



Fundamentals of Genomic Prediction and Data-Driven Crop Breeding (November 24-28, 2025)

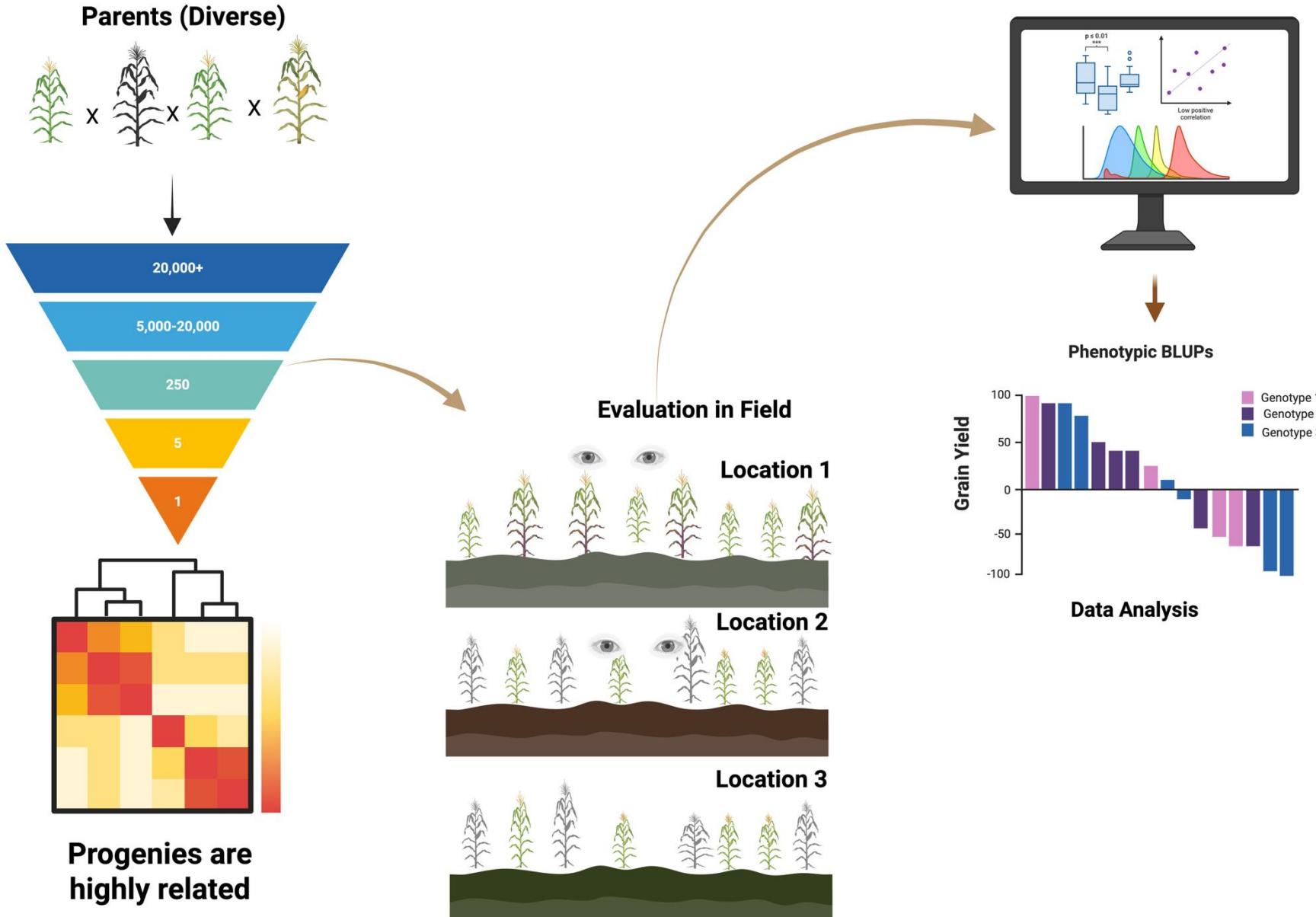
Fundamentals of Genomic Predictions

Module 2
November 25, 2025

**Waseem Hussain, Mahender Anumalla and Margaret
Catolos**

Rice Breeding Innovations Platform
IRRI

Traditional Breeding Phenotypic Selection



Assumption that all
Genotypes are
independent
(No Relationship)

***But that is not
True!***

*(Genotypes have a half-sib or full-sib relationship).
There is variance between the families, within families, But Also covariance (genetic similarity within families and across families)*

General Breeding Pipeline

Parental Pool

P1 x P2.....P3 x P4.....P1 x P4

Segregating Generations

F₂ - F₃ - F₄ - F₅

Fixed Generation

F₆

Multilocation Evaluation

F₇
(884 Genotypes)

Multi-location Evaluation

Final Selections

Seed System
(2-3)

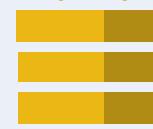
Phenotypic Selection

Option 1

Location 1
(884)

Location 2
(884)

Location 3
(884)



Field Evaluation across Locations

What we assume here

	G1	G2	G3	GN
G1	1	0	0	0	0	0
G2	0	1	0	0	0	0
G3	0	0	1	0	0	0
.	0	0	0	1	0	0
.	0	0	0	0	1	0
Gn	0	0	0	0	0	1

Perform Analysis



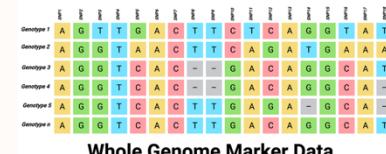
BLUEs and BLUPs

Genotypes are independent
(No-covariance)

Final Selections

Genomic Selection

All 884 GENOTYPES



Option 2

Limited Resources
Estimate Breeding Value
Perform Predictions
Dissect G x E better

Location 1
(254)

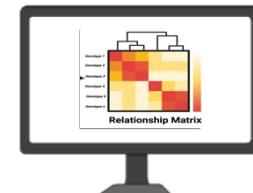
Location 2
(254)

Location 3
(254)



Field Evaluation across Locations

Perform Analysis

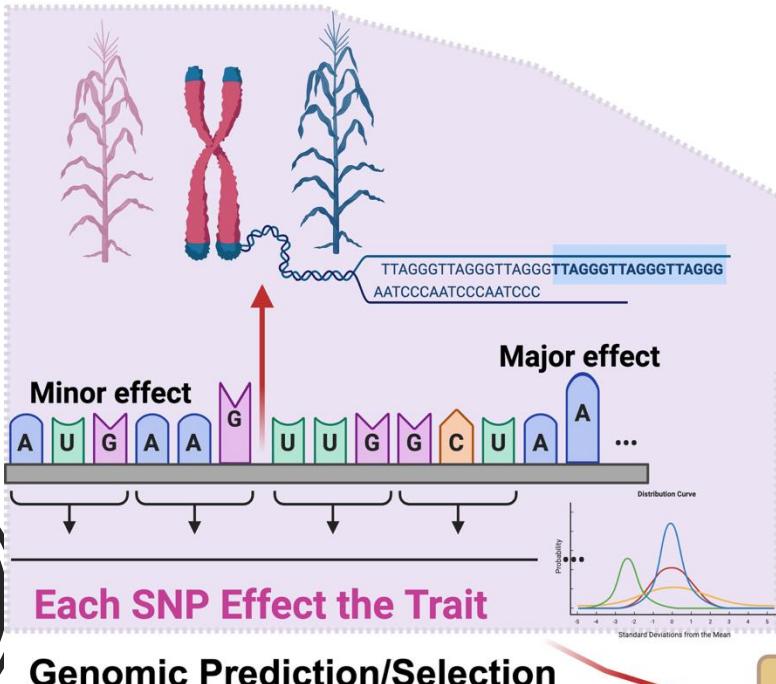


Leverage covariances
(Relative Information)
Genotypes are Related

(Predict un-tested)

BLUPs
Final Selections

Genomic Selection (Leveraging Relationship)

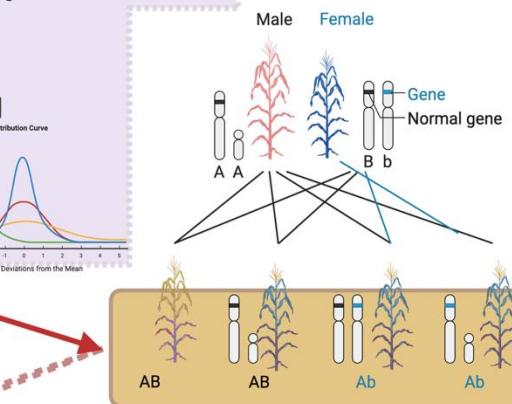


Genomic Prediction/Selection

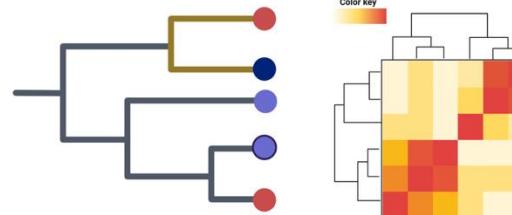
Significant SNPs
Only Few Effect Trait
(Missing Heritability)

MAS

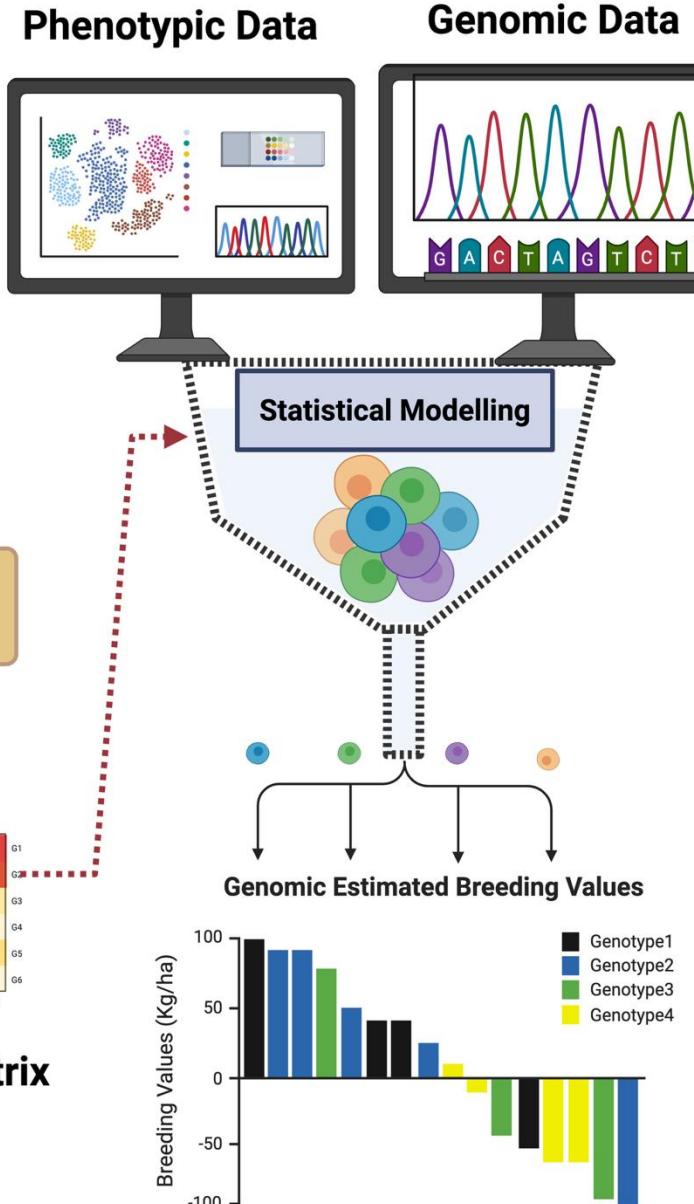
Chromosome



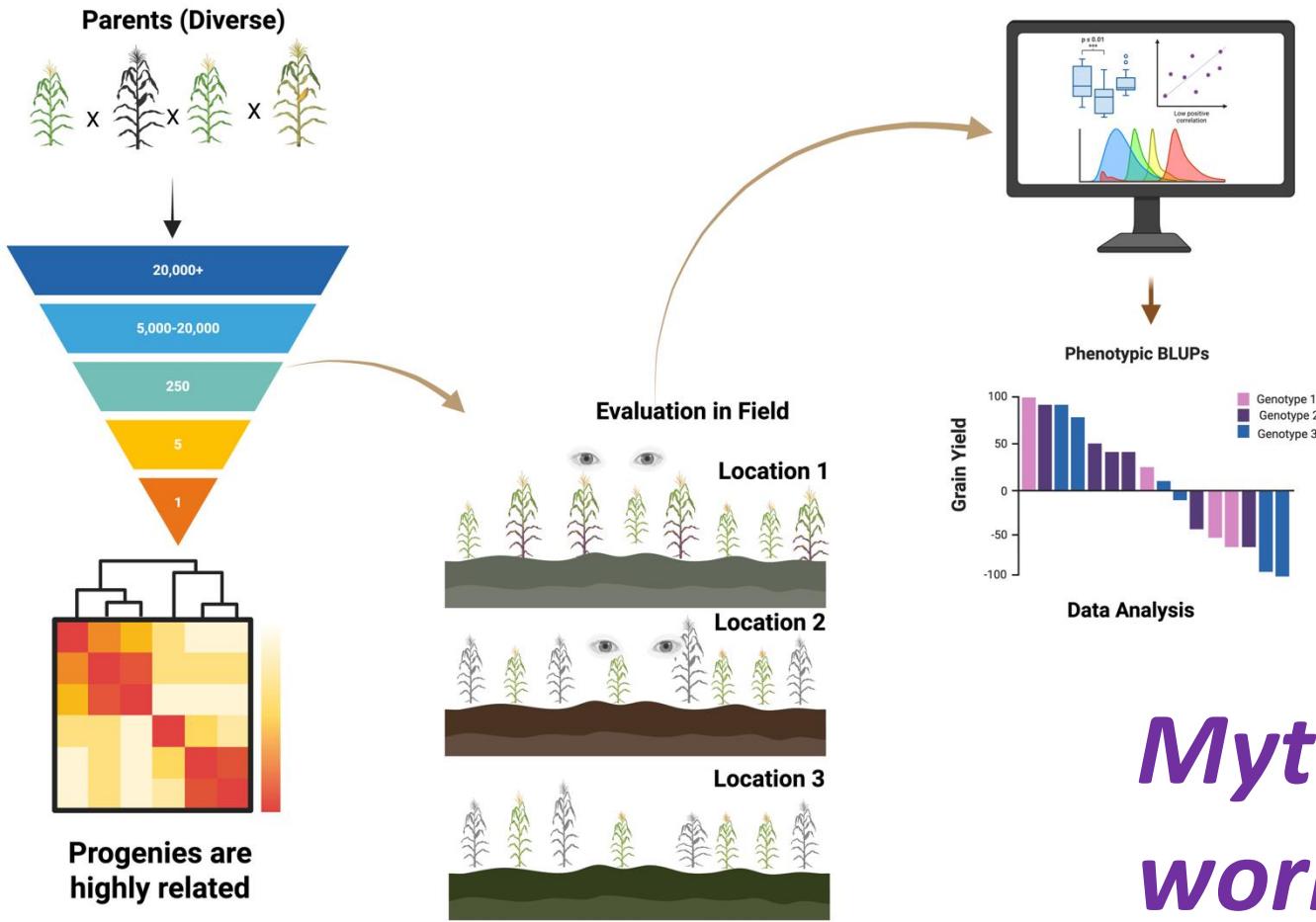
Phenotype (e.g Yield)



Leverage Relationship Matrix



Phenotypic Selection vs MAS vs Genomic Selection



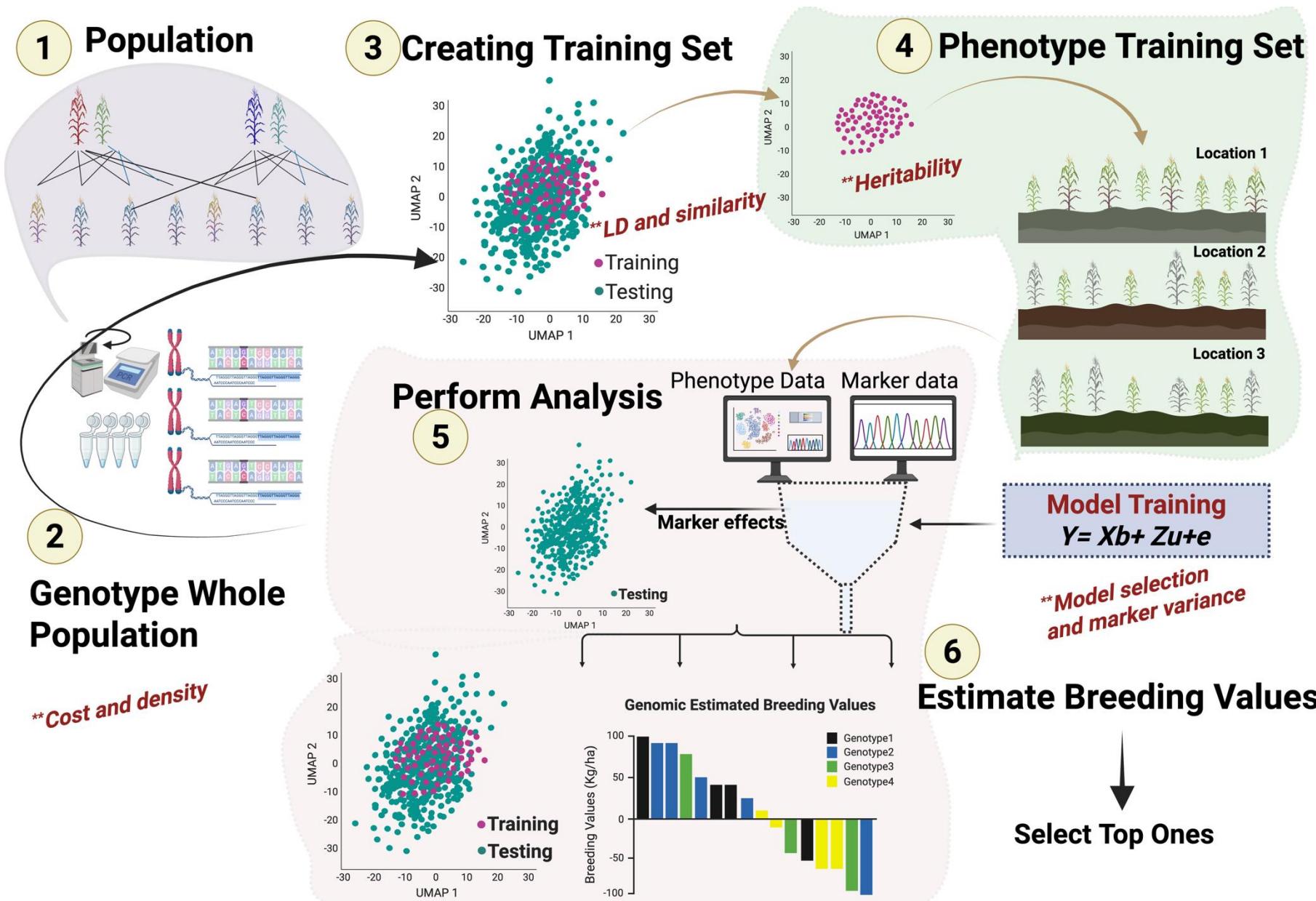
Phenotypic selection: No Relationship

MAS: No relationship

Genomic Selection: Relationship is Foundation

Myth 1: Genomic Prediction works in all Populations like MAS

How Genomic Prediction Works



Evaluate Prediction Performance

Predictive correlation (r) = $\text{cor}(\text{True value}, \text{Predicted value})$

Divide the data into training and testing set

Training set $\in (y_{\text{training}}, X_{\text{training}})$

Testing set $\in (y_{\text{testing}}, X_{\text{testing}})$

$$y_{\text{training}} = X_{\text{training}}\beta_{\text{training}} + \varepsilon_{\text{training}}$$

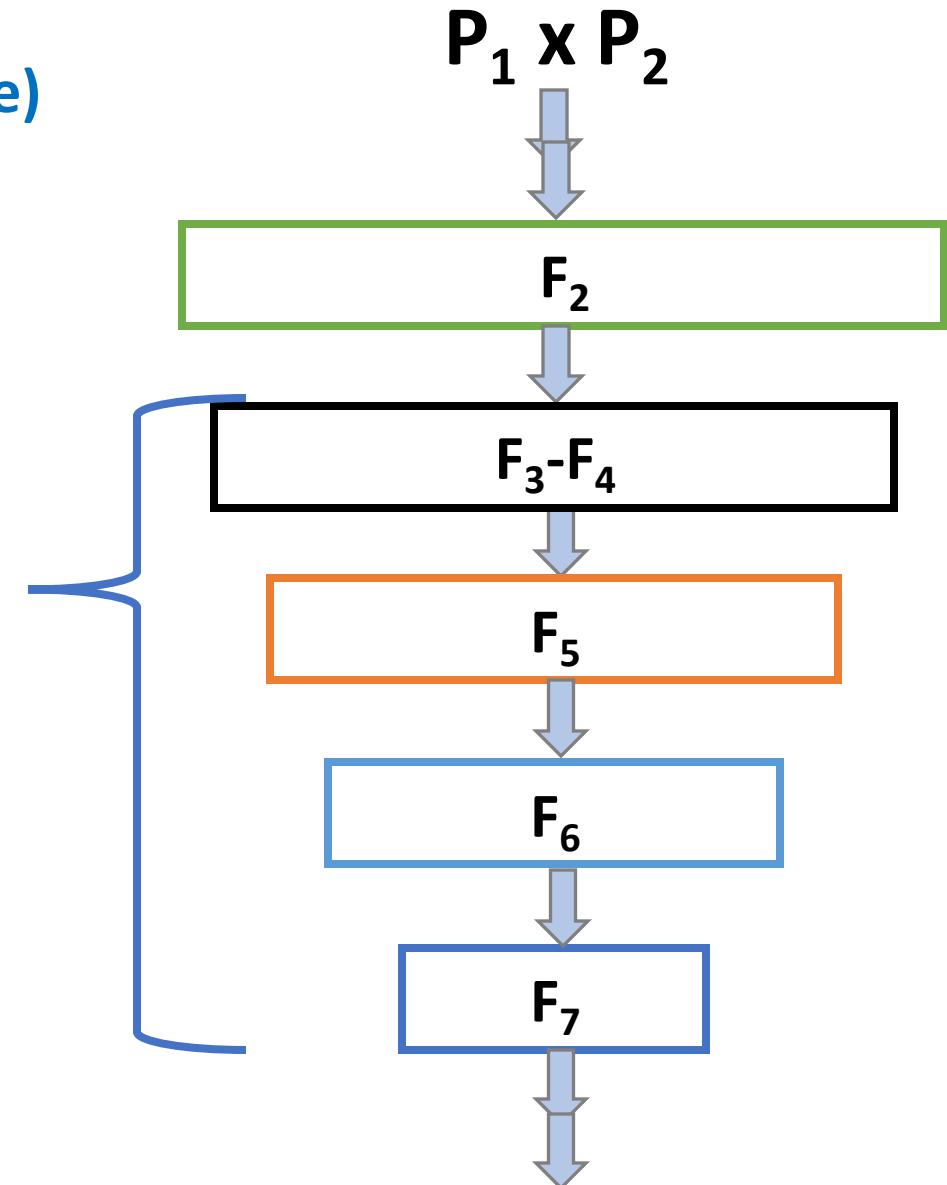
Perform cross-validation

$$\hat{y}_{\text{testing}} = X_{\text{testing}}\beta_{\text{testing}}$$

$$= \text{cor}(y_{\text{testing}}, \hat{y}_{\text{testing}})$$

$$= \text{cor}(y_{\text{testing}}, X_{\text{testing}}\beta_{\text{testing}})$$

Breeding Pipeline

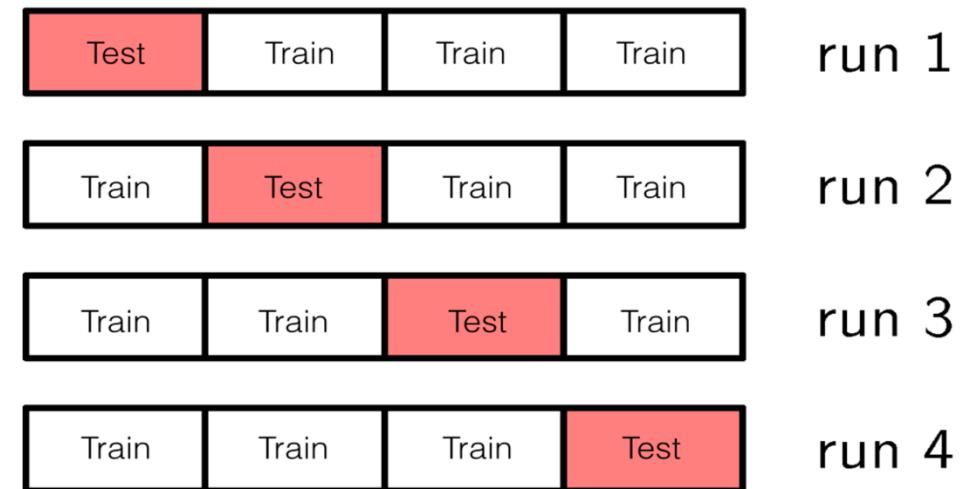


Cross-validation

(evaluate prediction performance)

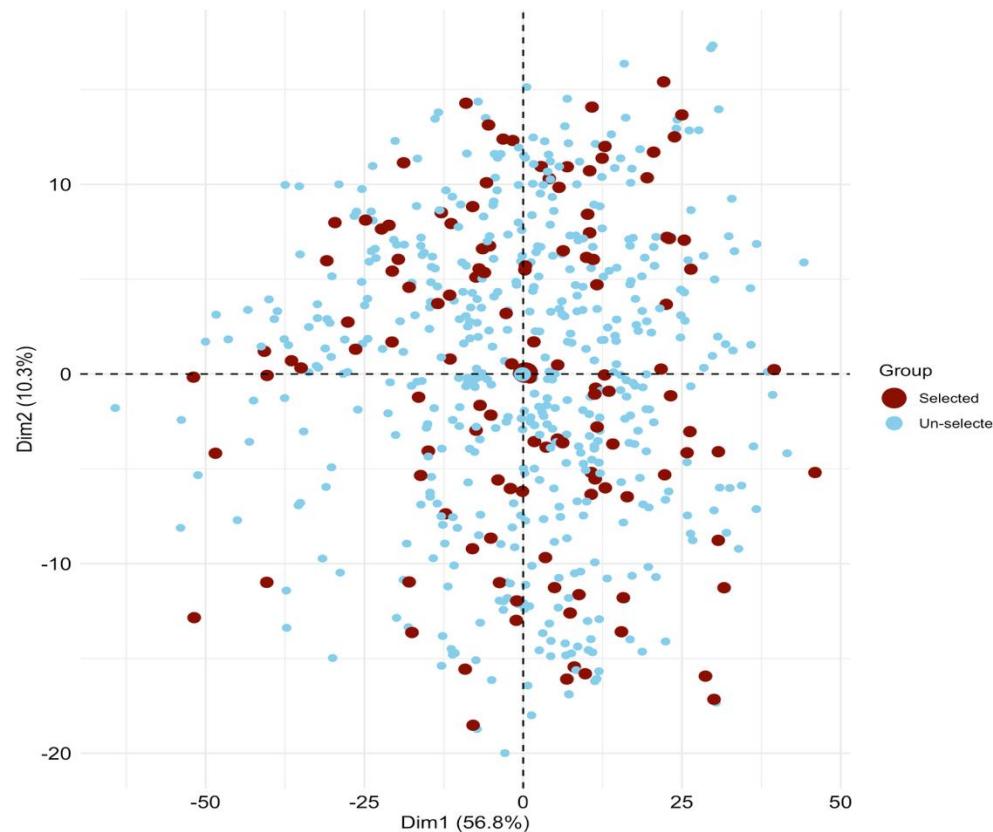
- ❖ Take model uncertainty into account
- ❖ Divide data into training and testing sets
- ❖ Train the model in the training set
- ❖ Evaluate predictive performance in the testing set
- ❖ Predictive correlation: $r = \text{cor}(y, y_{predicted})$

K-fold Cross-validation



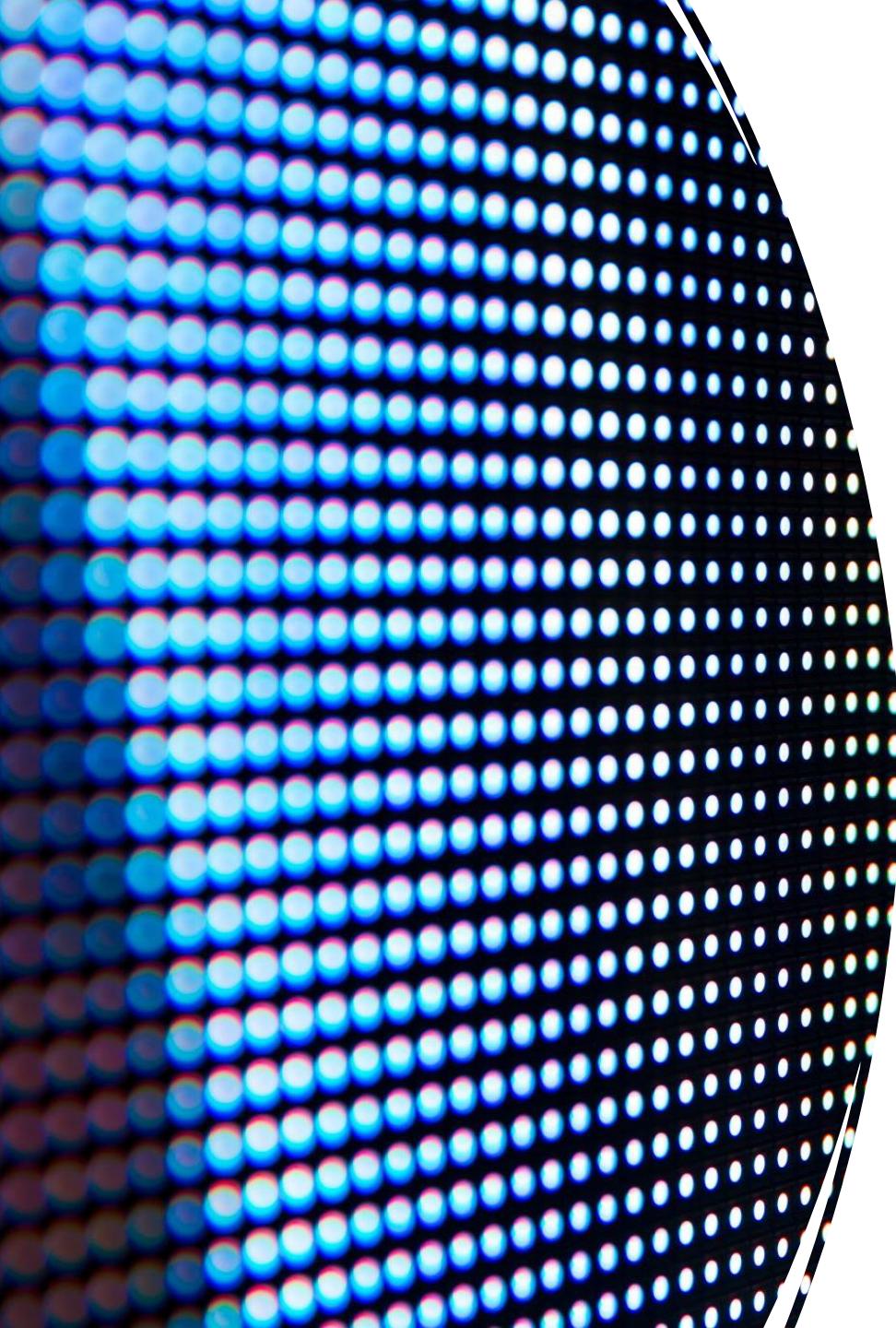
How to Design The Training Set

- Maximizes the relationship between training and testing set
- Key for success of genomic selection



Various Approaches to Design Training Set

- ❖ **Prediction error variance (PEV)** means: *Minimize the error variance*
- ❖ **Critical Difference (CD)** means: *Minimize error variance and Relationship*
- ❖ **K-means algorithm**: *Based on G matrix and similarity*



Literature to Read on Matrix

- <https://courses.lumenlearning.com/wmopen-collegealgebra/chapter/introduction-matrices-and-matrix-operations/>
- <https://courses.lumenlearning.com/boundless-algebra/chapter/introduction-to-matrices/>
- [https://math.libretexts.org/Bookshelves/Applied_Mathematics/Book%3A_Applied_Finite_Mathematics_\(Sekhon_and_Bloom\)/02%3A_Matrices/2.01%3A_Introduction_to_Matrices](https://math.libretexts.org/Bookshelves/Applied_Mathematics/Book%3A_Applied_Finite_Mathematics_(Sekhon_and_Bloom)/02%3A_Matrices/2.01%3A_Introduction_to_Matrices)
- <https://byjus.com/math/singular-matrix/>
- <https://www.onlinemathlearning.com/singular-matrix.html>

**Genomic Selection is
Simply a tool to
Supplement the
Breeding Pipeline**



**Thank You
Questions**