Review of quantitative genetics

...

Evangelina López de Maturana & Oscar González-Recio

Something you need to carefully look at, or that may impair your GWP

Something to do, or that optimizes your GWP

Don’t. Discourage to use this.

Smart tip. Something that makes the trick.

Advanced. Something to dive in.

Challenges

What you need to know from this lecture

Basic concepts

How genes and environment modulate the phenotype

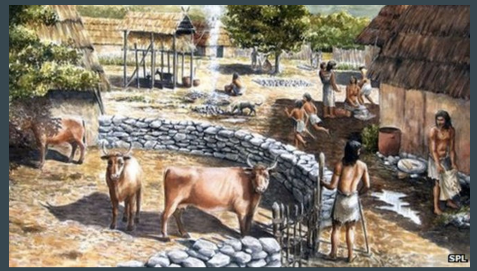
What is genomic heritability and how it affects GWP

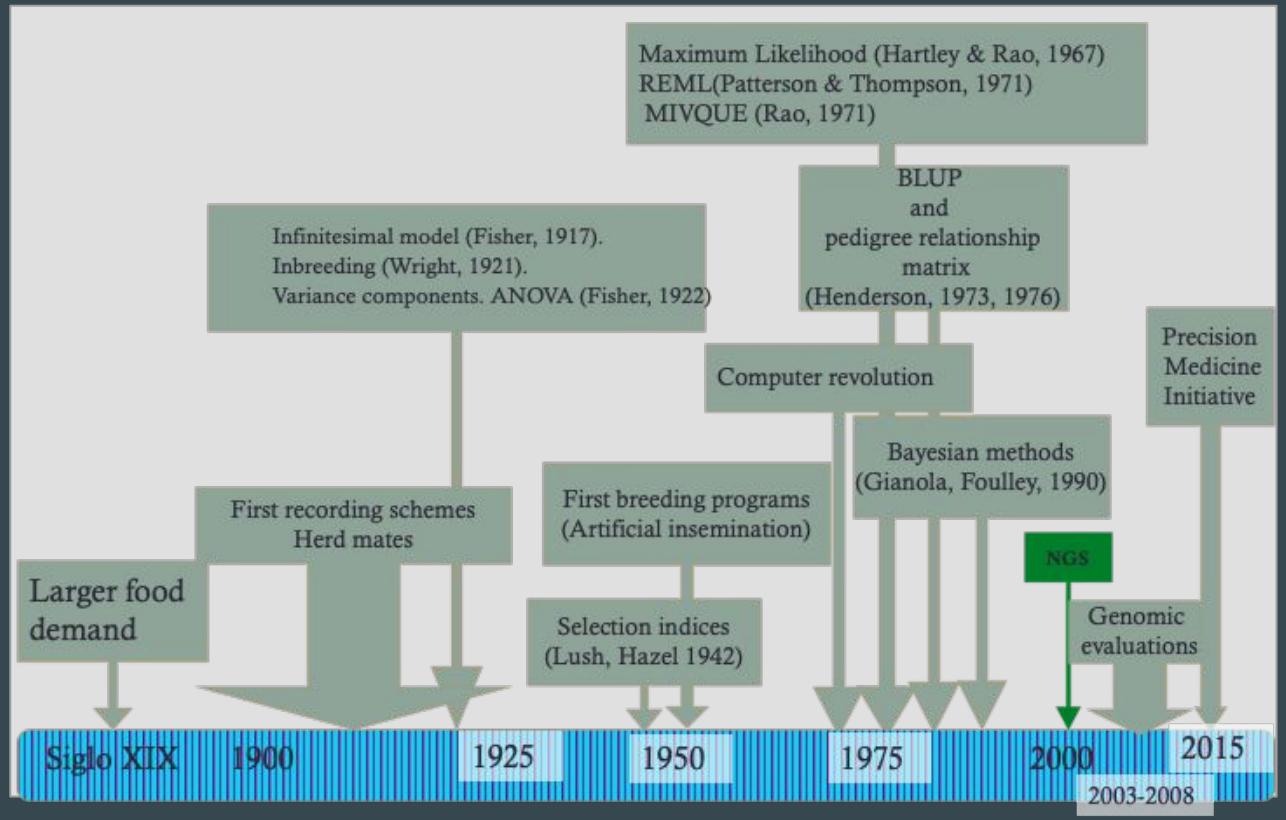
Interpret what a GWP implies

Genotyping strategies

Phenotyping strategies

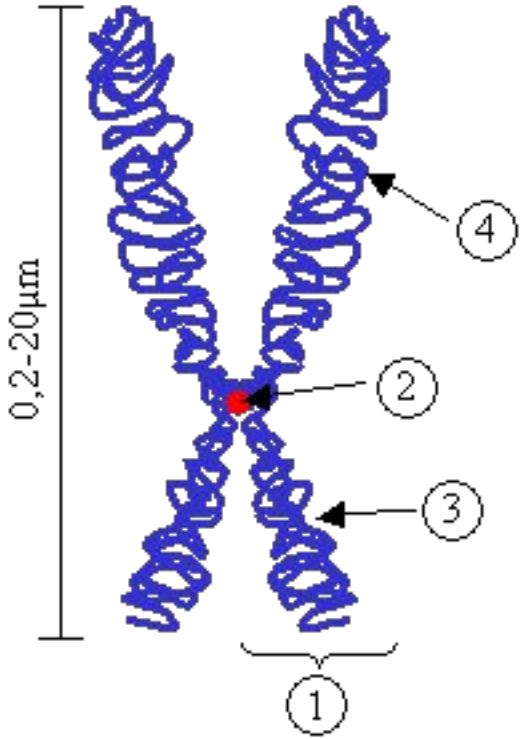
A bit of history





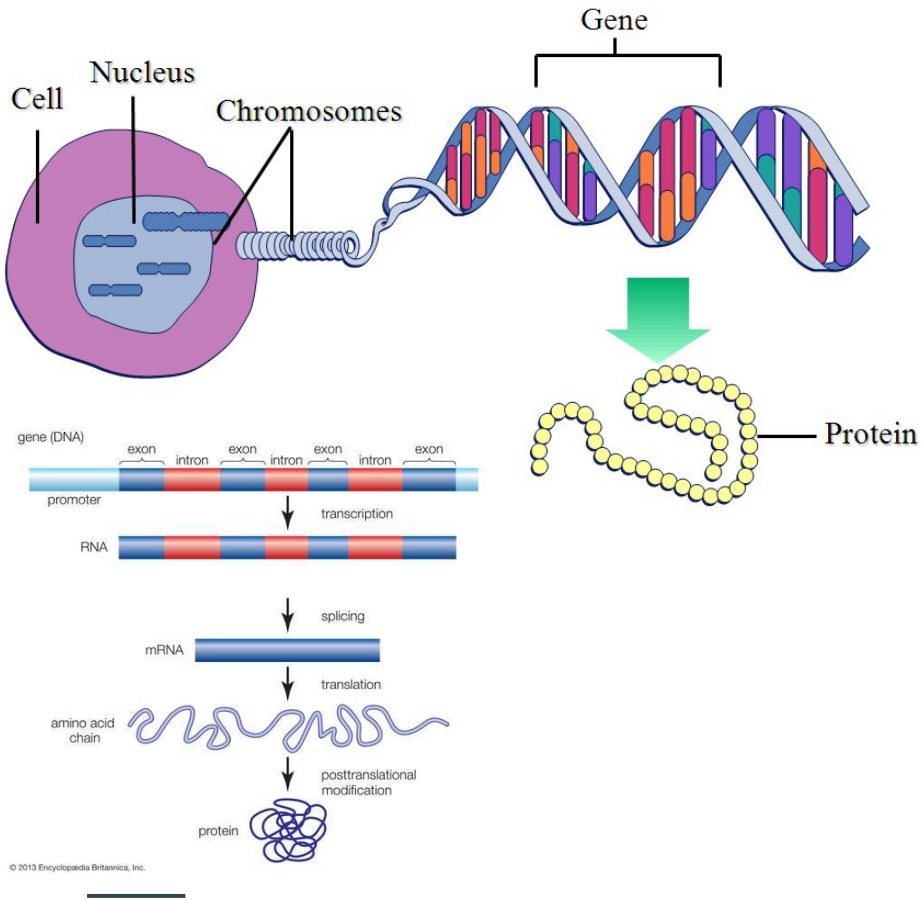
Locus, loci

A specific physical location of a gene, DNA sequence or genetic marker on a chromosome; like a genetic street address



Gene

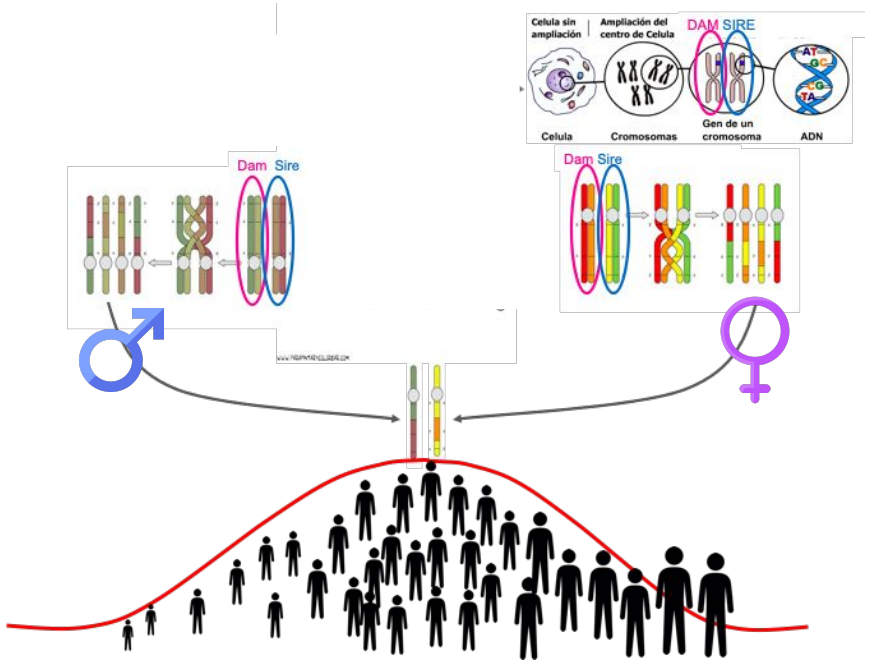
Gene, unit of hereditary information that occupies a fixed position (locus) on a chromosome. Genes achieve their effects by directing the synthesis of proteins.



Genome-wide prediction

Mendelian effect

Deviation from the expected parent average



$$a_i = \frac{1}{2}a_p + \frac{1}{2}a_m + \phi_i$$

$$\phi_i \sim N(0, \sigma_a^2 / d_i)$$

Genome-wide prediction

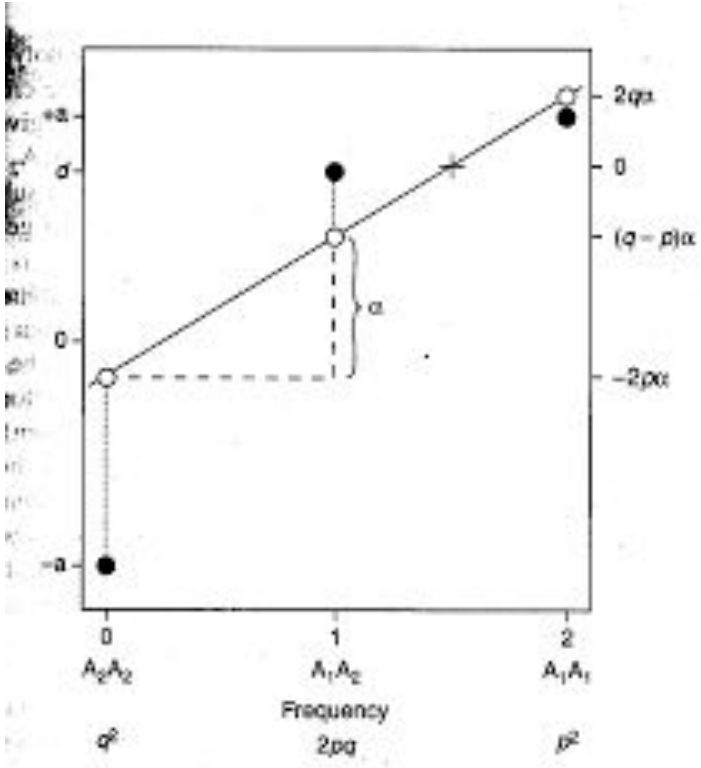
Pedigree index

Parent average

$$1/2\ EBV_{sire} + 1/2\ EBV_{dam}$$

⚠ Allele substitution effect

The effect that the presence of a copy of an allele has on the phenotype (regarding the reference allele).



f(A)= mean(Aa)-mean(aa)

Pleiotropy

the phenomenon in which a single locus affects two or more apparently unrelated phenotypic traits

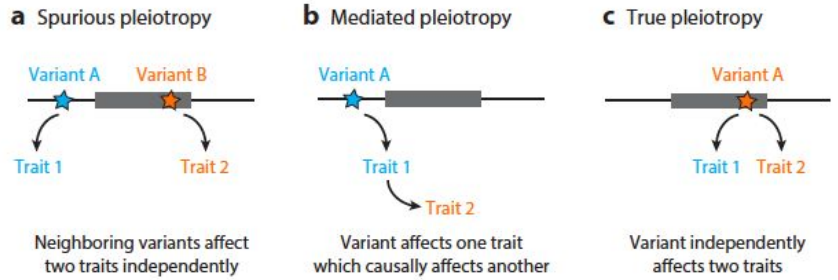


Figure 4

Diagrams illustrating (*a*) spurious pleiotropy, in which two neighboring, separately causal variants (*blue* and *orange stars*) are mistakenly inferred to be pleiotropic because they cannot be statistically distinguished; (*b*) mediated pleiotropy, in which a variant is statistically associated with two traits because it has a causal effect on one trait that in turn causally impacts another; and (*c*) true pleiotropy, in which a single unambiguous causal variant is separately biologically causal for two independent traits.

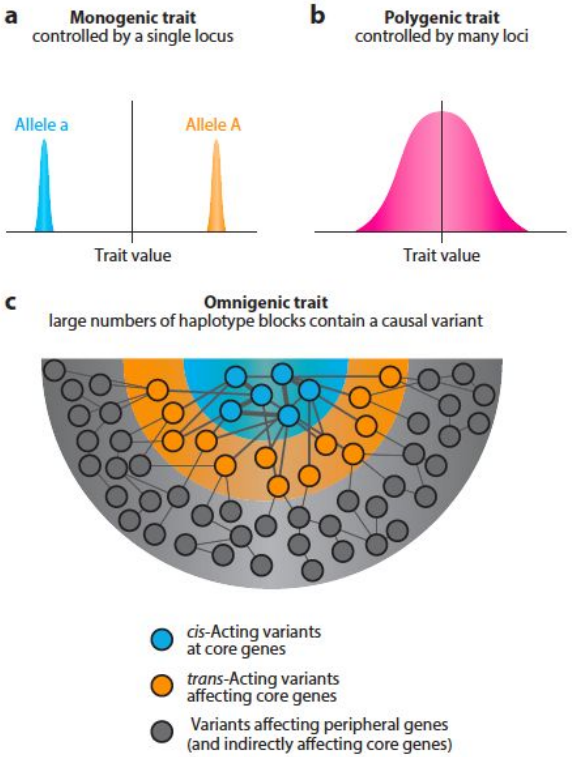


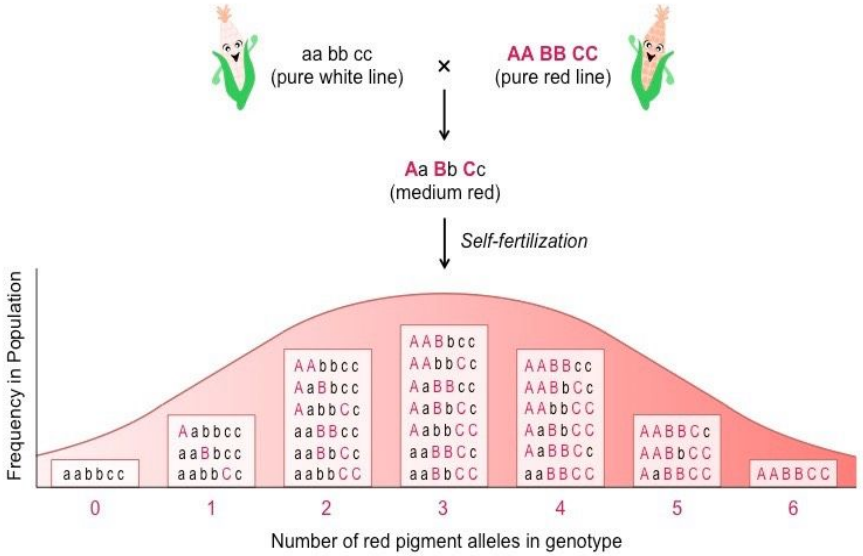
Figure 1

(*a*) A bimodal trait distribution for a monogenic trait controlled by a single genetic locus, as compared to (*b*) a continuous trait distribution for a polygenic trait controlled by many genetic loci. (*c*) Schematic of one possible architecture for an omnigenic trait, in which several large-effect *cis*-acting and many smaller-effect *trans*-acting variants modulate a set of core genes, as does a much larger ensemble of *cis*- and *trans*-acting variants impacting peripheral genes that only indirectly modulate the phenotype.

Genome-wide prediction

Infinitesimal model

A quantitative trait is influenced by an infinitely large number of genes, each of which makes an infinitely small (infinitesimal) effect, as well as by environmental factors. Random sampling of alleles at each gene produces a continuous, normally distributed phenotype in the population (at least around the average of that of the individual's parents).

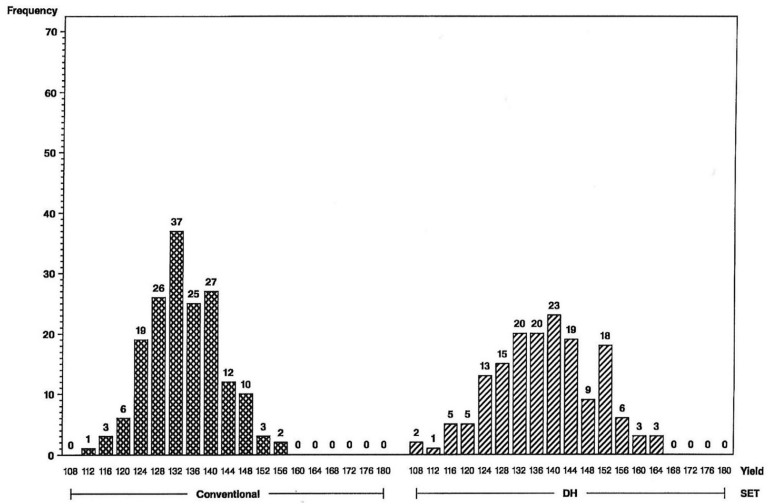


Probability of j major alleles in k biallelic loci $\Rightarrow$

$$\Rightarrow \binom{2k}{j} = \left(\frac{1}{2}\right)^{2k} \frac{2k!}{j!\,(2k-j)!}$$

Genetic variance

Phenotype deviation from the mean phenotype caused by the combination of alleles inherited from parentals and these alleles independent effects on the specific phenotype

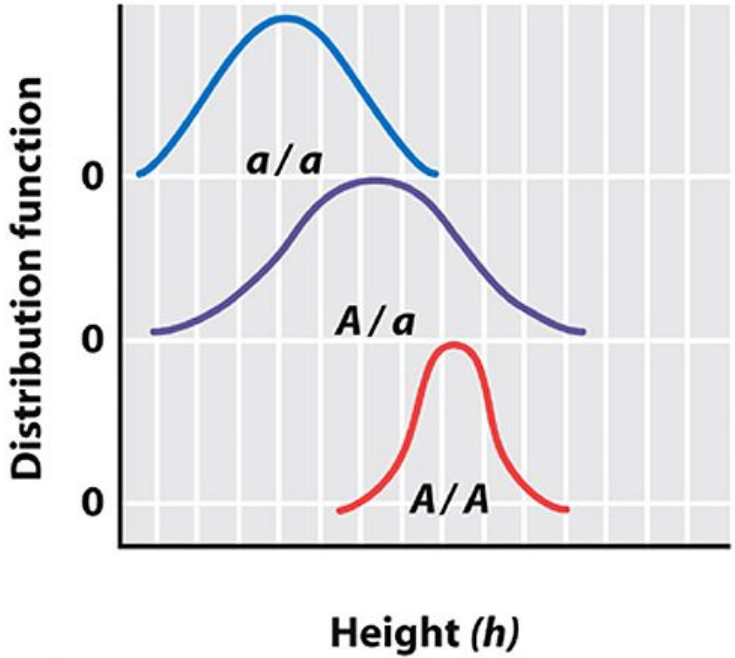


Genome-wide prediction

Phenotype decomposition

Phenotype is affected by genetic (additive +dominance +epistasis), environment and their interactions.

$$P = G + E$$



Heritability

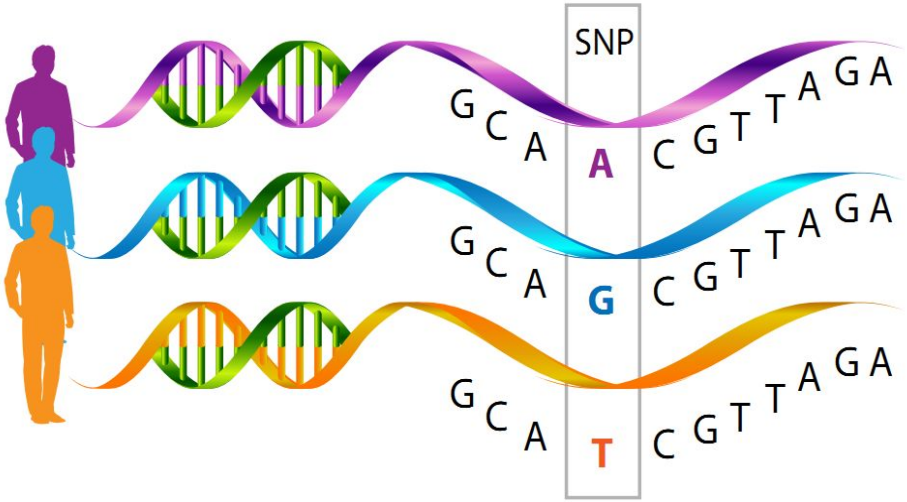
The amount of phenotypic (observable) variation in a population that is attributable to individual genetic differences

“Narrow and broad sense”

$$H^2 = \frac{V_g}{V_g + V_e}$$

Genetic marker

DNA sequence with a known location on a chromosome that can be used to identify individuals or species. It can be described as a variation (which may arise due to mutation or alteration in the genomic loci) that can be observed



Marker variance

Phenotype deviation from the mean phenotype caused by the inheritance of a particular allele from parentals and this allele's independent effect on the phenotype

Notation	Variance component	Genotype coding
V_A	$2pq[a + d(p - q)]^2$	$x_A \in \{0, 1, 2\}$
V_D	$(2pqd)^2$	$x_D \in \{0, 2p, 2(p - q)\}$
V'_D	$\frac{4pq^2}{1+q}(a + dq)^2$	$x'_D \in \{0, 2, 2\}$
V'_A	$\frac{2p^2q}{1+q}(a - d)^2$	$x'_A \in \{0, \frac{1-q}{1+q}, \frac{-2q}{1+q}\}$
V''_{AA}	computed numerically	$x''_{AA} \in (x_{A,1} - 1)(x_{A,2} - 1)$

Genomic variance

The amount of variance explained by marker effects (<genetic variance, because of incomplete LD with QTLs or missingness)

$$
\begin{aligned}
Var(\beta' x_i) &= \beta' Cov(x_i, x_i') \beta \\
&= \beta' \Sigma_x \beta \\
&= \alpha' \Sigma_{zx} \Sigma_x^{-1} \Sigma_x \Sigma_x^{-1} \Sigma_{xz} \alpha \\
&= \alpha' \Sigma_{zx} \Sigma_x^{-1} \Sigma_{xz} \alpha
\end{aligned}
$$

Genomic heritability

The proportion of variance of a trait that can be explained (in the population) by a linear regression on a set of markers

$$h_g^2 = \frac{\sigma_g^2}{\sigma_y^2} = \frac{\sigma_a^2}{\sigma_y^2}\frac{\sigma_g^2}{\sigma_a^2} = h^2\frac{\sigma_g^2}{\sigma_a^2}$$

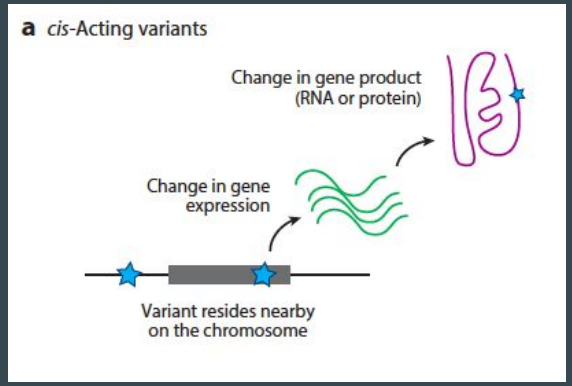
$h^2_g <= h^2$

Missing heritability

The problem of missing heritability, that is to say the gap between heritability estimates from genotype data and heritability estimates from twin data

The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. Brendan Maher shines a light on six places where the missing loot could be stashed away.



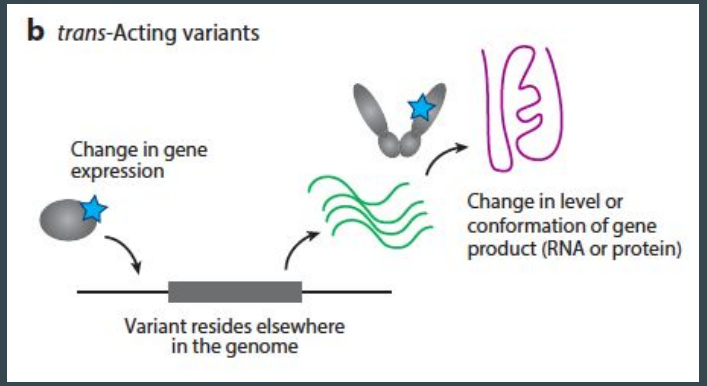
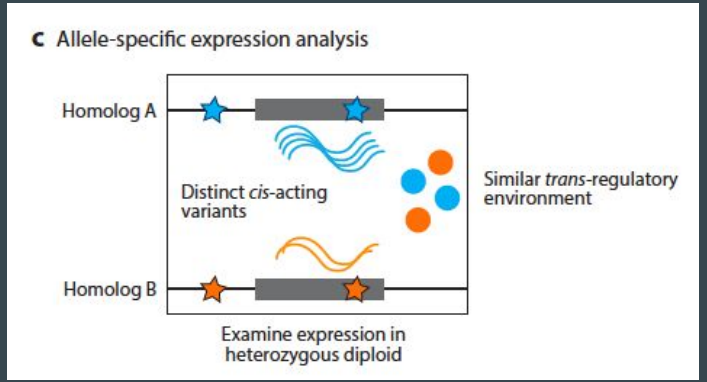
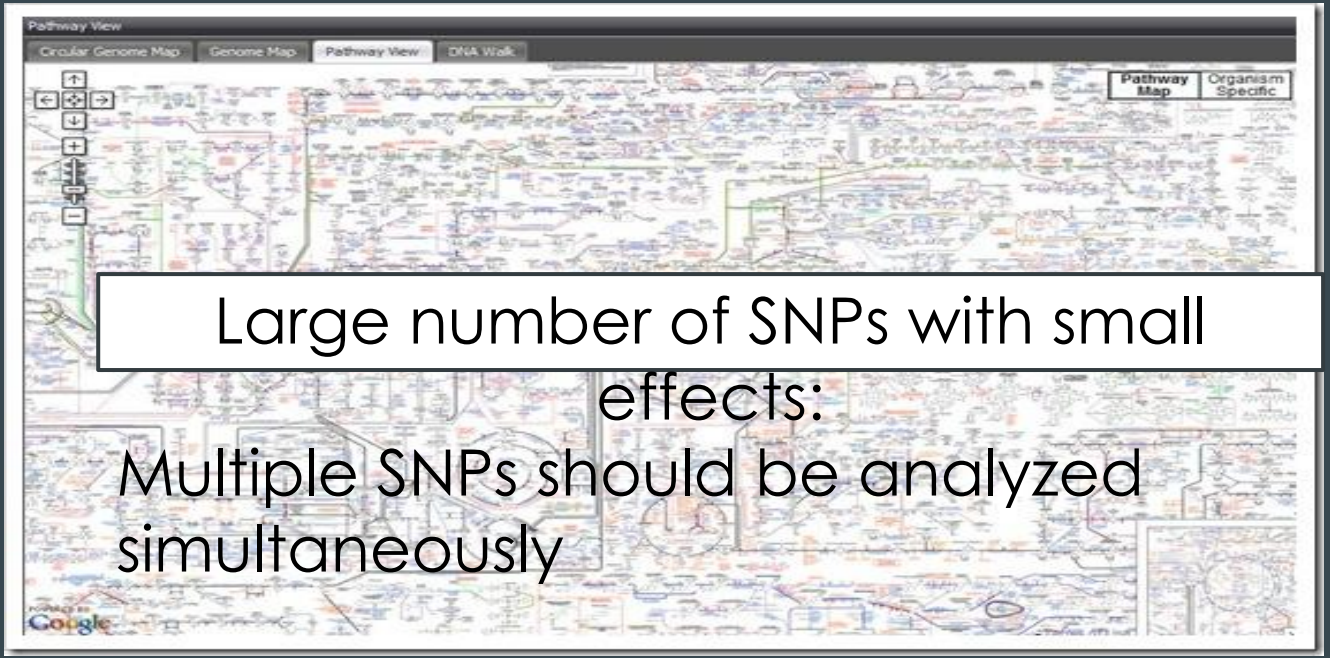


Figure 2

(*a*) *cis*-Acting variants that impact the expression of a gene immediately proximal on the chromosome. (*b*) *trans*-Acting variants that impact a gene product originating from a distal genetic locus. (*c*) Schematic of an experimental design to measure allele-specific mRNA levels. Due to the presence of both parental alleles in the F_0 heterozygote, *cis*-acting regulatory activity is inferred from differential expression of the messenger RNA attributable to one of the two homologous loci. This is because both homologs exist in an essentially equivalent *trans*-regulatory environment; any difference in abundance must therefore be due to a nearby *cis*-acting variant.



Jakobson and Jarosz, 2020



Large number of SNPs with small effects:

Multiple SNPs should be analyzed simultaneously

Reference population

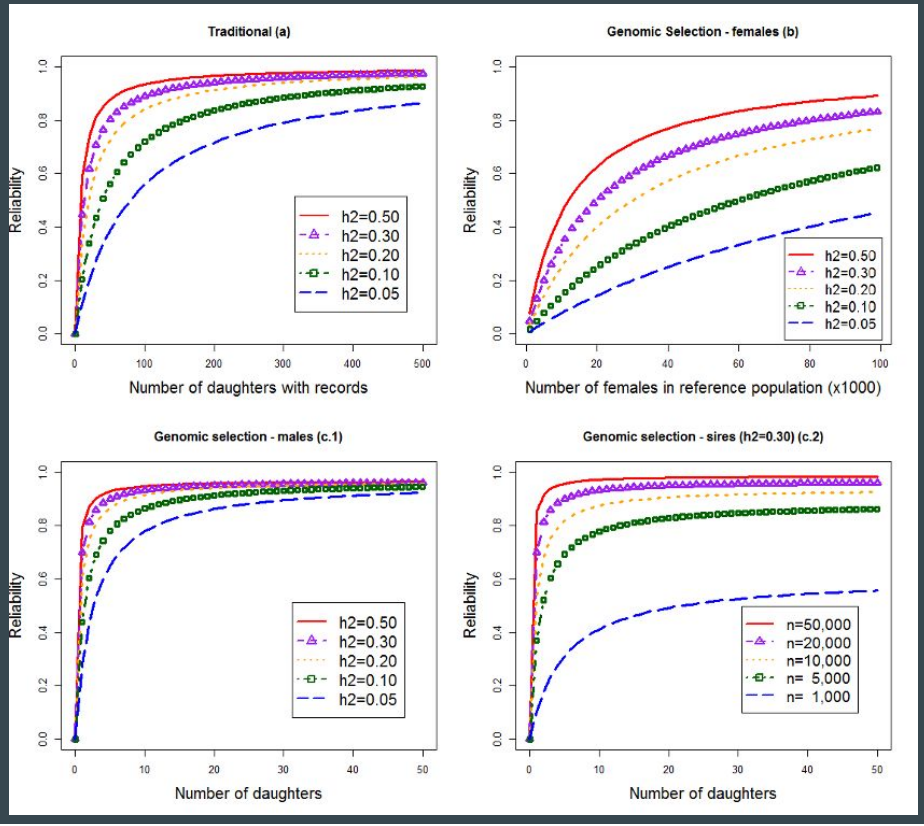
A Reference population is needed in Genome-wide prediction to train the statistical models

- Genotypes and phenotypes
- Statistical association between genotypes and phenotypes
  - Covariates
  - Genomic relationship
- Genomic predictions may be achieved in individuals without phenotype (but w/ genotypes)

Strategies for phenotyping

Cost-benefit function

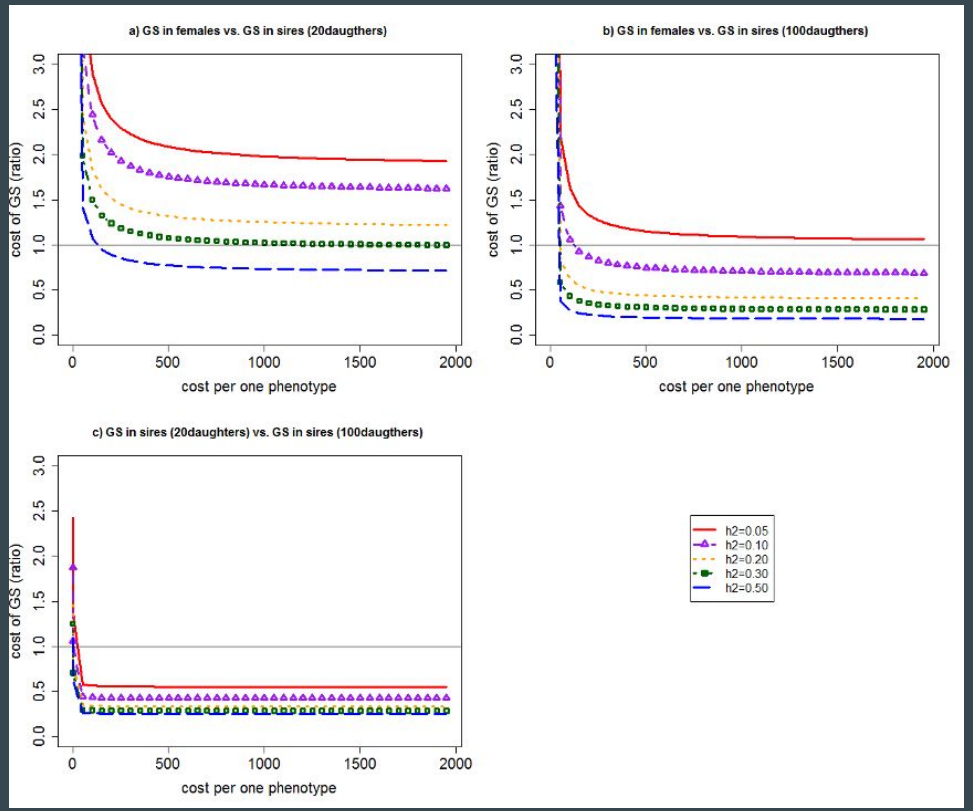
- How much phenotyping cost vs how much prediction accuracy is gained (trait dependent)



Strategies for phenotyping

Cost-benefit function

- Cost of phenotype influences the genotyping strategy (parents vs individuals)



Strategies for phenotyping

Cost-benefit function

- Traits easy to measure
  - Use progeny tests and genotype parents
  - Establish a routine phenotype recording
- Traits difficult to measure and expensive
  - Use individual phenotype and record
  - Experimental conditions

⚠️(This rule of thumb may not work in human medicine, depending on the importance of the trait)

Strategies to obtain genetic marker information in populations

→Whole Genome Sequencing (WGS)

→Genotyping chips (SNPchips)

→Restriction site-associated DNA sequencing (RADseq)

→Genotype by low-pass sequencing (skim-Seq)

Strategies to obtain genetic marker information in populations

→Whole Genome Sequencing (WGS)

- Next generation sequencing
- Wide range of genetic variants (SNP, Indels, CNV, Structural Variants)
- High cost

Strategies to obtain genetic marker information in populations

→Genotyping arrays

- Most used / Widely implemented
- Large variety (species and densities)
- Most available from sequencing services /labs
- SNP + short indels
- Biallelic markers

Strategies to obtain genetic marker information in populations

→Genotyping arrays

Microarray
For Research Use Only
DNA

BovineHD DNA Analysis Kit

This comprehensive genome-wide bovine genotyping array kit features over 777,000 SNPs, and is compatible with any breed of beef or dairy cattle.

Microarray
For Research Use Only
DNA

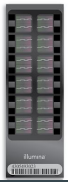
BovineLD Genotyping BeadChip

Extend genomic selection to the entire herd with this expert-designed genotyping array featuring scalable content at an economical price.

Microarray
For Research Use Only
DNA

BovineSNP50 v3 DNA Analysis BeadChip

This BeadChip microarray provides high density multi-sample bovine genotyping for genome characterization of major dairy and beef cattle breed types.



OvineSNP50 DNA Analysis Kit

This sheep microarray features over 54,241 evenly spaced SNP probes for genome-wide association studies, genome-wide selection, and genetic merit determination.Read More...

The BeadChip was developed in collaboration with leading ovine researchers from AgResearch, Baylor UCSC, CSIRO, and the USDA as part of the International Sheep Genomics Consortium. It features over 54,241 evenly spaced probes that target single nucleotide polymorphisms (SNPs).

Microarray
For Research Use Only
DNA

GGP Equine Arrays 65k

Select desired breed traits and verify animal pedigree with the GeneSeek Genomic Profiler (GGP) equine microarray.

Microarray
For Research Use Only
DNA

CanineHD Whole-Genome Genotyping BeadChip 170k

This array enables genotyping of any domestic dog breed, and offers ample SNP density for within-breed association and CNV studies.

GGP Porcine HD Array

This genome-wide porcine genotyping array is ideal for marker-assisted selection and prediction applications. Array content includes:

- 70,000 SNPs for all major porcine breeds
- Average marker spacing of ~42 kb
- 20 key causative mutations

GGP Porcine LD Array

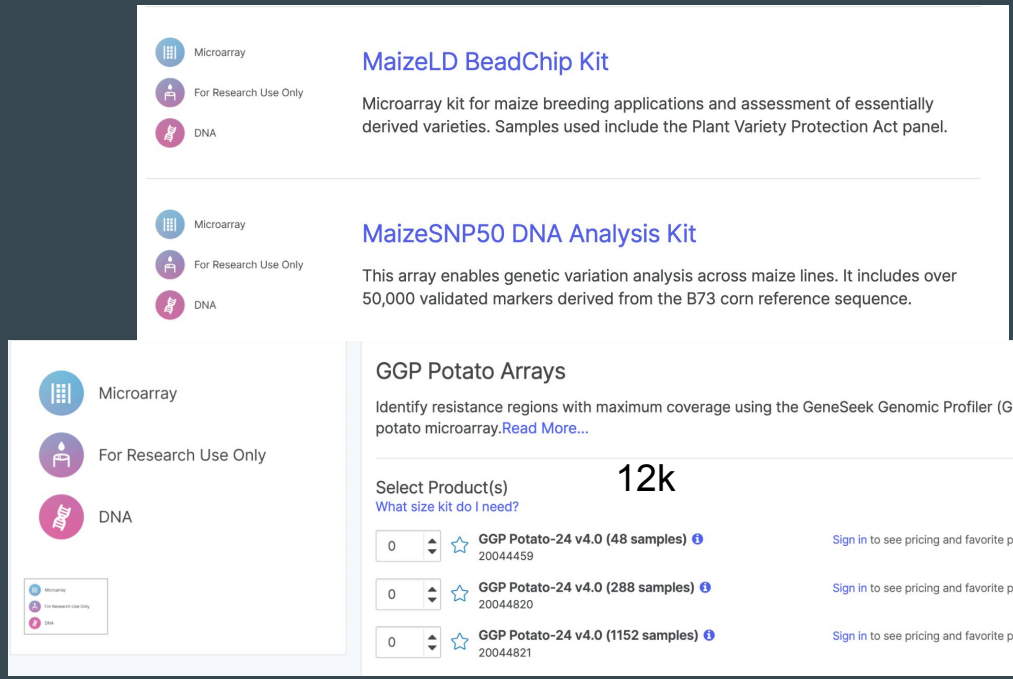
This array is designed for marker-assisted selection, Illumina PorcineSNP60 imputation, GGP Porcine HD imputation, and prediction applications. Array content includes:

- More than 10,000 SNPs for all major porcine breeds
- Average marker spacing of ~250 kb
- ~20 important causative mutations

Genome-wide prediction

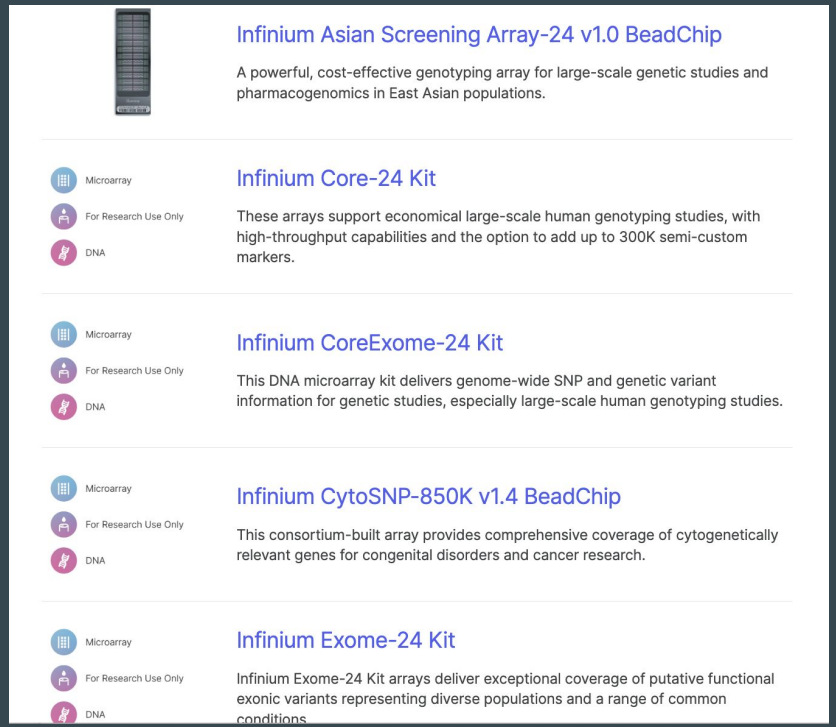
Strategies to obtain genetic marker information in populations

→Genotyping arrays



Strategies to obtain genetic marker information in populations

→Genotyping arrays



Infinium Asian Screening Array-24 v1.0 BeadChip

A powerful, cost-effective genotyping array for large-scale genetic studies and pharmacogenomics in East Asian populations.

Microarray
For Research Use Only
DNA

Infinium Core-24 Kit

These arrays support economical large-scale human genotyping studies, with high-throughput capabilities and the option to add up to 300K semi-custom markers.

Microarray
For Research Use Only
DNA

Infinium CoreExome-24 Kit

This DNA microarray kit delivers genome-wide SNP and genetic variant information for genetic studies, especially large-scale human genotyping studies.

Microarray
For Research Use Only
DNA

Infinium CytoSNP-850K v1.4 BeadChip

This consortium-built array provides comprehensive coverage of cytogenetically relevant genes for congenital disorders and cancer research.

Microarray
For Research Use Only
DNA

Infinium Exome-24 Kit

Infinium Exome-24 Kit arrays deliver exceptional coverage of putative functional exonic variants representing diverse populations and a range of common conditions

Strategies to obtain genetic marker information in populations

→Restriction site-associated DNA sequencing (RADseq)

- Uses NGS
- Requires specific library preparation with specific restriction enzymes
- Most effective in organisms with well-characterized reference genomes
- Cost-effective and medium-throughput

Strategies to obtain genetic marker information in populations

→Restriction site-associated DNA sequencing (RADseq)

scientific reports

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › scientific reports › articles › article

Article | Open access | Published: 15 October 2020

Genomic predictions and genome-wide association studies based on RAD-seq of quality-related metabolites for the genomics-assisted breeding of tea plants

Hiroto Yamashita, Tomoki Uchida, Yasuno Tanaka, Hideyuki Katai, Atsushi J. Nagano, Akio Morita & Takashi Ikka ✉

Scientific Reports **10**, Article number: 17480 (2020) | Cite this article

BMC Genomics

Home About Articles Submission Guidelines Collections Join The Board Submit manuscript

Research | Open access | Published: 19 May 2023

Restriction site-associated DNA sequencing technologies as an alternative to low-density SNP chips for genomic selection: a simulation study in layer chickens

Florian Herry, Frédéric Hérault, Frédéric Lecerf, Laëtitia Lagoutte, Mathilde Doublet, David Picard-Druet, Philippe Bardou, Amandine Varenne, Thierry Burlot, Pascale Le Roy & Sophie Allais ✉

BMC Genomics **24**, Article number: 271 (2023) | Cite this article

Genome-wide prediction

Strategies to obtain genetic marker information in populations

→Restriction site-associated DNA sequencing (RADseq)

Highly dependent on the restriction enzyme

GEBV correlation <0.50 with HD SNP arrays

BMC Genomics

Home About Articles Submission Guidelines Collections Join The Board Submit manuscript

Research | Open access | Published: 19 May 2023

Restriction site-associated DNA sequencing technologies as an alternative to low-density SNP chips for genomic selection: a simulation study in layer chickens

Florian Herry, Frédéric Hérault, Frédéric Lecerf, Laëtitia Lagoutte, Mathilde Doublet, David Picard-Druet, Philippe Bardou, Amandine Varenne, Thierry Burlot, Pascale Le Roy & Sophie Allais

BMC Genomics **24**, Article number: 271 (2023) | Cite this article

Table 6 Pearson correlations between true "Full_HD" GEBVs and imputed HD GEBVs based on ancestry for the 67 G1 breeders, according to each enzyme used for egg weight (EW), eggshell colour (ESC), eggshell strength (ESS) and albumen height (AH).

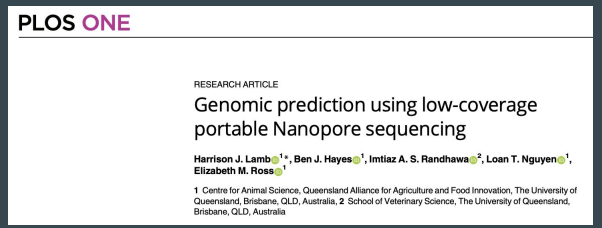
	Number of SNPs	EW	ESC	ESS	AH
EcoRI	1,797	0.3774	0.2962	0.3420	0.4261
TaqI	4,126	0.4476	0.2453	0.3906	0.4478
TaqI_PstI	11,193	0.4740	0.2442	0.3869	0.4684
AvaII	12,453	0.4681	0.2430	0.3859	0.4794
PstI	14,390	0.4664	0.2450	0.3953	0.4689
HD SNP chip	300,028	0.4713	0.2460	0.3940	0.4802

The line HD SNP chip corresponds to the Pearson correlation between true "Full_HD" GEBVs and true HD GEBVs based on ancestry for the 67 G1 breeders.

Strategies to obtain genetic marker information in populations

→Genotype by low-pass sequencing (Skim-Seq)

- Low cost
- Non targeted sequencing (needs imputation)
- NGS (Illumina or ONT)
- Minimum coverage ranges between 0.5x and 4x depending on population and sequencing method

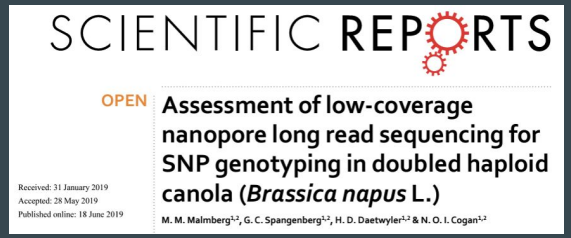

PLOS ONE

RESEARCH ARTICLE

Genomic prediction using low-coverage portable Nanopore sequencing

Harrison J. Lamb[1]*, Ben J. Hayes[1], Imtiaz A. S. Randhawa[2], Loan T. Nguyen[1], Elizabeth M. Ross[1]

1 Centre for Animal Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD, Australia, 2 School of Veterinary Science, The University of Queensland, Brisbane, QLD, Australia

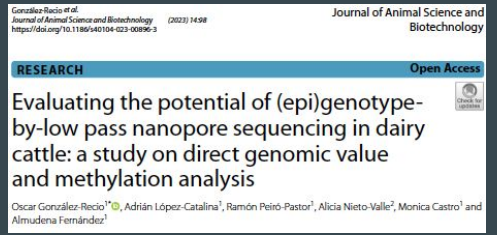

SCIENTIFIC REPORTS

OPEN

Assessment of low-coverage nanopore long read sequencing for SNP genotyping in doubled haploid canola (*Brassica napus* L.)

Received: 31 January 2019
Accepted: 28 May 2019
Published online: 18 June 2019

M. M. Malmberg[1,2], G. C. Spangenberg[1,2], H. D. Daetwyler[1,2] & N. O. I. Cogan[1,2]


González-Recio et al. Journal of Animal Science and Biotechnology (2023) 14:98
https://doi.org/10.1186/s40104-023-00896-3

Journal of Animal Science and Biotechnology

RESEARCH

Open Access

Evaluating the potential of (epi)genotype-by-low pass nanopore sequencing in dairy cattle: a study on direct genomic value and methylation analysis

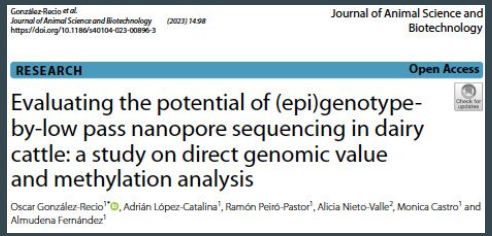
Oscar González-Recio[1*], Adrián López-Catalina[1], Ramón Peiró-Pastor[1], Alicia Nieto-Valle[2], Monica Castro[1] and Almudena Fernández[1]

Strategies to obtain genetic marker information in populations

→Genotype by low-pass sequencing (Skim-Seq)

Imputation accuracy improves with

sequencing depth →


González-Recio et al.
Journal of Animal Science and Biotechnology (2023) 14:98
https://doi.org/10.1186/s40104-023-00896-3

Journal of Animal Science and Biotechnology

RESEARCH Open Access

Evaluating the potential of (epi)genotype-by-low pass nanopore sequencing in dairy cattle: a study on direct genomic value and methylation analysis

Oscar González-Recio[1*], Adrián López-Catalina[1], Ramón Peiró-Pastor[1], Alicia Nieto-Valle[2], Monica Castro[1] and Almudena Fernández[1]

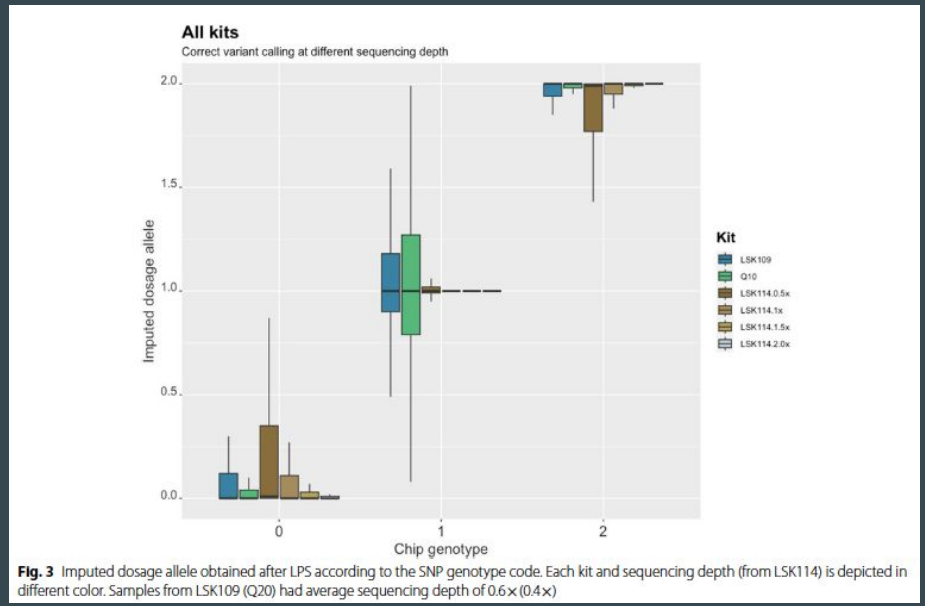


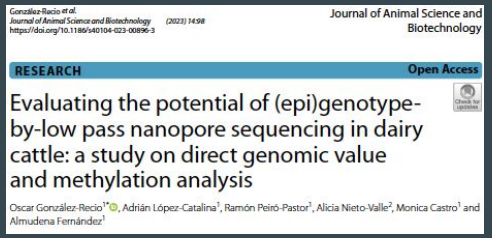
Fig. 3 Imputed dosage allele obtained after LPS according to the SNP genotype code. Each kit and sequencing depth (from LSK114) is depicted in different color. Samples from LSK109 (Q20) had average sequencing depth of 0.6× (0.4×)

Genome-wide prediction

Strategies to obtain genetic marker information in populations

→Genotype by low-pass sequencing (Skim-Seq)

High correlation (>0.98) with latest technology

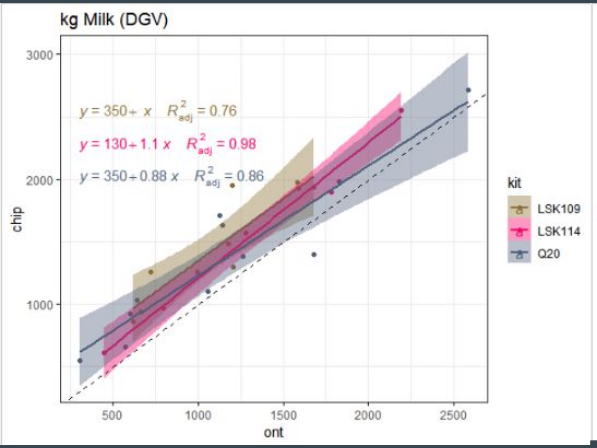

González-Recio et al.
Journal of Animal Science and Biotechnology (2023) 14:98
https://doi.org/10.1186/s40104-023-00896-3

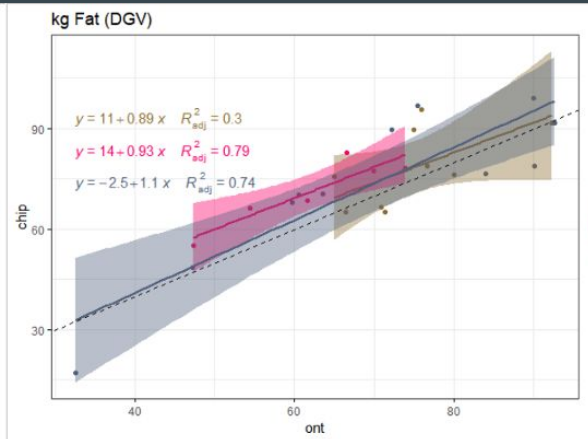
Journal of Animal Science and Biotechnology

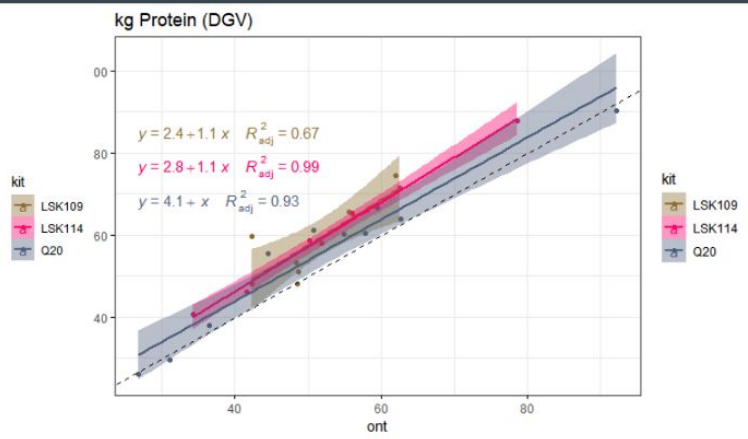
RESEARCH Open Access

Evaluating the potential of (epi)genotype-by-low pass nanopore sequencing in dairy cattle: a study on direct genomic value and methylation analysis

Oscar González-Recio[1*], Adrián López-Catalina[1], Ramón Peiró-Pastor[1], Alicia Nieto-Valle[2], Monica Castro[1] and Almudena Fernández[1]





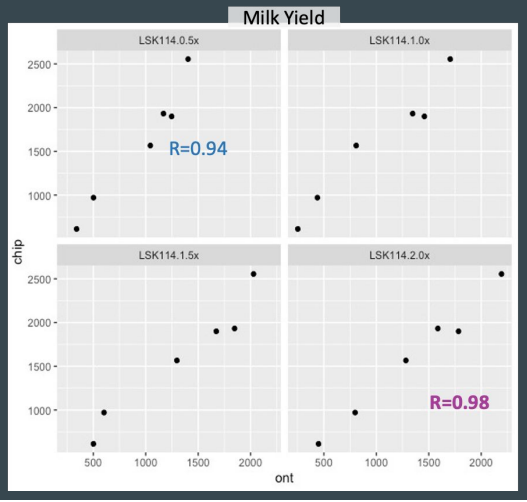


Strategies to obtain genetic marker information in populations

→Genotype by low-pass sequencing (Skim-Seq)

High correlation (>0.98) with latest technology

and seq-depth>2x.



González-Recio et al.
Journal of Animal Science and Biotechnology (2023) 14:98
https://doi.org/10.1186/s40104-023-00896-3
Journal of Animal Science and Biotechnology
RESEARCH
Open Access

Evaluating the potential of (epi)genotype-by-low pass nanopore sequencing in dairy cattle: a study on direct genomic value and methylation analysis

Oscar González-Recio[1*], Adrián López-Catalina[1], Ramón Peiró-Pastor[1], Alicia Nieto-Valle[2], Monica Castro[1] and Almudena Fernández[1]

Genome-wide prediction

RECAP

Assume gaussian distribution on phenotypes (... subsequently residuals)

Why variance is important

Inference is different from prediction.

Genetic architecture challenge

How to establish a reference population: Phenotyping strategies & Genotyping strategies