

Genomic Prediction Methods



SNP & VARIATION SUITE

Steve Hystad - Field Application Scientist

CIOReview

20 most promising
Biotech Technology
Providers

pharma
TECH OUTLOOK

Top 10 Analytics
Solution Providers

Gartner.

Hype Cycle for
Life sciences

Golden Helix – Who We Are



Golden Helix is a global bioinformatics company founded in 1998.



- Variant Calling
- Filtering and Annotation
- Clinical Reports
- CNV Analysis
- Pipeline: Run Workflows

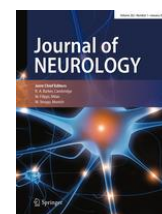
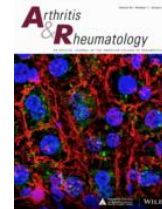
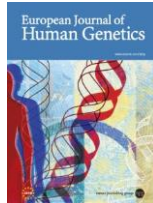
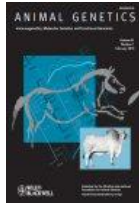
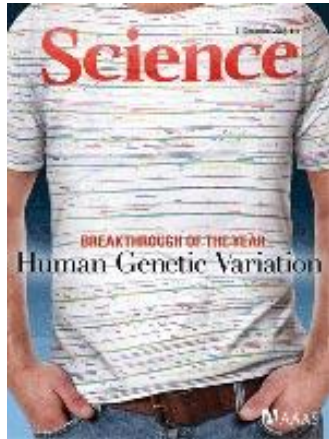


- Variant Warehouse
- Centralized Annotations
- Hosted Reports
- Sharing and Integration

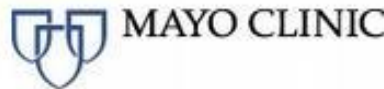


- CNV Analysis**
- GWAS
- Genomic Prediction
- Large-N-Population Studies
- RNA-Seq
- Large-N CNV-Analysis

Cited in over 1100 peer-reviewed publications



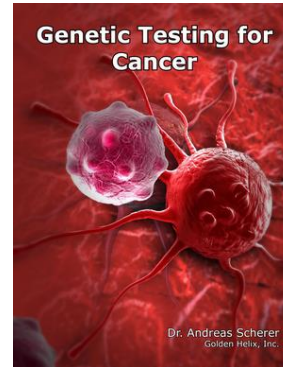
Over 350 customers globally





When you choose a Golden Helix solution, you get more than just software

- REPUTATION
- TRUST
- EXPERIENCE



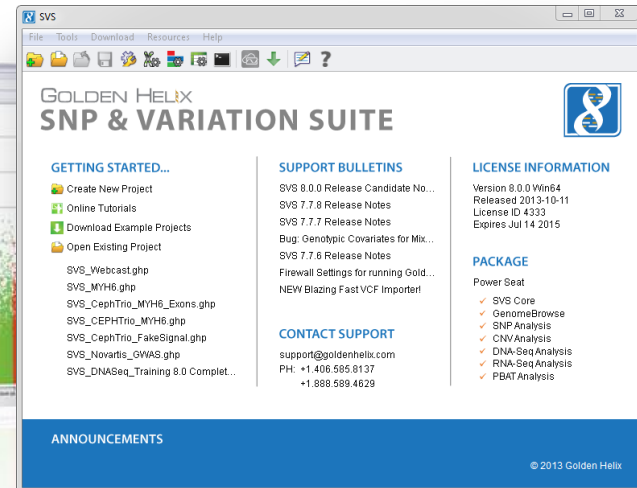
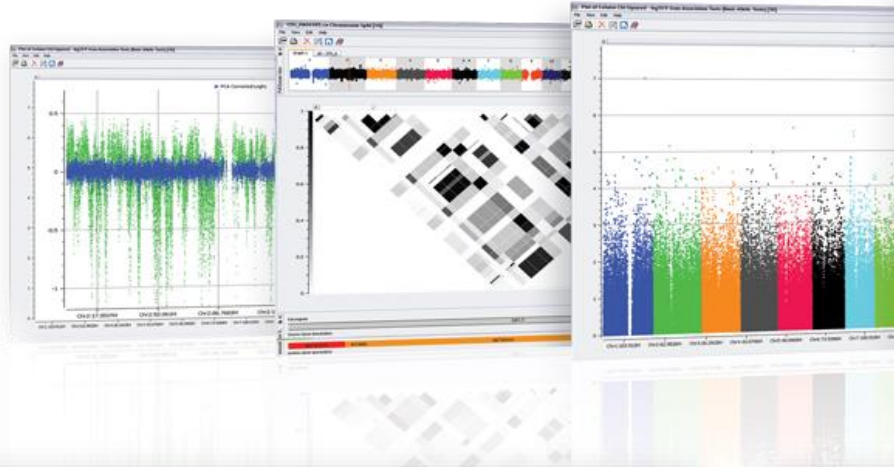
- INDUSTRY FOCUS
- THOUGHT LEADERSHIP
- COMMUNITY

- TRAINING
- SUPPORT
- RESPONSIVENESS



- INNOVATION and SPEED
- CUSTOMIZATIONS

SNP & Variation Suite (SVS)



Core Features

- Powerful Data Management
- Rich Visualizations
- Robust Statistics
- Flexible

Applications

- Genotype Analysis
- DNA sequence analysis
- CNV Analysis
- RNA-seq differential expression



1 CNV Analysis

2 NGS Based CNV Analysis – How it works

3 NGS Based CNV Demo

4 Use Case – Cardio Cohort

CNVs Polymorphisms



- **CNVs are key contributors to intra-species genetic variation.**
- **Humans – drivers of specific cancers and associated with various diseases.**
 - EGFR Exon 19 deletion common in lung cancer.
 - PIK3CA Amplification in breast cancer
 - Autism Spectrum Disorder (ASD)
- **Animals - CNVs associated with bovine health and production traits¹**
 - Beta-defensin gene families.
- **Plants**
 - 400 CNVs Maize inbred Mo17 compared to *teosinte*²
 - 641 identified CNVs that distinguished two rice cultivars, Nipponbare (*O. sativa* ssp. *japonica*) and Guang-lu-ai 4 (*O. sativa* ssp. *indica*)³
 - Several 100 unique CNVs found to three soybean cultivars⁴
 - Flowering time and plant height (*Vrn-A1* & *Ppd*)

¹Fadista et al. 2010 – BMC Genomics

²Springer et al 2009 – PLoS Genetics

³Yu et al 2011 – BNC Genomics

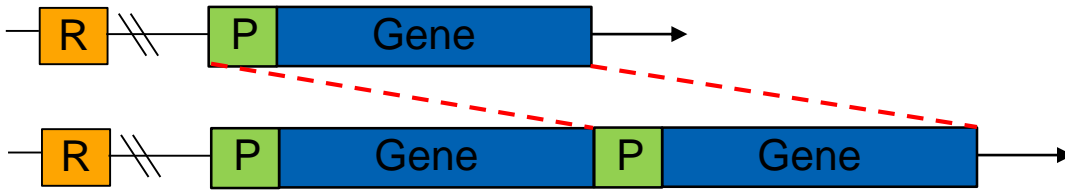
⁴Swanson-Wagner et al 2010 – Genome Res

CNVs Polymorphisms

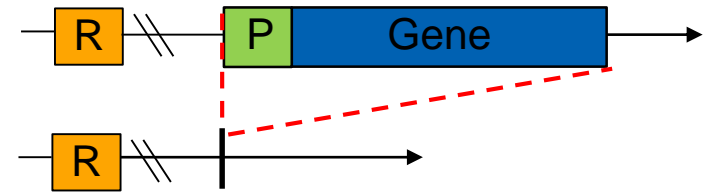


- CNVs overlapping a gene may alter the expression level of the gene by virtue of changing the number of functional copies.

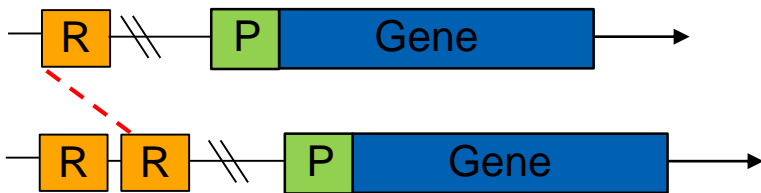
Tandem Gene Duplication



Complete Gene Deletion

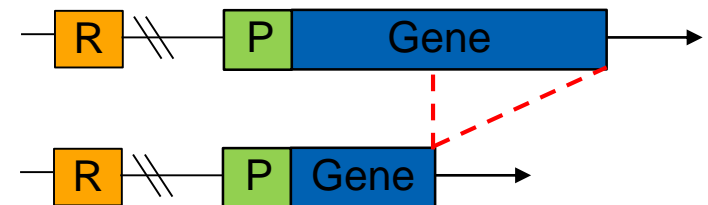


Duplication of Enhancer Sequences



Elevated Level of Transcript

Partial Gene Deletion



Decreased level of transcript



- Genomic prediction is a key focus for agrigenomics
- Growing world population requires improved food production
 - 3B in 1960
 - 7.3B today
 - 9.6B projected in 2050

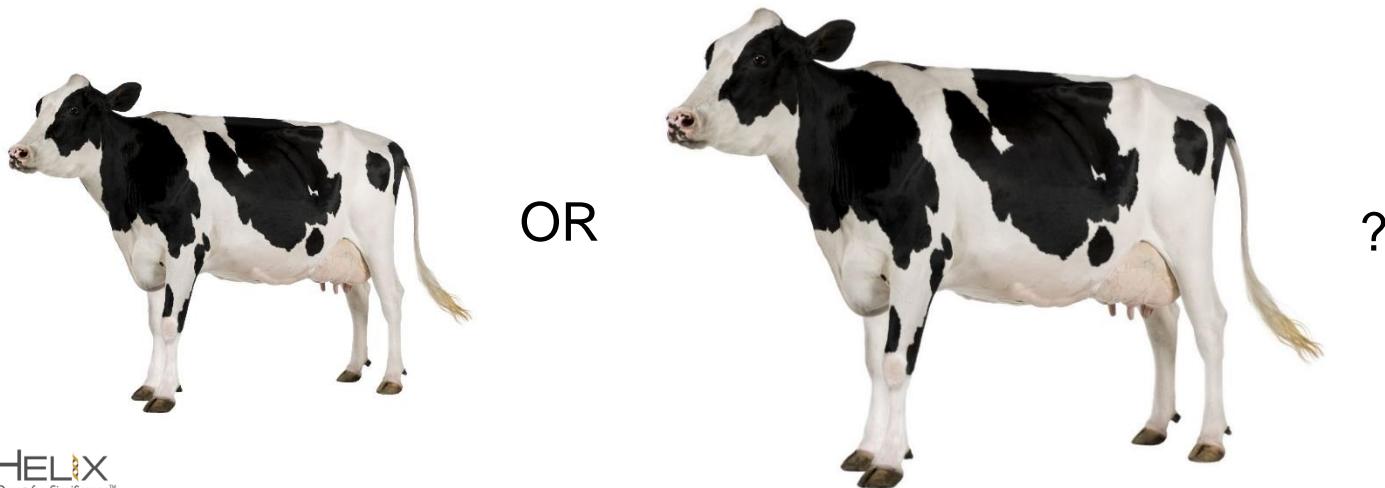
Source: Wikipedia

Country	Population 2010	Population 1990	Growth (%) 1990–2010
<i>World</i>	6,895,889,000	5,306,425,000	30.0%
China	1,341,335,000	1,145,195,000	17.1%
India	1,224,614,000	873,785,000	40.2%
United States	310,384,000	253,339,000	22.5%
Indonesia	239,871,000	184,346,000	30.1%
Brazil	194,946,000	149,650,000	30.3%
Pakistan	173,593,000	111,845,000	55.3%
Nigeria	158,423,000	97,552,000	62.4%
Bangladesh	148,692,000	105,256,000	41.3%
Russia	142,958,000	148,244,000	-3.6%
Japan	128,057,000	122,251,000	4.7%

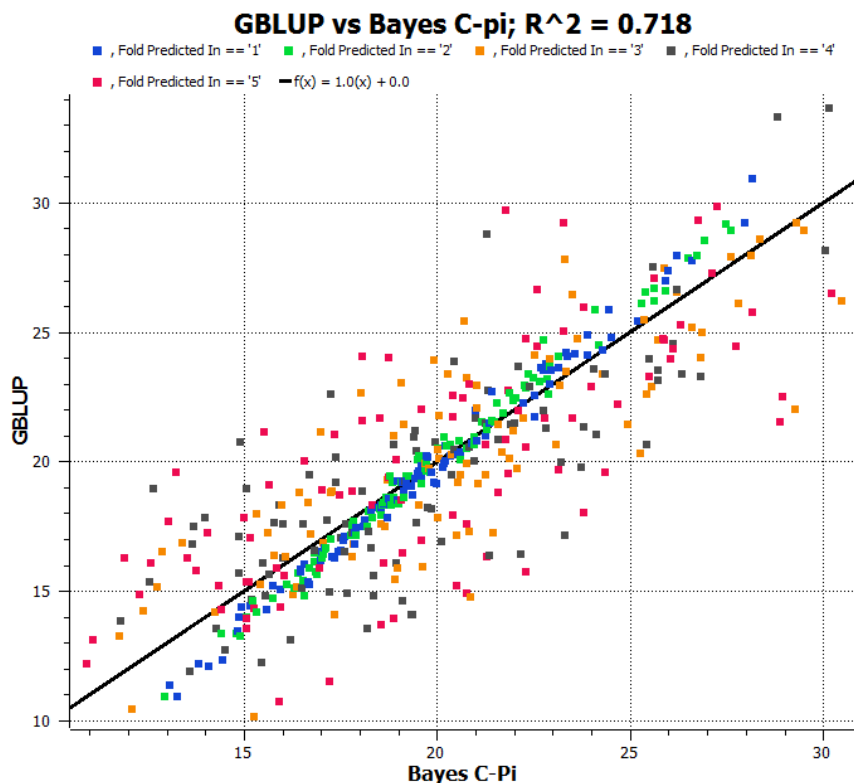
Why Use Genomic Prediction?



- **Calculate breeding value (gEBV) for all subjects in a population**
 - May be more accurate than breeding selection based only on pedigree and trait data
- **Predict breeding values for subjects with unknown phenotypes**
 - May avoid costly and lengthy field trials
 - May not always be possible to measure the phenotype
- **Identify genetic markers with best predictive power for a trait**
 - Assist in development of predictive tests and other assays



Genomic Prediction Methods Available in SVS



■ GBLUP

- Assumes all loci contribute to the phenotype

■ Bayes C

- Estimates effects of gene loci together with parameters required to define probability distribution over events
- Prior probability that any SNP will have no effect fixed

■ Bayes C-pi

- Prior probability that any SNP will have no effect unknown and allowed to vary



- **Assumes all loci contribute to phenotype**
- **Incorporates genomic relationship matrix (GRM) in mixed linear model framework to account for relatedness among samples**
- **Calculates allele substitution effect (ASE) for each SNP**
- **Computes estimated breeding values (GEBV) and predicted phenotypes for all samples**
- **Also calculates:**
 - Pseudo-heritability of trait
 - Genetic component of trait variance
 - Error component of trait variance

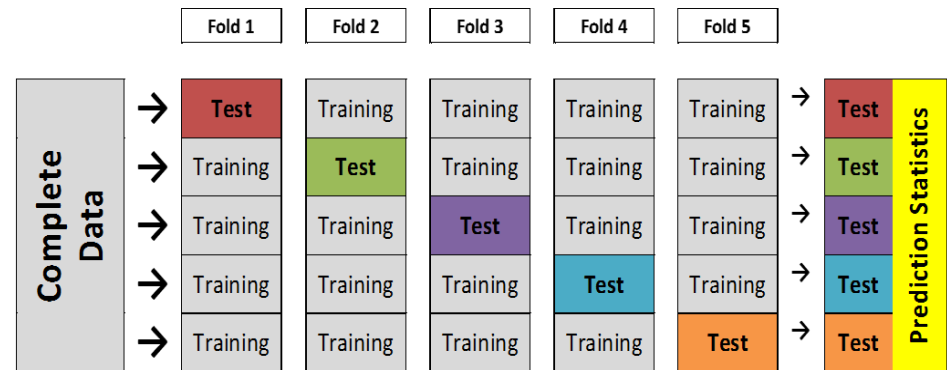


- Bayesian methods predict effects of loci together with parameters required to define probability distribution over effects
- Gibbs sampling (MCMC) used to obtain parameter estimates
- The π (pi) parameter is the prior probability that any SNP will have no effect.
- Pi is fixed in Bayes C, typically at $\pi=0.9$
- Pi is considered unknown and allowed to vary by Bayes C-pi
- Both methods return ASE, gEBV, other parameters of final model
- SVS implementation incorporates GRM

K-Fold Cross-Validation

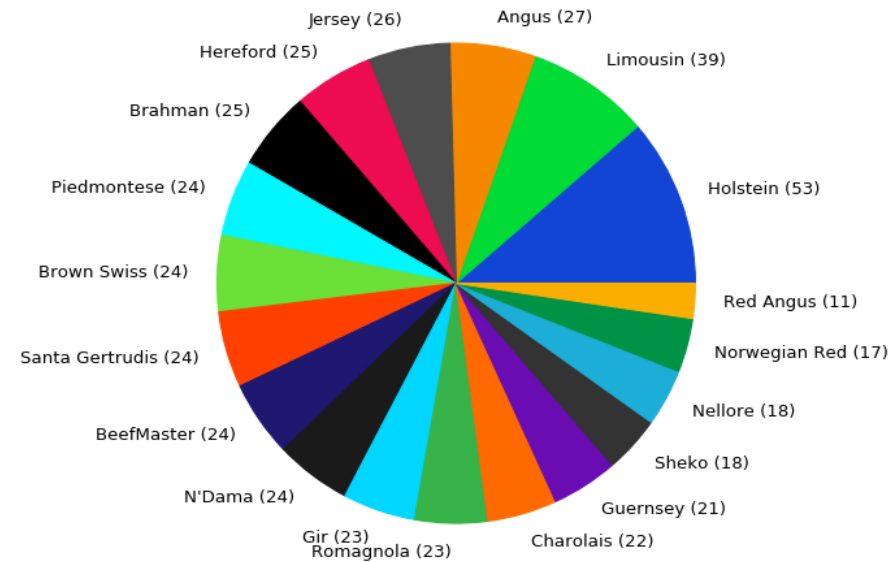


- Build a model that can be applied to new genetic data to predict a phenotype
- Cross Validation makes it possible to assess the performance built from a given reference/training dataset.
- Can be used with GBLUP, Bayes C, Bayes C-pi
- Requires all samples have a phenotype value
- Can include covariates





- **472 Bos taurus cattle from Bovine HapMap project**
 - 422 samples: training set
 - 50 samples: validation set (no phenotypes)
- **Illumina 50k genotypes**





GOLDEN HELIX

SNP & VARIATION SUITE

Use Case – Improving Beef Quality



■ Dr. Raluca Mateescu

- Department of Animal Sciences at University of Florida
- Journal of Animal Science 90, 4248-4255 (2012)
 - Link to slides
http://goldenhelix.com/media/pdfs/webcasts/GoldenHelix_July2015_Handout.pdf
 - Link to webcast
https://www.youtube.com/watch?v=e6Czycr_DnE

■ Develop tools to select for:

- Nutritious beef
- Tasty beef
- Improved production

■ Focus of Research

- Overall: Beef healthfulness
- Phenotypes measured:
 - **WB-Steak Shear Force**
 - Tenderness
 - Juiciness
 - Connective tissue
 - Beef Flavor

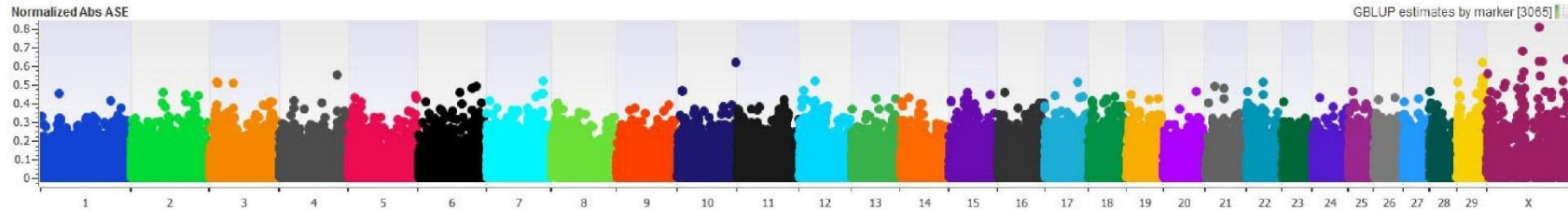
Use Case – Improving Beef Quality



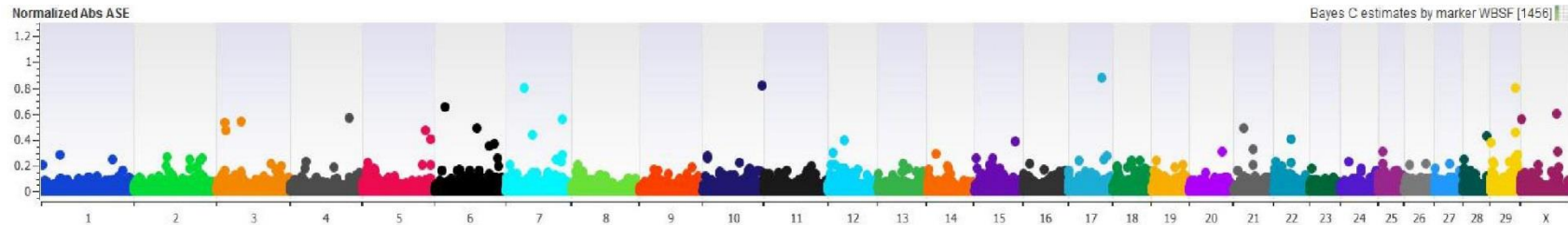
■ Genomic Prediction results – Shear Force

- GBLUP and Bayes C-pi
- Manhattan plot with Allele Substitution Effects

GBLUP



Bayes C-pi

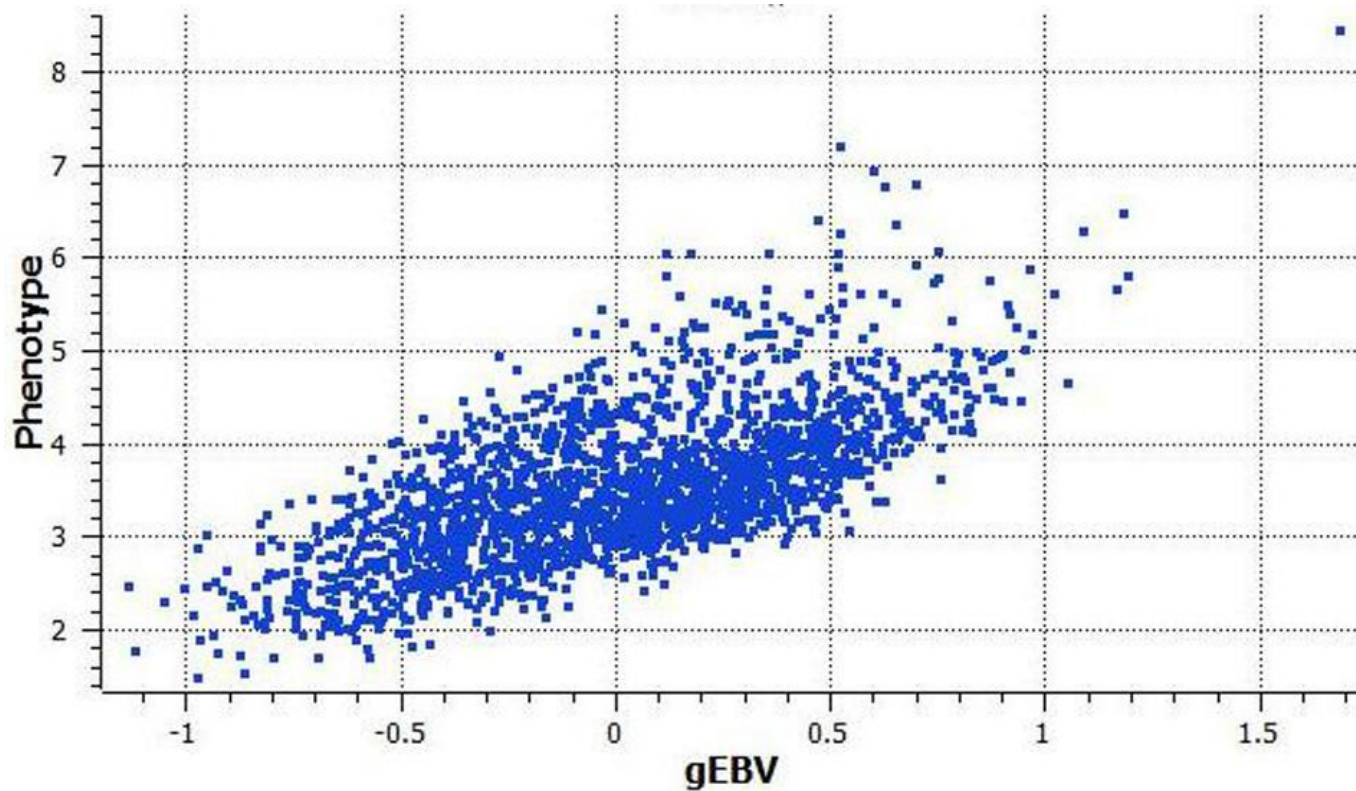


Use Case – Output from GBLUP



■ Correlations for Shear Force

- Actual phenotypes vs Estimated Breeding value (gEBV)



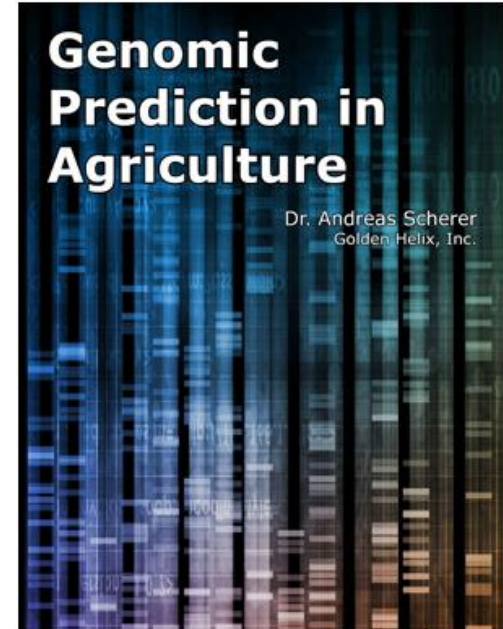


■ Outcome of the study

- Identified markers associated with beef quality
 - *SAPS3*: modulates protein phosphatase catalytic subunits
 - *CAPN1*: modulates proteolysis of cytoskeletal remodeling and signal transduction
 - *CHI3L2*: involved in cartilage biogenesis
 - *CA10*: catalyzes reversible hydration of carbon dioxide
 - *GPHN*: involved in membrane protein-cytoskeleton interactions



- **Genomic Prediction ebook**
- **Golden Helix at PAG 2018 – San Diego!**
 - Grand Exhibit Hall
 - Come see demos and ask us questions!
- **SVS package upgrade!**
 - NGS-based CNV calling for large sample size & association tests





Questions or more info:

- Email info@goldenhelix.com
- Request an evaluation of the software at www.goldenhelix.com
- Check out our abstract competition!

