The American Journal of Human Genetics

Supplemental Data

Two-Variance-Component Model Improves Genetic Prediction in Family Datasets

George Tucker, Po-Ru Loh, Iona M. MacLeod, Ben J. Hayes, Michael E. Goddard, Bonnie Berger, and Alkes L. Price

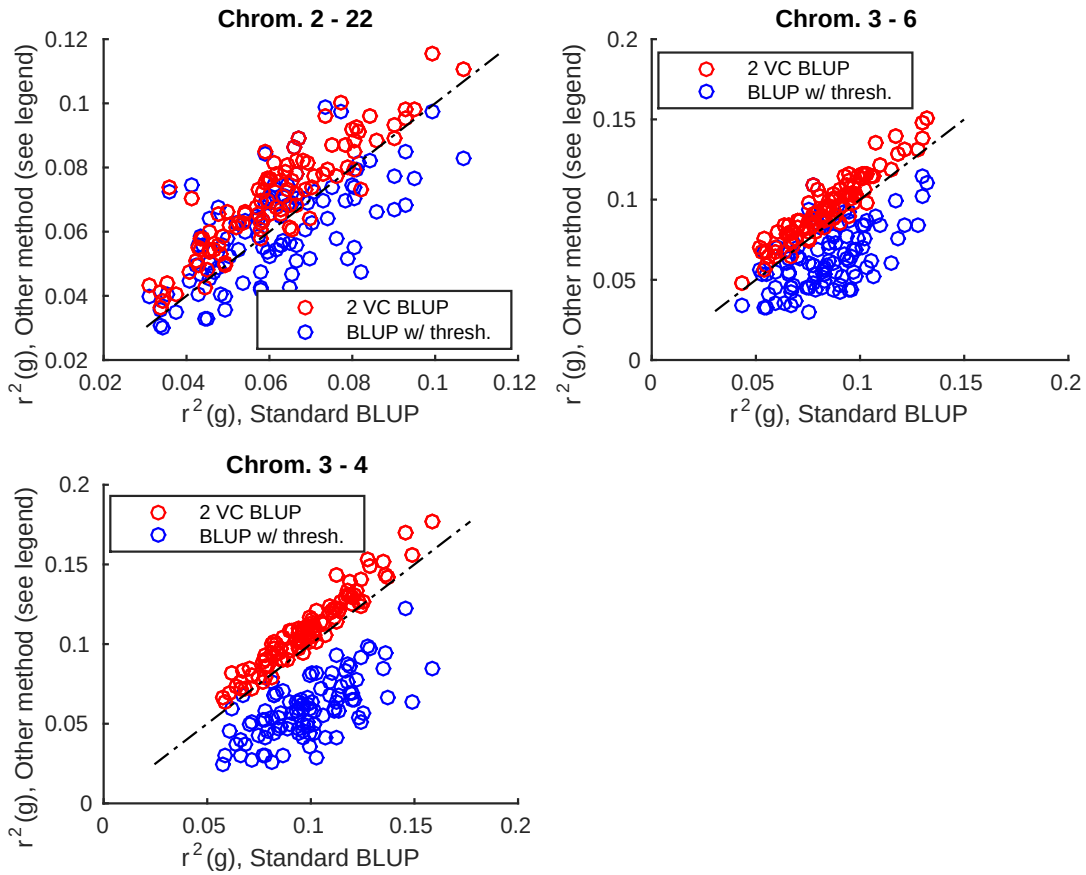

Figure S1. Comparison of prediction $r^2(g)$ estimates for simulated phenotypes with CARe genotypes. We compare two variance component BLUP (2 VC BLUP) and BLUP using the thresholded matrix (BLUP w/ thresh) vs. standard BLUP. Plotted values correspond to the 100 random $90/10$ train/test data splits summarized in Table 1a.

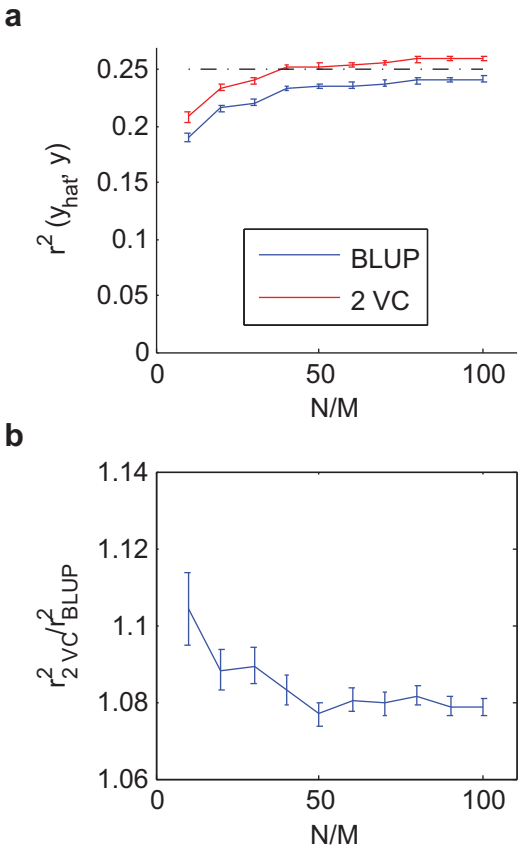

Figure S2. Performance of two variance component prediction vs. standard BLUP on simulated genotypes and phenotypes. We simulated genotypes for sets of sib-pairs (relatedness = 0.5) using the simulation procedure described in Material and Methods with M=100 SNPs and N/M=10,20,...,100. We simulated phenotypes with h_g^2=0.25 and h^2=0.5. We computed predictions using both standard BLUP and two variance component prediction for 10% of the data, using the remaining 90% for training. We ran standard BLUP using the genetic relationship relationship; for the two variance component approach, we included the true pedigree as a second variance component and also assumed the ratio of variance parameters was known to be equal to $(h^2 - h_g^2)/h_g^2 = 1$. These results are therefore an upper bound for the performance of the two variance components approach we have described in this manuscript (which approximates the true pedigree with a thresholded GRM and estimates the ratio of variance parameters). Plotted curves are means over 100 simulation replicates; error bars, s.e.m.

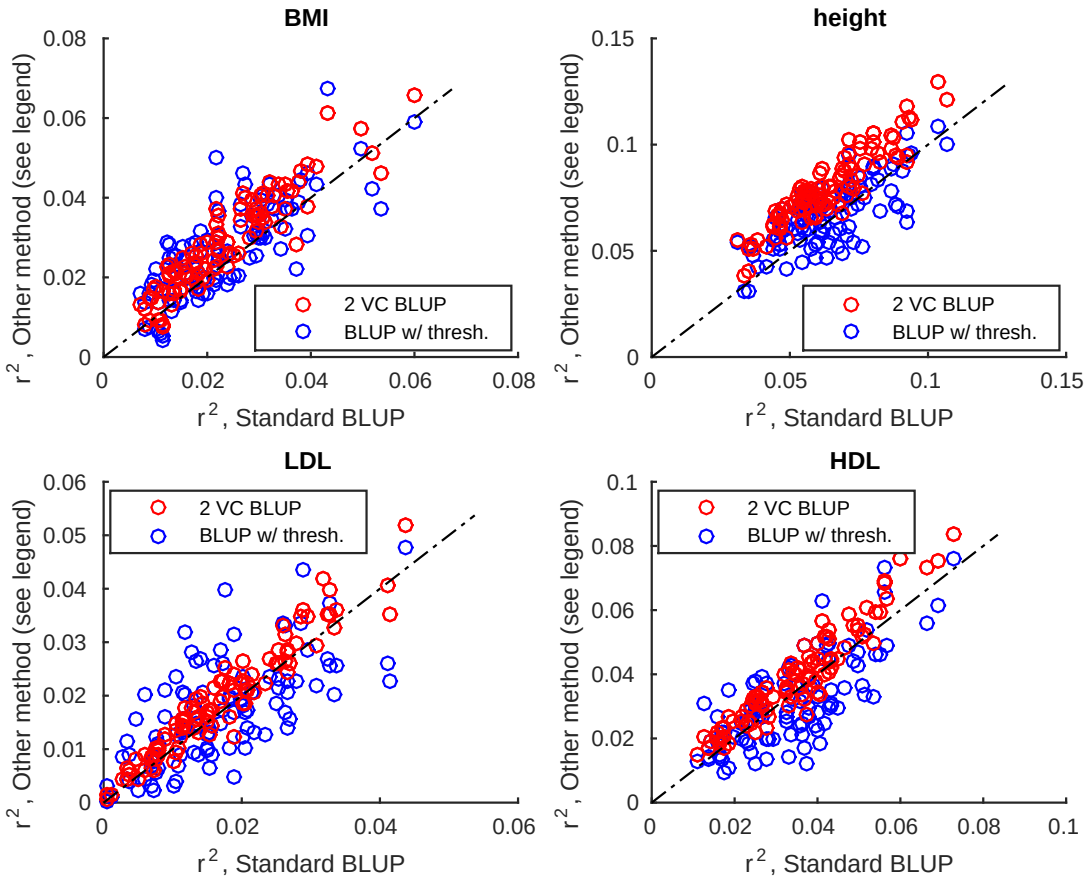

Figure S3. Comparison of prediction r^2 estimates for CARe phenotypes. We compare two variance component BLUP (2 VC BLUP) and BLUP using the thresholded matrix (BLUP w/ thresh) vs. standard BLUP. Plotted values correspond to the 100 random $90/10$ train/test data splits summarized in Table 2a.

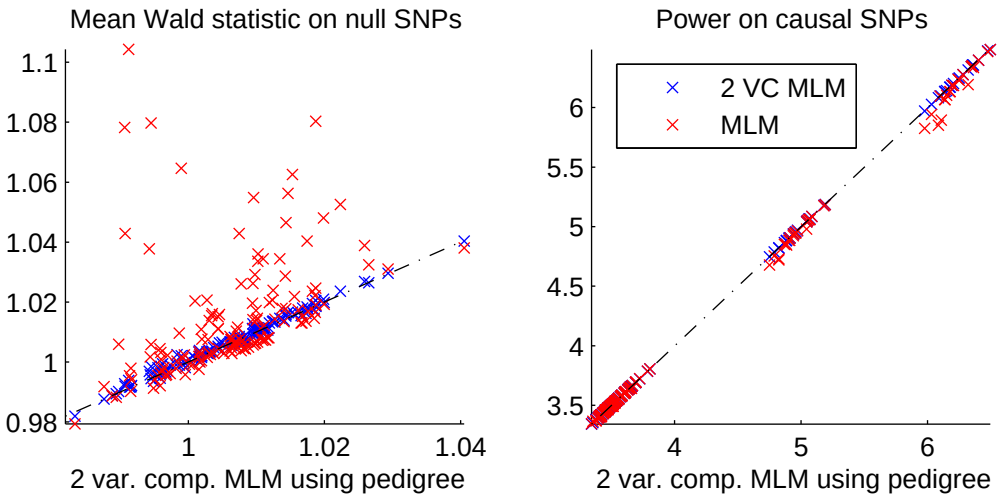

Figure S4. Inflation and power for mixed models on simulated genotypes and phenotypes. Over a range of simulation parameters, we plot our metrics for inflation (left) and power (right) of association testing using the standard mixed model (MLM) and the two variance component model (2 var. comp. MLM) against a two variance component model that replaces the thresholded GRM with the true pedigree matrix. Each plotted point corresponds to a simulation parameter setting (i.e., choice of N/M, h^2, h^2_g, and NS) plotted in Figure 1. Plotted values are means over 50 simulations. The two variance component method produces near-identical results whether using the thresholded GRM or the true relatedness matrix, whereas standard MLM association produces inflated statistics in many cases and sometimes suffers decreased power.

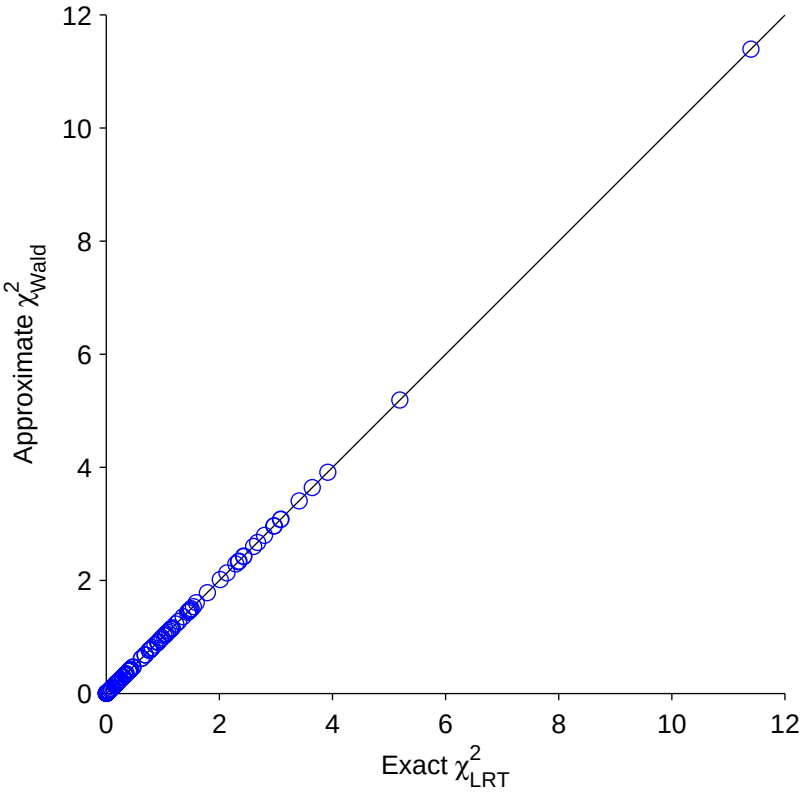

Figure S5. Comparison of approximate vs. exact two variance component association test statistics for CARe height phenotype. We computed exact likelihood ratio test statistics at 110 random SNPs (5 per chromosome) under the two variance component model and compared them to the approximate statistics we computed genome-wide. (The exact likelihood ratio test is computationally expensive, as it requires performing full maximum likelihood computations independently at each SNP.) We observed that our approximate method produced near-identical results (r^2 = 0.999997).

Table S1. Heritability parameters for simulations using CARe and FHS genotypes

(a) CARe genotypes

Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP	
	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$
Chrom. 2 - 22	0.380 (0.004)	0.507 (0.004)	0.225 (0.006)	0.285 (0.007)
Chrom. 3 - 6	0.323 (0.003)	0.491 (0.004)	0.238 (0.004)	0.256 (0.006)
Chrom. 3 - 4	0.303 (0.003)	0.482 (0.004)	0.244 (0.003)	0.242 (0.005)

(b) FHS genotypes

Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP	
	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$
Chrom. 2 - 22	0.440 (0.002)	0.495 (0.002)	0.250 (0.006)	0.247 (0.005)
Chrom. 3 - 6	0.393 (0.002)	0.489 (0.002)	0.243 (0.003)	0.248 (0.003)
Chrom. 3 - 4	0.364 (0.002)	0.475 (0.002)	0.231 (0.003)	0.248 (0.002)

Phenotypes were simulated to have $h^2 = 0.5, h^2_g = 0.25$, and heritability parameters were estimated using a random 90% of samples as training data. Reported values are mean prediction r^2 and s.e.m. over 100 independent simulations (in which phenotypes were re-simulated and train/test splits resampled). BLUP w/ thresh. denotes BLUP prediction using the thresholded relationship matrix instead of the standard approach of using the GRM (denoted simply "BLUP").

Table S2. Prediction accuracy for simulations using CARe genotypes with no untyped causal SNPs

	Prediction $r^2(g)$		
Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP
Chrom. 2 - 22	0.097 (0.002)	0.053 (0.002)	0.097 (0.002)
Chrom. 3 - 6	0.172 (0.003)	0.055 (0.002)	0.172 (0.003)
Chrom. 3 - 4	0.244 (0.003)	0.058 (0.002)	0.243 (0.003)

Phenotypes were simulated to have $h^2 = 0.5, h^2_g = 0.5$ (i.e., no untyped causal SNPs, so the two variance component model is expected to achieve no gain). Prediction $r^2(g)$ was measured using a random 90% of samples as training data and the remaining 10% as test data. Reported values are mean prediction $r^2(g)$ and s.e.m. over 100 independent simulations (in which phenotypes were re-simulated and train/test splits resampled). BLUP w/ thresh. denotes BLUP prediction using the thresholded relationship matrix instead of the standard approach of using the GRM (denoted simply "BLUP"). Prediction $r^2(g)$ denotes r^2 between predicted phenotypes and true genetic components of the simulated phenotypes.

Table S3. Prediction accuracy and heritability parameters for CARe simulations with LD between typed and untyped SNPs

	BLUP	2 VC BLUP		BLUP	2VC BLUP
h^2_{typed}	$\hat{h}^2_g$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$r^2(g)$	$r^2(g)$
0	0.348 (0.004)	0.188 (0.005)	0.321 (0.006)	0.064 (0.002)	0.077 (0.002)
0.05	0.362 (0.003)	0.213 (0.005)	0.291 (0.007)	0.065 (0.002)	0.077 (0.002)
0.1	0.378 (0.003)	0.244 (0.005)	0.263 (0.007)	0.068 (0.002)	0.078 (0.002)
0.15	0.392 (0.004)	0.271 (0.006)	0.232 (0.007)	0.073 (0.002)	0.080 (0.002)
0.2	0.409 (0.004)	0.296 (0.006)	0.213 (0.006)	0.073 (0.002)	0.079 (0.002)
0.25	0.415 (0.003)	0.310 (0.006)	0.193 (0.007)	0.082 (0.002)	0.086 (0.002)

We modified our simulations based on real CARe genotypes (Table 1a) to include LD between typed and untyped SNPs by setting typed SNPs to be the 90% of CARe SNPs with highest MAF and untyped SNPs to be the 10% of CARe SNPs with lowest MAF. (The MAF cutoff corresponding to this split was 5.4%.) As in our simulations without LD between typed and untyped SNPs, we simulated phenotypes with a total heritability of h^2 = 0.5; in these new simulations, we varied the fraction of variance directly explained by typed SNPs (h^2_{typed} = 0, 0.05, 0.1, 0.15, 0.2, or 0.25), with the rest of the heritable variance ($h^2 - h^2_{typed}$) explained by untyped SNPs. We note that the fraction of variance attributed to typed SNPs (by both standard BLUP and 2VC BLUP) exceeds h^2_{typed} in these simulations because of the LD between typed and untyped SNPs; we varied h^2_{typed} from 0 to 0.25 for this reason. For the setting of h^2_{typed} in which 2VC BLUP partitions variance roughly equally between the GRM and thresholded GRM components (h^2_{typed} = 0.1)—matching the partitioning in our simulations without LD (Table S1a)—we observed that 2VC BLUP achieved an increase in prediction accuracy (0.068 $\rightarrow$ 0.078) similar to our simulations without LD between typed and untyped SNPs (0.062 $\rightarrow$ 0.071, Table 1a).

Table S4. Heritability parameters for CARe and FHS phenotypes

(a) CARe heritability parameters

Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP	
	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$
BMI	0.336 (0.002)	0.468 (0.002)	0.148 (0.002)	0.321 (0.004)
height	0.673 (0.002)	0.953 (0.002)	0.364 (0.002)	0.591 (0.003)
LDL	0.339 (0.002)	0.432 (0.003)	0.219 (0.003)	0.216 (0.004)
HDL	0.512 (0.002)	0.666 (0.003)	0.299 (0.003)	0.366 (0.004)

(b) FHS heritability parameters

Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP	
	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$
BMI	0.435 (0.001)	0.474 (0.001)	0.217 (0.002)	0.256 (0.002)
height	0.823 (0.001)	0.878 (0.001)	0.436 (0.002)	0.441 (0.002)

(c) CARe heritability parameters using genome-wide significant SNPs as fixed effect covariates

Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP	
	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$
BMI	0.336 (0.002)	0.470 (0.002)	0.143 (0.002)	0.326 (0.004)
height	0.672 (0.002)	0.953 (0.002)	0.363 (0.002)	0.592 (0.003)
LDL	0.339 (0.003)	0.452 (0.003)	0.195 (0.004)	0.258 (0.005)
HDL	0.503 (0.002)	0.662 (0.003)	0.291 (0.003)	0.370 (0.004)

(d) FHS heritability parameters using genome-wide significant SNPs as fixed effect covariates

Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP	
	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$
BMI	0.432 (0.001)	0.472 (0.001)	0.210 (0.002)	0.261 (0.002)
height	0.822 (0.001)	0.877 (0.001)	0.436 (0.002)	0.440 (0.002)

Heritability parameters are means over 100 random 90%-subsamples corresponding to the train/test splits used to estimate prediction r^2.

Table S5. Prediction accuracy for CARe and FHS phenotypes (1 – MSE)

(a) CARe prediction

	1 – MSE			1 – MSE relative to BLUP (s.e.)	
Phenotype	BLUP	BLUP w/ thresh.	2 VC BLUP	BLUP w/ thresh.	2 VC BLUP
BMI	0.022	0.025	0.027	+14% (9%)	+19% (5%)
height	0.061	0.065	0.078	+7% (6%)	+22% (3%)
LDL	0.014	0.015	0.017	+1% (16%)	+13% (5%)
HDL	0.032	0.030	0.036	-8% (11%)	+12% (4%)

(b) CARe prediction using genome-wide significant SNPs as fixed effect covariates

	1 – MSE			1 – MSE relative to BLUP (s.e.)	
Phenotype	BLUP	BLUP w/ thresh.	2 VC BLUP	BLUP w/ thresh.	2 VC BLUP
BMI	0.021	0.024	0.026	+15% (9%)	+20% (6%)
height	0.060	0.064	0.076	+6% (6%)	+22% (3%)
LDL	0.035	0.037	0.038	+4% (7%)	+7% (3%)
HDL	0.048	0.046	0.052	-6% (7%)	+8% (3%)

We normalized mean square errors by dividing by phenotypic variance and computed the mean of 1 – MSE over 100 random $90/10$ train/test data splits. Relative performance values reported are ratios of means minus 1; standard errors are estimated as standard deviations of per-split differences in 1 – MSE (over the random 10% test sets) divided by $\sqrt{10}$ (to account for the 10x larger sample size of the full data set; see Material and Methods). BLUP w/ thresh. denotes BLUP prediction using the thresholded relationship matrix instead of the standard approach of using the GRM (denoted simply "BLUP").

Table S6. Prediction accuracy for simulations with 25% untyped individuals using CARe and FHS genotypes

(a) CARe genotypes

	Prediction $r^2(g)$		
Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP
Chrom. 2 - 22	0.061 (0.002)	0.060 (0.002)	0.068 (0.002)
Chrom. 3 - 6	0.080 (0.002)	0.064 (0.002)	0.089 (0.002)
Chrom. 3 - 4	0.089 (0.002)	0.058 (0.002)	0.098 (0.002)

(b) FHS genotypes

	Prediction $r^2(g)$		
Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP
Chrom. 2 - 22	0.226 (0.003)	0.225 (0.003)	0.235 (0.003)
Chrom. 3 - 6	0.240 (0.003)	0.228 (0.003)	0.260 (0.003)
Chrom. 3 - 4	0.257 (0.003)	0.232 (0.003)	0.282 (0.003)

Phenotypes were simulated to have $h^2 = 0.5, h_g^2 = 0.25$, and prediction $r^2(g)$ was measured using a random 90% of samples as training data and the remaining 10% as test data. Reported values are mean prediction $r^2(g)$ and s.e.m. over 100 independent simulations (in which phenotypes were re-simulated and train/test splits resampled). BLUP w/ thresh. denotes BLUP prediction using the thresholded relationship matrix instead of the standard approach of using the GRM (denoted simply "BLUP"). Prediction $r^2(g)$ denotes r^2 between predicted phenotypes and true genetic components of the simulated phenotypes.

Table S7. Prediction accuracy for CARe and FHS phenotypes with 25% untyped individuals

(a) CARe prediction

	Prediction r^2			Prediction r^2 relative to BLUP (s.e.)	
Phenotype	BLUP	BLUP w/ thresh.	2 VC BLUP	BLUP w/ thresh.	2 VC BLUP
BMI	0.024	0.027	0.028	+12% (8%)	+15% (5%)
height	0.064	0.067	0.076	+4% (5%)	+16% (3%)
LDL	0.017	0.017	0.019	-1% (13%)	+7% (5%)
HDL	0.035	0.032	0.038	-9% (9%)	+8% (4%)

(b) FHS prediction

	Prediction r^2			Prediction r^2 relative to BLUP (s.e.)	
Phenotype	BLUP	BLUP w/ thresh.	2 VC BLUP	BLUP w/ thresh.	2 VC BLUP
BMI	0.103	0.104	0.106	+0.8% (2.2%)	+3.0% (1.1%)
height	0.344	0.342	0.352	-0.5% (1.1%)	+2.4% (0.5%)

(c) CARe prediction using genome-wide significant SNPs as fixed effect covariates

	Prediction r^2			Prediction r^2 relative to BLUP (s.e.)	
Phenotype	BLUP	BLUP w/ thresh.	2 VC BLUP	BLUP w/ thresh.	2 VC BLUP
BMI	0.023	0.027	0.028	+12% (8%)	+16% (5%)
height	0.064	0.066	0.076	+4% (5%)	+16% (3%)
LDL	0.036	0.036	0.037	+0% (6%)	+4% (3%)
HDL	0.052	0.049	0.055	-6% (6%)	+5% (3%)

(d) FHS prediction using genome-wide significant SNPs as fixed effect covariates

	Prediction r^2			Prediction r^2 relative to BLUP (s.e.)	
Phenotype	BLUP	BLUP w/ thresh.	2 VC BLUP	BLUP w/ thresh.	2 VC BLUP
BMI	0.104	0.105	0.107	+0.9% (2.2%)	+3.0% (1.1%)
height	0.344	0.342	0.352	-0.6% (1.1%)	+2.3% (0.5%)

Prediction r^2 values are means over 100 random $90/10$ train/test data splits. Relative performance values reported are ratios of means minus 1; standard errors are estimated as standard deviations of per-split differences in r^2 (over the random 10% test sets) divided by $\sqrt{10}$ (to account for the 10x larger sample size of the full data set; see Material and Methods). BLUP w/ thresh. denotes BLUP prediction using the thresholded relationship matrix instead of the standard approach of using the GRM (denoted simply "BLUP").

Table S8. Heritability parameters for simulations with 25% untyped individuals using CARe and FHS genotypes

(a) CARe genotypes

Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP	
	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$
Chrom. 2 - 22	0.421 (0.004)	0.507 (0.004)	0.228 (0.007)	0.280 (0.008)
Chrom. 3 - 6	0.355 (0.004)	0.489 (0.004)	0.239 (0.004)	0.251 (0.006)
Chrom. 3 - 4	0.328 (0.003)	0.482 (0.004)	0.244 (0.004)	0.240 (0.005)

(b) FHS genotypes

Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP	
	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$
Chrom. 2 - 22	0.457 (0.002)	0.495 (0.002)	0.253 (0.006)	0.242 (0.006)
Chrom. 3 - 6	0.417 (0.002)	0.489 (0.002)	0.240 (0.004)	0.250 (0.003)
Chrom. 3 - 4	0.386 (0.002)	0.475 (0.002)	0.229 (0.003)	0.249 (0.003)

Phenotypes were simulated to have $h^2 = 0.5, h^2_g = 0.25$, and heritability parameters were estimated using a random 90% of samples as training data. Reported values are mean prediction r^2 and s.e.m. over 100 independent simulations (in which phenotypes were re-simulated and train/test splits resampled). BLUP w/ thresh. denotes BLUP prediction using the thresholded relationship matrix instead of the standard approach of using the GRM (denoted simply "BLUP").

Table S9. Heritability parameters for CARe and FHS phenotypes with 25% untyped individuals

(a) CARe heritability parameters

Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP	
	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$
BMI	0.372 (0.002)	0.468 (0.002)	0.142 (0.005)	0.326 (0.006)
height	0.781 (0.003)	0.953 (0.002)	0.374 (0.004)	0.580 (0.004)
LDL	0.371 (0.003)	0.432 (0.003)	0.234 (0.006)	0.201 (0.007)
HDL	0.563 (0.003)	0.666 (0.003)	0.311 (0.006)	0.354 (0.007)

(b) FHS heritability parameters

Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP	
	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$
BMI	0.448 (0.001)	0.474 (0.001)	0.221 (0.004)	0.252 (0.004)
height	0.843 (0.001)	0.878 (0.001)	0.448 (0.005)	0.427 (0.005)

(c) CARe heritability parameters using genome-wide significant SNPs as fixed effect covariates

Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP	
	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$
BMI	0.372 (0.002)	0.468 (0.002)	0.140 (0.005)	0.328 (0.006)
height	0.780 (0.003)	0.953 (0.002)	0.373 (0.004)	0.582 (0.004)
LDL	0.376 (0.003)	0.446 (0.003)	0.213 (0.006)	0.234 (0.007)
HDL	0.559 (0.003)	0.662 (0.003)	0.314 (0.006)	0.347 (0.007)

(d) FHS heritability parameters using genome-wide significant SNPs as fixed effect covariates

Observed SNPs	BLUP	BLUP w/ thresh.	2 VC BLUP	
	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$	$\hat{h}^2_g$	$\hat{h}^2_{>0.05}$
BMI	0.448 (0.001)	0.474 (0.001)	0.219 (0.004)	0.254 (0.004)
height	0.843 (0.001)	0.878 (0.001)	0.448 (0.005)	0.427 (0.005)

Heritability parameters are means over 100 random 90%-subsamples corresponding to the train/test splits used to estimate prediction r^2.

Table S10. Type I error of standard and two-variance-component mixed model association statistics in CARe and FHS simulations

(a) CARe genotypes

Observed SNPs	Standard mixed model		Two variance components	
	α=0.01	α=0.0001	α=0.01	α=0.0001
Chrom. 3 - 22	1.05e-02	1.13e-04	1.00e-02	1.02e-04
Chrom. 3 - 6	1.10e-02	1.26e-04	1.01e-02	1.07e-04
Chrom. 3 - 4	1.11e-02	1.30e-04	1.01e-02	1.07e-04
Chrom. 22	1.14e-02	1.35e-04	1.05e-02	1.16e-04

(b) FHS genotypes

Observed SNPs	Standard mixed model		Two variance components	
	α=0.01	α=0.0001	α=0.01	α=0.0001
Chrom. 3 - 22	1.12e-02	1.40e-04	1.01e-02	1.09e-04
Chrom. 3 - 6	1.28e-02	1.84e-04	1.04e-02	1.20e-04
Chrom. 3 - 4	1.40e-02	2.11e-04	1.07e-02	1.25e-04
Chrom. 22	1.81e-02	3.57e-04	1.21e-02	1.50e-04

Type I error of Wald statistics on candidate null SNPs for simulations with CARe or FHS genotypes and a trait with $h^2 = 0.5, h_g^2 = 0.25$ (see Table 3 for details). Reported values are aggregated over 100 simulations testing null SNPs on chromosome 2 (63,077 SNPs for CARe, 34,608 SNPs for FHS).

Table S11. Prediction accuracy for a range of dairy cattle traits

Trait	Num. of Records Training (Validation)	Prediction r^2 BLUP	BLUP w/ pedigree	2 VC BLUP
Fat Yield	8820 (1053)	0.360	0.169	0.359
Milk Yield	8820 (1053)	0.490	0.267	0.506
Protein Yield	8820 (1053)	0.442	0.265	0.453
Teat Length	2500 (360)	0.312	0.203	0.315
Temperament	5543 (734)	0.110	0.053	0.109
Fertility	8428 (838)	0.225	0.122	0.225

We analyzed four dairy cattle traits from the Holstein breed using a data set previously described in [59]. We added two further phenotypes which had been recorded for the same animals: temperament score and teat length (both have an influence on the ease of milking). Animals had 632,002 SNP genotypes. We tested three BLUP prediction methods: standard mixed model BLUP, BLUP using the pedigree, and two variance component BLUP using both the GRM and pedigree. (We used the full pedigree relationship matrix in place of the thresholded GRM because extensive pedigree records are typically available in dairy cattle.) We analyzed corrected phenotypes including both progeny tested bulls and cows with repeat records in a weighted analysis (with weights calculated from the effective number of records per animal) as described in [59]. We implemented the analyses using ASReml software [60]. The training/test data split (approximately 90/10) was based on a date of birth cutoff, with the youngest bulls used for the test set. Only bull data was used for the test set because their phenotypes (progeny test with $\geq$20 daughter records) are considerably more accurate than those of cows.

We did not observe a consistent advantage in prediction accuracy using the two variance component model compared to standard BLUP. Possible reasons for the difference between cattle and human results are:

1. In Holstein dairy cattle, linkage disequilibrium (LD) decays much more slowly with physical distance between variants compared to humans because of their recent sharp decline to a very small effective population size [61].

2. The recent very small effective population size results in fewer rare variants segregating compared to some human populations which have relatively large effective population size and have undergone recent expansion. This means there is likely to be a lower proportion of rare causal variants in cattle compared to human populations.

3. The training population is very closely related to the test population.

These three factors combined suggest that similarly dense SNP genotypes may more accurately track the variance due to causal mutations within a single cattle breed compared to the human data in this study. We might still therefore expect some improvement in accuracy from the 2 VC model in cattle if using a less dense SNP chip (e.g., 50K) or mixed breed analysis because a more significant proportion of causal mutations may not be in high LD with typed SNPs.