

*Genomic Prediction Workshop - Ames 2015*

# Introduction to Genomic Prediction

Dorian Garrick

Lush Endowed Chair in Animal Breeding & Genetics

dorian@iastate.edu

## Genomics

▼ ————— *Dictionary* —————

**ge•no•mics** |jē'nōmiks, -'näm- , |

plural noun [ treated as sing. ]

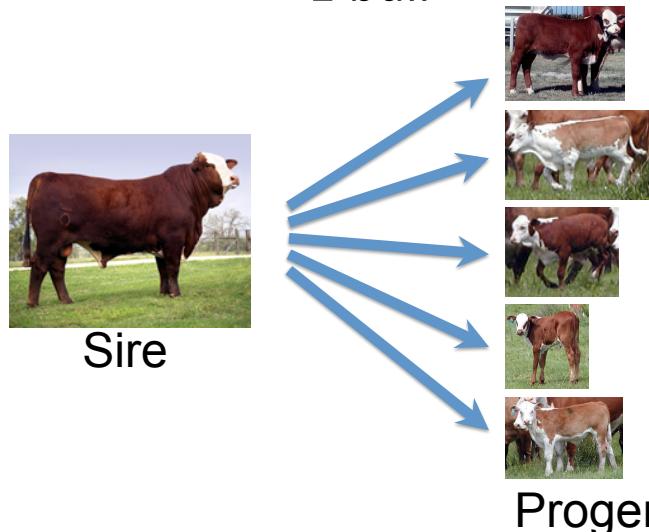
the branch of molecular biology concerned with the structure, function, evolution, and mapping of genomes.

ORIGIN 1980s: from *genome* '**the complete set of genes present in an organism**' + *-ics*.

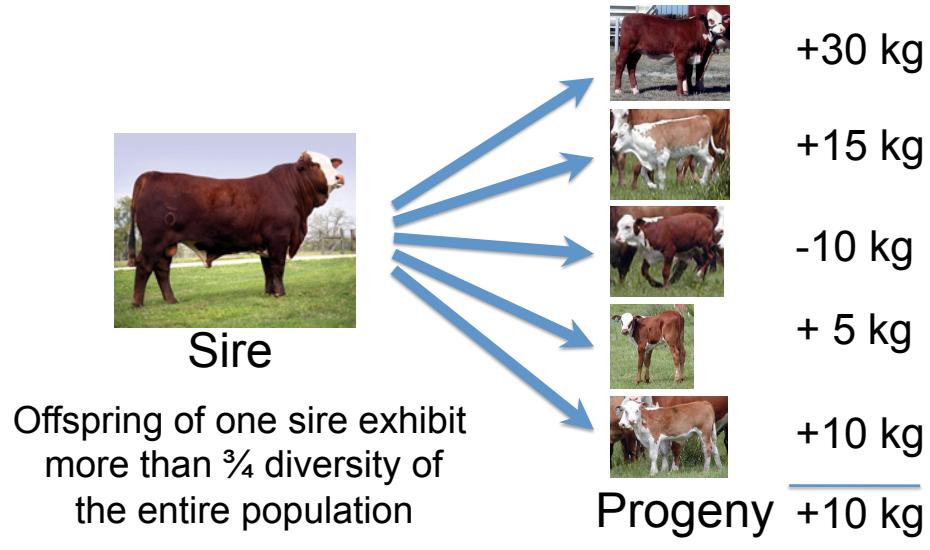
## Genomic Prediction

- Ranking candidates for selection using knowledge of the “complete set of genes” *along with conventional pedigree and performance information*
  - Using everything we’ve got to obtain the most accurate EPD/EBV (at as young an age as possible)

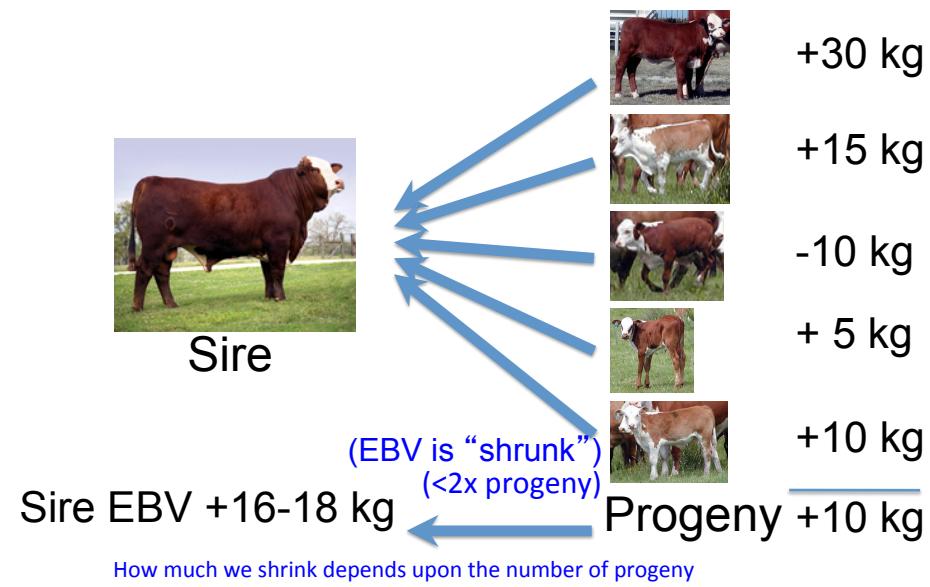
Suppose we generate 100 progeny on  
1 bull



## Performance of the Progeny



## We Learn about Parents from Progeny



EBVs on widely-used old sires are accurate



Sire

With enough progeny,  
this is usually close to  
the bulls true EBV/EPD  
(not surprisingly!)

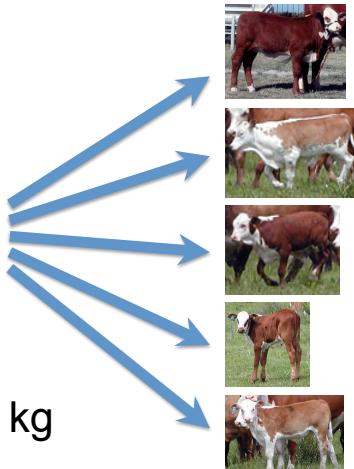
Sire EBV +16-18 kg

Suppose we generate new progeny



Sire

Sire EBV +16-18 kg



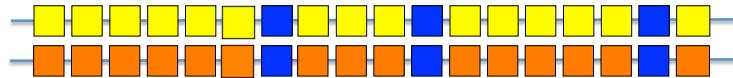
Expect them  
to be 8-9 kg  
heavier than  
those from an  
average sire

Some will be more  
others will be less  
but we can't tell  
which are better  
without "buying"  
more information

Progeny

## Chromosomes are a sequence of base pairs

Part of 1 pair  
of chromosomes



Cattle usually have 30 pairs of chromosomes  
One member of each pair inherited from the sire, one from the dam  
Each chromosome has about 100 million base pairs (A, G, T or C)  
About 3 billion describe the animal

- █ Blue base pairs represent genes
- █ Yellow represents the strand inherited from the sire
- █ Orange represents the strand inherited from the dam

A common error is the  
substitution of one base pair  
for another  
Single Nucleotide Polymorphism  
(SNP)

- Errors in duplication
- Most are repaired
  - Some will be transmitted
  - Some of those may influence performance
    - Some will be beneficial, others harmful

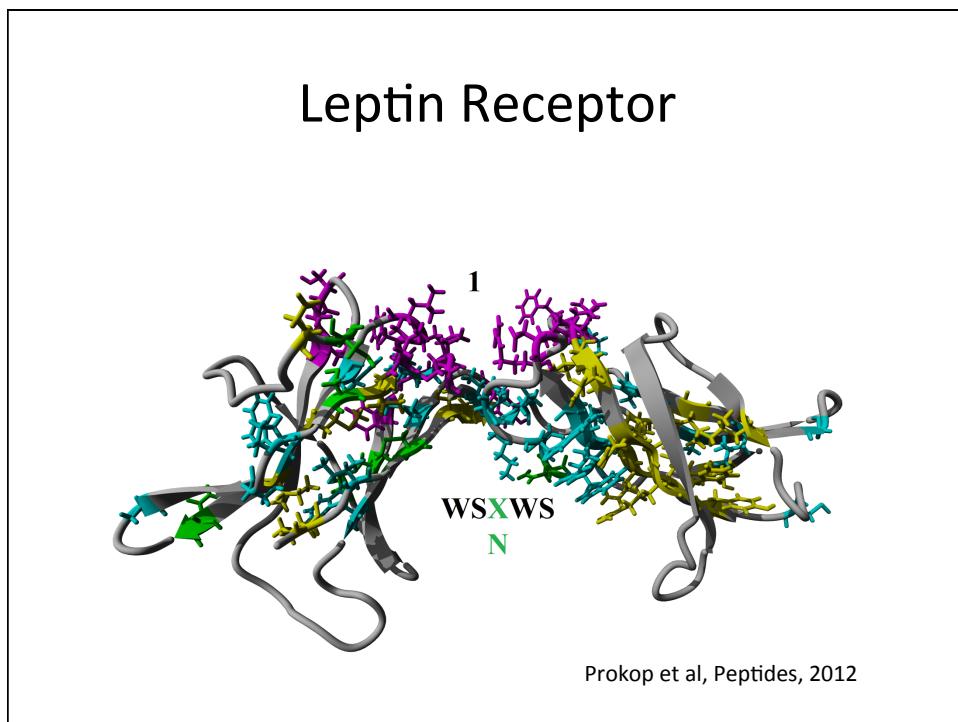
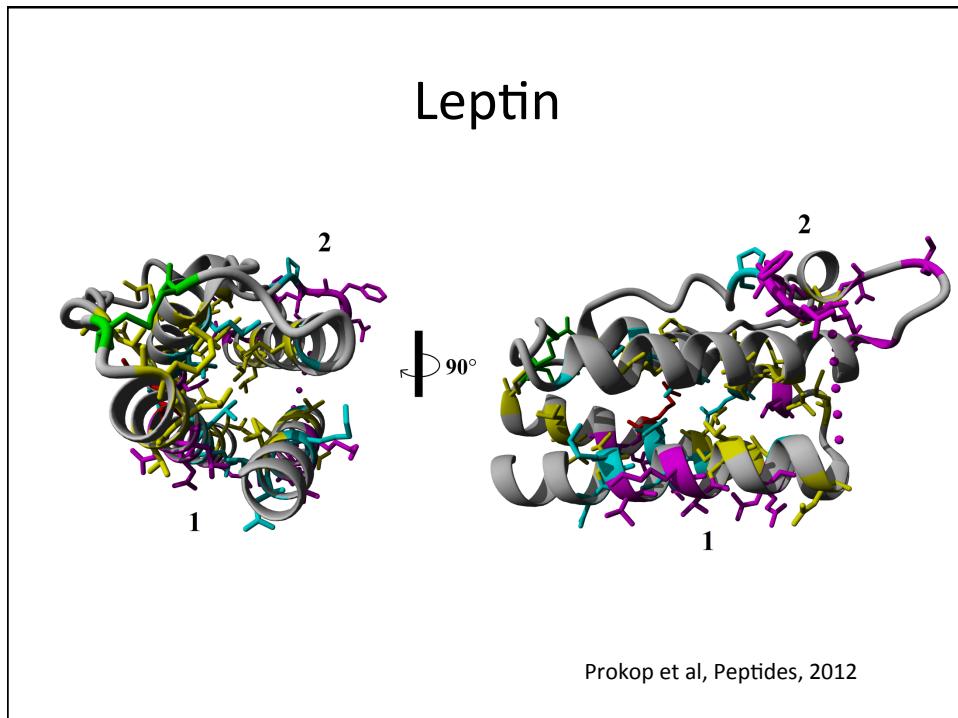
- Inspection of whole genome sequence
- Demonstrate historical errors
  - And occasional new (de novo) mutations

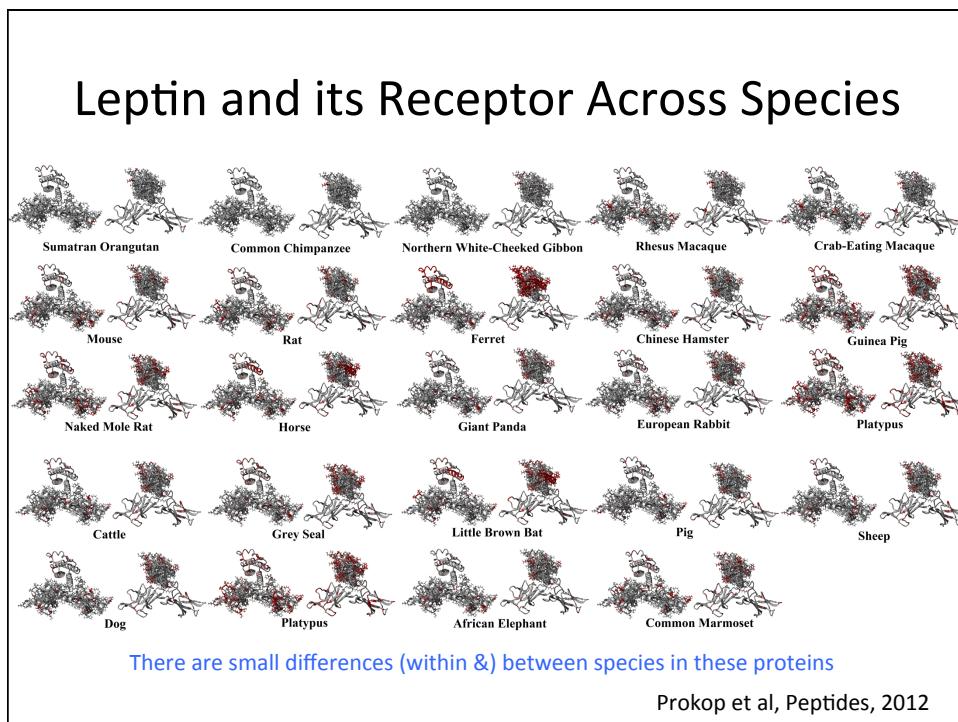
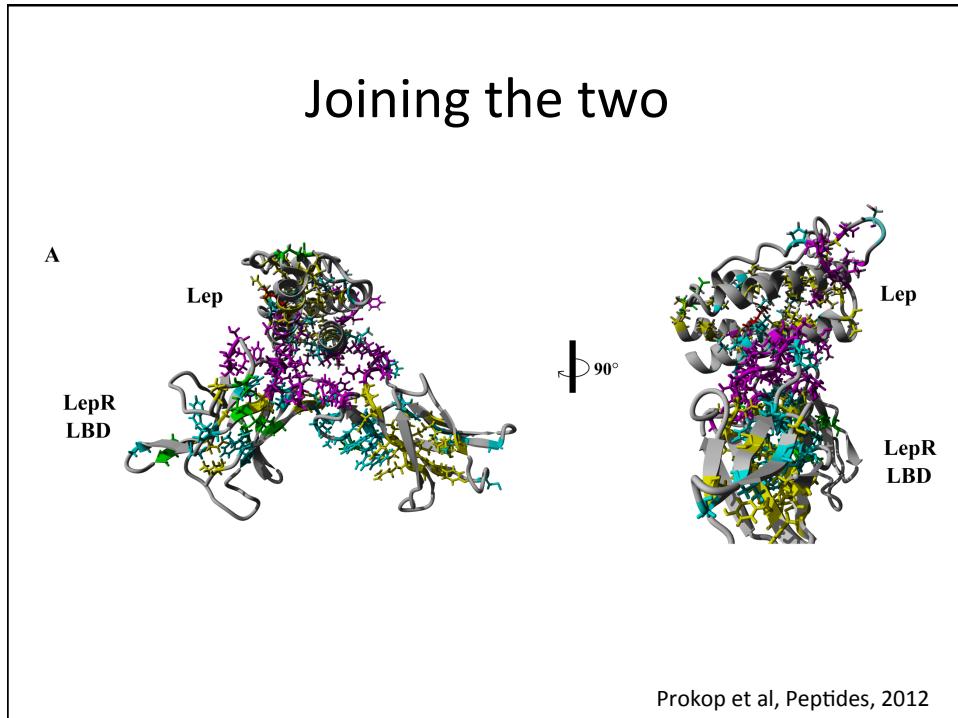
## Mutations

- Could cause complete loss-of-function of the gene (ie the gene is “broken”)
  - These can sometimes be catastrophic when an individual is homozygous and carries 2 copies of the broken gene
    - For examples DUMPS, Citrullinemia, BLAD, etc

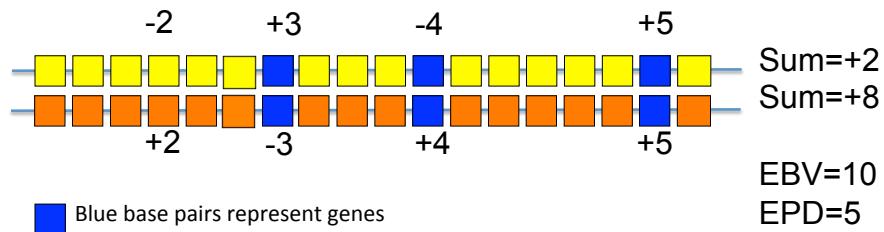
## Mutations

- Could cause complete loss-of-function of the gene (ie the gene is “broken”)
- Could increase or decrease expression level
- The variant might change amino acid sequence to cause subtle changes to the shape of the protein products making them function a little better or a little worse
  - Natural or artificial selection will favour the variants that improve fitness in that particular climatic and environmental circumstance



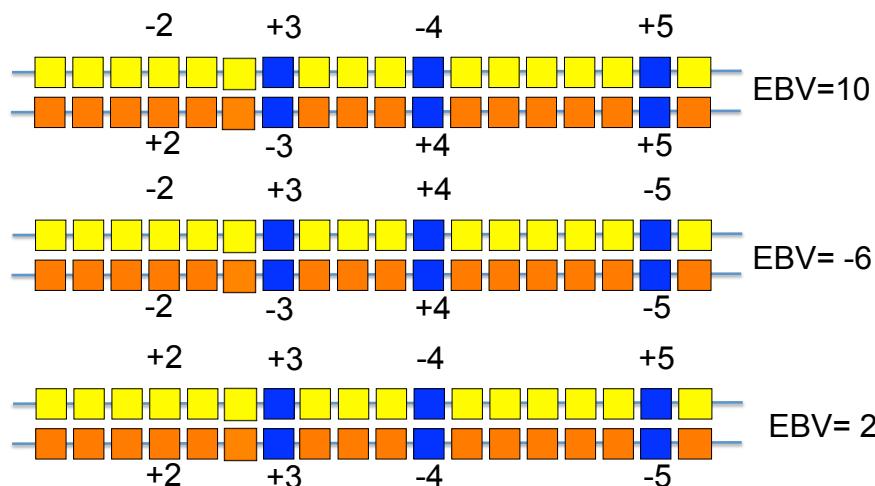


## EBV is sum of the Gene Effects

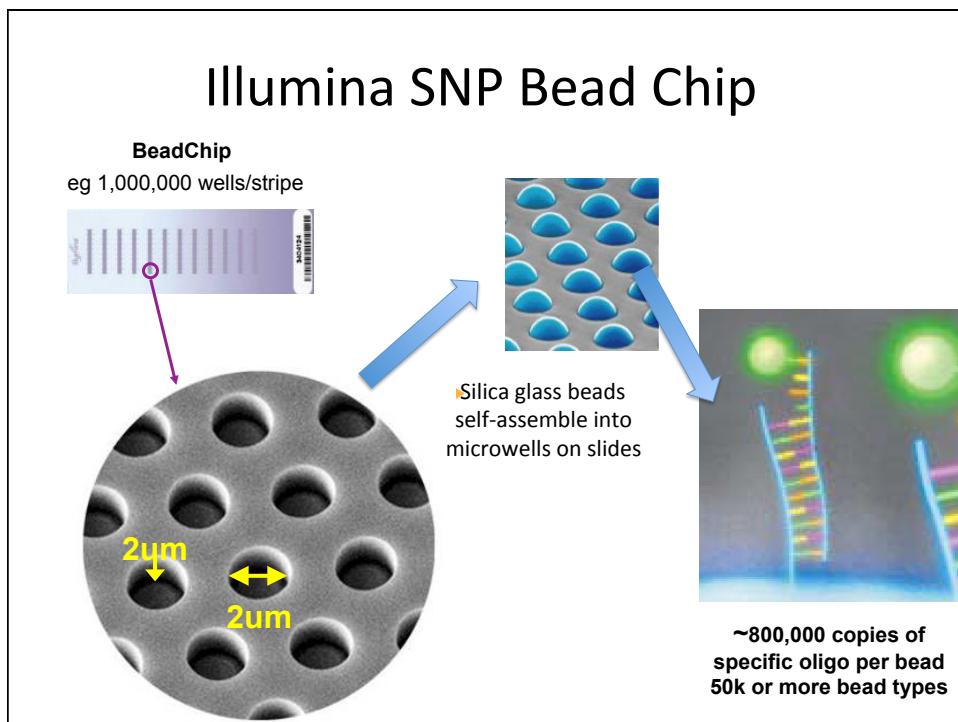
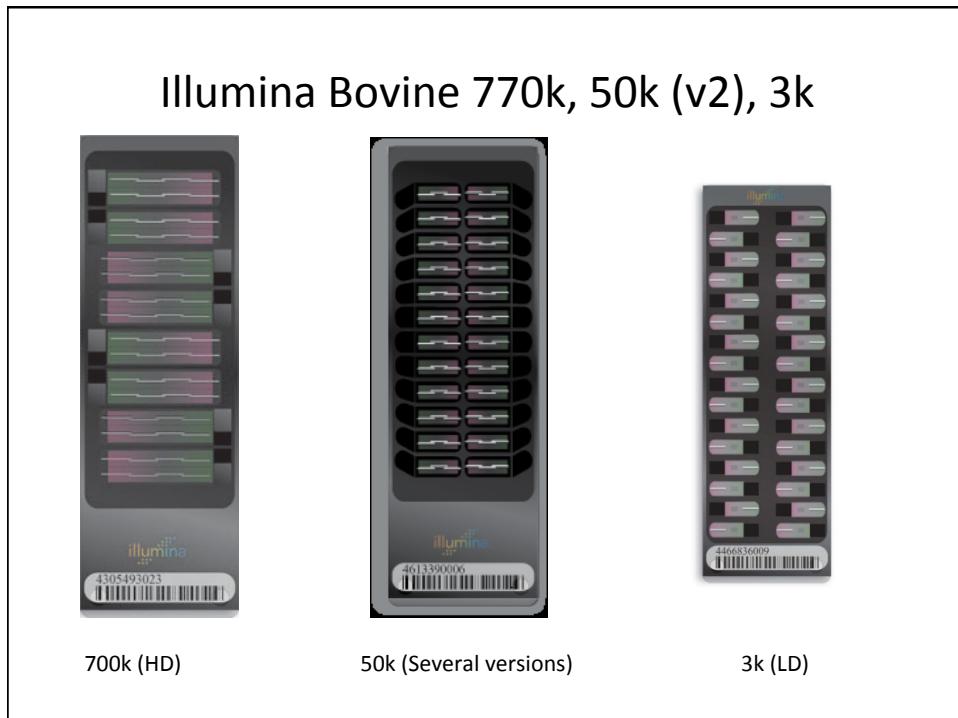


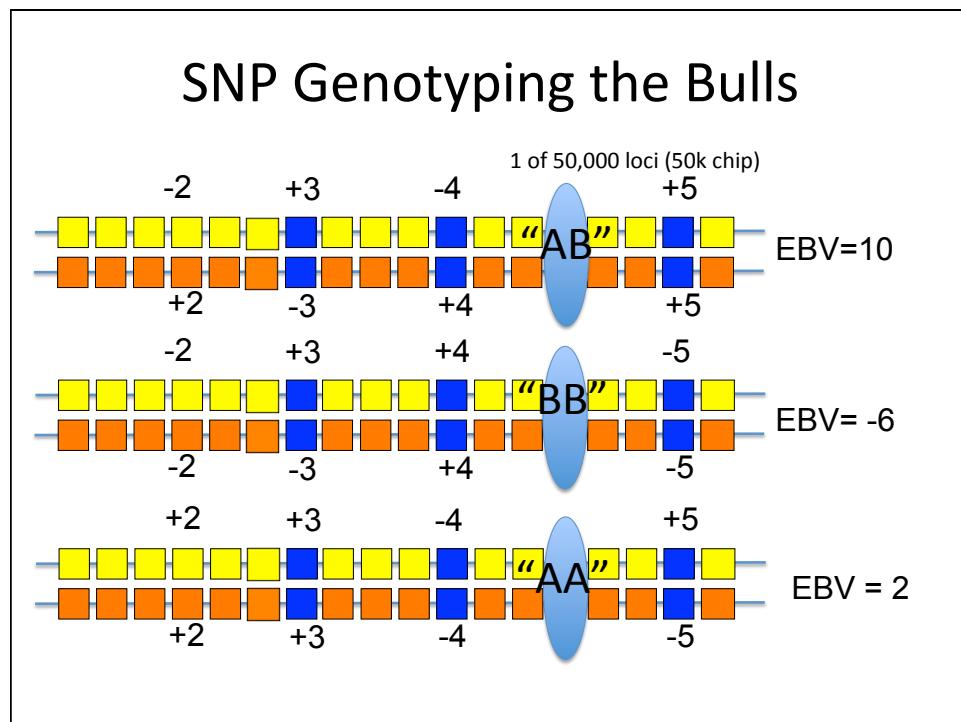
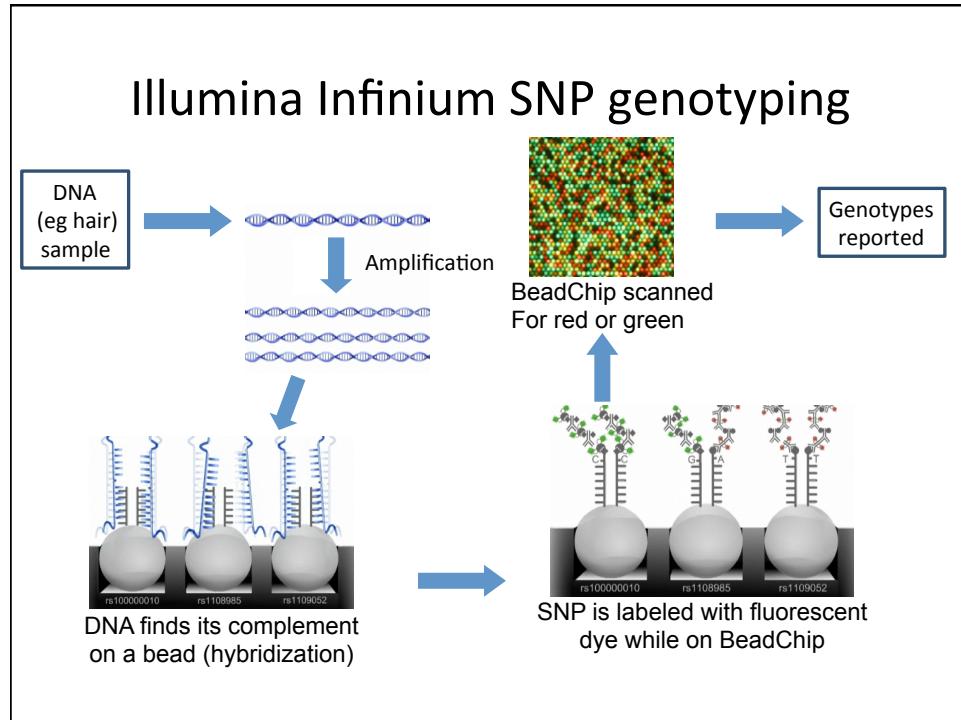
EPD is HALF the sum of the gene effects

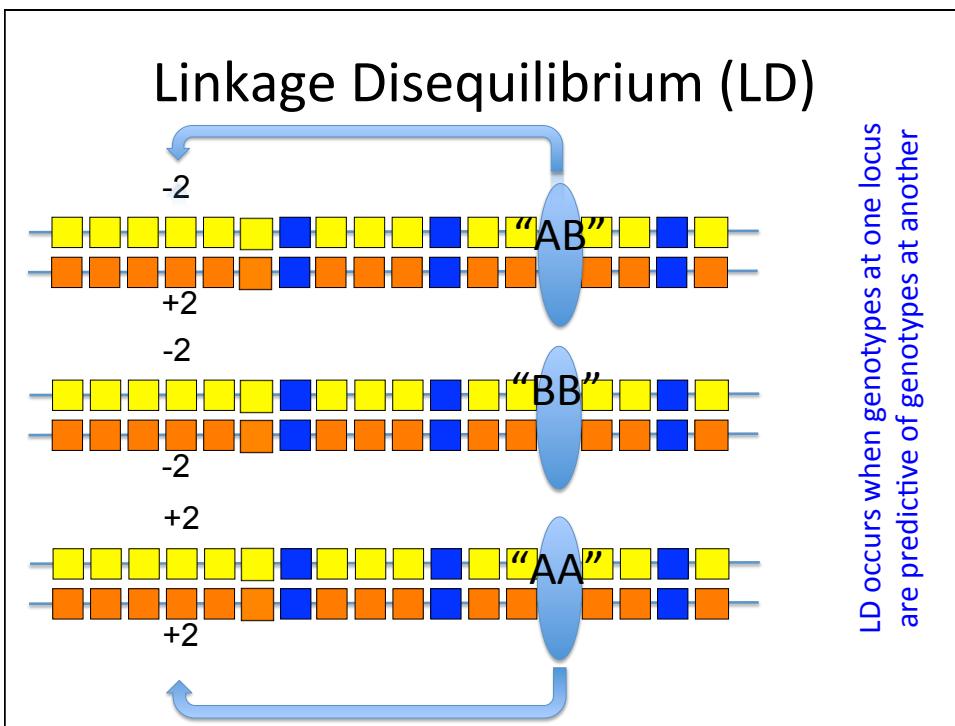
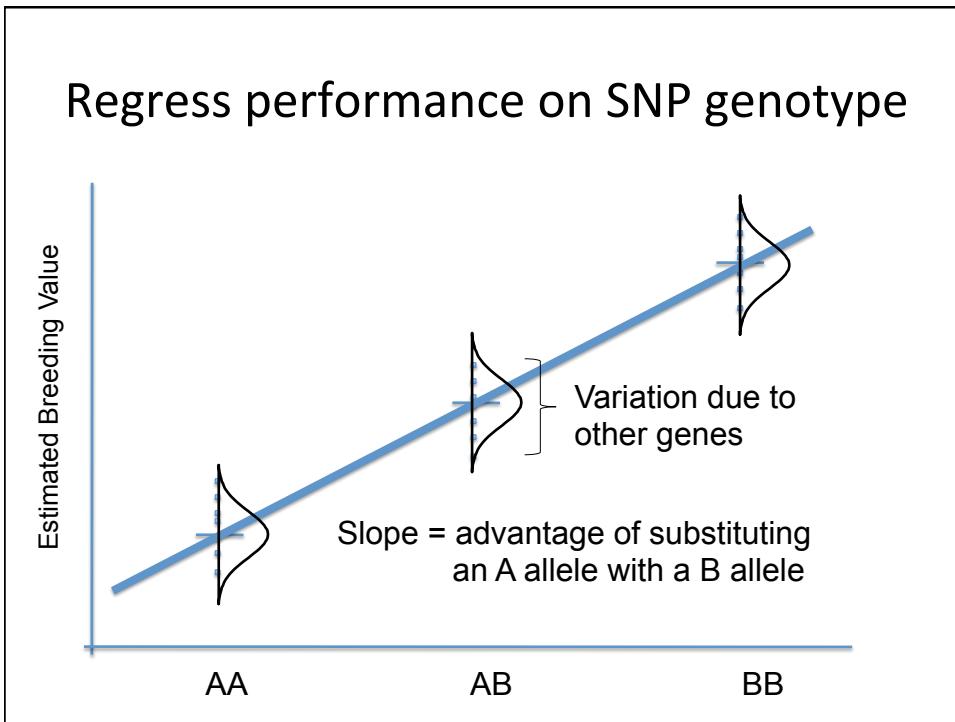
## Consider 3 Bulls

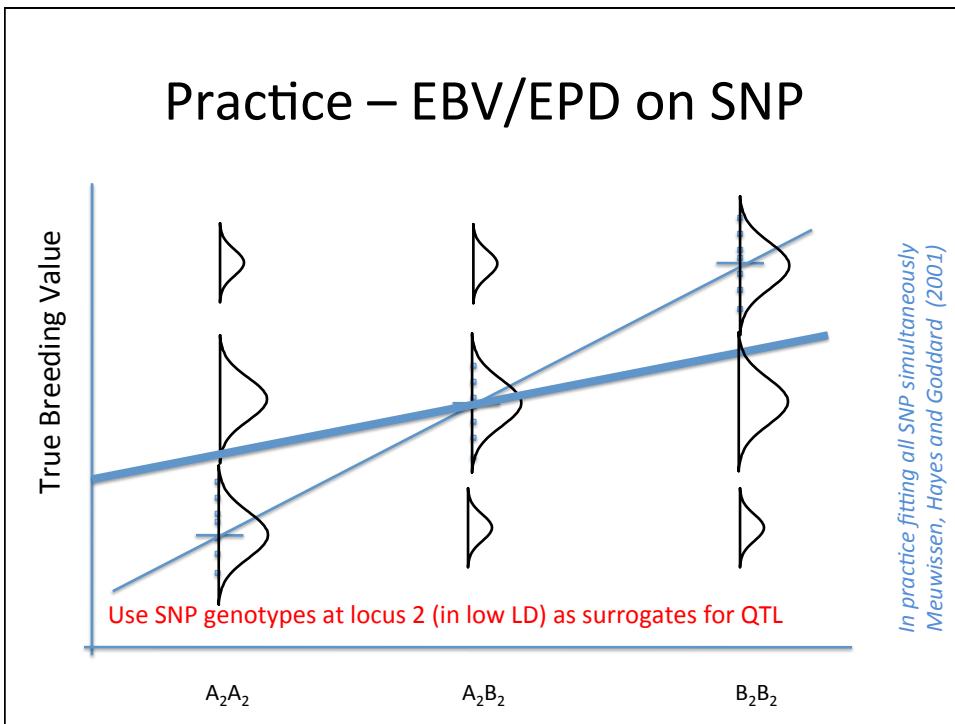
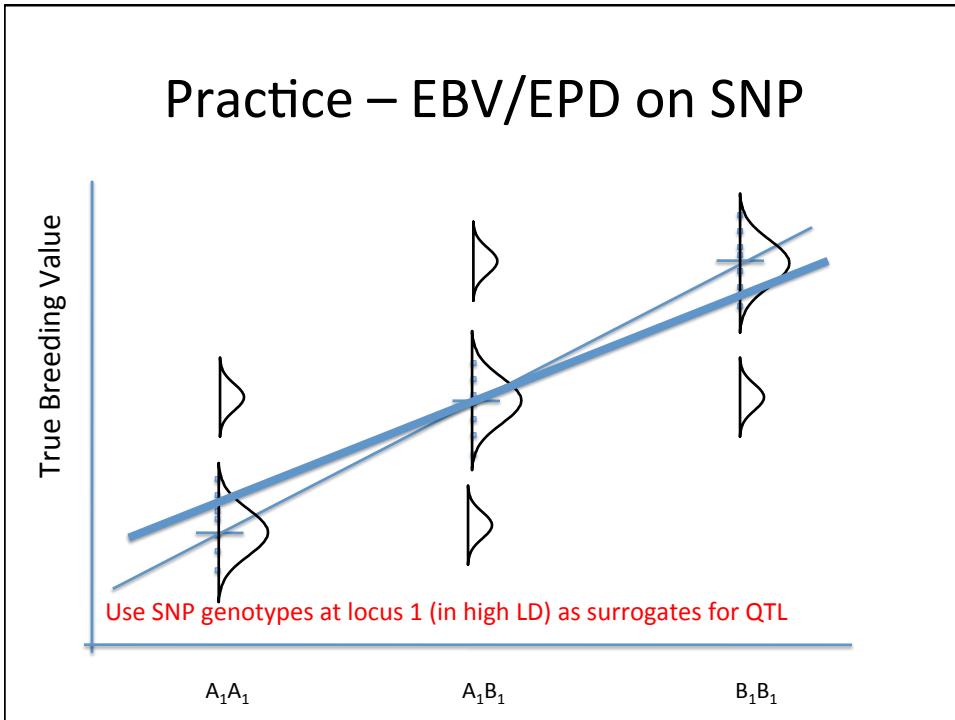


Below-average bulls will have some above-average alleles and vice versa!









[www.23andme.com](http://www.23andme.com)

Decreased Risk ?

NAME	CONFIDENCE	YOUR RISK	AVG. RISK	COMPARED TO AVERAGE
Alzheimer's Disease	★★★	4.9%	7.2%	0.69x

**Marker Effects**

Only significant, validated GWAS findings used in prediction

[www.23andme.com](http://www.23andme.com)

- Coronary Heart Disease

**Marker Effects**

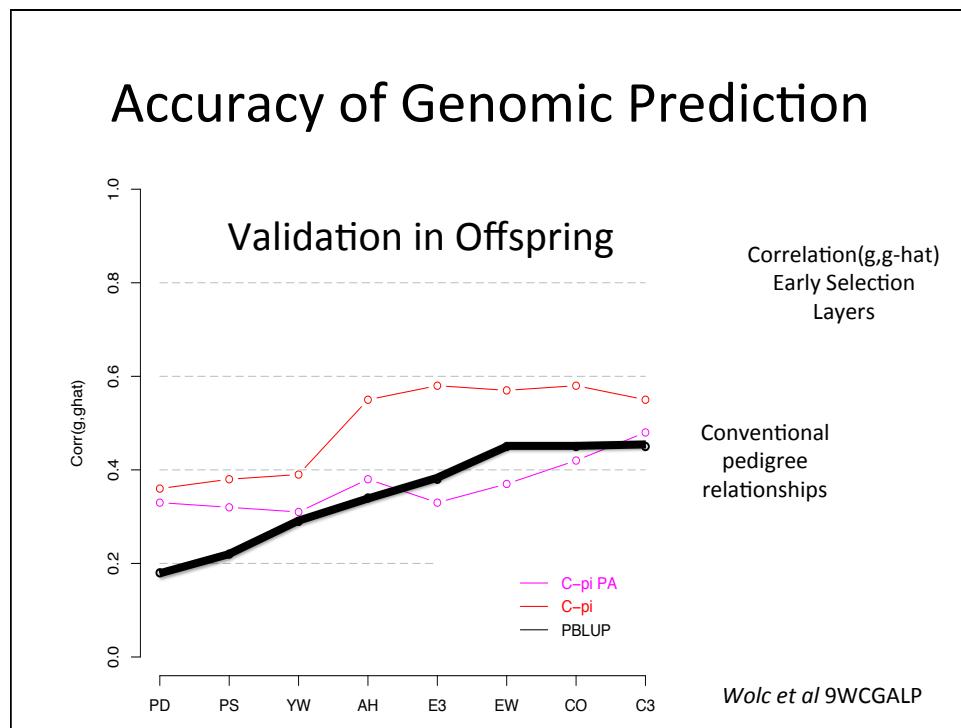
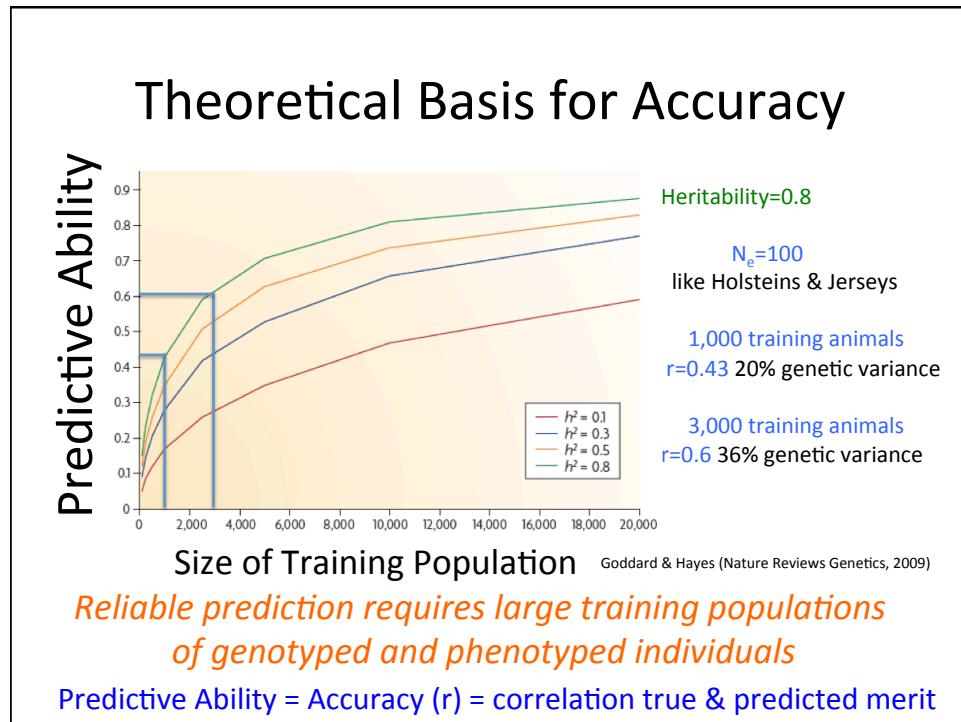
Each bar represents a different risk QTL allele  
(mouseover shows the allele and links to the research publications)  
QTL=Quantitative Trait Locus

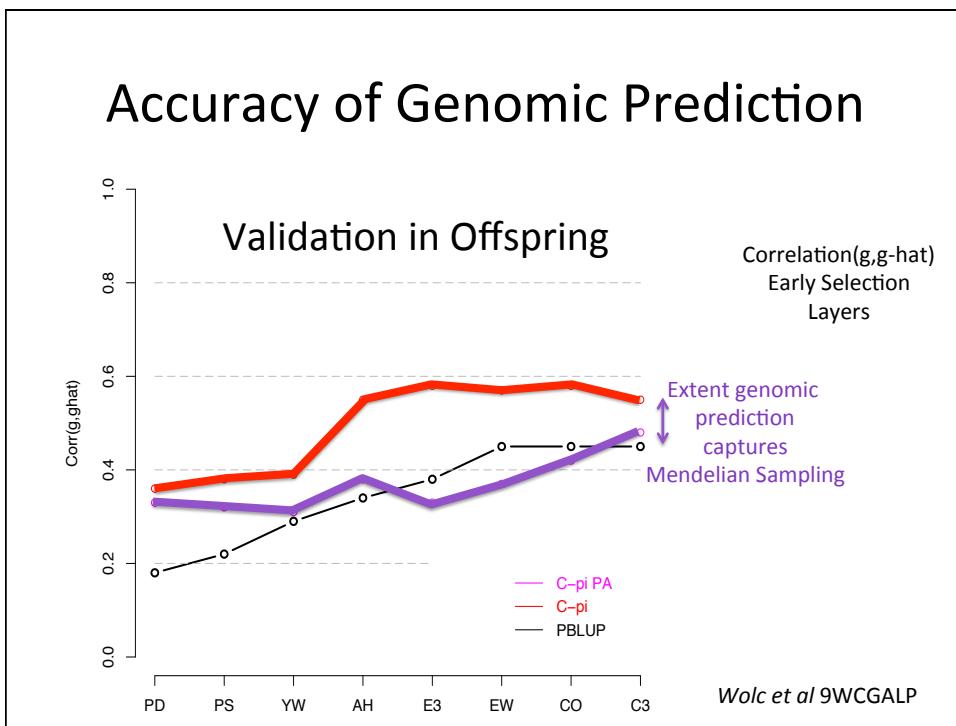
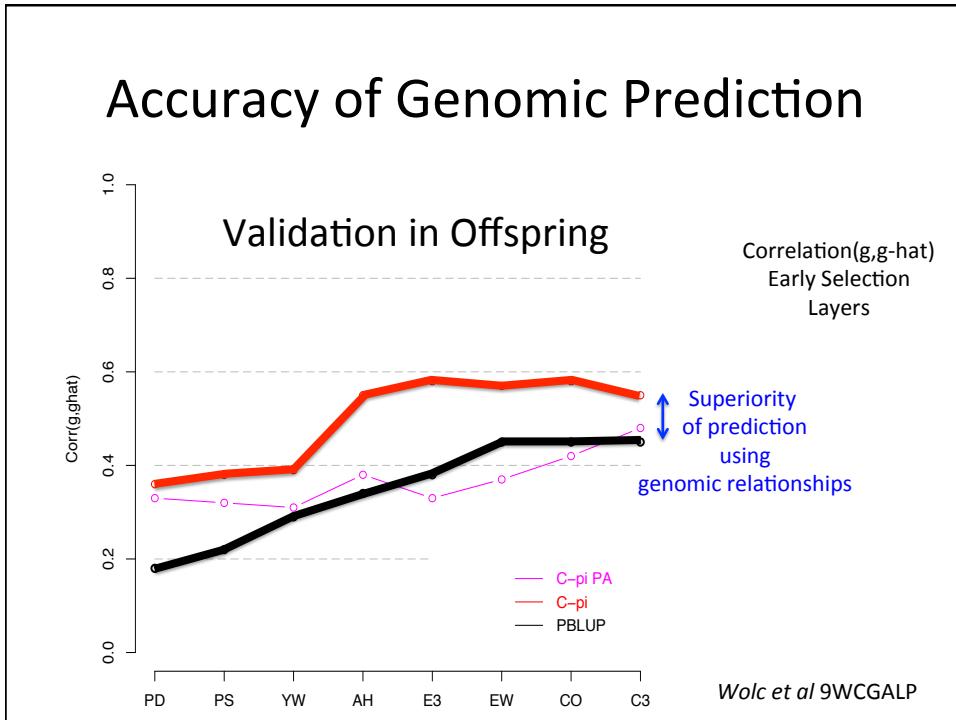
Only significant, validated GWAS findings used in prediction

39-56 % Attributable to Genetics

Dorian Garrick  
**55.0 out of 100**  
men of European ethnicity who share Dorian Garrick's genotype will develop Coronary Heart Disease between the ages of 45 and 79.

Average  
**46.8 out of 100**  
men of European ethnicity will develop Coronary Heart Disease between the ages of 45 and 79.





## Genome-Wide Association Studies (GWAS)

- Use a historical population of bulls and cows with EBV information that have been genotyped with 50k panels
- Derive an EBV for every chromosome fragment (we call this **training**), and find the regions with biggest effects

Cut genome into 2,700 1Mb windows

#SNPs	%Var	Cum%Var	map_pos	
11	7.10	7.10	7_93	Regions with biggest effects
28	3.70	10.80	20_4	
22	1.34	12.14	13_58	
22	1.23	13.37	26_34	
9	0.92	14.29	6_29	
25	0.89	16.09	4_75	
26	0.79	16.88	4_114	
23	0.65	17.53	2_121	
17	0.61	18.14	18_55	
25	0.60	18.74	8_88	

Angus Birth Weight

## Major Regions for Birth Weight

Genetic Variance %

Chr_mb	Angus	Hereford	Limousin	Simment al	Gelbvieh
7_93	7.10	5.85	0.02	0.18	0.02
6_38-39	0.47	8.48	5.90	16.3	4.75
20_4	3.70	7.99	0.07	1.53	0.03
14_24-26	0.42	0.01	0.71	3.05	8.14

Some of these same regions have big effects on one or more of weaning weight, yearling weight, marbling, ribeye area, calving ease

## Iowa State University (ISU)

- A land-grant institution with responsibilities for research, teaching and extension
  - Such activities have been applied to genetic improvement of animals since 1930's when Iowa State Professor, Dr JL Lush, wrote the first textbook on animal breeding
  - That tradition continues just as strongly today as we research the role of genomics for improvement

## Summary

- Genomics will increase accuracy of evaluation
  - The technology is **starting to mature** but works better in some traits and breeds than in others
  - It works better with **greater** amounts of **data**
  - Genomic prediction will **get more accurate** than it is today if we continue to undertake research
- This workshop will explain the statistical basis for methods of genomic prediction and GWAS