|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Years | Locations | Varieties | Phenotypes |
| Raw CDBN data | 1950 – 2018 | 80 + | 544 + | 25 – 152 |
| This analysis | 1981 – 2015 | 75 | 544 in mixed linear models;  312 in genome-wide association | 22 |

**Supplementary Table 1.** Summary of the Cooperative Dry Bean Nursery Dataset of common bean (*Phaseolus vulgaris*) phenotypes and the subset used in the present analysis.

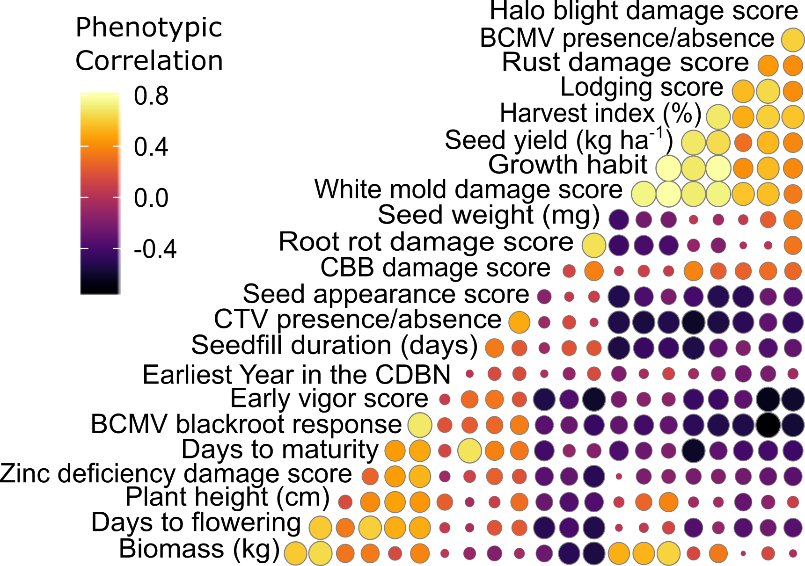
**Supplementary Table 2.** Excel File. Location information, genotyped Cooperative Dry Bean Nursery germplasm information, and corrected phenotype medians for 22 phenotypes used for each entry for genome-wide association on common bean (*Phaseolus vulgaris*).

|  |  |
| --- | --- |
|  | Principle Components |
| Seed yield (kg ha-1) | 2 |
| Seed weight (mg) | 7 |
| Days to maturity | 0 |
| Days to flowering | 0 |
| Seedfill duration (days) | 0 |
| Lodging score | 2 |
| Harvest index (%) | 2 |
| Plant height (cm) | 0 |
| Biomass (kg) | 0 |
| Growth habit | 2 |
| Seed appearance score | 0 |
| CBB damage score | 0 |
| Rust damage score | 0 |
| Early vigor score | 2 |
| White mold damage score | 2 |
| CTV presence/absence | 0 |
| Halo blight damage score | 0 |
| Blackroot BCMV response | 0 |
| BCMV presence/absence | 0 |
| Root rot damage score | 1 |
| Zinc deficiency damage score | 2 |
| Earliest Year in the CDBN | 0 |

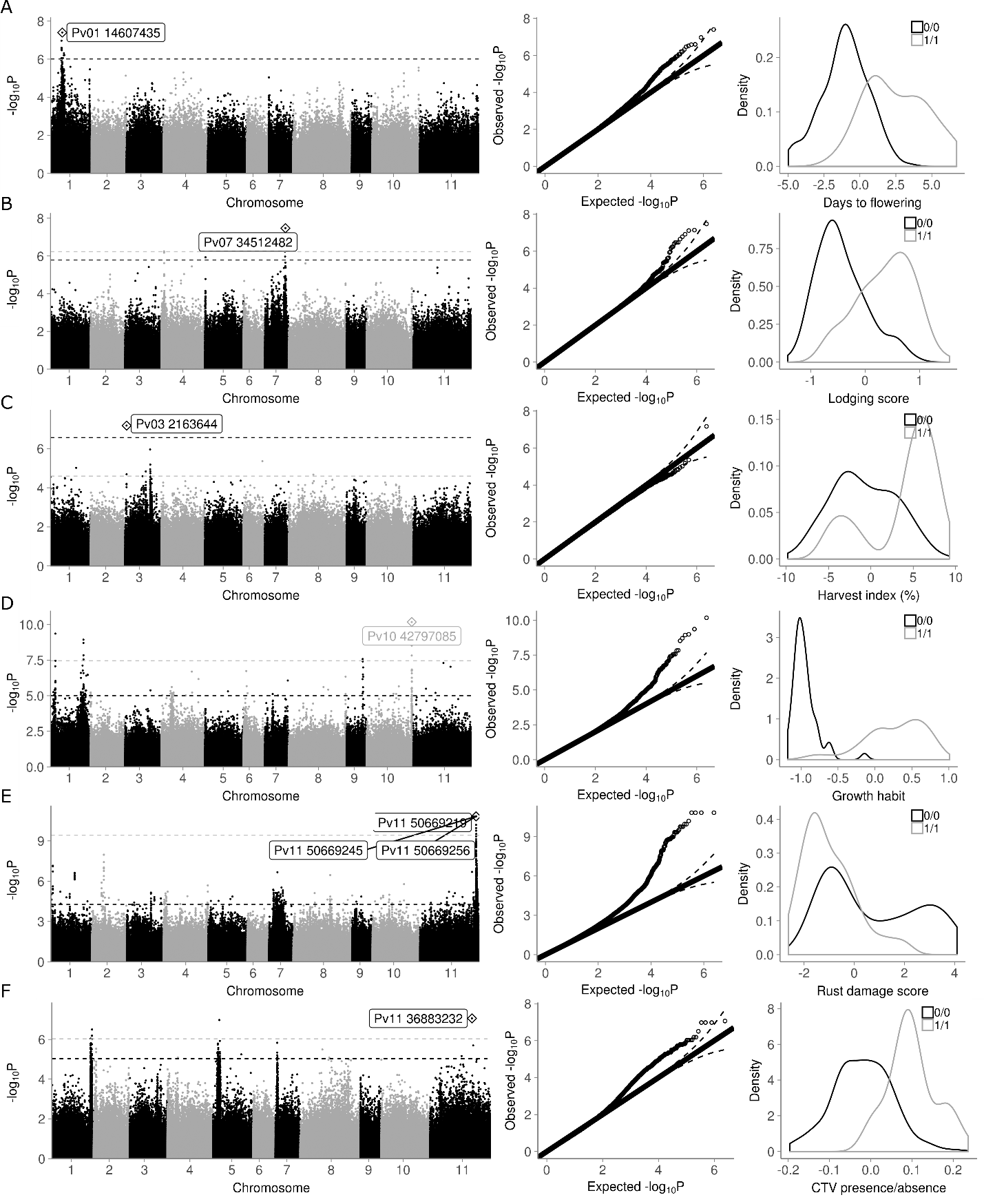
**Supplementary Table 3.** Number of principle components that maximized the Bayesian Information Criterion for model selection in GAPIT, for each set of BLUPs derived from phenotypes in the Cooperative Dry Bean Nursery dataset.

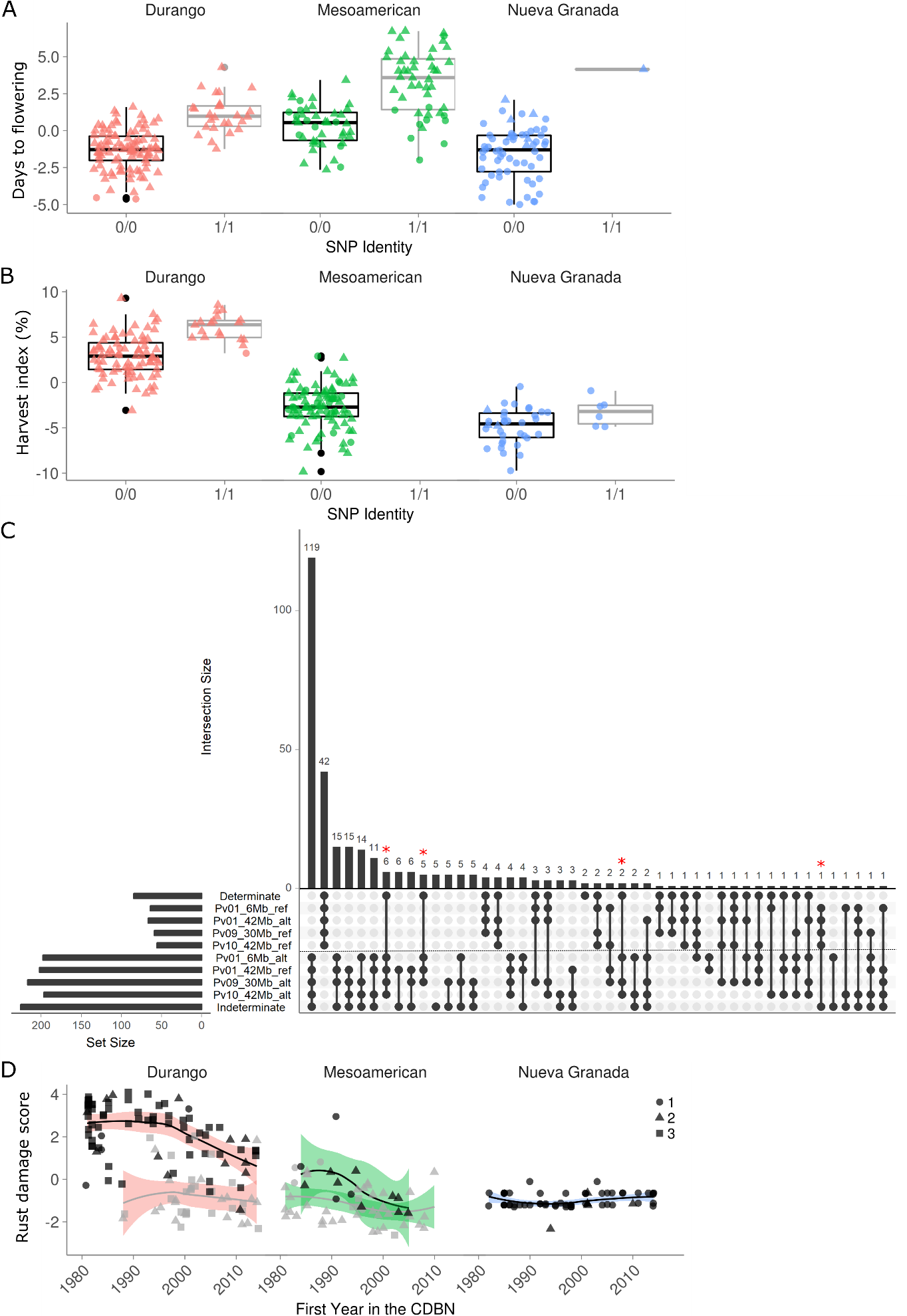
**Supplementary Table 4.** Excel File. Associations from single phenotype genome-wide association significant using a Benjamini-Hochberg false discovery rate threshold of 10%. Separate tabs of the document are associations for separate phenotypes.

**Supplementary Table 5.** Excel File. Associations from the multivariate shrinkage analysis significant using a local false sign rate threshold of 5%.

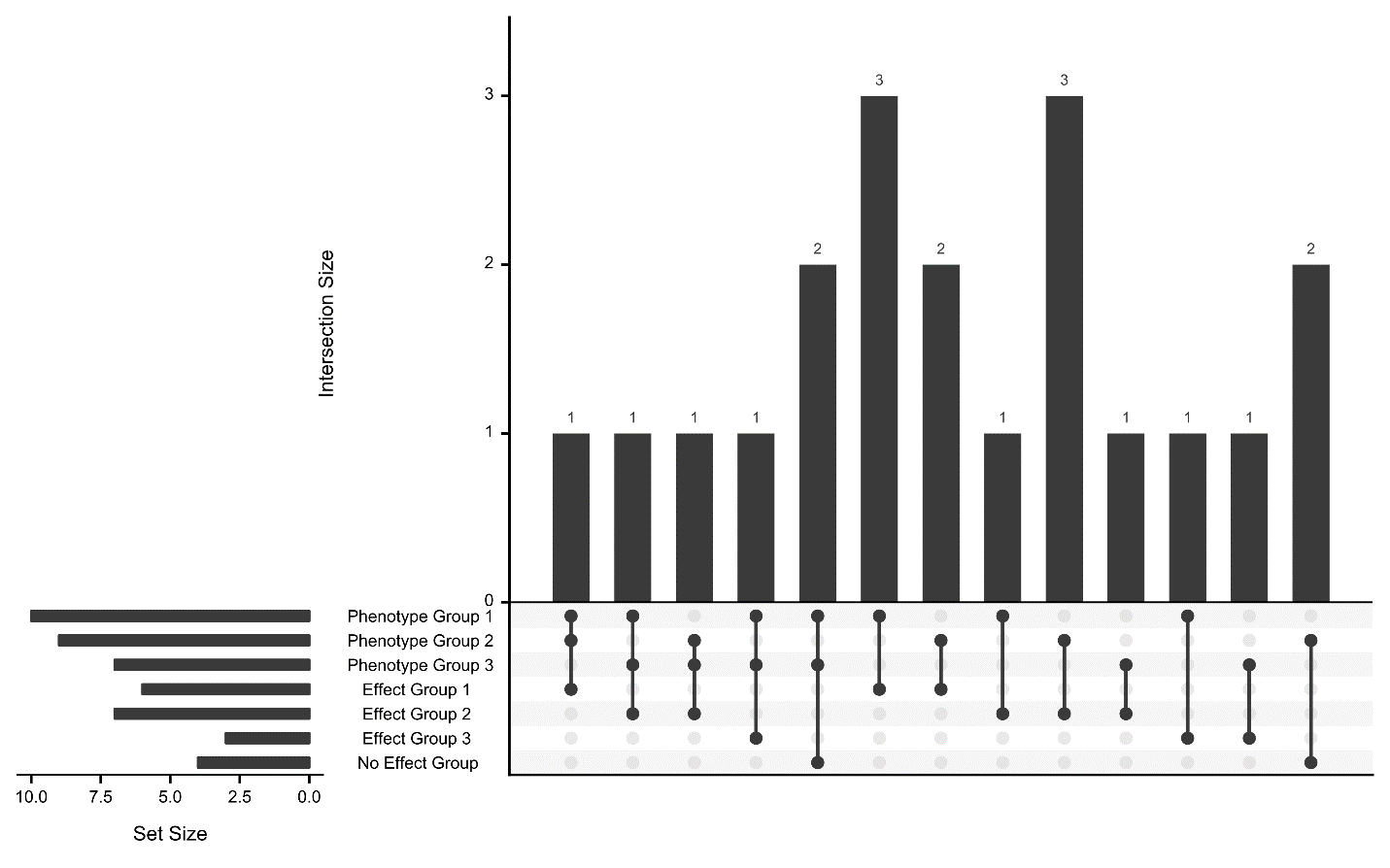


**Supplementary Figure 1.** Correlations between best linear unbiased predictors (BLUPs) for each phenotyped entry in the Cooperative Dry Bean Nursery. Color and circle size indicate the size of the correlation coefficient between the BLUPs for two phenotypes.

**Supplementary Figure 2.** Genomic associations with six additional phenotypes with associations above a Benjamini-Hochberg false discovery rate correction. The dashed lines are the cutoff values for peak significance. Single-nucleotide polymorphisms above the Benjamini-Hochberg false discovery rate are above the black line, and those in the 0.001 percentile are above the grey line. In the third column, black represents the presence of the reference allele for the top SNP labeled in the first column, and grey represents the presence of the alternate allele. A) Manhattan plot, Quantile-quantile (Q-Q) plot, and distribution of BLUPs for days to flowering from the Cooperative Dry Bean Nursery (CDBN) data. B) Manhattan, Q-Q, and distribution of BLUPs for lodging score from the CDBN data. C) Manhattan, Q-Q, and distribution of BLUPs for harvest index (%) from the CDBN data. D) Manhattan, Q-Q, and distribution of BLUPs for growth habit from the CDBN data. E) Manhattan, Q-Q, and distribution of BLUPs for rust (*Uromyces appendiculatus*) damage score from the CDBN data. F) Manhattan, Q-Q, and distribution of BLUPs for curly top virus (CTV) presence/absence from the CDBN data.

****

**Supplementary Figure 3.** Specific effects of top associations for four phenotypes in the Cooperative Dry Bean Nursery (CDBN) dataset of common bean (*Phaseolus vulgaris*). Orange represents genotypes from the Durango race of common bean, green represents genotypes from the Mesoamerican race, and blue represents genotypes from the Nueva Granada race of common bean. A) Best linear unbiased predictors (BLUPs) for days to flowering for each genotype separated by the top association for days to flowering and by common bean race. B) BLUPs for harvest index for each genotype, separated by the top association for harvest index and by common bean race. C) Upset plot for the four top associations for growth habit with determinacy information, plotted for each genotype in the CDBN. Bars show the number of genotypes with that phenotype or SNP (horizontal), or that combination of phenotype and SNPs (vertical bars). Set membership is indicated by the connected circles. Red stars indicate genotypes with an unexpected combination of phenotype and four genomic associations. D) BLUPs for rust (*Uromyces appendiculatus*) damage score shown separated by allele from the top association for rust (black and grey lines), by common bean race, and by the first year that genotype was phenotyped in the CDBN.

****

**Supplementary Figure 4.** Upset plot showing group membership overlap between three phenotypic correlation and three effect size correlation groups (from Figure S1 and Figure 3c, respectively). Phenotype groups had high phenotypic correlations in BLUPs for those phenotypes within the group. Effect groups had many associations with the same sign and similar magnitudes of effects within the group. There was no consistent pattern in membership between phenotypic correlation and effect size correlation groups.

**Supplementary Note**

*Phenotypic data processing*

Location and year (L\*Y) combinations that measured seed yield (in kg ha-1 when the seed was at 12% moisture, hereafter SY) four or fewer times had SY removed from the dataset. Biomass in kg ha-1 was converted from the total plant dry weight in lbs acre-1 measured when the plants were at 12% moisture, and harvest index (in %) was seed yield (kg ha-1) divided by biomass (kg ha-1). Outlier L\*Y where biomass was much larger than typical were removed for both biomass and harvest index. L\*Y combinations with outliers for seed weight (in mg, measured for 100 seed) were dropped. The number of days to flowering, typically scored when 50% of plants had at least one flower, was coalesced with the phenotype “days to bloom 50%”, and values of 30 or less were removed. Here, coalesce means to combine, keeping only the first of the two phenotypes in the (rare) cases when both phenotypes were scored for the same variety in the same L\*Y. Days to maturity, the number of days when 50% of plants had at least one dry pod, was coalesced with days to harvest. Seedfill duration, in days, was calculated as the number of days to maturity minus the number of days to flowering. This value was coalesced with the phenotype “seedfill duration”. Plant height (in cm) was coalesced with canopy height; to remove L\*Y that measured plant length, L\*Y with plant height values of 90 cm or greater were removed. Canopy height is typically measured to the top node in the canopy of a stand of plants, while plant length is the stretched aboveground length of a single plant to the top node with an open leaf. For prostrate indeterminate varieties in particular, these two quantities can be markedly different.

Six non-disease phenotypes were created from score and/or evaluation data. Growth habit was separated into two phenotypes, because some L\*Y scored growth habit on the International Center for Tropical Agriculture (CIAT) classification scale where 1 was determinate, 2 was upright indeterminate, and 3 was prostrate indeterminate, while some L\*Y also scored the presence and length of a vine (Aart van Schoonhoven, 1987).The CIAT classification was kept as growth habit. An early vigor score was created by coalescing early vigor and emergence scores, both scored in the first two weeks of growth on a 1-9 scale where 1 was best and 9 was worst (Aart van Schoonhoven, 1987). A seed appearance score was created by coalescing three phenotypes scored in different L\*Y: seed appearance score, seed appearance desirability, and seed appearance desirability score. SA was an integrated value of multiple traits including seed color fading, seed fill, and general seed appearance. Some L\*Y scored SA on a 1-9 scale where 1 was best, some on a 1-5 scale where 1 was best, and some on a 1-7 scale where 7 was best. These scales were converted to a 1-3 scale where 1 was best. Five lodging scores and evaluation phenotypes were combined to create a lodging score on a 1-5 scale, where 1 was best. Two zinc deficiency scores and two zinc evaluations were combined to create a zinc deficiency damage score.

Eight disease phenotypes were created from score and evaluation data. For common bacterial blight (*Xanthomonas campestris* pv*. phaseoli*), one score, one evaluation, and two percentage scales of plant surface area affected were combined to create a common bacterial blight damage score (CBB damage score) on a 1-9 scale, where 1 was resistant and 9 was susceptible. Three scores and two percentage scales of plant surface area damage by rust (*Uromyces appendiculatus* var*. appendiculatus*) were combined to create a rust damage score on the 1-9 CIAT scale (Aart van Schoonhoven, 1987), where 1 was resistant and 9 was susceptible. A score, a percent of plot infected, and an evaluation phenotype were combined for curly top virus (*Curtovirus bean curly top virus (CTV)*) to create a CTV presence/absence value, where 0 indicated the absence of disease and 1 its presence. Two phenotypes were created for bean common mosaic virus (*Potyvirus bean common mosaic virus (BCMV)*) from two virus evaluation phenotypes: a 0-1 scale indicating BCMV presence/absence, and a 0-1 scale indicating the absence or presence of the ‘blackroot’ or systemic necrosis reaction caused by BCMV in the presence of a dominant hypersensitive response gene (Blackroot BCMV response). For halo blight (*Pseudomonas syringae* pv*. phaseolicola*), a halo blight damage score on a scale of 1-5, where 1 is resistant and 5 is susceptible, was created from six halo blight phenotypes, three scores and three percentages. Root rot, caused by *Fusarium solani* f. *phaseoli*, was made into a root rot damage score on a 1-9 scale, where 1 is resistant and 9 susceptible. The root rot damage score was created from six root rot phenotypes, five scores and one evaluation. For white mold (*Sclerotinia sclerotiorum*), a white mold damage score on a 1-5 scale, where 1 is resistant and 5 susceptible, was created from five white mold phenotypes, three scores, one percent of the plot infected, and one evaluation (Soule *et al.*, 2011).