

# DATA *Nugget*

## **Supplemental Activity: A plant breeder's quest to improve perennial grain**

Featured scientist: Hannah Stoll (she/her) from The University of Minnesota

*In the previous activity, you used data from Breeding Cycles 1-4 to look at whether different genotypes had different phenotypes and found that certain genotypes were associated with larger seeds. You will now explore how Hannah used genotype and phenotype data from those previous generations to make predictions about Breeding Cycle 5!*

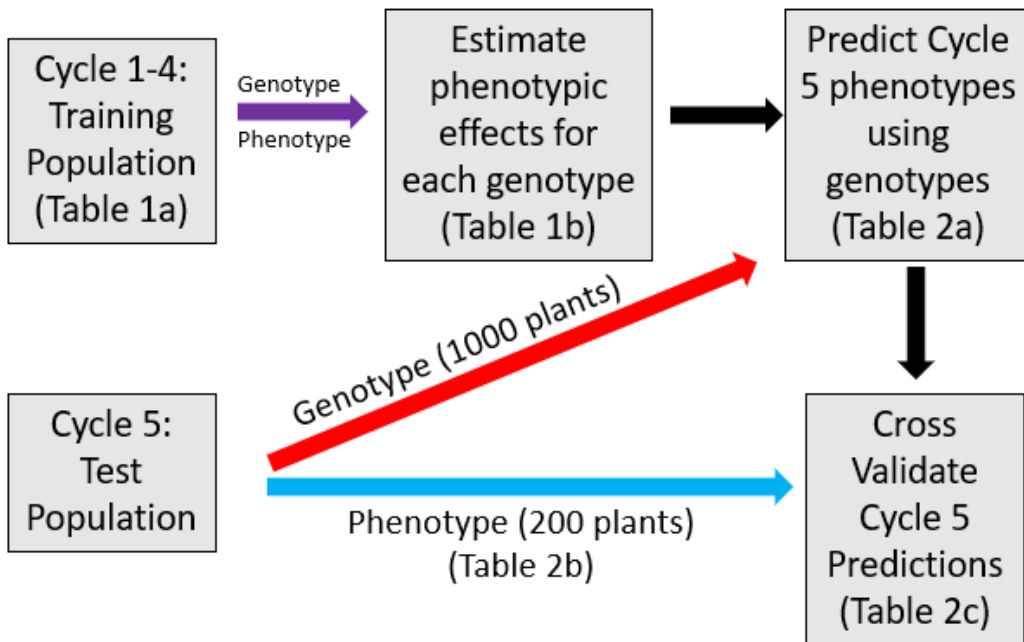
### Supplemental Research Background:

Breeders are interested in increasing seed size in Kernza<sup>®</sup>, and have done so up to this point selecting individuals based on phenotype, not genotype. This is an issue, since phenotyping plants requires scientists to grow thousands of plants, which is time-consuming and expensive. Hannah wants to know if breeders can instead rely on genotype to select the individuals that will contribute to future breeding cycles.

Hannah set out to evaluate whether she can accurately predict the phenotype of an individual plant from their genotype. She went back to data from Breeding Cycles 1-4. She used the average phenotypes for each genotype to make predictions about a new generation, Breeding Cycle 5. If her predictions were accurate, she would know that she could use genotype to predict phenotype!

Hannah looked at the resources and funds she had to collect data on Breeding Cycle 5. She saw that she had funds to genotype all 1,000 plants but could only phenotype a subset of 200 plants.

Hannah compared phenotyped individuals from Cycle 1-4 to the 200 phenotyped individuals from Cycle 5. If the predicted phenotype values from Cycle 1-4 for each genotype are accurate, then those values can be used to predict the phenotypes of the remaining 800 plants in Cycle 5, which she didn't have the resources to grow and obtain phenotypes directly.



Schematic of how a plant breeder may use both phenotypic and genotypic data in a Kernza® breeding program to make selections for future generations. Table 1a-1b reference the tables from the Data Nugget.

Scientific Question: How well does genotypic data from Breeding Cycles 1-4 predict phenotypes of Breeding Cycle 5 individuals?

Scientific Data:

**Use the data below to answer the scientific questions:**

*Remember: Hannah doesn't have the resources to phenotype all 1000 of her Cycle 5 plants, but she does have the resources to genotype all of them. Use the phenotypic values you calculated in the first activity (Table 1b) to predict seed area for this subset of Cycle 5 plants.*

Table 2a. Test Data: Fill in the genotypes and the associated predicted Cycle 5 phenotype.

Plant ID	Breeding Cycle	Seed Area (mm)	SNP 01	SNP 04	Genotype
1_2	5	5.57	A	G	AG
2_2	5	6.20	A	A	AA
3_2	5		T	A	
4_2	5		T	A	
5_2	5		A	A	
6_2	5		T	G	
7_2	5		A	C	
8_2	5		T	C	
9_2	5		A	C	
10_2	5		A	G	
11_2	5		A	A	
12_2	5		A	C	
13_2	5		T	G	
14_2	5		T	A	
15_2	5		T	C	
16_2	5		T	G	
17_2	5		A	G	
18_2	5		T	A	
19_2	5		T	A	
...	...	...	...	...	...
1000_2	5		T	C	

Hannah phenotyped 200 of the 1000 plants in the Cycle 5 population and calculated the following phenotypic values.

Table 2b. Test data calculated from phenotype and genotypes of 200 Cycle 5 plants.

Genotype	Average Seed Area (mm)	Sample Size	Standard Deviation	Standard Error
AA	6.14	30	0.01	0.08
AC	6.73	33	0.03	0.17
AG	5.53	28	0.02	0.10
TA	7.40	35	0.02	0.10
TC	7.78	41	0.02	0.15
TG	6.92	33	0.01	0.08

Now, calculate the difference between your predicted phenotypes (from Cycle 1-4 data) and the actual phenotypes from Cycle 5 plants.

Table 2c. Validation table.

Genotype	Cycle 1-4: Average Seed Area (mm)	Cycle 1-4: Standard Deviation	Cycle 5: Average Seed Area (mm)	Cycle 5: Standard Deviation	Difference between Cycle 1-4 Average and Cycle 5 Average
AA		0.14		0.08	
AC		0.19		0.17	
AG		0.15		0.10	
TA		0.20		0.10	
TC		0.15		0.15	
TG		0.15		0.08	

What data will you graph to answer the question?

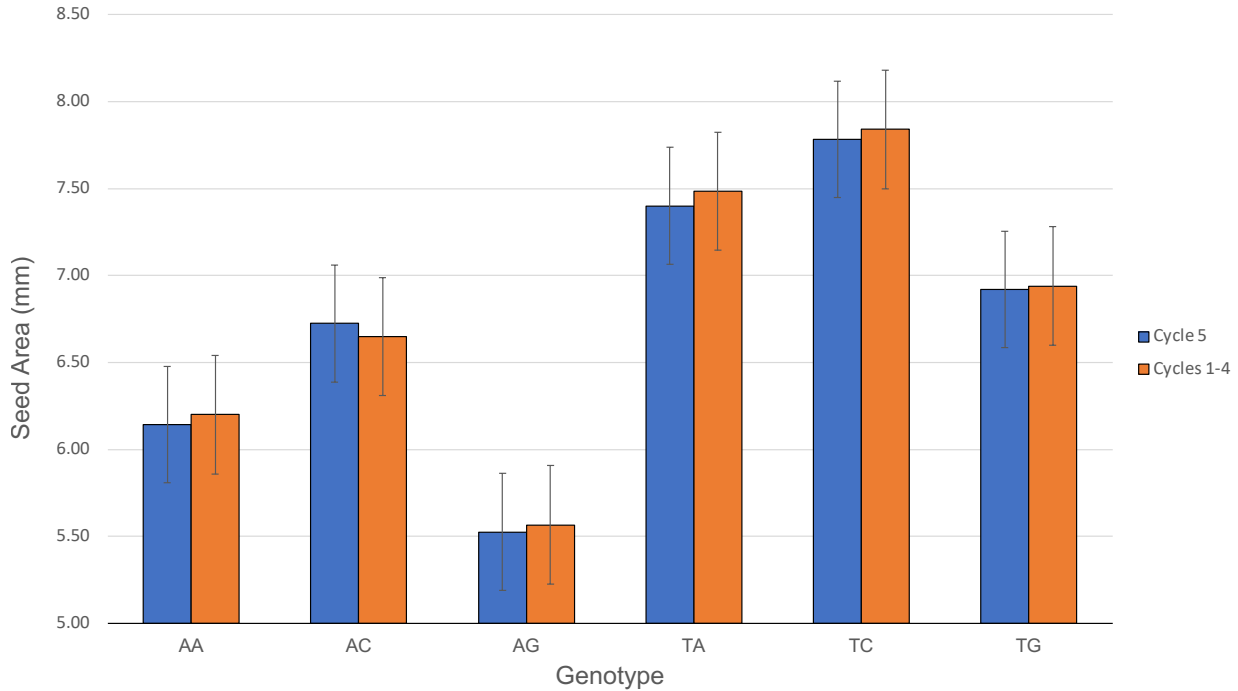
Independent variable: \_\_\_\_\_

\_\_\_\_\_

Dependent variable: \_\_\_\_\_

\_\_\_\_\_

*Below is a graph of the data:* Identify any changes, trends, or differences you see in your graph. Draw arrows pointing out what you see and write one sentence describing what you see next to each arrow.



***Interpret the data:***

Make a claim that answers the scientific question, how well does genotypic data from Breeding Cycles 1-4 predict phenotypes of Breeding Cycle 5 individuals?

What evidence was used to write your claim? Reference specific parts of the table or graph.

Name \_\_\_\_\_

Explain your reasoning and why the evidence supports your claim. Connect the data back to what you learned about how genotyping information can be leveraged in a Kernza® breeding program.

*Your next steps as a scientist:* Science is an ongoing process. What new question(s) should be investigated to build on Hannah's research? How do your questions build on the research that has already been done?