Fundamentals of Genomic Prediction and Data-Driven Crop Breeding (August 4-8, 2025)
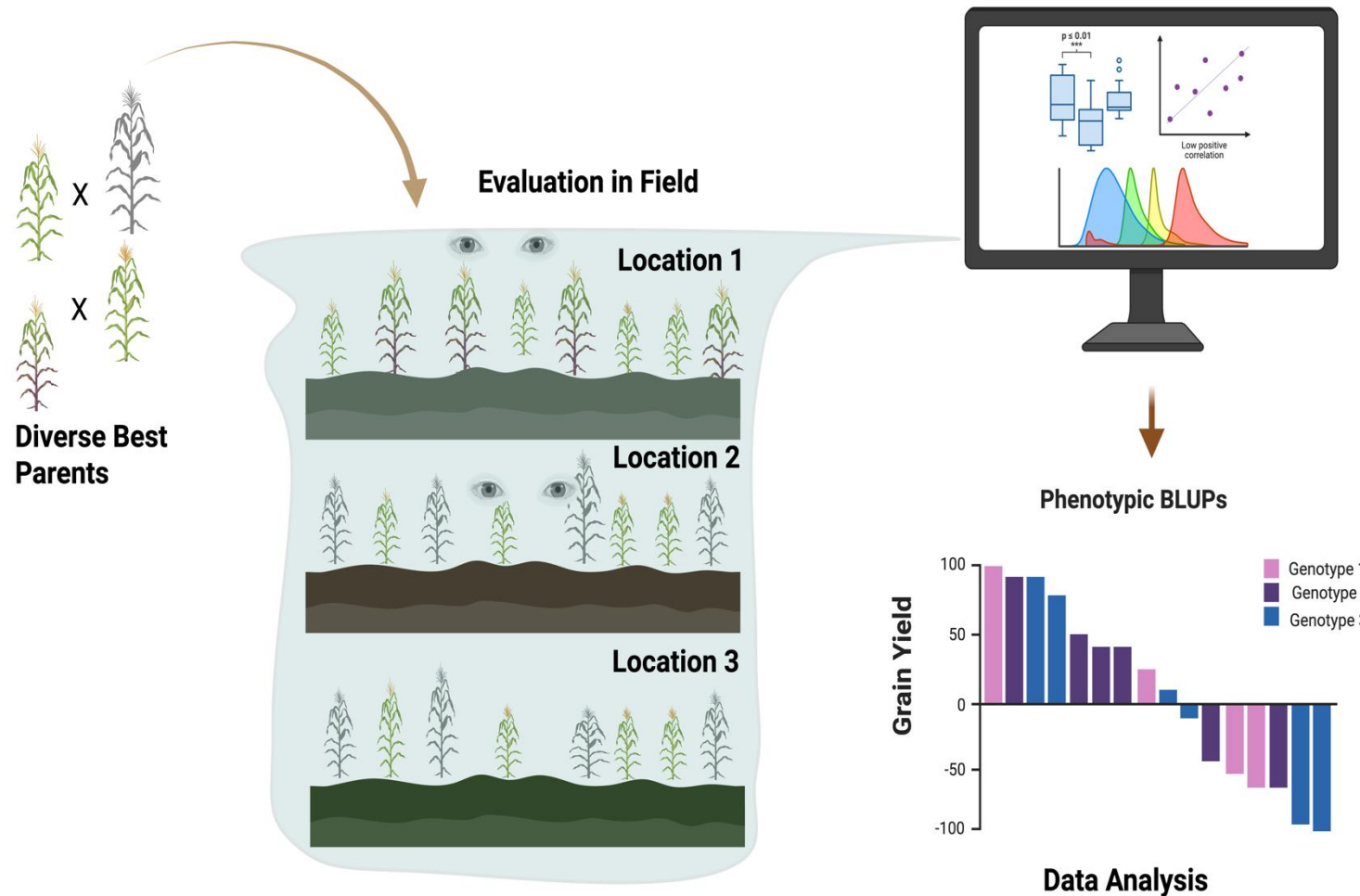
Fundamentals of Genomic Predictions

Module 2
August 5, 2025

Waseem Hussain and Mahender Anumalla
Rice Breeding Innovations Platform
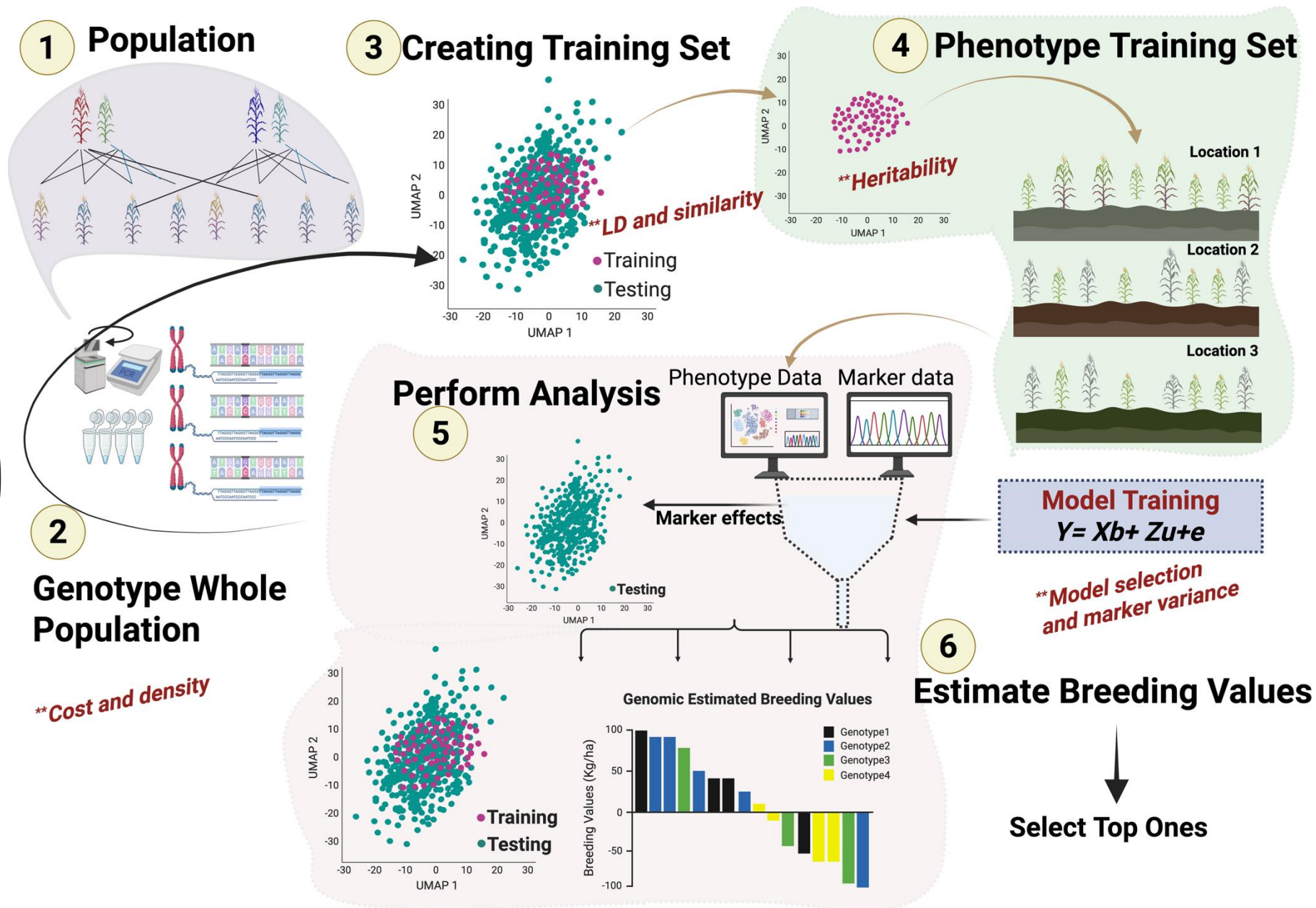**IRRI**

Traditional Breeding Phenotypic Selection

Genomic Selection (Leveraging Relationship)

Genomic Prediction/Selection

Minor effect

Major effect

Each SNP Effect the Trait

A U G A A G U U G G C U A A ...

Distribution Curve

Phenotype (e.g Yield)

Male   Female

Gene
Normal gene

A A   B b

AB   AB   Ab   Ab

Leverage Relationship Matrix

Color key

G1
G2
G3
G4
G5
G6

G1 G2 G3 G4 G5 G6

Significant SNPs

Only Few Effect Trait (Missing Heritability)

MAS

SNPs

Chromosome

Phenotypic Data

Genomic Data

G A C T A G T C T G

Statistical Modelling

Genomic Estimated Breeding Values

Breeding Values (Kg/ha)

100

50

0

-50

-100

Genotype1
Genotype2
Genotype3
Genotype4

Evaluate Prediction Performance

Predictive correlation (r)= cor(True value, Predicted value)

Divide the data into training and testing set

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

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

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

Perform cross-validation

$$\acute{y}_{testing} = X_{testing}\beta_{testing}$$

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

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

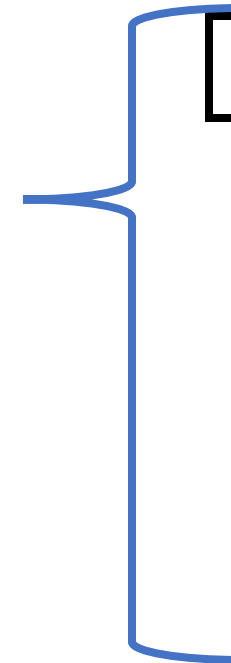**Breeding Pipeline**
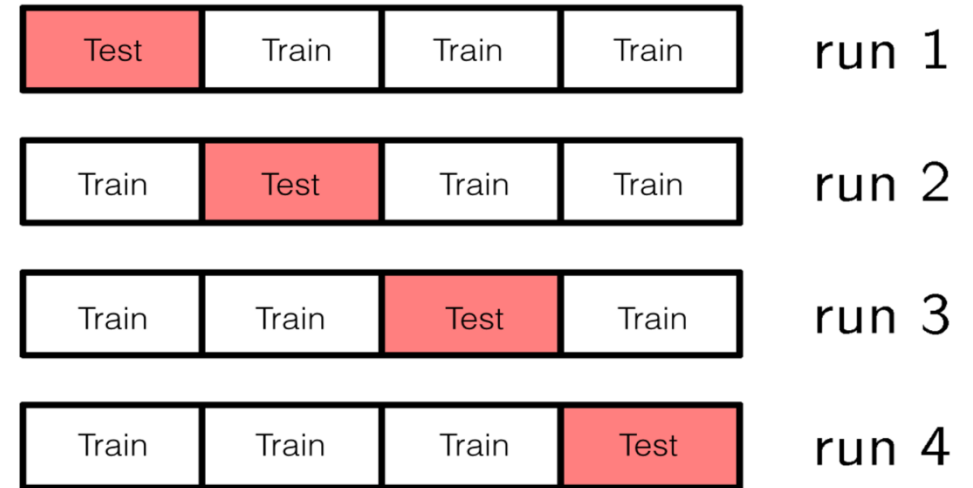
P_1 x P_2

F_2

F_3-F_4

F_5

F_6

F_7

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: $r = cor(y, y_{predicted})$

K-fold Cross-validation

Test	Train	Train	Train	run 1

Train	Test	Train	Train	run 2

Train	Train	Test	Train	run 3

Train	Train	Train	Test	run 4

How to Design The Training Set

➢ Maximizes the relationship between training and testing set

➢ Key for success of genomic selection
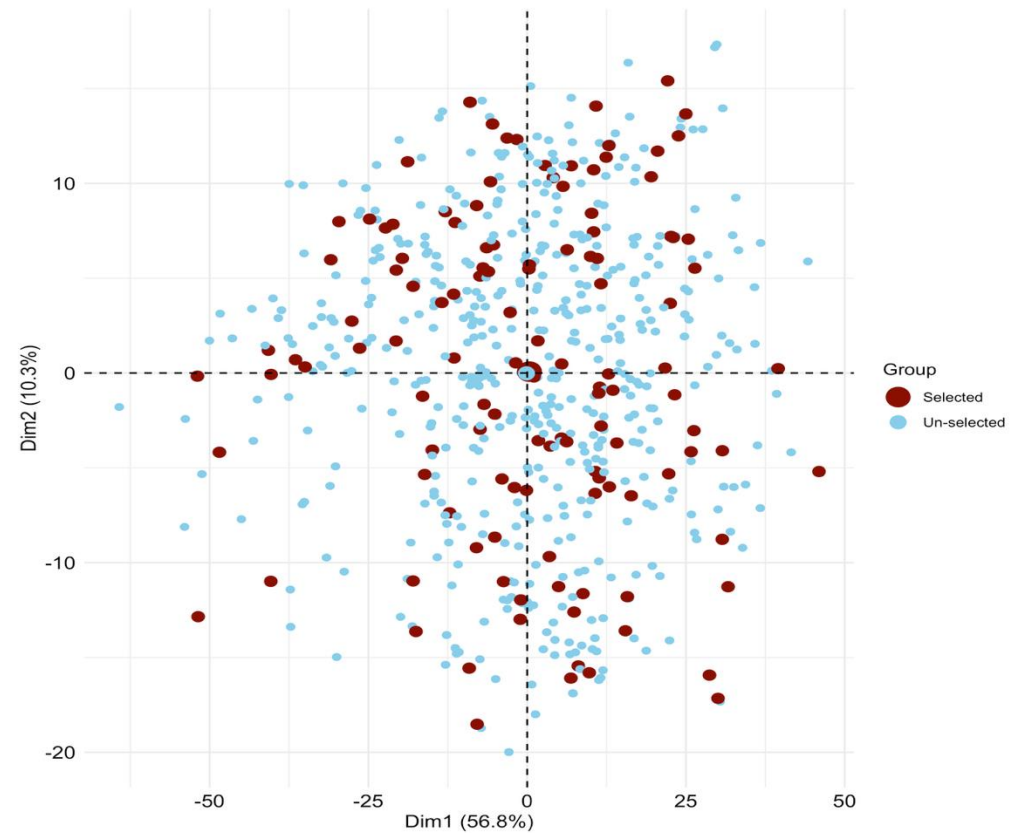
Various Approaches to Design Training Set

❖ **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*

Rincent et al. 2012; Rincent et al. 2012

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://byjus.com/maths/singular-matrix/

- https://www.onlinemathlearning.com/singular-matrix.html

Genomic Selection is Simply a tool to Supplement the Breeding Pipeline

Thank You

Questions