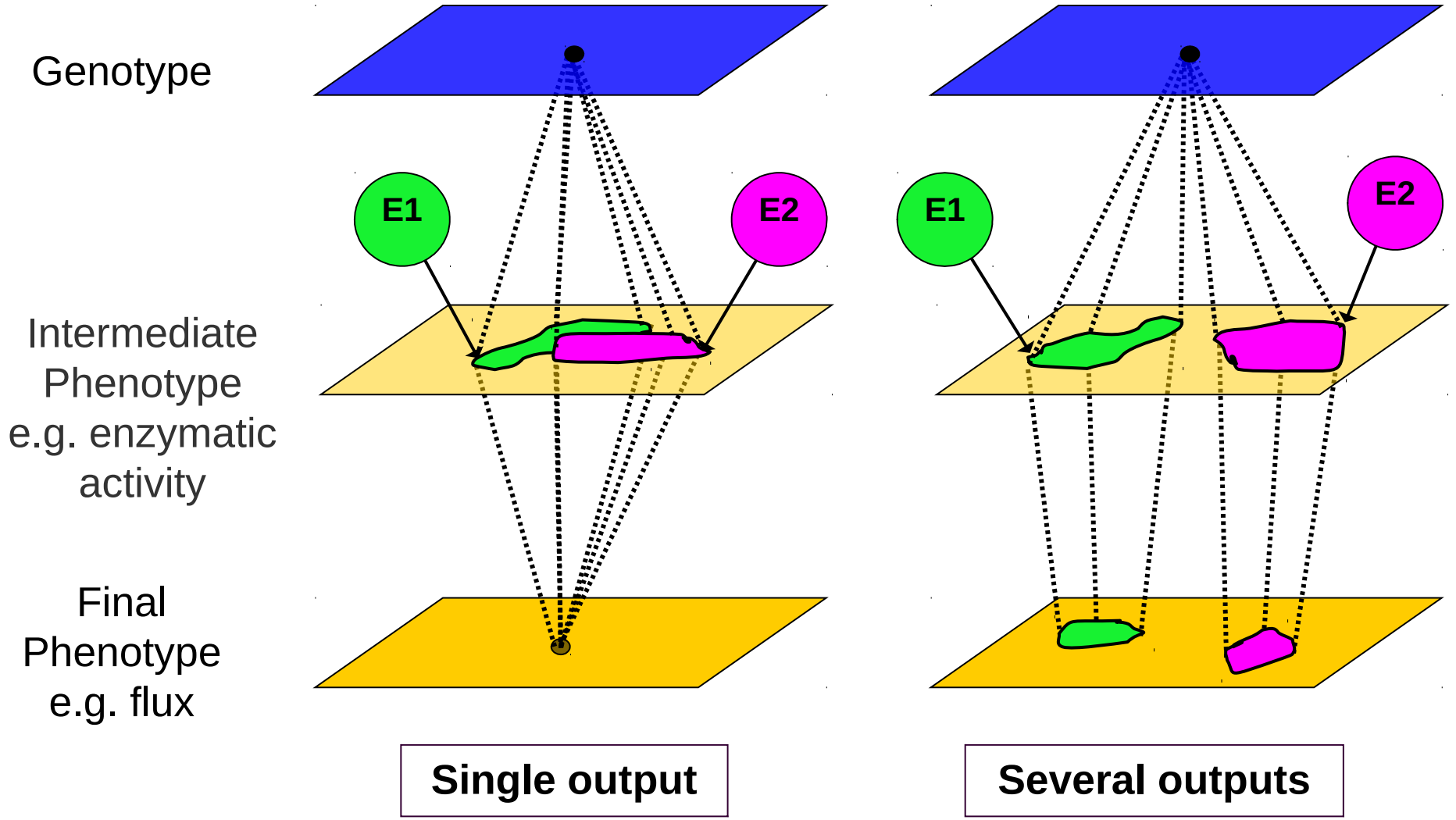
3) How much?

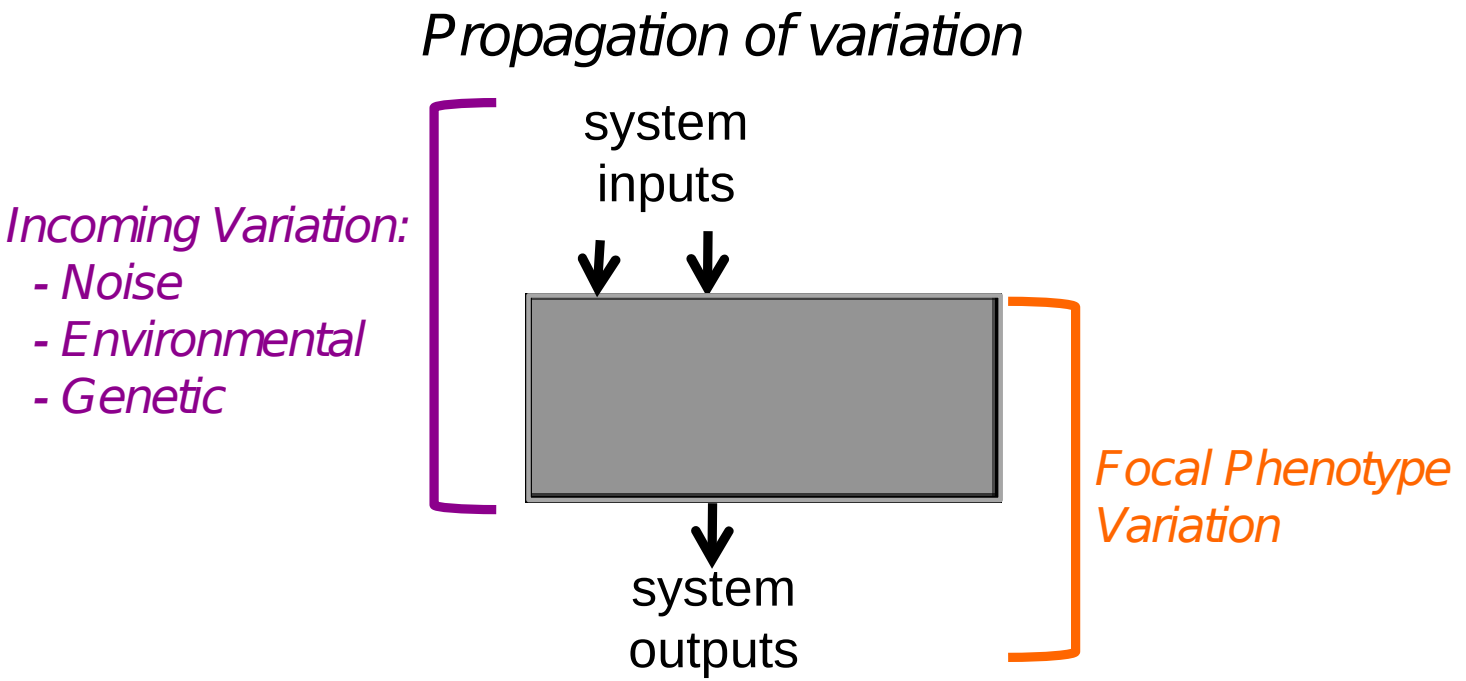
Different phenotypic metrics
Coefficient of variation: standard deviation/mean

Historically:
quantitative genetics (low variance, canalization)
physics/chemistry/engineering (robustness, buffering)

Canalization: mechanisms that make the system follow a certain trajectory

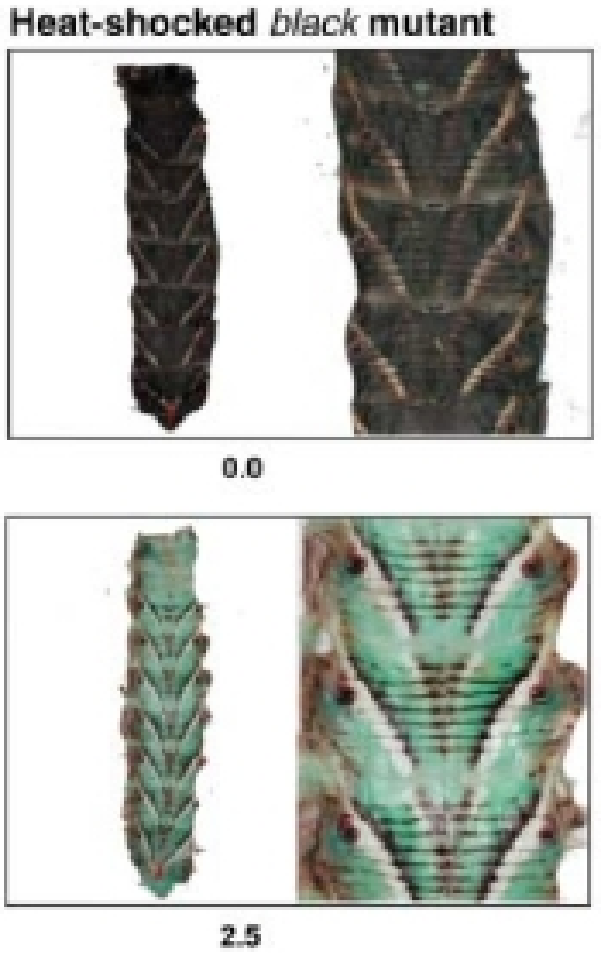
Trait plasticity versus invariance (robustness) at different levels of the genotype-phenotype map



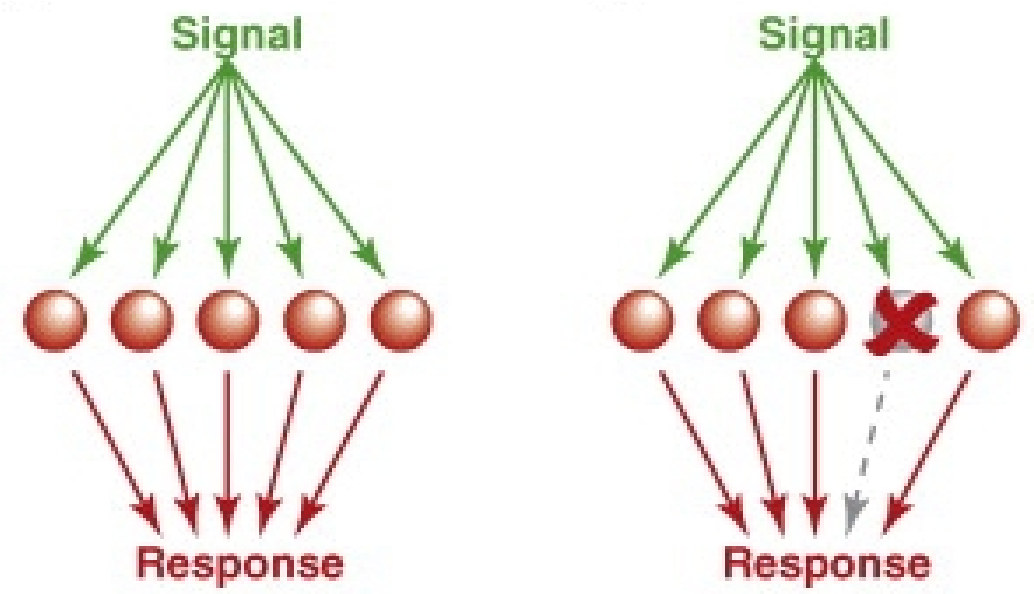

Propagation of variation
Incoming Variation:
- Noise
- Environmental
- Genetic
system inputs
system outputs
Focal Phenotype Variation

Causes of robustness

Non-linearity



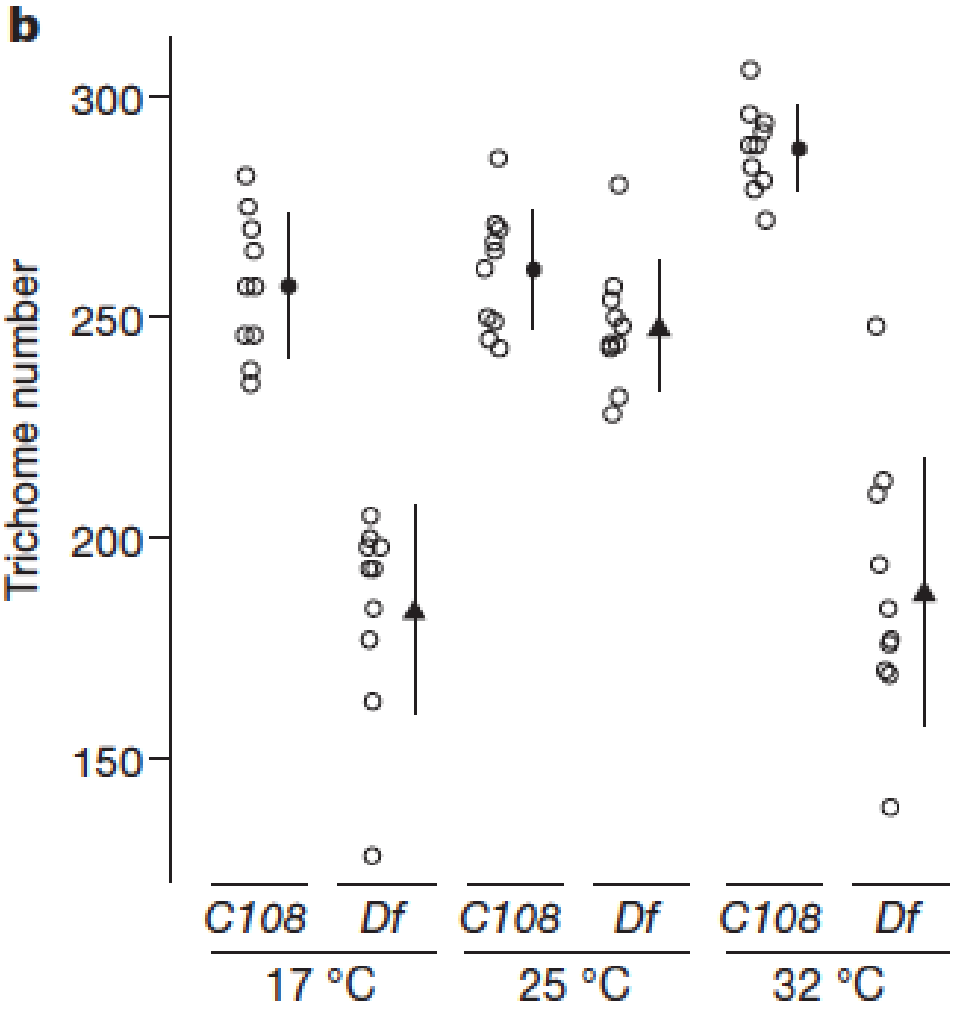
Redundancy

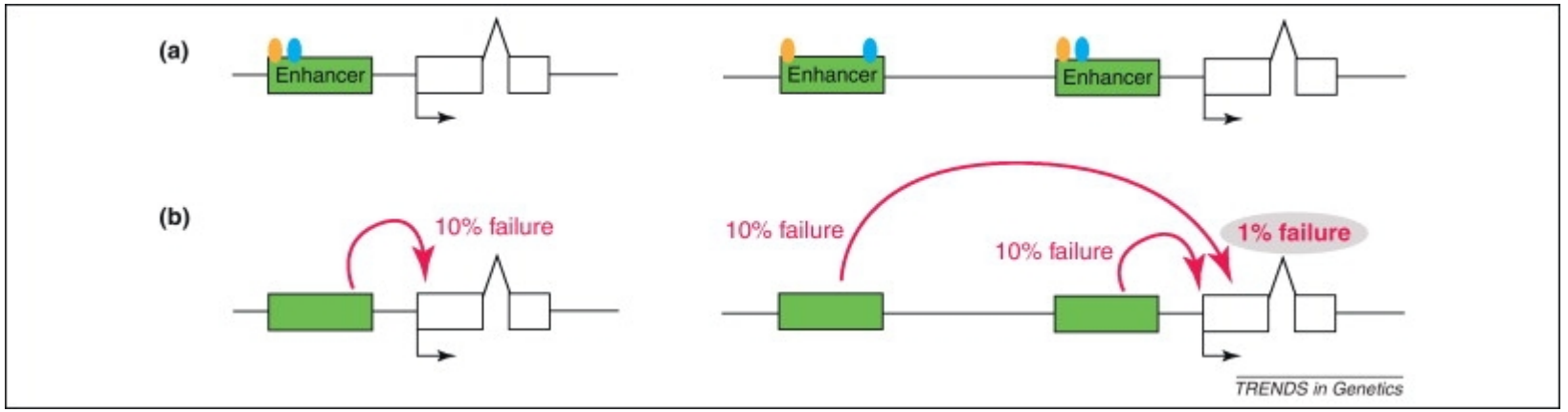


nature Vol 466|22 July 2010|doi:10.1038/nature09158

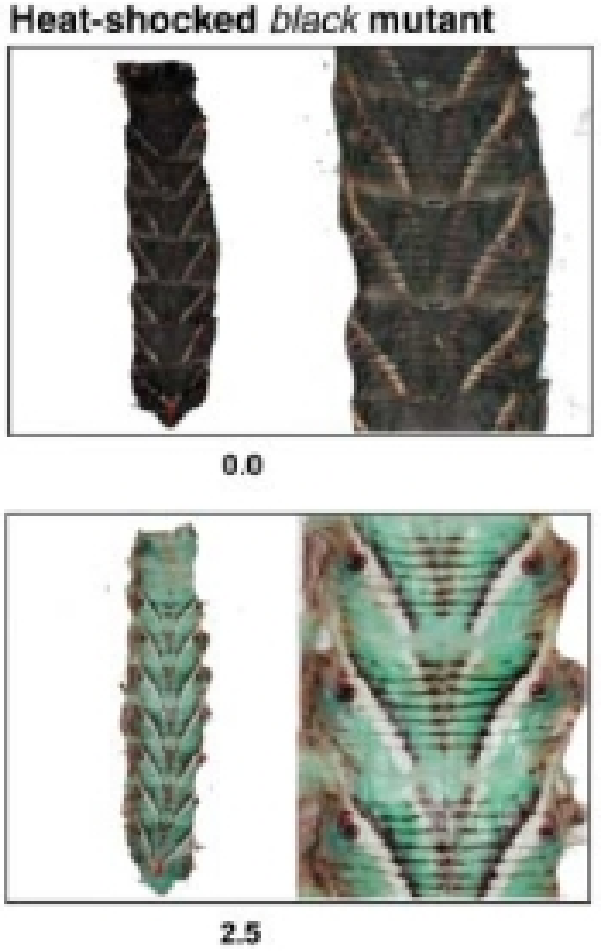
LETTERS

Phenotypic robustness conferred by apparently redundant transcriptional enhancers





Cryptic genetic variation



Cryptic genetic variation

First requires defining the *phenotype of interest*

Genetic variation that has no effect on phenotype of interest

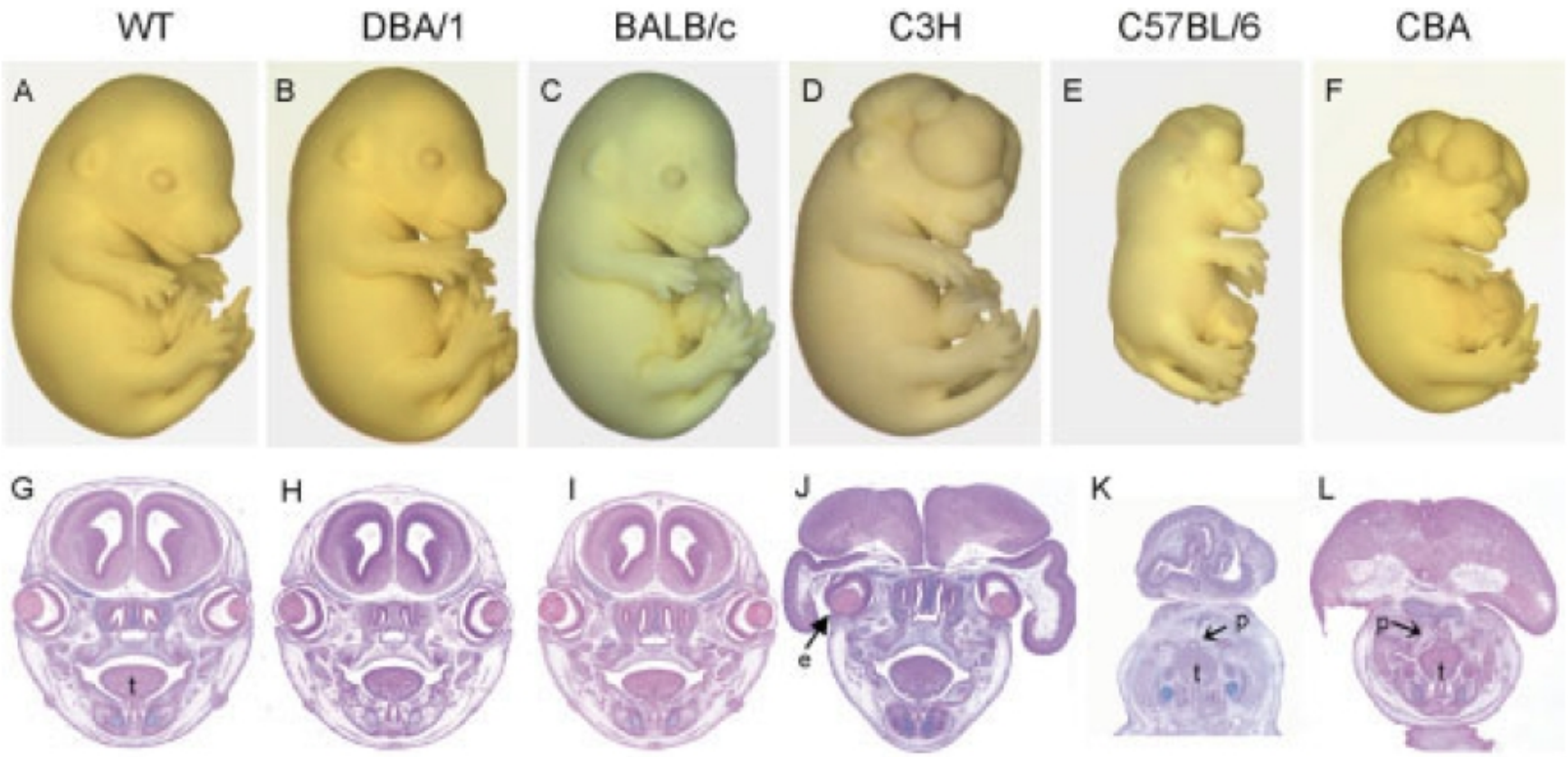
... but may be revealed *under some circumstances*
by its effect on this phenotype

Cryptic genetic variation (CGV) is defined as standing genetic variation that does not contribute to the normal range of phenotypes observed in a population, but that is available to modify a phenotype that arises after environmental change or the introduction of novel alleles.

Gibson & Dworkin *Nat Rev Gen* 2004

Expressivity of one mutation varies with wild genetic gackground

Tcof1/- heterozygote mice



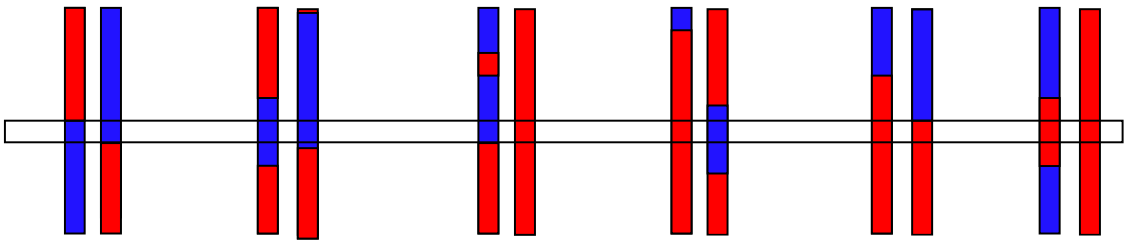
Dixon & Dixon *Dev Dyn* 2004

Epigenetics

Wild-type

Peloric

Linaria vulgaris

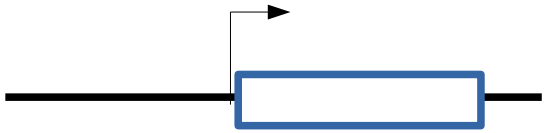


Cubas 1999 Nature

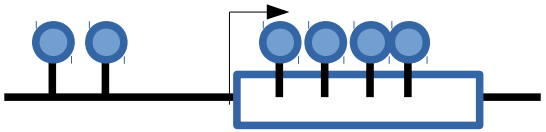
An epimutation

Wild-type

Peloric



CYCLOIDEA



CYCLOIDEA

Methylated DNA

Presence of CYCLOIDEA proteins

Absence of CYCLOIDEA proteins

Conclusion

Complexifications of the G-P map

Genetic Linkage

Epistasis

Supergene

Pleiotropy

GxE (introduction)

Large number of alleles

Noise

Robustness

Cryptic genetic variation

Epigenetics

What makes us different?

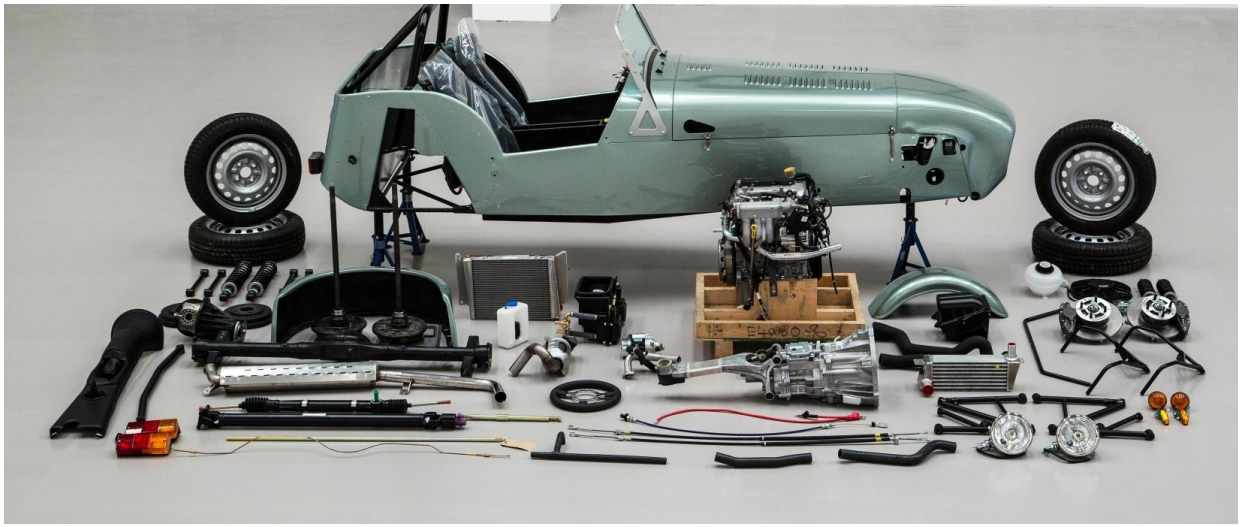
Genetics **Epigenetics** **Environment** **Stochasticity**

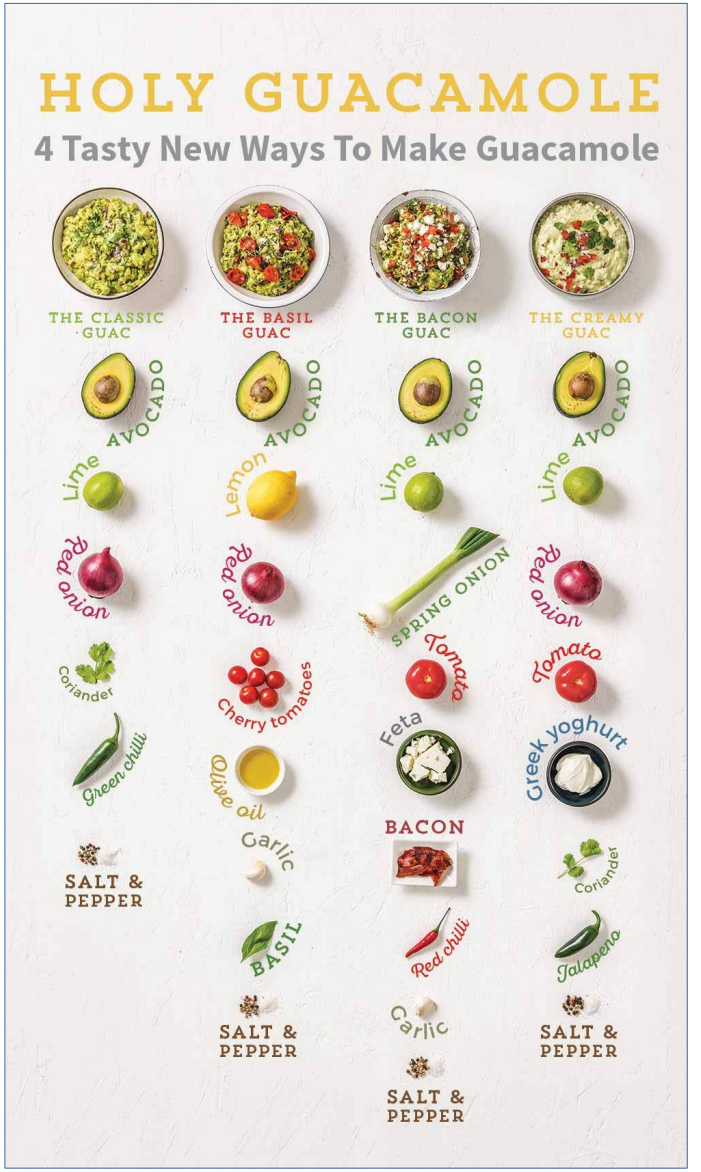
Heritable

Deterministic causes

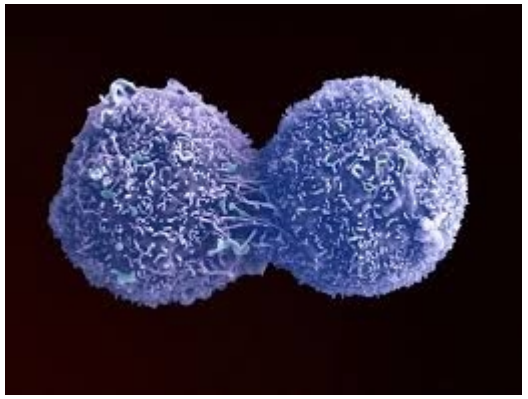
Interaction of all these parameters

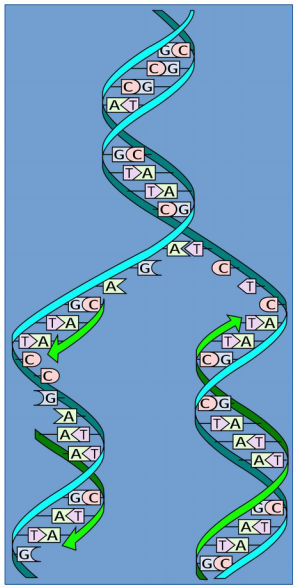
A living organism is not made by assembling pieces together





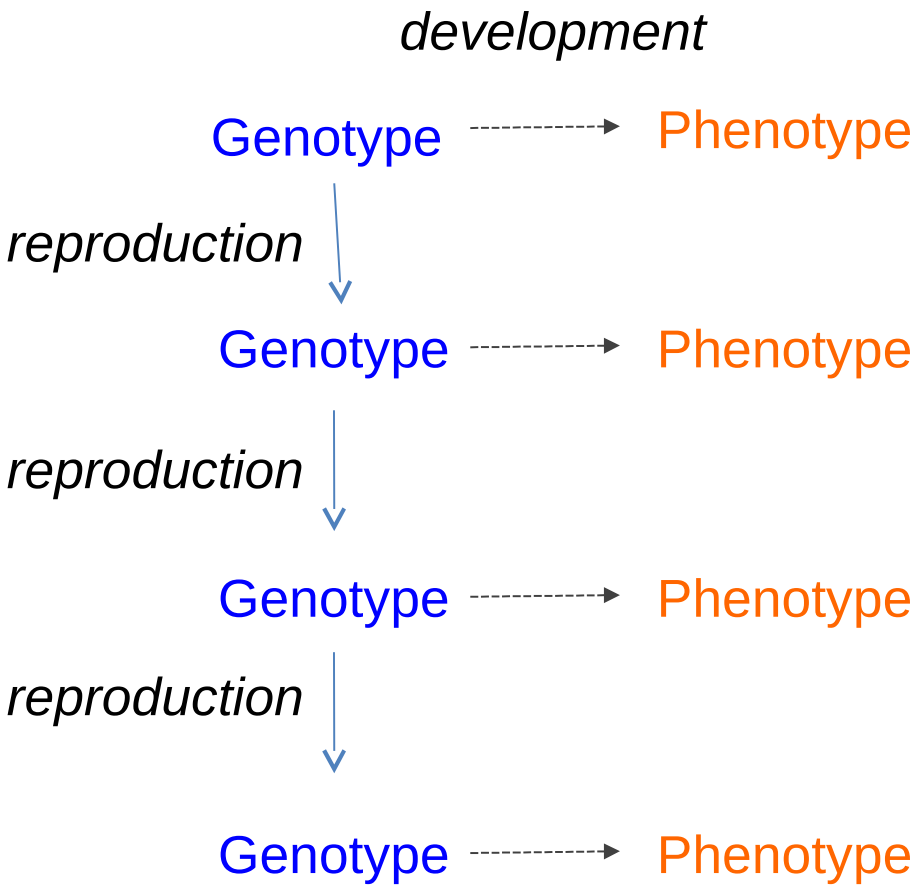
..but results from changes that occurred successively across evolutionary time





Orgogozo et al 2015 Frontiers Genetics

A simplistic view



Heritable traits are not always due to genes

The genotype does not determine entirely the phenotype

The genotype cannot replicate by itself

Genotype and phenotype imply variation