Marker Assisted Selection

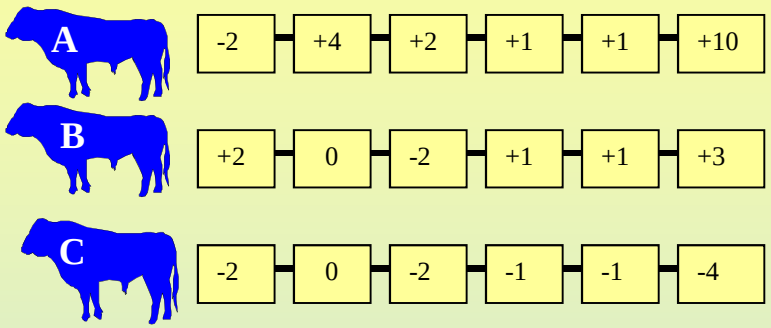
Using gene testing in livestock

- Parentage testing
- Marker Assisted Selection
- Marker Assisted Introgression
- Marker Assisted Conservation
- Development of transgenics

Selection for Quantitative Traits

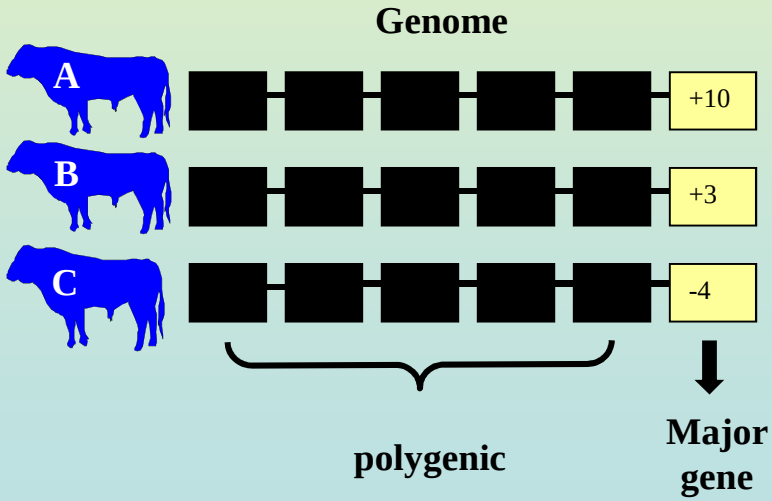
polygenes and major genes

True situation

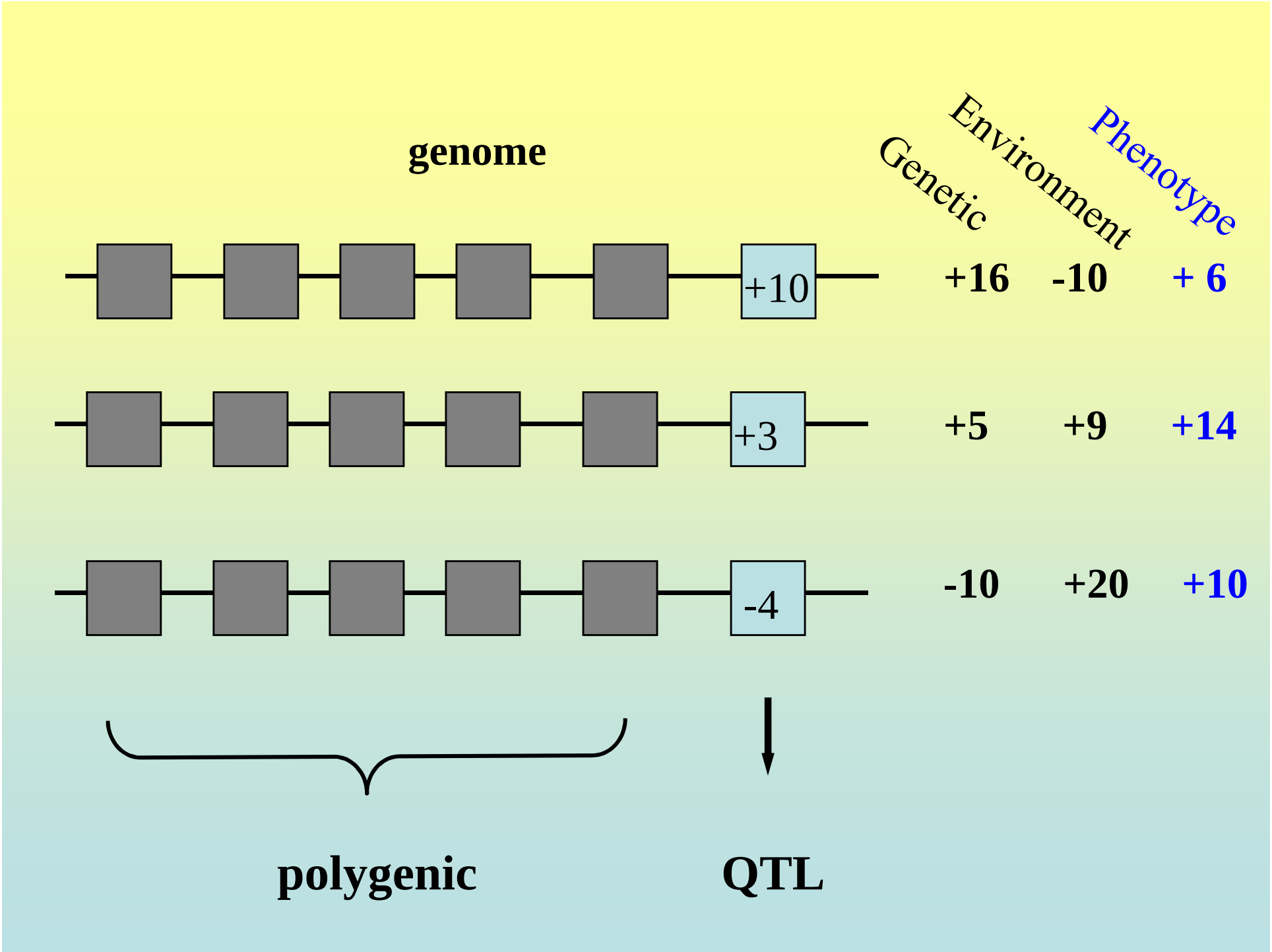


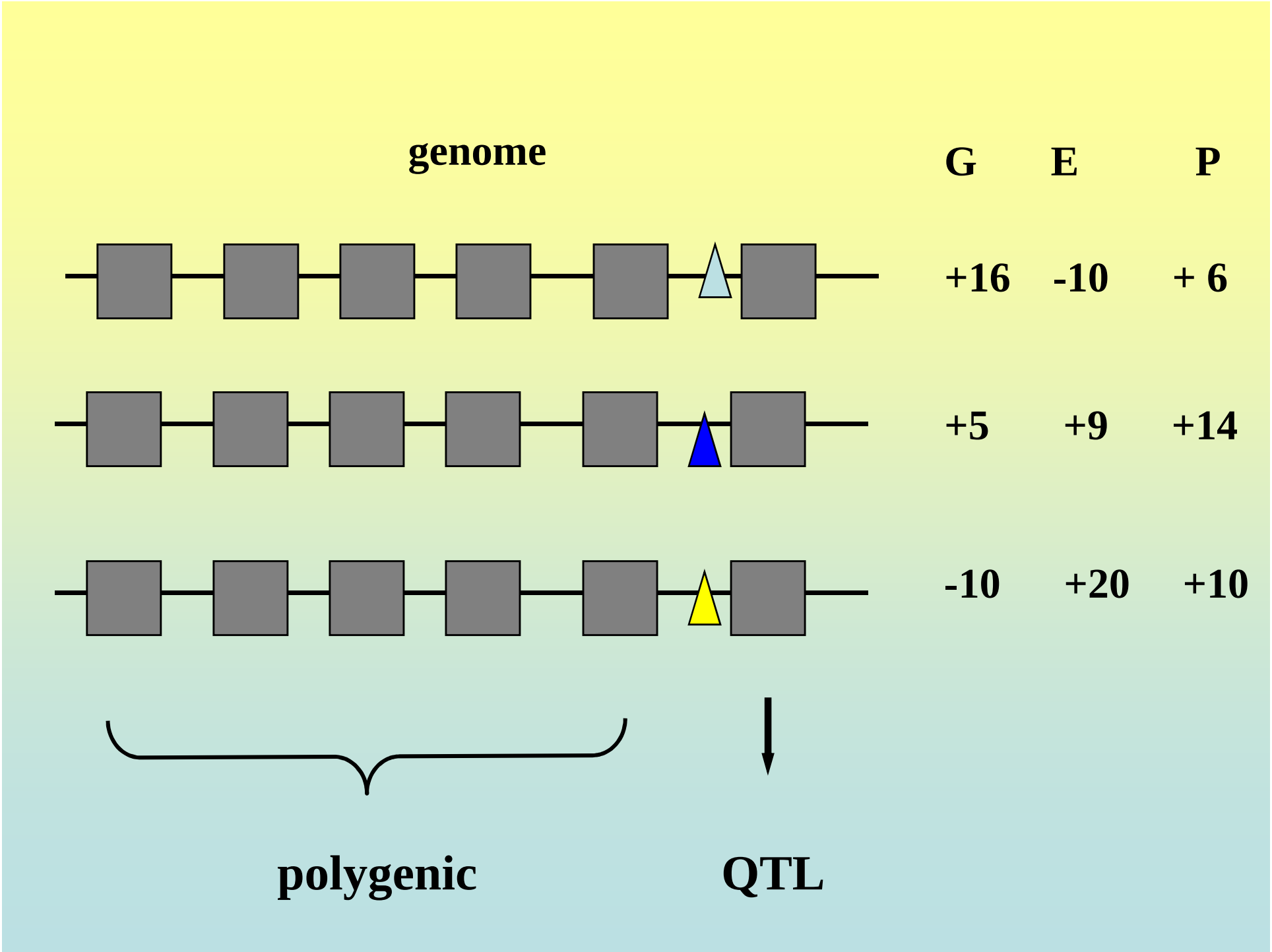
	Effects		
	Genetic	**Environment**	**Phenotype**
A	+16	-10	**+6**
B	+5	+9	**+14**
C	-10	+20	**+10**

Observed situation

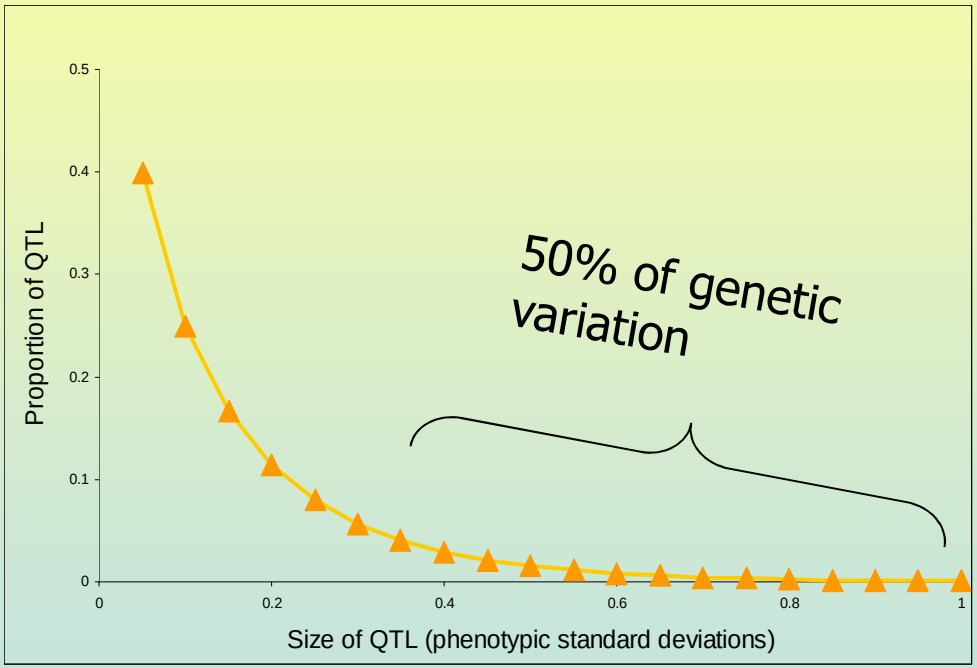


Phenotype
+6
+14
+10


genome
Genetic
Environment
Phenotype
+10
+16
-10
+ 6
+3
+5
+9
+14
-4
-10
+20
+10
polygenic
QTL


genome
G
E
P
+16
-10
+ 6
+5
+9
+14
-10
+20
+10
polygenic
QTL

The distribution of QTL effects

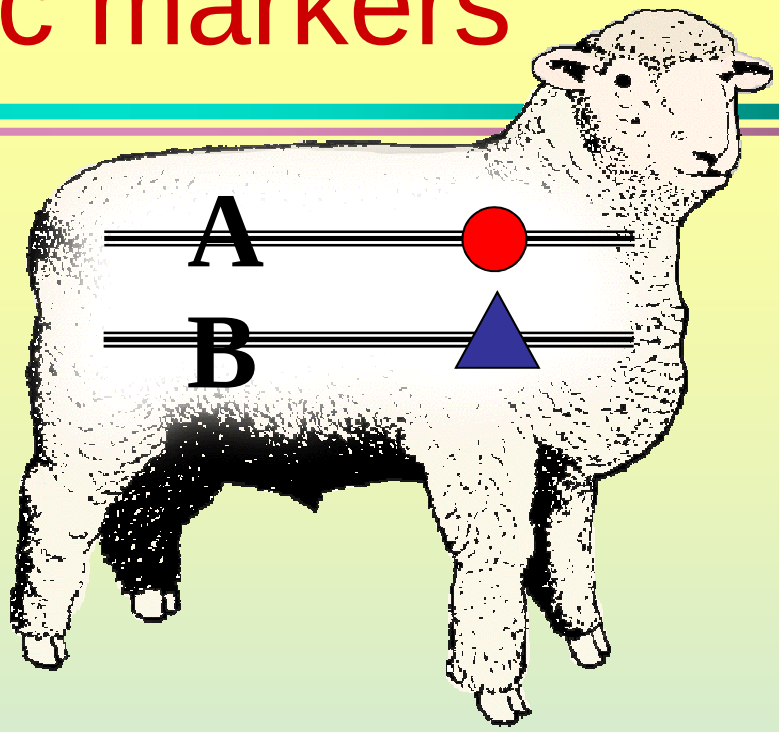


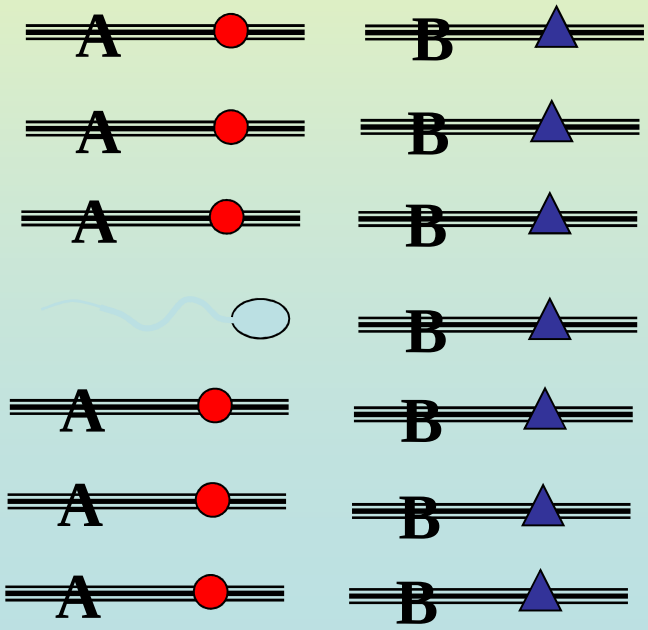
- Maybe 5-10 large QTL explain the majority of the genetic variance.
- Mapping experiments should be able to detect QTL as small as 0.2σp?

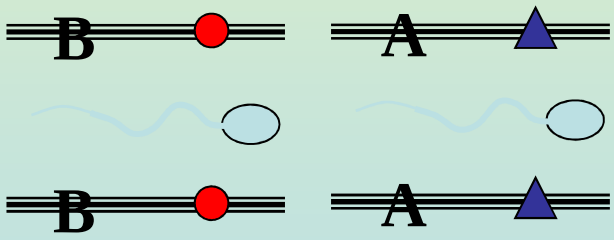
Many small QTL, few of large effect

Indirect genetic markers

Can select among offspring ...

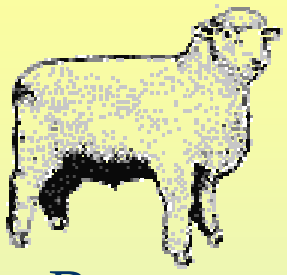




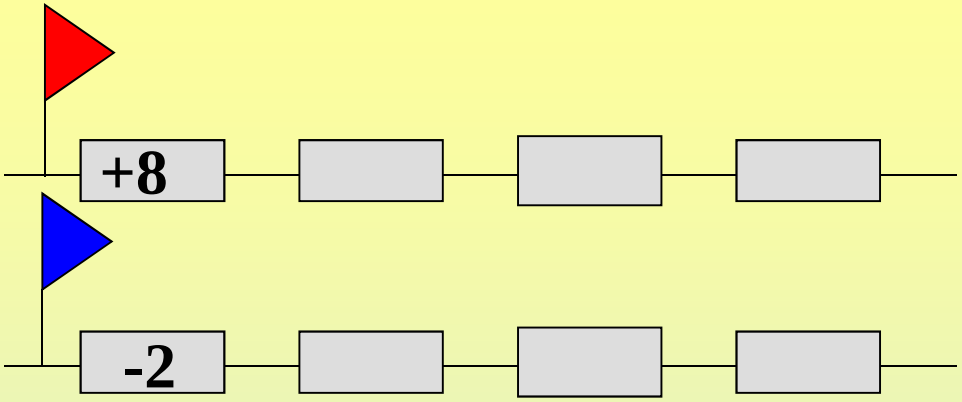


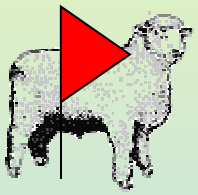
'recombinants'

Marker assisted selection

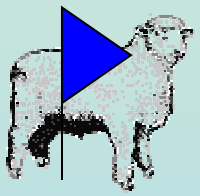


Ram





Progeny



Progeny

Marked gene from sire	Other genes from sire	Genes from dam
+8	?	?
-2	?	?

How important is the marker information?

depends on:

Size of QTL effect

Frequency of QTL alleles

Probability that an M-animal has indeed a Q-allele

Direct Markers

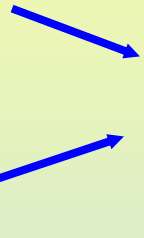
- No need for performance recording
- No extensive family testing
- Not very many examples (except the ‘obvious’)
- Not always guaranteed
  - false negatives (Double Muscling Example)
  - false positives (If not the true gene)

Linked Markers

- Need for performance recording within family
- Need for genotyping (2 generations)
- Linkage phase differs between families
- Need heterozygous parent (sire)
  - for marker genotype
  - for QTL genotype

Normal Genetic Evaluation

- Performance information
- Pedigree information

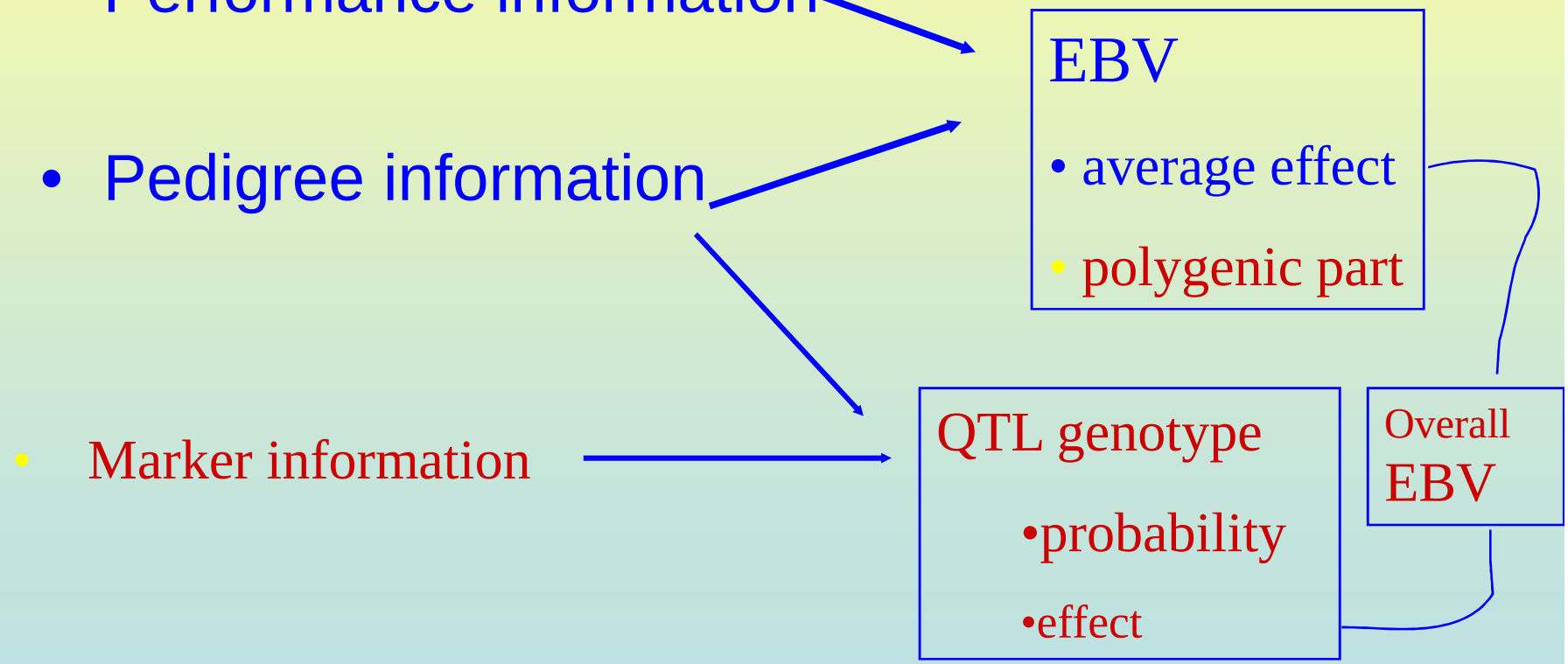


EBV

average effect

Genetic Evaluation with QTL

- Performance information
- Pedigree information
- Marker information



Effect of MAS on rate of genetic gain

	Selection *after* recording		Selection *before* recording	
	Gen 1	Gen 5	Gen 1	Gen 5
$h^2 = 0.11$, $V_{QTL}=0.33$	+21%	+6%	+45%	+23%
$h^2 = 0.27$, $V_{QTL}=0.33$	+9%	+2.3%	+38%	+15%
$h^2 = 0.27$, $V_{QTL}=0.11$	+1.3%	+1.3%	+8%	+6%

Meuwissen and Goddard, 1996

Conditions that are good for MAS

- Where heritability is low
  - e.g. fecundity
- Where the trait is sex limited.
  - e.g. milk production, fecundity
- Trait not measurable before first selection
  - e.g. milk production, longevity.
  - Most traits when using juvenile selection.
- Trait is difficult to measure.
  - e.g. disease resistance, recessive conditions, pigmented fibres, carcass traits

Discussion on simulation studies

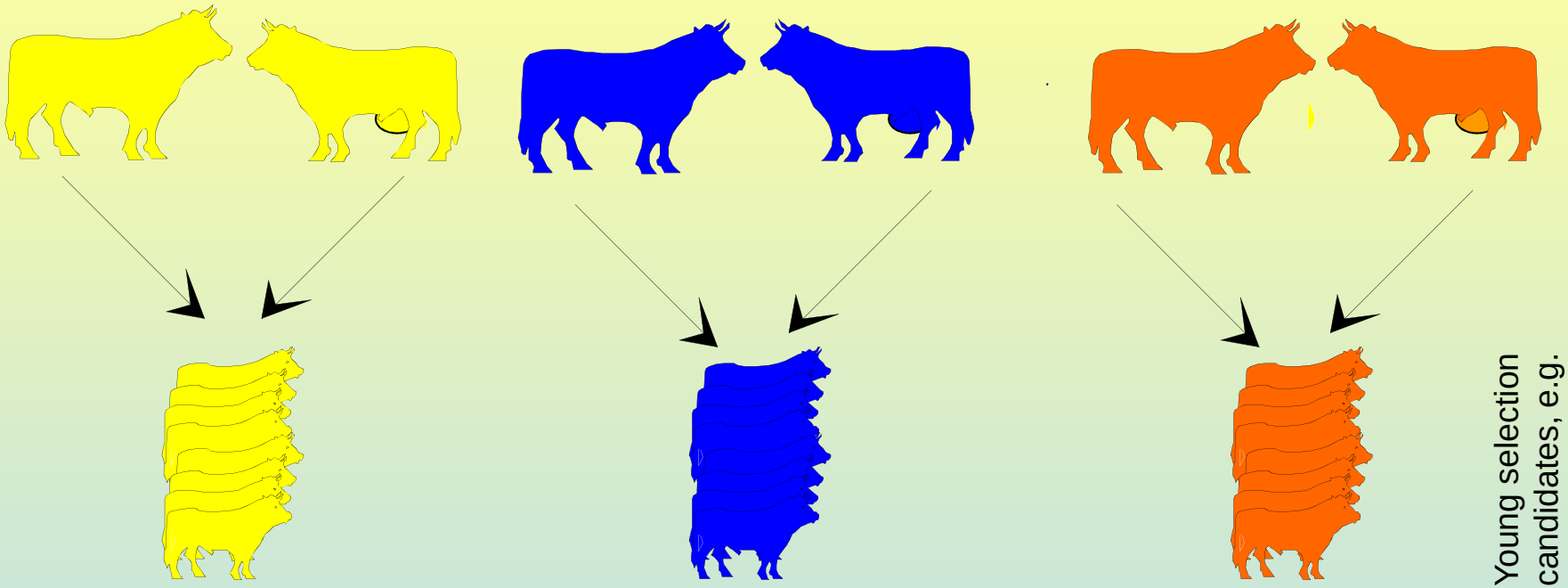
- They assume response in one trait
  - Need whole breeding objective context
- They assume abundant recording of pedigree and gene testing
  - Will we have cheap DNA testing available?
  - We can apply strategies to save on genotyping.
  - Some degree of phase-testing is needed
- They assume gene effects are known
  - Need monitoring by measurement

– Effect of background genes, environment, gene action?

Conclusion on MAS

- Effect on extra gain in breeding programs maybe limited to cases where
  - There are special genes with large effect
    - Disease resistance, Booroola, etc.
  - Breeding objective traits are difficult to measure
    - Some ‘retrospective measurement is needed’
  - When reproductive technologies are used

Between versus within family selection



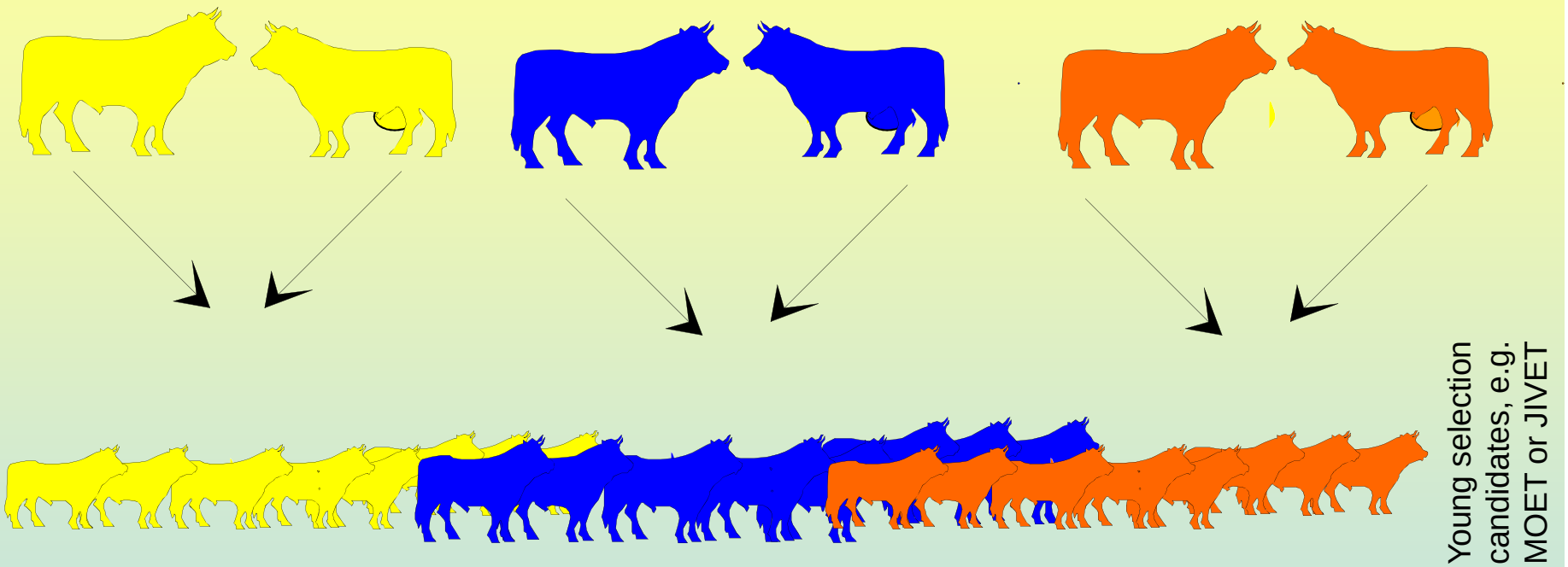
Young selection candidates, e.g. MOET or JIVET

No own information (performance or genotype):

Selection based on parent average

More between-family selection - ***more inbreeding***

Between versus within family selection



Own information (performance or genotype):

More variation within families

More within-family selection – ***less inbreeding***

MAS combined with reproductive technologies

- Genotype testing provides within family information
- Exploiting this variation allows genetic gain without jeopardizing inbreeding

Conclusions

- Marker assisted selection can have some benefit in quantitative trait selection
  - But genetic improvement should be driven by trait and pedigree recording
- Reproductive & gene technology are synergistic
- Main application of gene technologies for ‘special cases’
  - Large and special gene effects, disease resistance
- Gene testing most useful in selection across breeds
  - Introgression / genetic diversity

Whole Genome Association Studies (WGAS)

- A hype in science

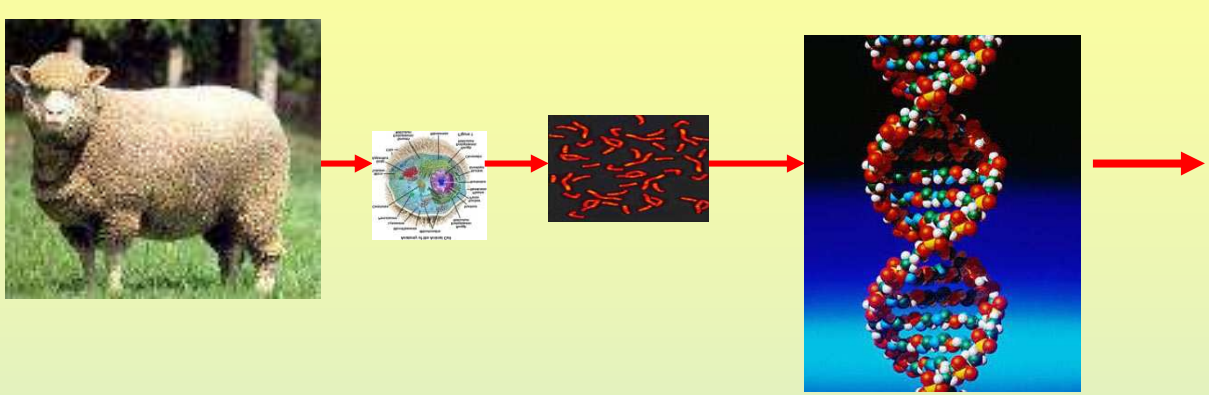
  “ Elucidating the inherited basis of genetic variation in human health and disease is one of the major scientific challenges of the twenty-first century”

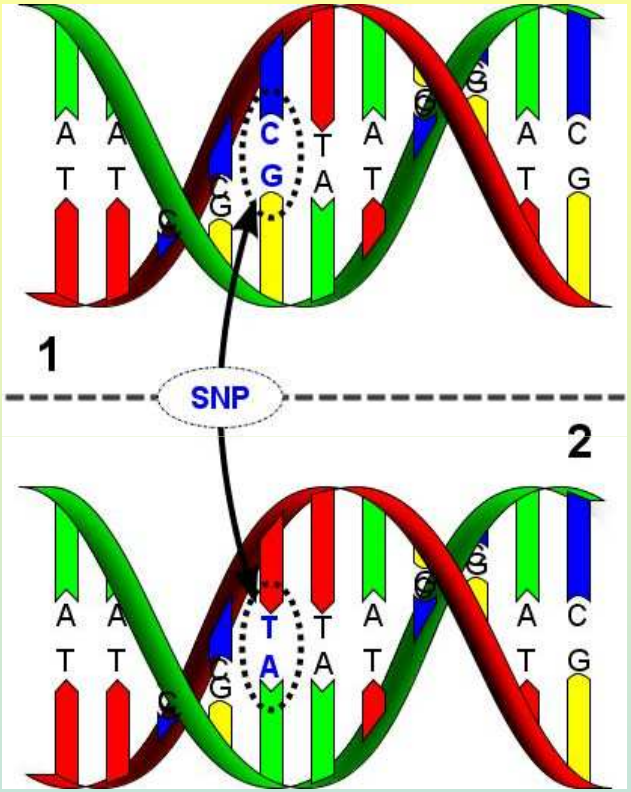
- Main purpose is to find causal variants of diseases

- Additionally find associations with ‘complex traits

  - *Predicting phenotypes* in: Medicine, Forensics, Breeding, Animal and Plant production

SNPs (Single Nucleotide Polymorphism)





DNA code shows subtle variants

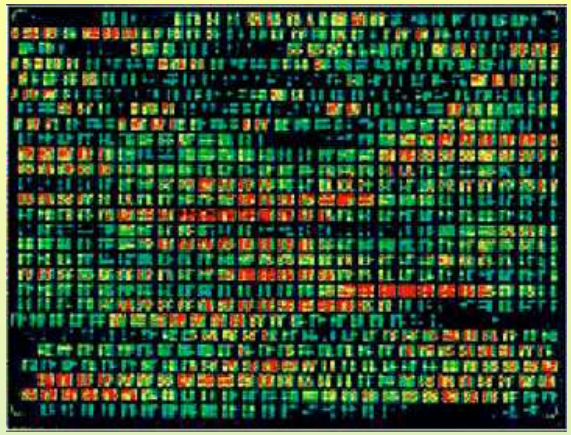
→The basis of genetic variation

These are genetic markers,

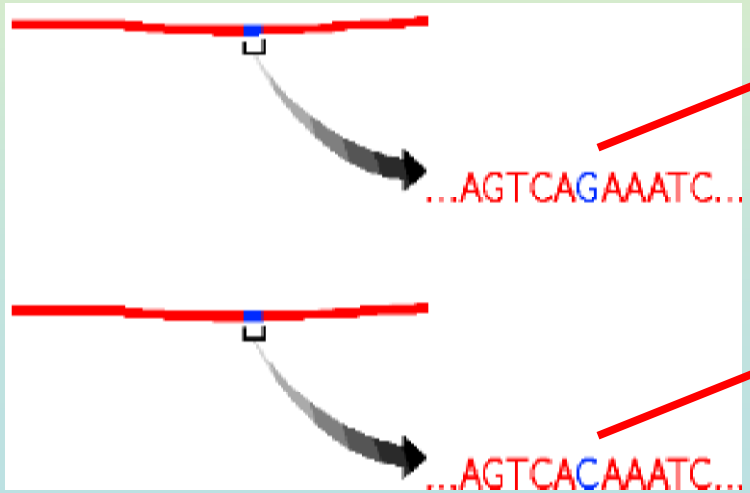
most common ones are called ‘SNPs’

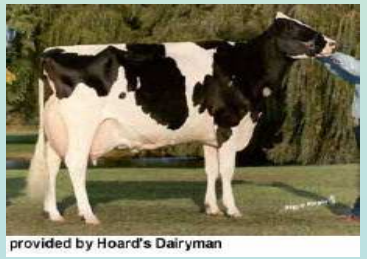
Every 1 out of 450 base pairs is a SNP (there are ~7 million!)

Whole Genome Association Studies



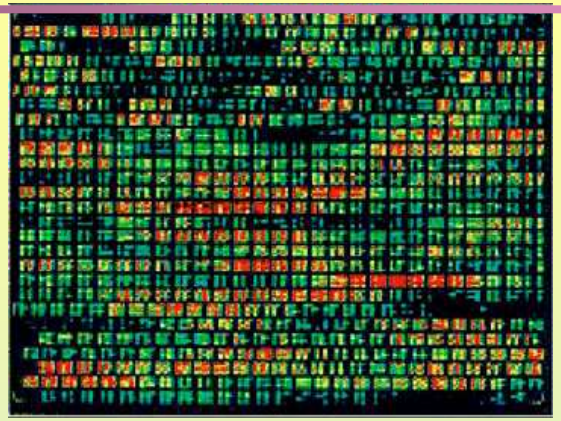
60,000 test for DNA differences, possibly predicting difference in characteristics (or BV)



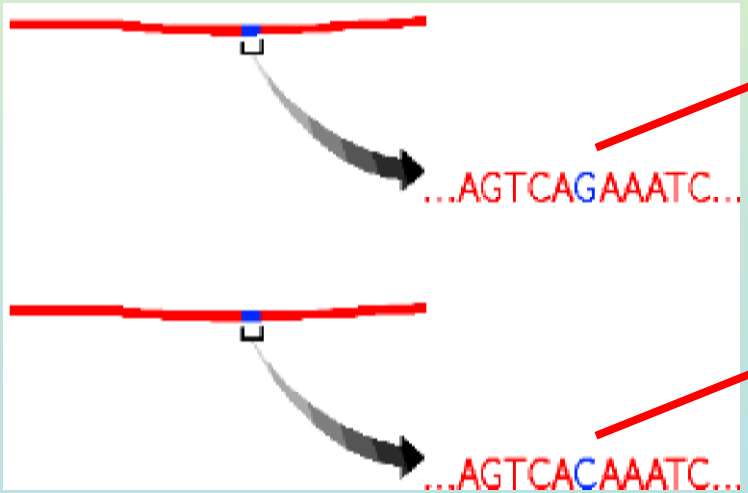


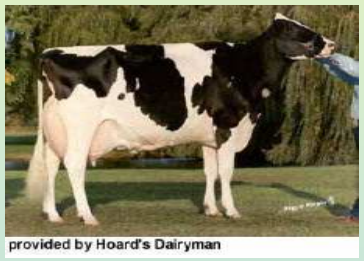
Either in coat colour……

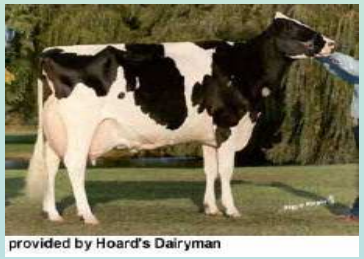
SNP chip

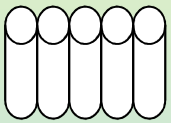


60,000 test for DNA differences, possibly predicting difference in characteristics (or BV)

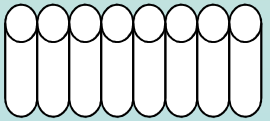




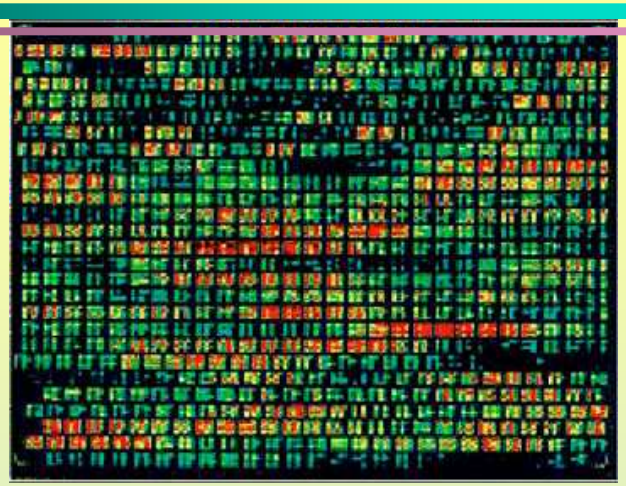




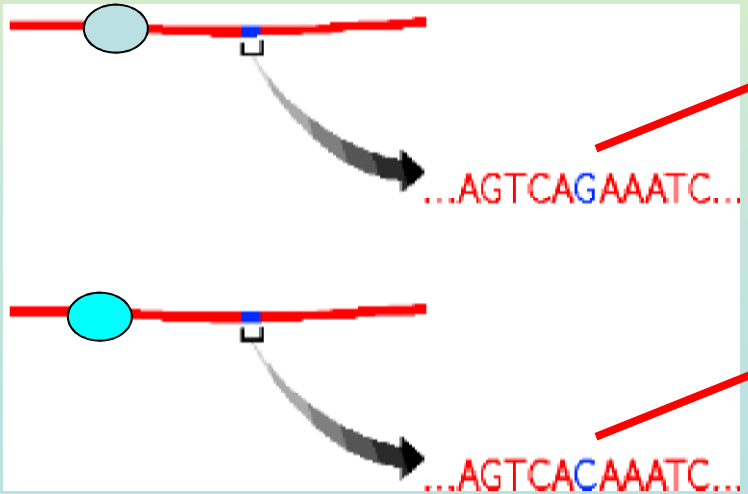
or in production

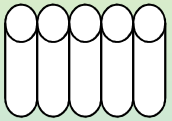


SNP chip

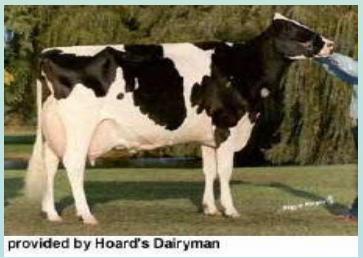


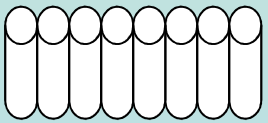
Would also pick up differences in genes close to the SNP marker



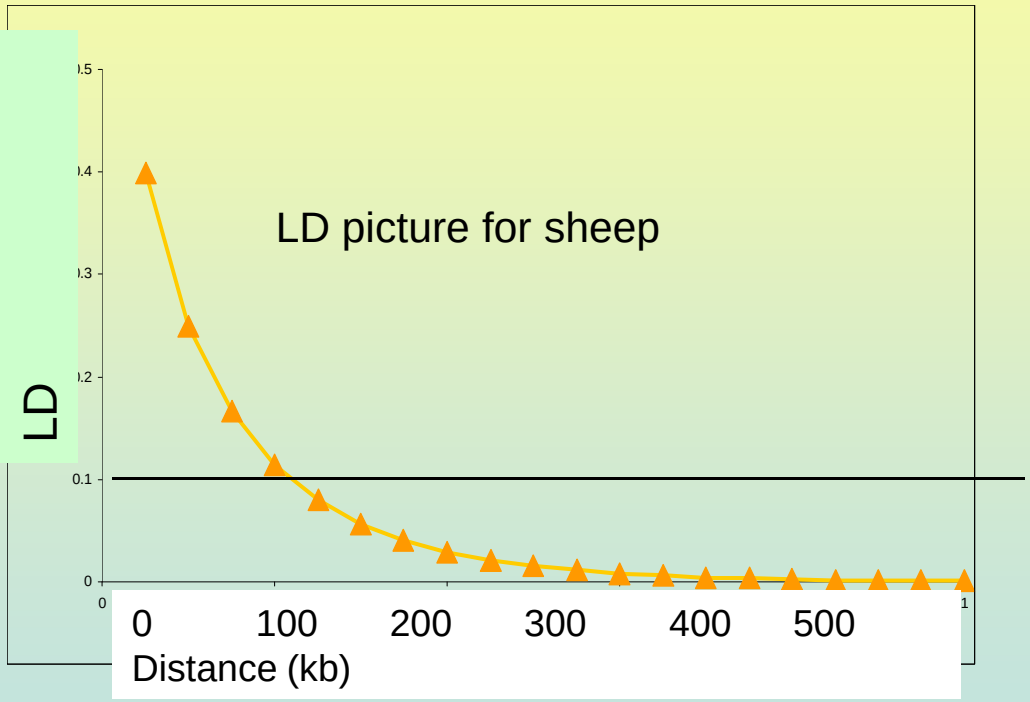


or in production





Why 57,000 SNPs in sheep SNP chip ?



Every marker in
Linkage Disequilibrium (LD)
with a polymorphism

- Threshold LD = 0.2
- → ~ 50k base pairs

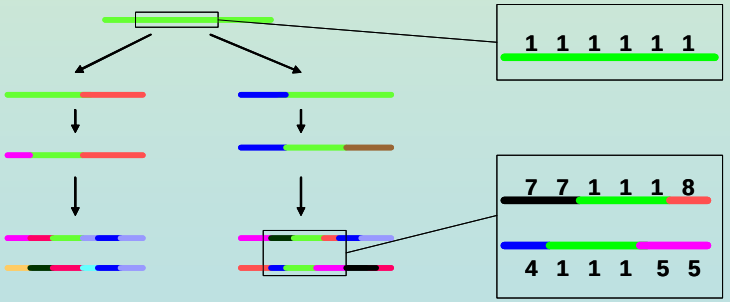
Conclusion: Need about 3 billion/ 50k = 60,000 markers

WGAS challenges

- Analyse 50,000 effects in a dataset with 2000 animals recorded
  - Many false positives
  - Stringent criteria (= low power)
  - How population specific are the predicted effects?

- Methods:
  - Fit SNP effect as random effects
  - BLUP, Bayesian methods, Model selection of SNPs
  - Use genomic relationships matrix

- Concepts
  - Are we predicting effects of genes, or simply contributions of ancestors?



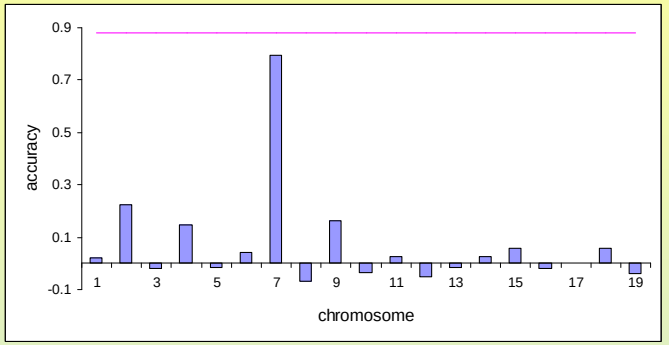
Potential Outcomes of WGAS

- Detection of causal variants
- Prediction of phenotype
- Prediction of breeding values
- Elucidate architecture of genetic variation

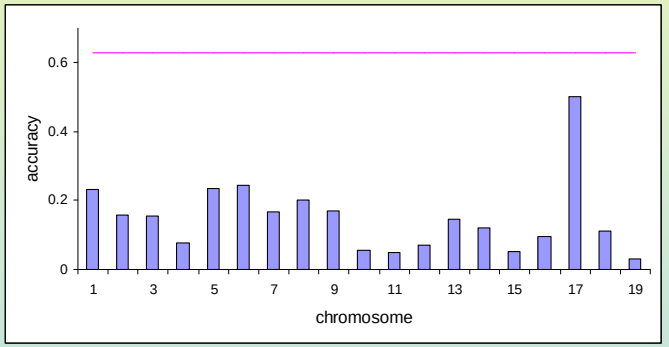
Revealing the architecture of genetic variation

Where are the causes of genetic variation?

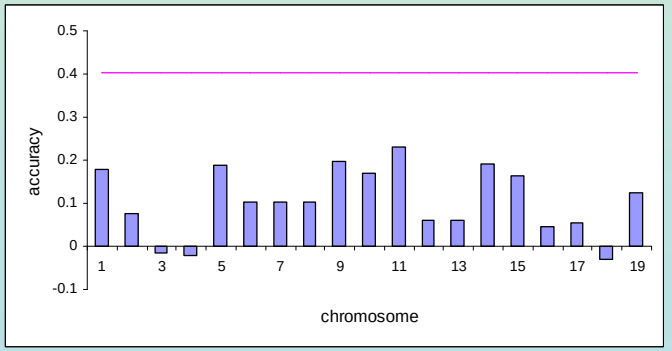
Coat colour



CD8%



MCH



Prediction of phenotype

Genomic data helps predicting phenotypes, even across families

Correlation between actual and predicted phenotypes (SD over 10 reps)

Model	Intra-family wise			Inter-family wise		
	Coat colour	%CD8	MCH	Coat colour	%CD8	MCH
BLUP (Ignoring genotypic data)	0.54 (0.02)	0.64 (0.02)	0.41 (0.01)	0.00	0.00	0.00
Fitting genotypic data and pedigree						
Model A	0.72 (0.02)	0.71 (0.02)	0.52 (0.02)	0.58 (0.06)	0.50 (0.05)	0.35 (0.07)
Model AD	0.89 (0.03)	0.73 (0.02)	0.55 (0.02)	0.87 (0.05)	0.58 (0.05)	0.36 (0.09)
Fitting genotypic data and ignoring pedigree						
Model A	0.65 (0.02)	0.65 (0.02)	0.46 (0.04)	0.54 (0.06)	0.51 (0.05)	0.33 (0.06)
Model AD	0.85 (0.04)	0.69 (0.02)	0.50 (0.04)	0.81 (0.08)	0.56 (0.06)	0.33 (0.09)

STATISTICAL GENETICS

Fitting phenotypes

> An alternative approach to determining gene variants that contribute to a particular trait is to group all SNPs together and ask whether they can predict a phenotype.

Analysing the results of genome-wide association studies is a painstaking effort — each SNP has to pass stringent significance thresholds to be regarded as a respectable candidate. An alternative approach to determining gene variants that contribute to a particular trait is to group all SNPs together and ask whether they can predict a phenotype. One such method, based on a Bayesian approach, has now been used to predict three mouse phenotypes. Similar approaches could be useful in other areas of medical genetics as well as in forensics and artificial selection in livestock.

Bayesian approaches are well suited to the prediction of phenotypes. The aim is not to test hypotheses but to estimate the effect of each SNP and to combine all the SNP effects into a prediction of phenotype that is as accurate as possible. In this paper, the authors have tested the feasibility of using a Bayesian approach called reversible jump Markov chain Monte Carlo (RJMCMC) on genome-wide SNPs to predict three phenotypes in heterogeneous stock mice — coat colour, the percentage of $CD8^+$ cells, and mean cellular haemoglobin (see the link for a description of how these mice were constructed).

The data came from four generations of mice, over 2,000 animals, and consisted of 10,000 SNPs as well as pedigree and phenotype information. Genetic models were developed based on the full genotypic data but using the phenotypes of only half the animals, and then they were validated by predicting phenotypes in the remaining half of the population. The models incorporated either additive effects only or a mixture of additive and dominance effects (the AD model).

Predictions were successful across all traits — accuracy ranged from 0.4 to 0.9 — with AD models being superior to additive-only models; for example, coat-colour predictions are 81% accurate under the AD model. More accurate predictions were obtained with traits, such as $CD8^+$ percentage, that are more heritable — that is, for which more of the trait variation between individuals actually depends on genetic factors. Phenotypes were predicted across families but also within families; in the latter case, predictions were enriched by pedigree information and therefore performed better.

Using genome-wide information gave a marked improvement in accuracy over using single SNPs or even entire chromosomes at a time. The high accuracy, computational efficiency and speed of the analysis method (this data set took 15 minutes to analyse) means that it could be adapted for use on additional traits and larger samples, and for other species and applications. This paper builds on previous work by the authors that demonstrated the use of dense SNP genotypes to predict genetic value in livestock and disease risk in humans.

Tanvita Casci

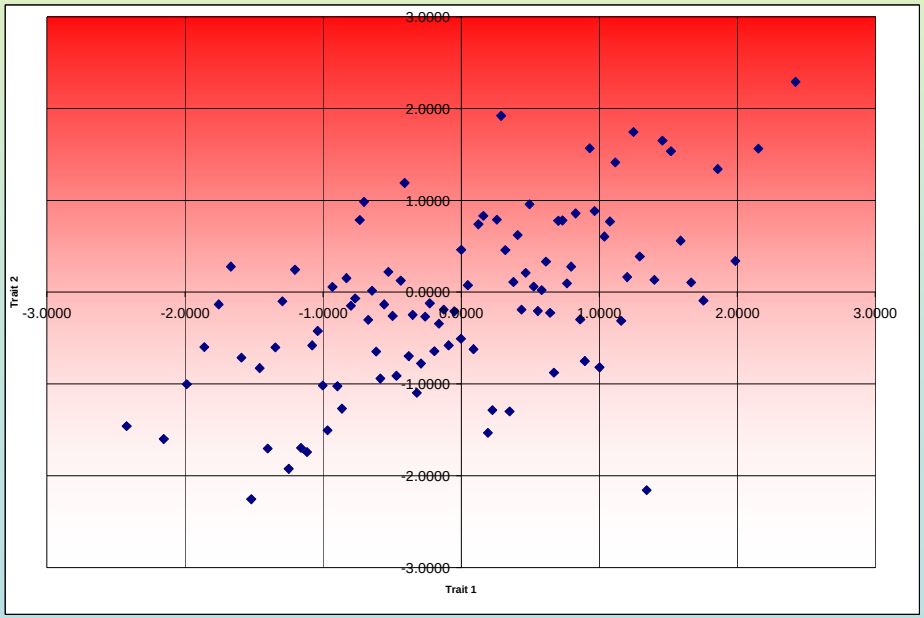
ORIGINAL RESEARCH PAPER Lee, S. H. et al. Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genet.* 4, e1000231 (2008)

WEB SITE

Heterogeneous stock mice: http://gsean.well.ox.ac.uk

Prediction of breeding value

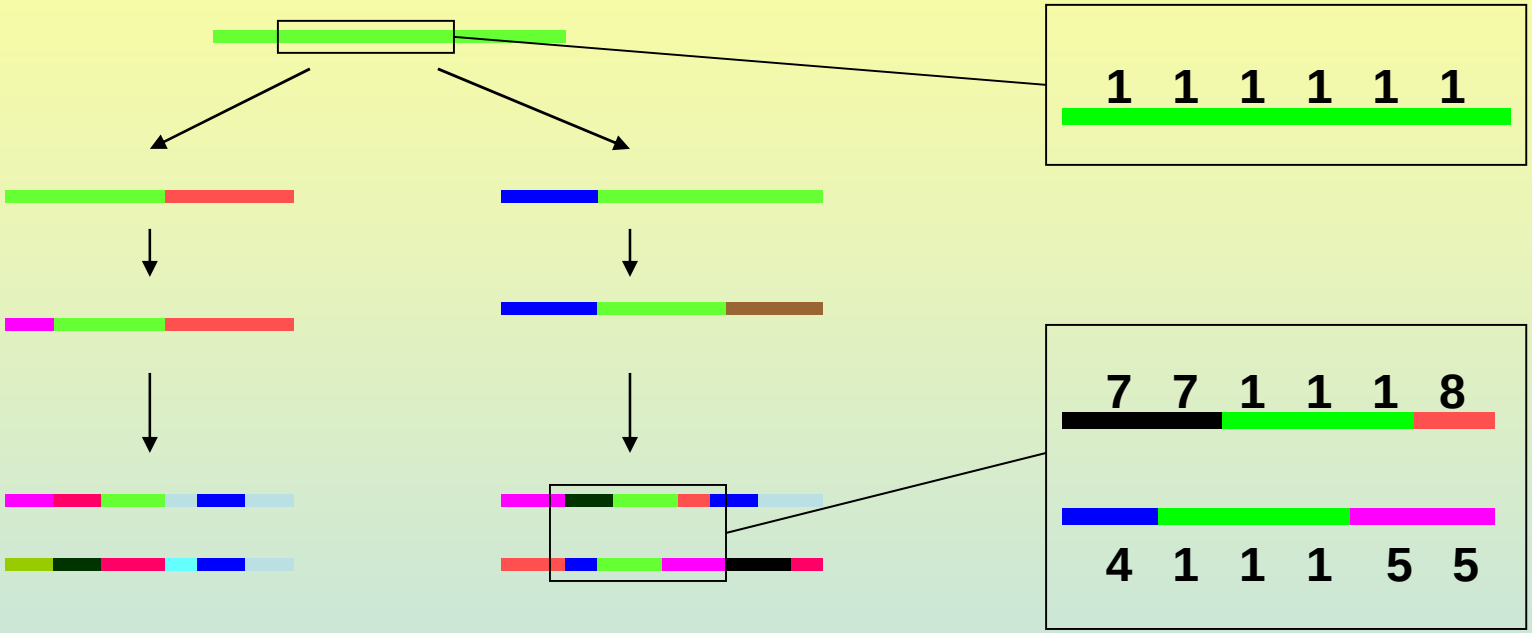
- 2007: Dairy CRC and DPIVic 798 Australian dairy bulls
- Reference group: Genotyped to predict effects
- Validation group: Genotyped and predict EBV and compare with progeny test



→ Correlation ~ 0.6

Genomic Selection

Dense markers allows to capture the right pieces of the genome

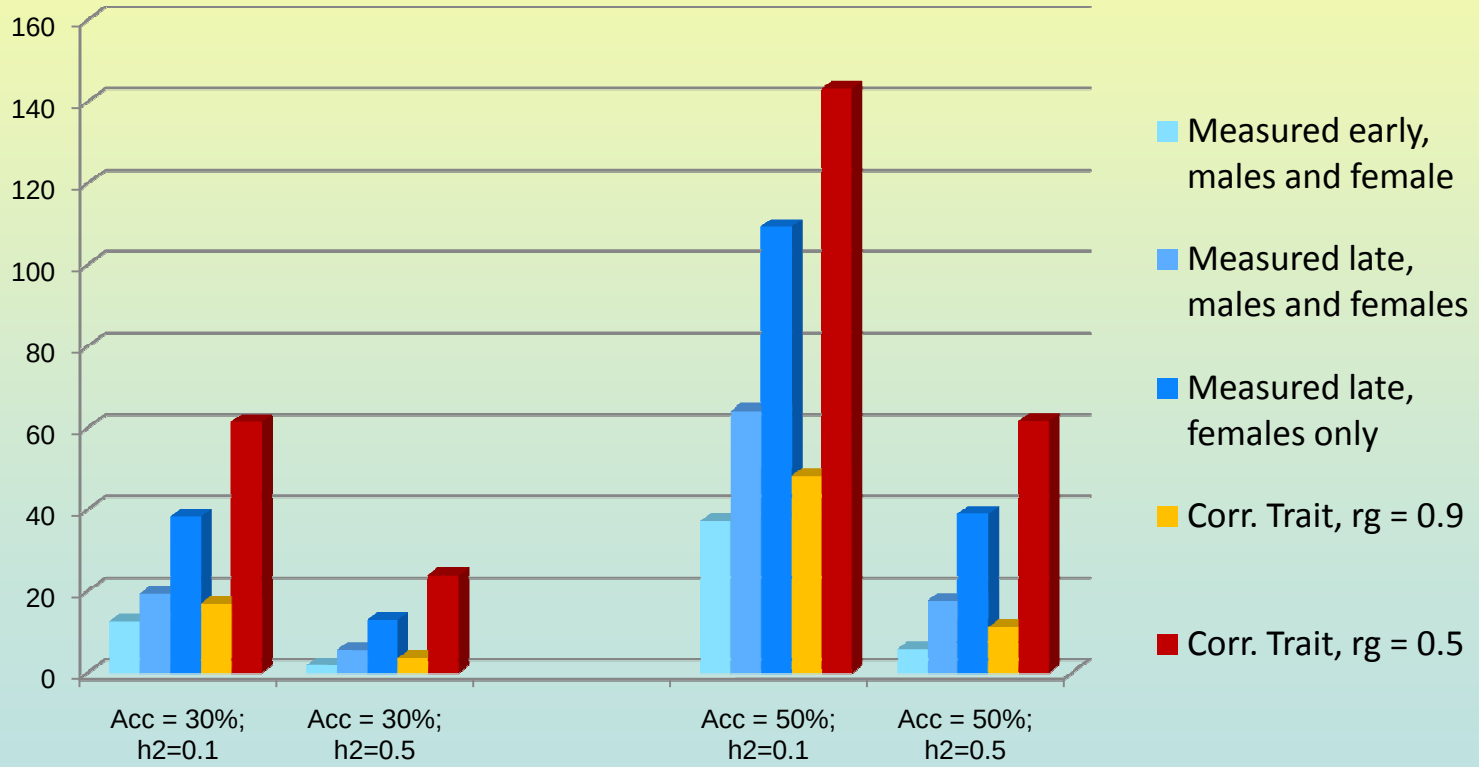


Allows to predict EBV of young animals accurately!

Back to the black box!

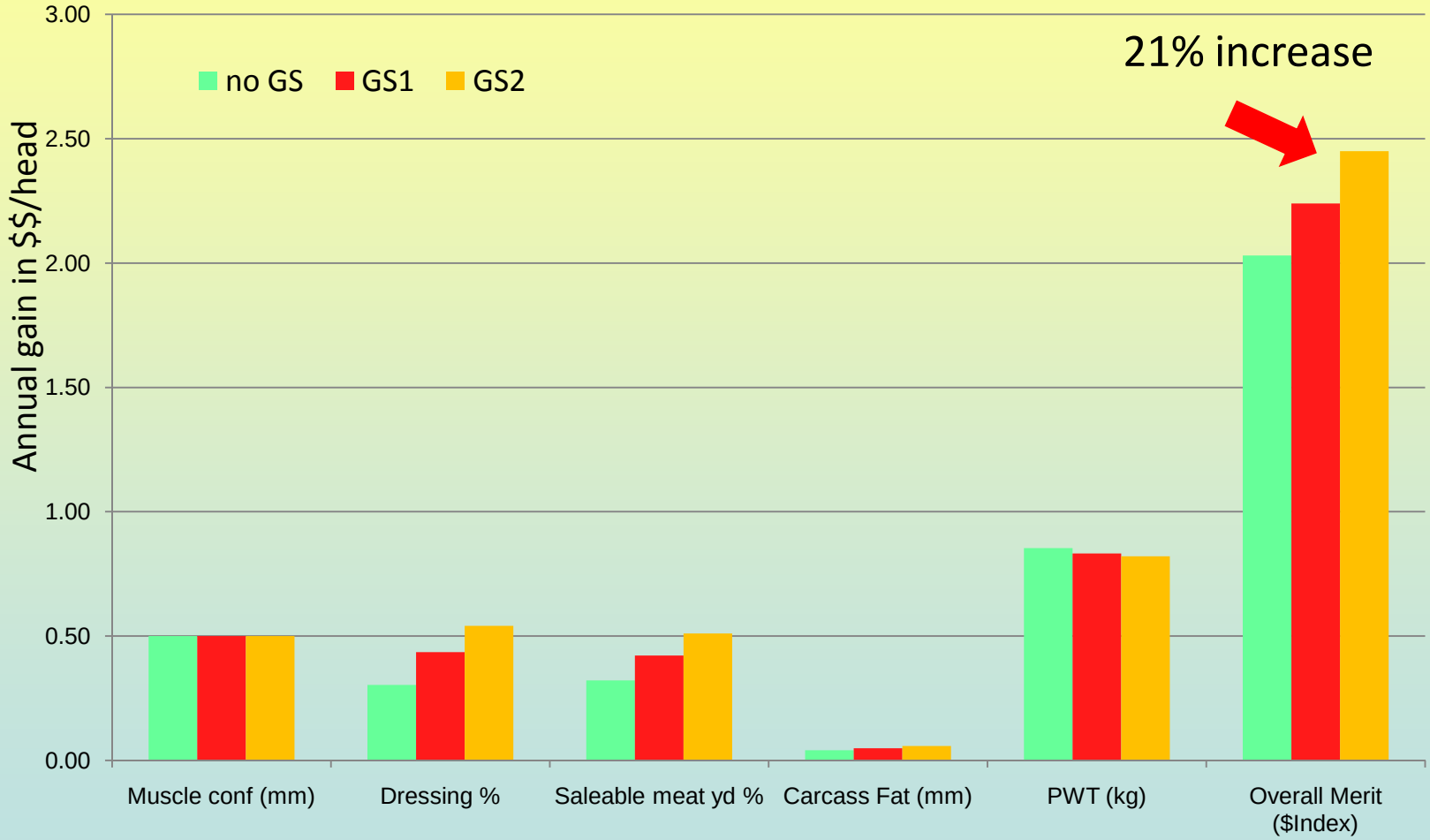
Percent increase in rate of genetic gain when using genomic selection

Selection on a single trait



Relative change of response for multiple traits – meat index

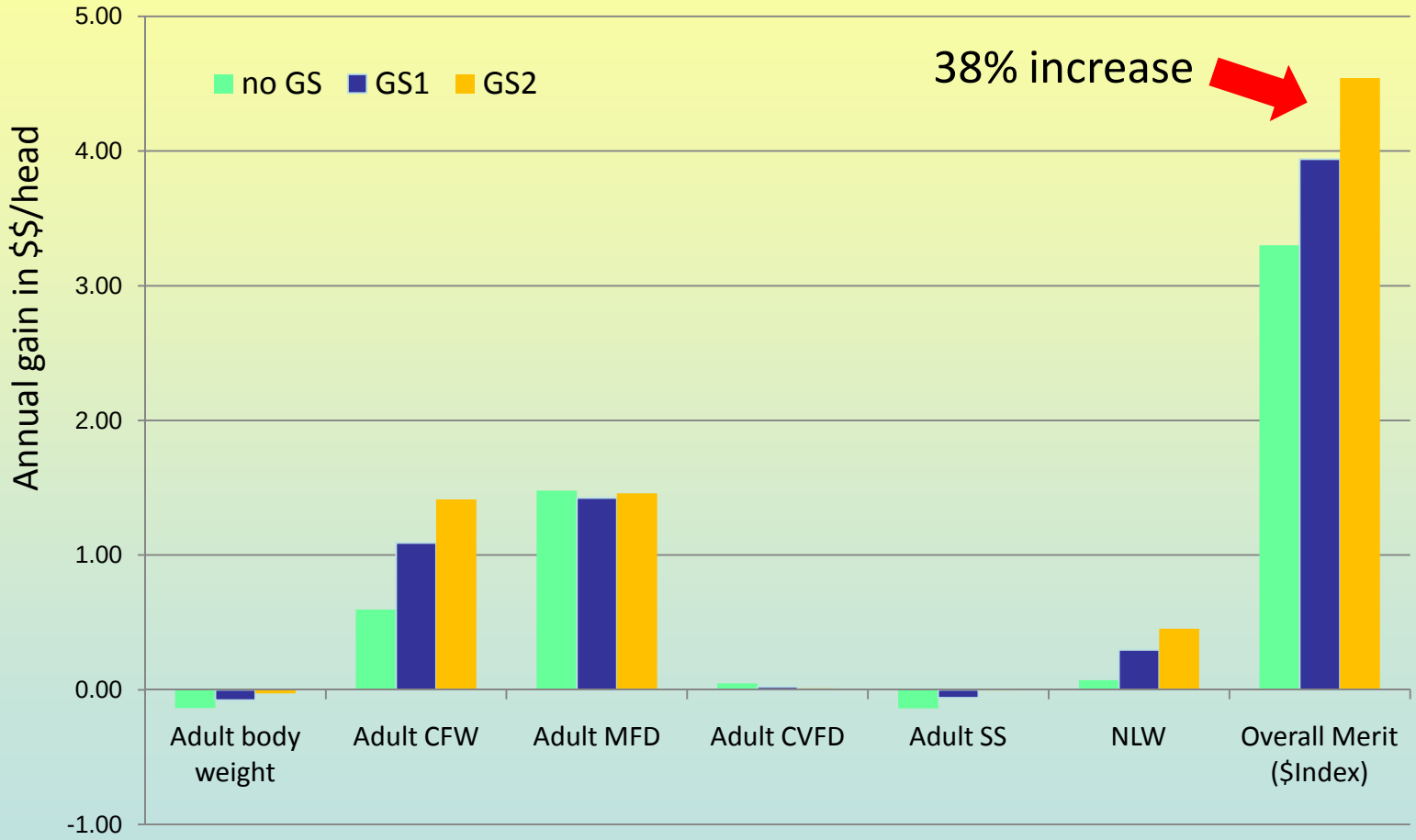
annual gain in $$



GS1: $V_{qtl}/V_A = \frac{1}{2}h^2$ → WGAS experiment needs ~1000 animals

GS2: $V_{qtl}/V_A \sim h^2$ → WGAS experiment needs ~2500 animals

Relative change of response for various traits – wool index

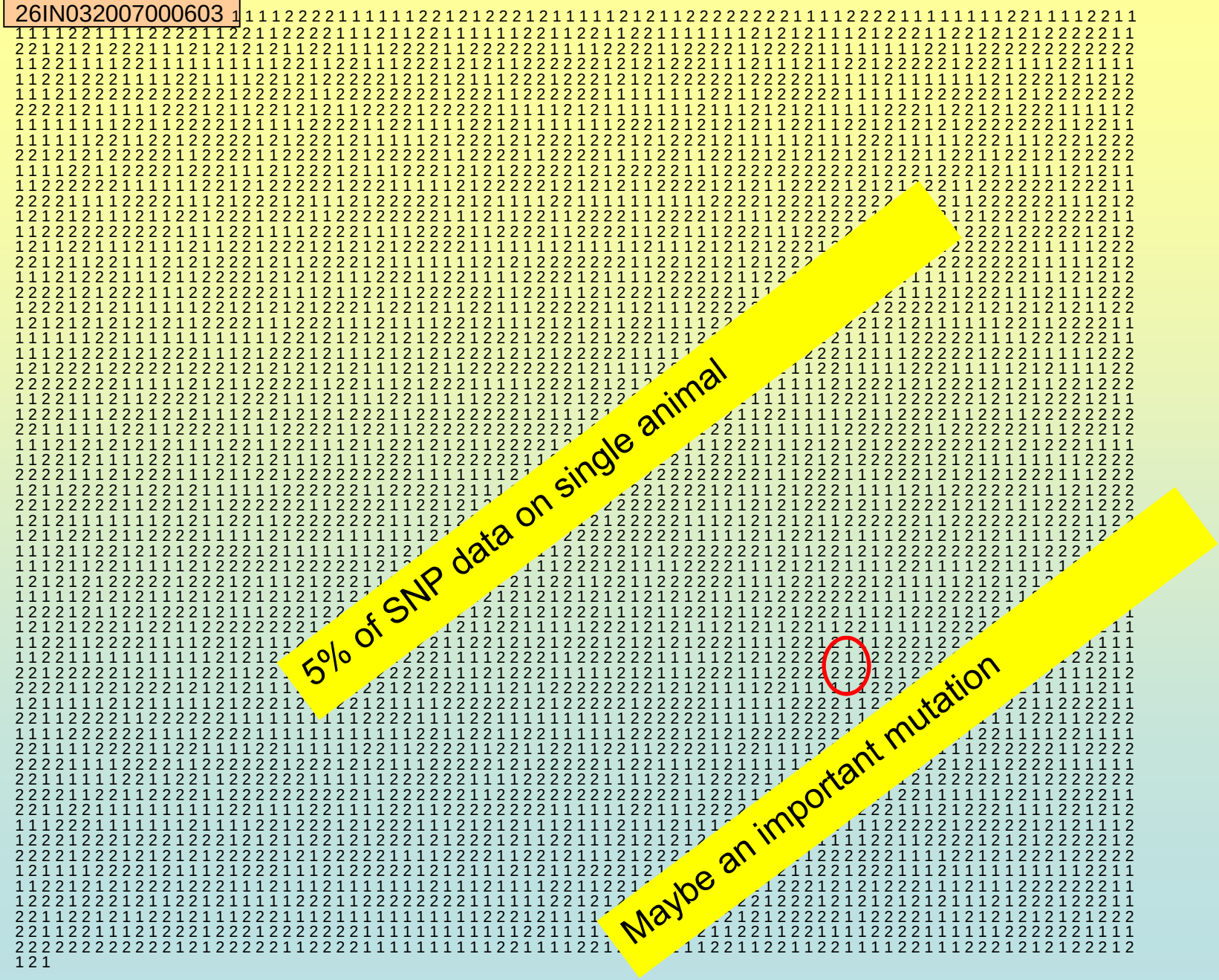


GS1: $V_{qtl}/V_A = \frac{1}{2}h^2 \rightarrow$ WGAS experiment needs ~1000 animals

GS2: $V_{qtl}/V_A \sim h^2 \rightarrow$ WGAS experiment needs ~2500 animals

Animal ID: Carcass weight = 24 kg

- 26IN032007000603



Conclusion

- SNP chips is likely to become a key tool for animal selection
- Beneficial for traits that are hard to measure (early in life)
- No pedigree information needed (?)