# SNP Allocation For Estimating Heritability (SAFE-$h^2$): A tool to explore genomic origins of phenotypes for estimation of SNP heritability using additive-only allelic effects or additive and non-additive allelic effects

Behrooz Darbani, Mogens Nicolaisen

***Department of Agroecology, Faculty of Technical Sciences, Aarhus University, DK-4200 Slagelse, Denmark***

Corresponding authors emails: bd@agro.au.dk/behroozdarbani@gmail.com, mn@agro.au.dk

**Keywords:** GWAS, SNP, heritability, additive effects, non-additive effects, false-positive SNPs

## Abstract

SNP heritability is a marked indication for genetic gains. Extreme polygenicity, *i.e.*, small effect-size causal SNPs, dictates as many SNPs as possible to be involved in heritability estimations. On the contrary, non-causal SNPs underestimate SNP heritability levels via negative contributions which is confirmed here on simulated and real datasets. Negative contributions to heritability are biologically absurd as polymorphic genomic positions can only have null or positive contributions to heritability. This paradox signifies a trade-off between the incorporation of unknown causal SNPs and the boycott of unknown non-causal SNPs for rigorous approximations of heritability. SAFE-$h^2$ (SNP Allocation For Estimating Heritability) is an application tool to overcome this dilemma by utilizing association *p*-values for SNP heritability profiling. SAFE-$h^2$ determines an association *p*-value threshold to filter out SNPs from its upper-bound as they contribute negatively to heritability estimations. Considering SNP heritability estimates which were calculated using all SNPs for 74 phenotypes in five plant and four animal species, SAFE-$h^2$ revealed up to 90 units (30 units on average) negative contribution imposed by SNPs from upper-bounds of *p*-value thresholds. SAFE-$h^2$ also provides a safe *p*-value interval to minimize false-positive SNP hits' contributions to heritability estimates. Furthermore, SAFE-$h^2$ captures non-additive effects through intra-locus allelic adjustments. The allelic adjustment algorithm of SAFE-$h^2$ captured additional phenotypic variance and improved the SNP heritability estimations for random phenotypes assigned to the human genotypes as well as artificial phenotypes associated, at different rates, with a biallelic locus. Likewise, SAFE-$h^2$ revealed considerable uncaptured phenotypic variance when relying on additive effects and improved the SNP heritability estimations up to 56 units by joint analysis of intra-locus additive and non-additive allelic effects for 50 phenotypes in four animal and three plant species. We found the EMMAX algorithm to be the most stable and accurate model for SNP heritability profiling by SAFE-$h^2$.

## Introduction

Single nucleotide polymorphism (SNP) heritability estimations depend on study designs (1, 2) and reflect the experimental power for maximum exploitation of phenotypic variances in genomic studies and breeding programs. Heritability estimates range from 0 to 1 and are often expressed in percentage units (0–100%). Genetic and phenotypic variances are required to estimate heritability. Phenotypic diversity is evolutionary achieved through genomic divergence. These diverging events can either impose intrinsic effects such as abolished protein function irrespective of environmental factors or have non-constitutive effects in response to certain environmental constraints. Therefore, phenotypic variances can exceed genetic variances in multi-environment experiments, and this is due to environmental influences, *i.e.*, phenotypic variations for every genotype among environments. In contrast, genetic variances are in theory expected to explain 100% of phenotypic variances, *i.e.*, 100% heritability, in single-environment experiments. However, estimates of SNP heritability are far below 100% in real single-environment experiments and this is because of experimental challenges highlighted later in this section. Adding extra environments to a given multi-environment experiment can alter the total phenotypic variance, the environmental variance, and thereby, the genetic variance.

Likewise, including new genotypes, those that have not already been included, changes total phenotypic and genetic variances in single-environment experiments. Therefore, there is no single true heritability for a given phenotype, and heritability estimates are in fact experiment-specific indicators to forecast the maximum genetic gains under given experimental conditions. It is also worth to notice that outperformed models for heritability estimation capture the most from phenotypic variations under an identical statistical power level.

In genetic and genomic studies, experimental observations that are under genetic governance can be defined as phenotypes. This is because genetic gains cannot be expected on observations that have no genetic root and these observations are in fact environmental characteristics and should not be considered as phenotypes. Most importantly, zero or near-to-zero SNP heritability can also be due to experimental defects, and these should not be mistaken for the true absence of SNP heritability. Inadequate genotypic coverage, partial genomic coverage, imprecise phenotypic scoring scales, assignment of non-causal SNPs into estimation models, exclusion of some causal SNPs from estimation models, and ignoring intra-locus non-additive allelic effects are experimental challenges affecting the heritability estimations. For example, we require $\approx 2^{100,000}$ different genotypes for precise estimation of genetic effects in a genome with 100,000 biallelic SNPs. Please note that this calculation cannot be affected by genomic linkage as the creation of mutations are independent. Lethal SNPs that are conditional on specific suppressor SNPs, are also very rare events. The imprecise phenotypic scoring scales can also be alleviated, *e.g.*, through defining of possible phenotypic subclasses for qualitative traits. In genomic studies, SNP heritability is widely estimated on additive effects and thereby, considered as narrow-sense heritability (3–6). Different models have been developed to estimate SNP heritability (7), for example, using the minor allele frequency (MAF) and the linkage disequilibrium based weighted SNP effects (8–11). Although the recent technological advancements provide up to several millions of variants, SNP heritability is still behind by capturing 30%–50% less phenotypic variance when compared to heritability estimates from family and twin studies (7, 12–17). To fill this gap, we can incorporate more variants to cover bigger fractions of genomes, add new genotypes, and define proper phenotype scoring systems. However, fulfilling all these practicalities cannot heal the gap sufficiently until a proper allocation of SNPs to estimation models and calling for both additive and non-additive intra-locus allelic effects. This study is accordingly focused on preventing the assignment of non-causal SNPs into estimation models while ensuring the incorporation of causal SNPs, minimizing false-positive SNP hits' contributions to heritability, and finally performing a joint analysis to capture additive and non-additive effects.

SNP heritability estimations that are based on all SNPs (causals and non-causals), suffer from genetic variance estimations that are not phenotype specific as every phenotype of interest is governed by a small fraction of genome or SNPs (causals) rather than a whole genome or all SNPs. In fact, family-based heritability estimations employ the true population statistic of genetic variance irrespective of phenotypes-of-interest. This parameter of genetic variance is constant across all polymorphic genomic positions and can be used for any phenotype of interest. For example, the genomes of monozygotic twins are 100% identical by descent while fraternal twins generally share 50% of their genomes (18). In contrast, SNP-derived genetic similarities and genetic variance are highly error-prone when applying the genome-wide genetic relatedness matrix to estimate SNP heritability. SNP heritability calculations are based on estimated genetic variances and thereby depend on SNPs, *i.e.*, various pools of SNP combinations lead to different levels of genetic similarities. In practice, this is caused through the genetic relationship matrix in which various pools of SNPs result in different inter-individual genetic similarity levels. As an example, we can consider a phenotype recorded in 1000 individuals being used in lab no.1 to estimate SNP heritability using 100k SNPs and in lab no.2 using 200k SNPs. Here, the inter-individual genetic similarity levels that are calculated within the genetic relationship matrices are different between the two labs even though they apply identical individuals and phenotypic observations. It is clear that genetic similarities get overestimated, and genetic variances shrunk by including SNPs with big association *p*-values in estimation models. Amon genotypes, SNPs with large association *p*-values are distributed more random than SNPs having small association *p*-values. Therefore, SNPs from upper-bound of a specific *p*-value threshold which is dataset dependent, lead to inflated genetic overlap degrees and thereby, an underestimated genetic variance. This genetic variance is not phenotype specific as it also represents non-related genomic regions. Taken together, non-causal SNPs overestimate inter-genotype genetic similarities by amplifying genetic overlap rates within genetic relationship matrices which subsequently lead to underestimation of SNP heritability levels. Apart from this, extreme polygenicity impose a potential risk of including non-causal SNPs into estimation models; polygenicity makes it essential to include as many variants as possible for exploiting a bigger fraction of SNPs having very small effects or SNPs having association *p*-values

being several orders of magnitude larger than the significance threshold of $10^{-8}$. This dilemma calls for development of an efficient SNP quota system for heritability models to facilitate the estimation of phenotype-specific genetic relationship matrix *via* excluding non-causal SNPs that have negative contributions rather than relying on a genome-wide relationship matrix. **Here, we introduce SAFE-$h^2$ that is equipped with a *p*-value based SNP and heritability profiling approach to exclude non-causal genomic regions that have biologically meaningless negative contributions to SNP heritability estimates. SAFE-$h^2$ also offers a safe *p*-value interval to minimize heritability contributions from false-positive SNP hits. Finally, it introduces a co-extraction algorithm for intra-locus non-additive allelic effects and intra-locus additive allelic effects to further advance the accuracy of SNP heritability estimates.**

## Results

Polymorphic positions such as SNPs can either be neutral or positive when incorporated into models to estimate SNP heritability. Causal and non-causal SNPs are unknown targets. Causal SNPs have positive or negative effects on phenotypes and thereby, can contribute positively to heritability. In contrast, non-causal SNPs have no effect on phenotypes and are expected to have no contribution to heritability. However, our results revealed negative contributions to heritability estimates imposed by SNPs having large association *p*-values. It is important to re-emphasize that negative contribution to heritability is biologically illogical and contributions to heritability, if any, should only be positive. Here, we propose an approach to eliminate non-causal SNPs that contribute negatively to heritability. It functions by defining an experiment and dataset-specific threshold for association *p*-values; SNPs from the upper-bound of *p*-value threshold are distributed randomly among genotypes in relation to phenotypic variations and, have biologically meaningless negative contributions to heritability estimates when they are merged with SNPs from the lower-bound of *p*-value threshold. Accordingly, we will henceforth call non-causals as SNPs with *p*-values from the upper-bound of data-specific *p*-value thresholds. We introduce SAFE-$h^2$ application tool which does SNP profiling and provides SNP heritability for *p*-value based SNP-clusters. SAFE-$h^2$ can utilize the heritability estimation models of EMMAX (19), LDAK-GCTA, LDAK-Thin (8), and GCTA-GREML (20). In this process, SAFE-$h^2$ determines the *p*-value threshold used for omitting SNPs with negative contributions to heritability. SAFE-$h^2$ provides the chance to focus on SNP heritability at the determined *p*-value threshold or any user defined lower *p*-values. SAFE-$h^2$ can also report a second *p*-value threshold at which false-positive SNP hits' contributions to heritability can be kept at its minimum level. False-positive SNP hits are not related to phenotypes but contribute to heritability estimations due to low statistical power. Finally, SAFE-$h^2$ offers a joint analysis approach for intra-locus allelic effects of additive and non-additive through an allelic adjustment strategy.

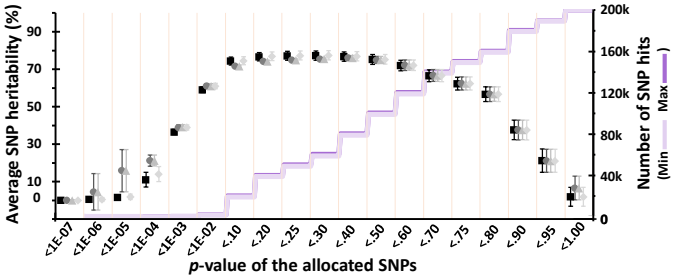
### *SAFE-$h^2$ prevents the illogical negative contributions of SNPs to heritability estimations*

To meet polygenicity and prevent negative contributions of non-causal SNPs to heritability, SAFE-$h^2$ employs association *p*-values and estimates SNP heritability using *p*-value-guided clusters of SNPs. The provided heritability profile facilitates the best approximation of SNP heritability by determining a *p*-value threshold. This threshold is used to exclude non-causal SNPs which otherwise overestimate genetic similarities within the genetic relationship matrix and thereby, impose negative contributions to heritability. To manifest SNP heritability profiles, SNP heritability estimates were first investigated in simulated genomic and phenotypic datasets. It involved two independent analyses on 1,000 and 10,000 genotypes each with 200k SNPs (MAF = 1.6%–10%). The independent phenotypes were random integers within certain ranges. Therefore, there were random chances of having small associations between the phenotypes and the SNPs which could lead to varying levels of heritability. By analyzing these datasets, we noticed a common pattern of negative contribution to SNP heritability imposed by the SNPs having large association *p*-values (Fig. 1a).

A similar pattern was also noticeable after increasing the power of experiment, *i.e.*, raising the number of genotypes from 1,000 to 10,000 genotypes (Fig. 1c). Increasing the power, as expected, led to larger phenotypic contributions by the SNPs having smaller *p*-values (Fig. 1a,c). Increasing the number of genotypes also shifted the *p*-value threshold towards a larger value at which the SNPs from its upper-bound had negative contributions and led to underestimation of SNP heritability (Fig. 1a,c). Nevertheless, the negative contribution by the SNPs with large association *p*-values did

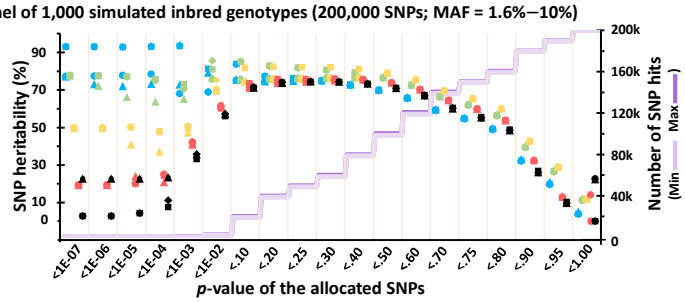
**a) Analysis of 10 artificial random phenotypes on a panel of 1,000 simulated inbred genotypes (200,000 SNPs; MAF = 1.6%–10%)**

| Random phenotypes | Range (N) | EMMAX | LDAK_GCTA | LDAK_Thin | GCTA_GREML |
|---|---|---|---|---|---|
| 1st | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 2nd | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 3rd | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 4th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 5th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 6th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 7th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 8th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 9th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 10th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |



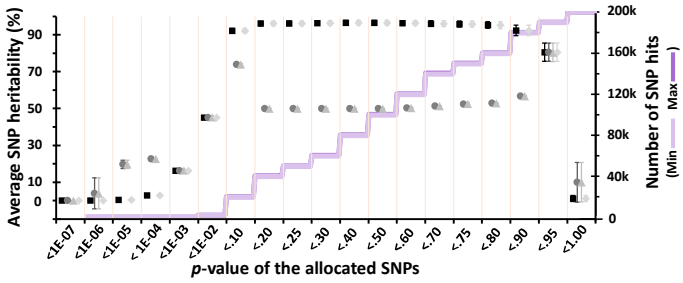

**b) Analysis of 5 artificial non-random phenotypes on a panel of 1,000 simulated inbred genotypes (200,000 SNPs; MAF = 1.6%–10%)**

| Phenotypes | The introduced non-random SNP position | | | | EMMAX | LDAK_GCTA | LDAK_Thin | GCTA_GREML |
|---|---|---|---|---|---|---|---|---|
| | Ref/Ref Genotype | Alt/Alt Genotype | | | | | | |
| | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | | | | |
| 1st | 40-50 (900) | 80-100 (1,000) | 6.995 | 93.005 | ■ | ● | ▲ | ◆ |
| 2nd | 40-60 (900) | 75-100 (100) | 22.389 | 77.611 | ■ | ● | ▲ | ◆ |
| 3rd | 40-70 (900) | 70-100 (100) | 50.267 | 49.733 | ■ | ● | ▲ | ◆ |
| 4th | 40-80 (900) | 60-100 (100) | 80.690 | 19.310 | ■ | ● | ▲ | ◆ |
| 5th | 40-90 (900) | 50-100 (100) | 97.170 | 2.830 | ■ | ● | ▲ | ◆ |



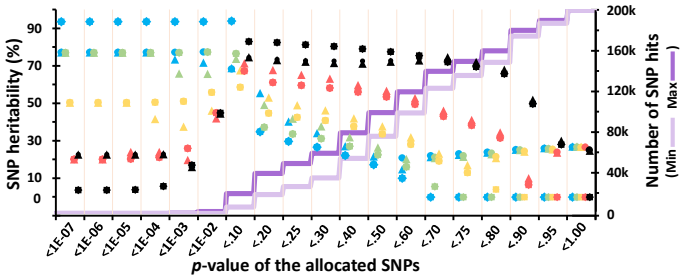

**c) Analysis of 10 artificial random phenotypes on a panel of 10,000 simulated inbred genotypes (200,000 SNPs; MAF = 1.6%–10%)**

| Random phenotypes | Range (N) | EMMAX | LDAK_GCTA | LDAK_Thin | GCTA_GREML |
|---|---|---|---|---|---|
| 1st | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 2nd | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 3rd | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 4th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 5th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 6th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 7th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 8th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 9th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 10th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |



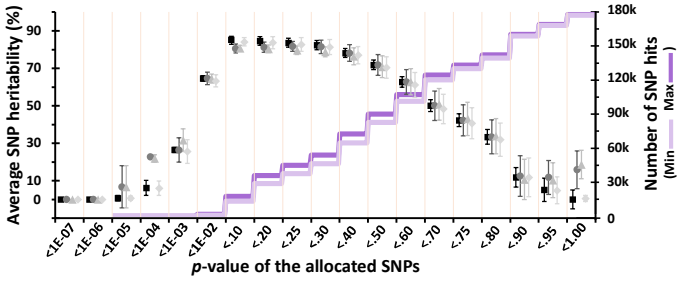

**d) Analysis of 5 artificial non-random phenotypes on a panel of 10,000 simulated inbred genotypes (200,000 SNPs; MAF = 1.6%–10%)**

| Phenotypes | The introduced non-random SNP position | | | | EMMAX | LDAK_GCTA | LDAK_Thin | GCTA_GREML |
|---|---|---|---|---|---|---|---|---|
| | Ref/Ref Genotype | Alt/Alt Genotype | | | | | | |
| | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | | | | |
| 1st | 40-50 (9,000) | 80-100 (1,000) | 6.605 | 93.395 | ■ | ● | ▲ | ◆ |
| 2nd | 40-60 (9,000) | 75-100 (1,000) | 22.738 | 77.262 | ■ | ● | ▲ | ◆ |
| 3rd | 40-70 (9,000) | 70-100 (1,000) | 48.937 | 51.063 | ■ | ● | ▲ | ◆ |
| 4th | 40-80 (9,000) | 60-100 (1,000) | 79.675 | 20.325 | ■ | ● | ▲ | ◆ |
| 5th | 40-90 (9,000) | 50-100 (1,000) | 96.309 | 3.691 | ■ | ● | ▲ | ◆ |



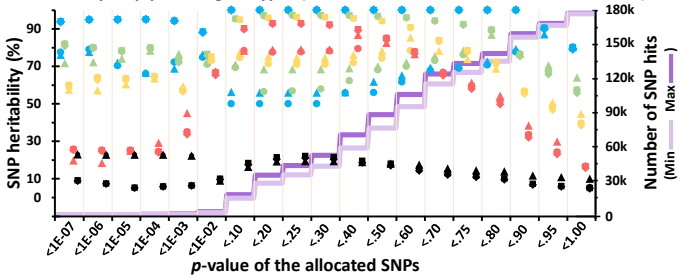

**e) Analysis of 10 artificial random phenotypes on a panel of 1,092 HapMap phase 1 genotypes (177,675 SNPs + 1 simulated SNP; MAF ≥1%)**

| Random phenotypes | Range (N) | EMMAX | LDAK_GCTA | LDAK_Thin | GCTA_GREML |
|---|---|---|---|---|---|
| 1st | 40-150 (1,092) | ■ | ● | ▲ | ◆ |
| 2nd | 40-150 (1,092) | ■ | ● | ▲ | ◆ |
| 3rd | 40-150 (1,092) | ■ | ● | ▲ | ◆ |
| 4th | 40-150 (1,092) | ■ | ● | ▲ | ◆ |
| 5th | 40-150 (1,092) | ■ | ● | ▲ | ◆ |
| 6th | 40-150 (1,092) | ■ | ● | ▲ | ◆ |
| 7th | 40-150 (1,092) | ■ | ● | ▲ | ◆ |
| 8th | 40-150 (1,092) | ■ | ● | ▲ | ◆ |
| 9th | 40-150 (1,092) | ■ | ● | ▲ | ◆ |
| 10th | 40-150 (1,092) | ■ | ● | ▲ | ◆ |




**f) Analysis of 5 artificial non-random phenotypes on a panel of 1,092 HapMap phase 1 genotypes (177,675 SNPs + 1 simulated SNP; MAF ≥1%)**

| Phenotypes | The introduced non-random SNP position | | | | | EMMAX | LDAK_GCTA | LDAK_Thin | GCTA_GREML |
|---|---|---|---|---|---|---|---|---|---|
| | Ref/Ref Genotype | Ref/Alt Genotype | Alt/Alt Genotype | | | | | | |
| | range (N) | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | | | | |
| 1st | 40-50 (982) | 80-100 (103) | 120-150 (7) | 5.704 | 94.296 | ■ | ● | ▲ | ◆ |
| 2nd | 40-60 (982) | 75-100 (103) | 112-150 (7) | 21.037 | 78.963 | ■ | ● | ▲ | ◆ |
| 3rd | 40-70 (982) | 70-100 (103) | 105-150 (7) | 41.501 | 58.499 | ■ | ● | ▲ | ◆ |
| 4th | 40-80 (982) | 60-100 (103) | 90-150 (7) | 72.099 | 27.901 | ■ | ● | ▲ | ◆ |
| 5th | 40-90 (982) | 50-100 (103) | 75-150 (7) | 91.416 | 8.584 | ■ | ● | ▲ | ◆ |




## Figure 1.



*Darbani B. & Nicolaisen M.* SAFE-$h^2$ applies phenotype-specific genetic variance for SNP heritability estimations.

**Fig. 1. The SNP heritability profiles of artificial random and non-random quantitative phenotypes computed on the simulated genomic variants and the variants of chr. 22 from the human HapMap phase 1 genotypes.** a–f) Initial GWAS was performed using EMMAX to realize association levels for the SNPs. The four different approaches of EMMAX, LDAK-GCTA, LDAK-Thin, and GCTA-GREML were then applied to estimate SNP heritability for the clusters of SNP hits captured using the association *p*-values. N represents the sample sizes, *i.e.*, number of genotypes. MAF stands for the minor allele frequency. a,c,e) The average SNP heritability estimates were obtained for 10 independent random phenotypes within identical ranges. The error bars represent 99% confidence intervals. b,d,f) The SNP heritability estimates were obtained for five artificial non-random phenotypes which were associated with a simulated non-random SNP position introduced within both the simulated and HapMap genomes. By ensuring additive-only effects, the non-random phenotypes were defined at different quantitative ranges among the genotypic groups of the introduced non-random SNP position, *i.e.*, Ref/Ref and Alt/Alt for the simulated inbred genotypes and Ref/Ref, Ref/Alt, and Alt/Alt for HapMap data. The intra-genotype assignment of the phenotypes, while being within their specific ranges, was still random. The inter-genotype variances of the phenotypes are shown as $\sigma_g^2$. $\sigma_e^2$ represents the intra-genotype variations. a–d) The analyses of inbred genotypes using GCTA-GREML were performed based on the "inbred" algorithm of GCTA.

not fade and including all SNPs into the estimation models shrank the heritability estimation levels sharply (Fig. 1c). In contrast to the increased number of genotypes (higher genotypic coverage), mimicking lower genomic coverage with analyzing limited number of SNP, as expected, decreased the chance of having random associations and resulted in lower levels of heritability (Extended Data Fig. 1). The most important was the similar patterns observed in these analyses in which the SNPs with very big association *p*-values contributed negatively to the heritability estimations. As mentioned, this was biologically illogical and contributions to heritability, if any, should only be positive. Further proof obtained by a similar analysis performed on the human HapMap phase 1 genomic data (21) including 1,092 genotypes and 177,675 SNPs on the chromosome 22 (MAF ≥1%). We also generated artificial phenotypes as random integers for HapMap genotypes and by heritability profiling, revealed similar patterns of negative contributions (Fig. 1e). Taken together, In all these three genomic datasets that were analyzed *via* the four different heritability estimation models, intense declines in the SNP heritability estimates were observed after allocating the SNPs with large association *p*-values (Fig. 1a,c,e).

By applying artificial non-random phenotypes, similar patterns were revealed on the same genomic datasets, *i.e.*, the simulated and HapMap chromosome 22 SNP datasets, each with an introduced non-random SNP position. The simulated non-random SNP position was introduced to be responsible, only by additive effects, for the artificial phenotypes. The phenotypic variance among the genotype groups of the introduced non-random SNP position was set from 2.8% to 94.3% (Fig. 1b,d,f). Therefore, SNP heritability estimations were predictable at the predefined levels or higher due to the all other SNPs that could also by chance explain additional phenotypic variance. Nevertheless, the most weakly-associated SNPs, as expected, had negative contributions and led to substantial underestimation of SNP heritability levels in all the scenarios and all the heritability estimation models of EMMAX (19), LDAK-GCTA, LDAK-Thin (8), and GCTA-GREML (20) (Fig. 1b,d,f). The SNP heritability estimations were far below the predefined heritability levels when all the SNPs were included in the models (Fig. 1b,d,f). As already discussed, the estimation of SNP heritability is a trade-off between causal and non-causal SNPs, with non-causals resulting in heritability underestimations through negative contributions. So, negative effects of non-causal SNPs on heritability estimations are anticipated to be weakened by including multiple highly associated SNPs. It is very rare and almost impossible to find multiple SNPs with high levels of contribution to phenotypic variances when analyzing limited number of SNPs such as 200k in real experiments. We, however, designed such rare scenarios and performed identical analyses as in Fig. 1b but this time in the presence of 12, 24, and 48 associated SNPs that were introduced within the genomic dataset (Extended Data Fig. 2). Again, there were similar negative contribution patterns imposed by the SNPs having large association *p*-values (Extended Data Fig. 2). As expected, the rate of negative contributions declined as the number of predesigned associated SNPs increased (Extended Data Fig. 2). However, the biologically meaningless negative contributions to heritability did not vanish even after introducing multiple associated SNPs among a limited number of random SNPs. Taken together, the analyses on simulated random and non-random phenotypes indicated that it is essential to roll out non-causal SNPs which impose negative contributions to heritability. We also noticed instability in SNP heritability estimations by LDAK-GCTA and LDAK-Thin as they did not correctly estimate the predefined heritability levels of 93.4%, 93%, 3.7%, and 2.8%

**Table 1. SAFE-$h^2$ improves the heritability estimations by preventing the negative contributions of weakly-associated genomic regions.**

| | | Number of Genotypes/Variants | SAFE-$h^2_{SNP}$ ([SAFE-$h^2_{SNP}$] - [$h^2_{SNP}$ based on all the SNPs]; [SAFE-$h^2_{SNP}$] - [$h^2_{SNP}$ based on the SNP hits with $p$-value < $10^{-5}$]) | | | | |
|---|---|---|---|---|---|---|---|
| | | | EMMAX | LDAK GCTA-model | LDAK Thin-model | GCTA-GREML[a] | Average |
| *Arabidopsis thaliana* phenotypes | CLN | 904 / 2,807,143 | **83.4%** (44.5%; 30.8%) | **75.4%** (36.4%; 22.8%) | **73.5%** (32.2%; 24.5%) | **83.4%** (44.5%; 30.8%) | **78.9%** (39.4%; 27.2%) |
| | FBE | 936 / 2,824,927 | **93.1%** (42.1%; 19.4%) | **77.0%** (26.0%; 3.3%) | **80.1%** (17.0%; 12.1%) | **93.1%** (42.1%; 19.4%) | **85.9%** (31.8%; 13.6%) |
| | DTF | 1,058 / 2,830,731 | **62.9%** (43.4%; 46.2%) | **62.8%** (43.4%; 46.1%) | **61.5%** (40.5%; 38.1%) | **62.9%** (43.4%; 46.2%) | **62.5%** (42.7%; 44.1%) |
| | FN | 409 / 2,628,158 | **62.2%** (62.2%; 53.0%) | **62.1%** (39.8%; 39.0%) | **65.9%** (44.2%; 42.8%) | **62.1%** (62.1%; 52.9%) | **63.1%** (52.1%; 46.9%) |
| | GR | 420 / 2,631,629 | **73.4%** (66.4%; 40.4%) | **73.4%** (66.4%; 40.4%) | **66.0%** (59.6%; 32.4%) | **73.4%** (66.4%; 40.4%) | **71.5%** (64.7%; 38.4%) |
| | RLN | 850 / 2,820,131 | **89.4%** (23.2%; 13.1%) | **80.9%** (14.7%; 4.6%) | **81.1%** (11.7%; 24.5%) | **89.4%** (23.2%; 13.1%) | **85.2%** (18.2%; 13.8%) |
| | RL | 328 / 2,621,362 | **71.5%** ( 71.5%; 49.6%) | **71.5%** (49.6%; 49.7%) | **72.6%** (59.8%; 44.6%) | **71.4%** (71.4%; 46.9%) | **71.8%** (63.1%; 48.4%) |
| | SL | 680 / 2,789,485 | **54.6%** (54.6%; 30.9%) | **54.6%** (54.6%; 31.0%) | **56.2%** (53.7%; 40.4%) | **54.6%** (54.6%; 30.9%) | **55.0%** (54.4%; 33.3%) |
| | As | 344 / 2,839,190 | **100%** (30.8%; 41.3%) | **78.8%** (9.8%; 20.2%) | **93.5%** (30.1%; 34.6%) | **99.5%** (30.3%; 40.9%) | **93.0%** (25.3%; 34.3%) |
| | Cd | 345 / 2,840,072 | **99.3%** (44.0%; 24.9%) | **79.4%** (24.1%; 5.0%) | **83.2%** (18.3%; 10.8%) | **99.3%** (44.0%; 24.9%) | **90.3%** (32.7%; 16.4%) |
| | Li | 345 / 2,840,072 | **97.6%** (56.2%; 69.8%) | **87.6%** (46.3%; 59.8%) | **84.8%** (49.9%; 55.2%) | **97.6%** (56.2%; 69.8%) | **91.9%** (52.2%; 63.7%) |
| | Mn | 345 / 2,840,072 | **100%** (45.5%; 65.3%) | **79.4%** (24.9%; 44.8%) | **84.5%** (31.4%; 50.6%) | **100%** (45.5%; 65.4%) | **91.0%** (36.8%; 56.5%) |
| | Zn | 345 / 2,840,072 | **100%** (67.4%; 35.2%) | **88.0%** (55.4%; 23.2%) | **81.5%** (56.4%; 35.4%) | **100%** (67.4%; 35.2%) | **92.4%** (61.7%; 32.3%) |
| | P | 345 / 2,840,072 | **96.8%** (44.1%; 46.5%) | **85.8%** (33.4%; 35.6%) | **80.0%** (32.3%; 44.0%) | **96.8%** (44.2%; 46.6%) | **89.8%** (38.5%; 43.2%) |
| Rice (*O. sativa*) phenotypes | DTF | 2,102 / 6,414,800 | **99.9%** (10.0%; 16.1%) | **90.0%** (0.0%; 18.1%) | **92.2%** (1.5%; 19.2%) | **99.9%** (11.7%; 17.8%) | **95.6%** (5.6%; 17.8%) |
| | GL | 2,099 / 6,411,145 | **99.9%** (14.0%; 16.0%) | **88.3%** (2.4%; 16.3%) | **91.3%** (3.1%; 17.3%) | **99.9%** (15.2%; 17.2%) | **94.9%** (8.6%;16.7%) |
| | GWE | 2,098 / 6,410,047 | **99.9%** (13.5%; 17.4%) | **88.8%** (2.4%; 16.6%) | **91.4%** (3.3%; 17.3) | **99.9%** (14.6%; 19.1%) | **95.0%** (8.4%; 17.6%) |
| | PL | 1,762 / 6,269,207 | **99.9%** (17.3%; 14.7%) | **86.9%** (4.2%; 14.4%) | **88.9%** (3.3%; 14.3%) | **99.9%** (17.8%; 14.7%) | **93.9%** (10.7%; 14.5%) |
| | GWI | 2,099 / 6,411,145 | **99.9%** (14.8%; 15.7%) | **87.6%** (2.4%; 15.4%) | **90.6%** (3.1%; 16.5%) | **99.9%** (15.9%; 16.8%) | **94.6%** (9.1%; 16.1%) |
| | SD (EC18)[b] | 1,998 / 6393347 | **99.9%** (54.3%; 4.6%) | **85.6%** (39.9%; 16.6%) | **88.3%** (29.7%; 20.1%) | **99.9%** (53.1%; 17.3%) | **93.5%** (44.2%; 14.6%) |
| | SD (EC12)[b] | 1,998 / 6,393,347 | **99.9%** (23.5%; 19.4%) | **88.4%** (11.9%; 15.11%) | **90.5%** (9.8%; 13.5%) | **99.9%** (23.8%; 19.9%) | **94.7%** (17.3%; 17.0%) |
| | SH | 1,743 / 6,254,739 | **99.9%** (16.4%; 13.7%) | **87.6%** (4.0%; 14.5%) | **89.5%** (3.1%; 15.0%) | **99.9%** (17.0%; 14.7%) | **94.3%** (10.1%; 14.4%) |
| | SPB | 1,422 / 6,131,325 | **99.9%** (35.3%; 2.2%) | **94.6%** (29.9%; 27.2%) | **98.3%** (26.9%; 30.0%) | **99.9%** (33.9%; 39.6%) | **98.2%** (31.5%;24.8%) |
| | SF | 2,100 / 6,411,911 | **99.9%** (21.2%; 27.7%) | **90.8%** (12.0%; 18.5%) | **92.5%** (10.4%; 16.2%) | **99.9%** (21.8%; 28.2%) | **95.8%** (16.3%; 22.6%) |
| **Average** | | | ..... (38.2%; 29.8%) | ..... (26.4%; 24.9%) | ..... (26.3%; 27.9%) | ..... (38.3%; 32.0%) | ..... **(32.3%; 28.7%)** |
| **Minimum** | | | ..... (10.0%; 2.2%) | ..... (0.0%; 3.3%) | ..... (1.5%; 10.8%) | ..... (11.7%; 13.1%) | ..... **(5.6%; 13.6%)** |
| **Maximum** | | | ..... (71.5%; 69.8%) | ..... (66.4%; 59.8%) | ..... (59.8%; 55.2%) | ..... (71.4%; 69.8%) | ..... **(64.7%; 63.7%)** |

CLN: cauline leaf number; FBE: days until the emergence of flowering buds; DTF: days until first open flower; FN: number of fruits per plant; GR: average growth rate; RLN: rossette leaf number; RL: root length; SL: stem length; As: arsenic content in leaf; Cd: cadmium content in leaf; Li: lithium content in leaf; Mn: manganese content in leaf; Zn: zinc content in leaf; P: phosphorus content in leaf; GL: grain length; GWE: grain weigth; PL: panicle length; GWI: grain width; SD (EC18): salt injury at EC of 18; SD (EC12): salt injury at EC of 12; SH: seedling height; SPB: secondary panicle branching; SF: spikelet fertility. [a]Analyzing inbred genotypes of rice and Arabidopsis was performed under the "inbred" algorithm of GCTA. [b]The electrical conductivity.

when the estimations were only based on the predesigned associated SNPs, *i.e.*, SNPs with $p < 10^{-7}$, and all the other SNPs were excluded (Fig. 1b,d, Extended Data Fig. 2). Here, the EMMAX and GCTA had better performances by obtaining the best approximates.

As the simulated analyses revealed, the SNP heritability was affected and underestimated by illogical negative contributions after including the SNPs with large association *p*-values. Therefore, it is essential to preclude the incorporation of these SNPs into heritability estimation models. These SNPs can be called non-causals as they are not related to phenotypes, *i.e.*, have random distribution among genotypes in relation to phenotypic variations and impose meaningless negative contributions to heritability. SAFE-$h^2$ prevents negative contributions to SNP heritability by defining a *p*-value threshold. This threshold excludes SNPs with negative contributions and thereby, prevents underestimation of genetic variance and SNP heritability. In fact, this *p*-value threshold is the best border that we can assume between causal and non-causal SNPs to estimate SNP heritability. Further validation was accomplished on public data from *Arabidopsis* and rice genotypes. The SNP heritability profiling was carried out for 14 different phenotypes from 328 to 1058 *Arabidopsis* genotypes (22). The heritability profiling was performed on 2,621k to 2,840k *Arabidopsis* SNPs (23, 24), depending on the phenotypes. The analyses also included 10 phenotypes observed in 1,422 to 2,102 rice genotypes (25) with 6,131k to 6,414k SNPs (25–27) varying among the phenotypes. The minimum MAF was set to 0.01 in all the analyses. As the bold observation of SNP heritability profiling in the *Arabidopsis* and rice datasets, the genomic contributions of the strongly associated SNPs were again neutralized by the weakly associated SNPs, *i.e.*, SNPs with large *p*-values (Table 1). This conclusion was in line with the observations on the simulated and HapMap chromosome 22 data (Fig. 1, Extended Data Fig. 1 and 2). Most importantly, the weakly associated SNPs with negative contributions masked the heritability estimates in all the examined *Arabidopsis* and rice phenotypes by 32.3 units on average and from 5.6 to 64.7 units for individual phenotypes by taking the four heritability estimation models into account (Table 1). The SNP heritability was also underestimated when it was only based on the SNP hits with small association *p*-values such as $<10^{-8}$ or even $<10^{-5}$. By calculating the SNP heritability at the *p*-value threshold, SAFE-$h^2$ improved the heritability estimates by 28.7 units on average and from 13.6 to 63.7 units for individual phenotypes, when compared to the SNP heritability which was estimated on the SNP hits with the association *p*-values less than $10^{-5}$ (Table 1). Finally, we used SAFE-$h^2$ to analyze 50 additional phenotypes from three plant and four animal species (Extended Data Table 1). In agreement with the previous results, these analyses also found the SNPs with large association *p*-values imposing up to 30 units of negative contributions to heritability estimations on average (Extended Data Table 1).

## *SAFE-h2 reports a safe p-value interval to minimize positive contributions of false-positive SNP hits to heritability estimations*

False-positive rates are dependent on the statistical power of experiments. It is also very difficult to secure considerable statistical power in GWAS. This is not only because of the high resource demanding nature of GWAS but also impracticality of near-to-complete genotypic coverage. In theory, we approximately expect $2^{(\text{number-of-SNPs})}$ possible genotypes. On the other side, we can also face a limited number of available unrelated genotypes that can be sampled from nature for experiments. Depending on experiments, different levels of false-positive rates are therefore expected. To bypass this challenge, SAFE-$h^2$ relies on its own innovative approach to exclude false-positive hits of SNPs. SAFE-$h^2$ reports a safe *p*-value interval which its upper-end acts as a new *p*-value threshold for heritability estimation with minimized contributions by false-positive SNP hits. For this, the range of phenotype-of-interest is used to simulate six independent random phenotypes for all the genotypes. Afterwards, SAFE-$h^2$ creates SNP heritability profiles for all the generated random phenotypes as well as the original phenotype-of-interest using the identical genomic dataset. Analyzing of the random phenotypes then reveals a safe *p*-value interval with minimum SNP heritability level at the lower-tail of heritability profile. The SNP heritability for the original phenotype can subsequently be estimated at the upper-end of this interval that acts as a *p*-value threshold to omit the false-positive SNP hits from the estimation model. Figure 2 illustrates the same analysis on four different genomic datasets from tomato, rice, foxtail millet, and *Arabidopsis*. We found the *p*-value threshold of $10^{-4}$ to estimate false-positive free SNP heritability for betaine and hesperetin contents in tomato (Fig 2a). This threshold turned to be $10^{-6}$ for flowering-bud emergence and zinc content in *Arabidopsis* (Fig. 2b). The SNP heritability estimations with minimized false-positive SNP contribution were also possible at the *p*-value thresholds of $10^{-6}$ and $10^{-5}$ for salt-injury and secondary-panicle branching in rice, respectively

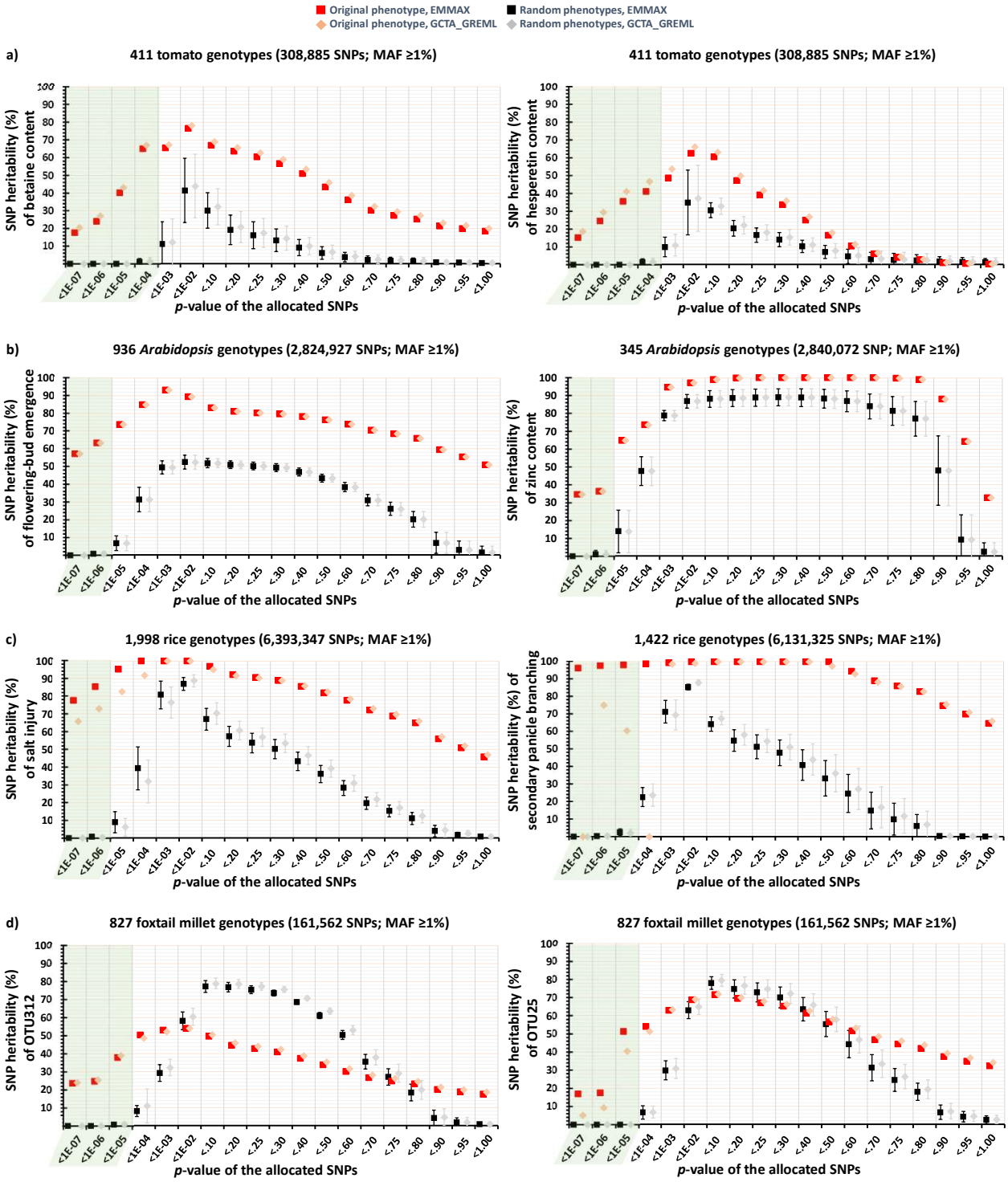

Original phenotype, EMMAX
Original phenotype, GCTA_GREML
Random phenotypes, EMMAX
Random phenotypes, GCTA_GREML
a)
411 tomato genotypes (308,885 SNPs; MAF ≥1%)
SNP heritability (%) of betaine content
411 tomato genotypes (308,885 SNPs; MAF ≥1%)
SNP heritability (%) of hesperetin content
b)
936 Arabidopsis genotypes (2,824,927 SNPs; MAF ≥1%)
SNP heritability (%) of flowering-bud emergence
345 Arabidopsis genotypes (2,840,072 SNP; MAF ≥1%)
SNP heritability (%) of zinc content
c)
1,998 rice genotypes (6,393,347 SNPs; MAF ≥1%)
SNP heritability (%) of salt injury
1,422 rice genotypes (6,131,325 SNPs; MAF ≥1%)
SNP heritability (%) of secondary panicle branching
d)
827 foxtail millet genotypes (161,562 SNPs; MAF ≥1%)
SNP heritability (%) of OTU312
827 foxtail millet genotypes (161,562 SNPs; MAF ≥1%)
SNP heritability (%) of OTU25
p-value of the allocated SNPs
<1E-07
<1E-06
<1E-05
<1E-04
<1E-03
<1E-02
<.10
<.20
<.25
<.30
<.40
<.50
<.60
<.70
<.75
<.80
<.90
<.95
<1.00

**Figure 2.**

**Fig. 2. SAFE-$h^2$ estimates the SNP heritability by minimizing the contributions from false-positive SNP hits.** a–d) The safe-intervals for *p*-value are highlighted in light green representing the area with minimum contributions to heritability by false-positive SNP hits. The two different approaches of EMMAX and GCTA-GREML were applied to estimate the SNP heritability for the clusters of SNP hits captured using the association *p*-values. MAF stands for the minor allele frequency. The average estimations are illustrated for six independent random phenotypes (simulated within the range of original phenotype) and the error bars represent standard deviations among the six estimations.

(Fig. 2c). Finally, the analysis determined the threshold of $10^{-5}$ for two independent operational taxonomic units of root-microbiome in foxtail millet (Fig. 2d). In all these examples, the SNP heritability estimations using all SNPs were also below and underestimated when compared to the SNP heritability estimations obtained after minimizing the chance for false-positive contributions (Fig. 2). It was also apparent that the level of negative contribution is a trade-off between causal and non-causal SNPs and thereby, can fluctuate among phenotypes (Table 1, Extended Data Table 1, Fig. 2).

## *SNP heritability estimation by joint-analysis of intra-locus allelic effects of additive and non-additive*

The genetic scoring system for SNP heritability estimations is designed to predict additive allelic effects. In fact, linear models can extract intra-locus non-additive allelic effects partially and thereby some levels of non-additive allelic effects always remain uncaptured. For example, the GCTA-GREMLd model has been introduced to address dominance effects by heritability estimations (28). Here, SAFE-$h^2$ introduces an allelic adjustment strategy to extract non-additive effects including dominance, overdominance, and heterosis-like effects while relying on linear models. Fig. 3a and 3b illustrate examples of different intra-locus allelic effects where the exponential model performs better than the linear approach in the presence of non-additive allelic effects. The linear models, when powered by allelic adjustments that are specific to the type of non-additive effects, can also secure the same levels of genetic variance captured by the exponential model (Fig. 3c–f). SAFE-$h^2$ uses the same algorithm to find the best fitted linear model among adjusted allele-sets for every SNP position and phenotype of interest. The method includes allelic adjustments, *i.e.*, allele conversions or replacements, specific for either of the dominance, overdominance, or heterosis-like effects (see Fig. 3a–f). For every SNP position, SAFE-$h^2$ examines different scenarios, *i.e.*, additive (original genotypes), adjusted genotypes for dominance effects, adjusted genotypes for overdominance effects, and adjusted genotypes for heterosis-like effects, to find the smallest association *p*-value to the phenotype of interest. Before heritability profiling, combinatory genotype datasets can then be built based on the selected adjustments for every SNP position. Using the allelic adjustments, SAFE-$h^2$ improved the SNP heritability estimations for the 10 artificial random phenotypes which was assigned to 1,092 HapMap 1 genotypes (chr.22 SNPs, see Fig. 1e). This resulted in up to 9.7 units improvement when compared to the additive-only strategy (Fig. 3g). On the same datasets, the GCTA-GREMLd showed no improvements over its additive based model of GCTA-GREML (Fig. 3g). A similar pattern, although at a lower rate, was also observed on 10,000 simulated genotypes which were derived from 1,092 HapMap 1 genotypes (Fig. 3h). On this dataset, GREMLd led to 3.46 units underestimation on average when compared to the additive approach (Fig. 3h). The allele adjustment strategy of SAFE-$h^2$ was strong enough to leave no space for further improvements in heritability estimations even after increasing the population size from 1,092 to 10,000 (Fig. 3i). On the contrary, the additive and GREMLd based estimates were improved by population enlargement (Fig. 3i). The SAFE-$h^2$ performance on SNP heritability estimation by joint analysis of additive and non-additive allelic effects was also examined using artificial non-random phenotypes (Table 2). These six sets of artificial phenotypes (simulated phenotypes within specific ranges, also see Fig. 1f) were associated with a biallelic locus at different rates and through different genetic effects including additive, dominance, overdominance, and heterosis-like effects (Table 2). For this example (*i.e.*, Table 2), the biallelic locus was the sole SNP position within the analyzed genomic dataset. When the real genetic effects were additive on the phenotypes, the SAFE-$h^2$ estimates of SNP heritability by capturing various combinations of additive and non-additive allelic effects were identical to the SNP heritability estimates obtained on the additive-only effects (Table 2). When the real genetic effects were non-additive, the allelic adjustments for non-additive effects by SAFE-$h^2$ improved the heritability estimations using all the four models, *i.e.*, EMMAX, GCTA-GREML, LDAK-Thin, and LDAK-GCTA (Table 2). GCTA-GREML and LDAK were however

**Figure 3.**



*Darbani B. & Nicolaisen M.* SAFE-$h^2$ applies phenotype-specific genetic variance for SNP heritability estimations.

**Fig. 3. Allele adjustments capture intra-locus non-additive allelic effects.** a–f) Demonstration of four biallelic loci; locus with A/a alleles having additive effect, locus with B/b alleles having dominance effect, locus with C/c alleles having overdominance effect, and locus with D/d alleles having heterosis-like effect. The best linear-fits are highlighted by bold and underlined $R^2$. The adjusted alleles are in italic. Single arrows indicate the allele adjustment by conversion and double arrows indicate allele adjustment by replacement. a) Illustration of allelic effects in a biallelic scenario. Dominance and overdominance effects can be on either alternative or reference alleles (Alt. or Ref.) which are differentiated by continues and dashed extrapolations. b) The artificial phenotype no. 3 that was assigned to the 1,092 human HapMap phase 1 genotypes (see Fig. 1f), and its derivatives are used to represent the additive and non-additive effects, respectively. Continues lines represent the most fitted models; for the locus with A/a alleles this was linear. Low-fit linear approximates (dashed lines) are also shown for dominance, overdominance, and heterosis-like effects in the other loci with alleles B/b, C/c, and D/d, respectively. c) Allelic adjustment for the dominance effects of Ref alleles (A, B, C, & D) leads to an improved linear-fit only on the locus having the B allele with dominance effect. d) Allelic adjustment for the overdominance effects of Ref alleles (A, B, C, & D) leads to an improved linear-fit only on the locus having the C allele with overdominance effect. e) Allelic adjustment for the heterosis-like effects leads to an improved linear-fit on the locus with D/d alleles with heterosis-like effect. f) A combinatory genotype dataset contains every SNP position in either of original or one of the adjusted states determined by having the best linear-fit, *i.e.*, bigger coefficient of determination $R^2$. g) SAFE-$h^2$ estimates of SNP heritability for the 10 artificial random phenotypes assigned to the 1,092 HapMap 1 genotypes on chr. 22 (see Fig. 1e) while considering additive-only effects, additive plus non-additive effects by allele adjustment, and additive plus dominance effects by GREMLd model. The *p*-values are from two-tailed paired t-tests comparing additive-only heritability estimates with either of additive plus non-additive combinations. h) SAFE-$h^2$ estimates of SNP heritability for the 10 artificial random phenotypes (at the same range as in G and Fig. 1e) assigned to the simulated 10,000 genotypes derived from the chr. 22 of 1,092 HapMap genotypes. Again, SAFE-$h^2$ profiling was performed by applying the three approaches of additive-only effects, additive plus non-additive effects by allele adjustment, and additive plus dominance effects by GREMLd model. The *p*-values are from two-tailed paired t-tests comparing the SNP heritability estimates on additive-only effect with the estimates from either of additive plus non-additive effects combinations. i) Comparing the results for 1,092 and 10,000 genotypes. The error bars represent 95% confidence intervals. j) SAFE-$h^2$ estimates of SNP heritability under additive-only model, combined additive and non-additive models, and GREMLd model for different phenotypes in plant and animal species. The analyses of tomato and foxtail millet genotypes using GCTA-GREML model were performed based on the "inbred" algorithm of GCTA. The average inbreeding coefficients (Fhat) for tomato and foxtail millet genotypes were > 0.8. g–j) Add: heritability based on additive effects, AddDom: heritability based on additive and dominance effects, AddDomOD: heritability based on additive, dominance and overdominance effects, AddDomODHet: heritability based on additive, dominance, overdominance, and heterosis-like effects, GREMLd: heritability estimated by GREMLd model, MAF: minor allele frequency, OUT: operational taxonomic unit.

insecure to capture additive-only or additive plus non-additive allelic effects at different association levels (Table 2). SAFE-$h^2$ obtained the best approximates of the genetic variances, *i.e.*, here equivalent to the broad-sense heritability, when applying the EMMAX model (Table 2). This performance was noticeable by considering the type of allelic effects being analyzed for estimating SNP heritability as well as the type of real allelic effects being responsible for the phenotypes (Table 2). SAFE-$h^2$ led to better estimates of SNP heritability by capturing 6.22% extra broad-sense heritability on the allelic effects of dominance, 16.22% extra broad-sense heritability on the allelic effects of overdominance, and 24.31% extra broad-sense heritability on the heterosis-like effects on average. The efficiency of SAFE-$h^2$ was further challenged using 10 published datasets (29–38), including 50 phenotypes observed on three plants and four animal species (Fig. 3j, Extended Data Fig. 3). Again, SAFE-$h^2$ was able to extract extra intra-locus non-additive allelic effects at different rates depending on the dataset and the phenotype of interest (Fig. 3j, Extended Data Fig. 3). These results also confirmed the outperformance of SAFE-$h^2$ allelic adjustment model versus GREMLd (Fig. 3j, Extended Data Fig. 3). As seen for the simulated datasets, SAFE-$h^2$ led to the most stable performance and large improvements when analyzing the real datasets (Fig. 3j, Extended Data Fig. 3). It is important to emphasize that the overestimation of SNP heritability is only possible by positive contributions of false-positive SNP hits and this depends on statistical power levels. Under an identical statistical power level, *i.e.*, genomic and genotypic coverage rates, the best performed method for heritability estimation is therefore the model that captures the most from phenotypic variations.

## Table 2: SNP heritability estimations on additive or additive + non-additive effects of a biallelic locus.

| Phenotypes | | | | | | SAFE-$h^2$ estimates of SNP-heritability | | | | | | | | | | | | | | | | GCTA GREMLd -model |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ref/Ref Genotype | Ref/Alt Genotype | Alt/Alt Genotype | | | using EMMAX model | | | | using GCTA GREML-model | | | | using LDAK GCTA-model | | | | using LDAK Thin-model | | | | |
| Data_Genetic effect | range (N) | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | A[a] | AD[b] | ADO[c] | ADOH[d] | A[a] | AD[b] | ADO[c] | ADOH[d] | A[a] | AD[b] | ADO[c] | ADOH[d] | A[a] | AD[b] | ADO[c] | ADOH[d] | |
| 1st_Additice | 40-50 (982) | 80-100 (103) | 120-150 (7) | 5.7 | 94.3 | 94.29 | 94.29 | 94.29 | 94.29 | 93.91 | 93.91 | 93.91 | 93.91 | 77.22 | 77.22 | 77.22 | 77.22 | 77.22 | 77.22 | 77.22 | 77.22 | 93.92 |
| 1st_Dominance | 40-50 (982) | 120-150 (103) | 120-150 (7) | 2.16 | 97.84 | 92.8 | ▲ 97.83 | ▲ 97.83 | ▲ 97.83 | 92.77 | ▲ 95.74 | ▲ 95.74 | ▲ 95.74 | 77.16 | ▲ 77.53 | ▲ 77.53 | ▲ 77.53 | 77.16 | ▲ 77.53 | ▲ 77.53 | ▲ 77.53 | 97.89[e] |
| 1st_Overdominance | 40-50 (982) | 120-150 (103) | 80-100 (7) | 2.35 | 97.65 | 84.77 | ▲ 95.9 | ▲ 97.65 | ▲ 97.65 | 84.29 | ▲ 92.12 | ▲ 95.48 | ▲ 95.48 | 85.21 | ▼ 77.35 | ▼ 77.51 | ▼ 77.51 | 85.21 | ▼ 77.35 | ▼ 77.51 | ▼ 77.51 | 97.77[e] |
| 1st_Heterosis-like | 40-50 (982) | 120-150 (103) | 40-50 (7) | 6.11 | 93.89 | 70.51 | ▲ 86.47 | ▲ 91.72 | ▲ 93.86 | 69.34 | ▲ 76.16 | ▲ 84.93 | 94.14[e] | 70.77 | ▲ 76.61 | ▲ 77.02 | ▲ 77.19 | 70.77 | ▲ 76.61 | ▲ 77.02 | ▲ 77.19 | 94.24[e] |
| 2nd_Additive | 40-60 (982) | 75-100 (103) | 112-150 (7) | 21.04 | 78.96 | 78.86 | 78.86 | 78.86 | 78.86 | 77.7 | 77.7 | 77.7 | 77.7 | 78.84 | 78.84 | 78.84 | 78.84 | 78.84 | 78.84 | 78.84 | 78.84 | 77.53 |
| 2nd_Dominance | 40-60 (982) | 112-150 (103) | 112-150 (7) | 12.75 | 87.25 | 82.1 | ▲ 87.23 | ▲ 87.23 | ▲ 87.23 | 81.2 | ▼ 77.34 | ▼ 77.34 | ▼ 77.34 | 82.22 | ▼ 76.67 | ▼ 76.67 | ▼ 76.67 | 82.22 | ▼ 76.67 | ▼ 76.67 | ▼ 76.67 | 87.57[e] |
| 2nd_Overdominance | 40-60 (982) | 112-150 (103) | 75-100 (7) | 14.7 | 85.3 | 71.58 | ▲ 82.7 | ▲ 85.13 | ▲ 85.13 | 70.33 | ▲ 70.49 | ▲ 74.44 | ▲ 74.44 | 71.74 | ▲ 82.69 | ▲ 76.5 | ▲ 76.5 | 71.74 | ▲ 82.69 | ▲ 76.5 | ▲ 76.5 | 86.02[e] |
| 2nd_Heterosis-like | 40-60 (982) | 112-150 (103) | 40-60 (7) | 8.1 | 91.9 | 70.56 | ▲ 85.58 | ▲ 90.28 | ▲ 91.89 | 69.36 | ▲ 74.78 | ▲ 82.54 | 92.26[e] | 70.79 | ▲ 76.54 | ▲ 76.91 | ▲ 77.04 | 70.79 | ▲ 76.54 | ▲ 76.91 | ▲ 77.04 | 92.35[e] |
| 3rd_Additive | 40-70 (982) | 70-100 (103) | 105-150 (7) | 41.5 | 58.5 | 58.34 | 58.34 | 58.34 | 58.34 | 56.67 | 56.67 | 56.67 | 56.67 | 58.32 | 58.32 | 58.32 | 58.32 | 58.32 | 58.32 | 58.32 | 58.32 | 56.31 |
| 3rd_Dominance | 40-70 (982) | 105-150 (103) | 105-150 (7) | 15.47 | 84.53 | 79.48 | ▲ 84.51 | ▲ 84.51 | ▲ 84.51 | 78.45 | ▼ 73.15 | ▼ 73.15 | ▼ 73.15 | 79.57 | ▲ 84.49 | ▲ 84.49 | ▲ 84.49 | 79.57 | ▲ 84.49 | ▲ 84.49 | ▲ 84.49 | 84.91[e] |
| 3rd_Overdominance | 40-70 (982) | 105-150 (103) | 70-100 (7) | 15.16 | 84.84 | 72.05 | ▲ 82.65 | ▲ 84.76 | ▲ 84.76 | 70.82 | ▼ 70.41 | ▲ 73.88 | ▲ 73.88 | 72.21 | ▲ 82.63 | ▲ 84.74 | ▲ 84.74 | 72.21 | ▲ 82.63 | ▲ 84.74 | ▲ 84.74 | 85.57[e] |
| 3rd_Heterosis-like | 40-70 (982) | 105-150 (103) | 40-70 (7) | 16.44 | 83.56 | 62.59 | ▲ 76.85 | ▲ 81.56 | ▲ 83.52 | 61.12 | ▲ 62.39 | ▲ 69.24 | 84.19[e] | 62.73 | ▲ 76.83 | ▲ 81.54 | ▲ 83.5 | 62.73 | ▲ 76.83 | ▲ 81.54 | ▲ 83.5 | 84.38[e] |
| 4th_Additive | 40-80 (982) | 60-100 (103) | 90-150 (7) | 72.1 | 27.9 | 27.34 | 27.34 | 27.34 | 27.34 | 26.01 | 26.01 | 26.01 | 26.01 | 27.33 | 27.33 | 27.33 | 27.33 | 27.33 | 27.33 | 27.33 | 27.33 | 25.07 |
| 4th_Dominance | 40-80 (982) | 90-150 (103) | 90-150 (7) | 33.36 | 66.64 | 62.6 | ▲ 66.6 | ▲ 66.6 | ▲ 66.6 | 61.01 | ▼ 49.91 | ▼ 49.91 | ▼ 49.91 | 62.61 | ▲ 66.59 | ▲ 66.59 | ▲ 66.59 | 62.61 | ▲ 66.59 | ▲ 66.59 | ▲ 66.59 | 67.29[e] |
| 4th_Overdominance | 40-80 (982) | 90-150 (103) | 60-100 (7) | 32.11 | 67.89 | 56.78 | ▲ 65.73 | ▲ 67.72 | ▲ 67.72 | 55.15 | ▼ 48.93 | ▼ 51.63 | ▼ 51.63 | 56.82 | ▲ 65.71 | ▲ 67.7 | ▲ 67.7 | 56.82 | ▲ 65.71 | ▲ 67.7 | ▲ 67.7 | 69.16[e] |
| 4th_Heterosis | 40-80 (982) | 90-150 (103) | 40-80 (7) | 35.4 | 64.6 | 51.1 | ▲ 60.96 | ▲ 63.84 | ▲ 64.5 | 49.34 | ▼ 43.82 | ▼ 47.32 | 65.64[e] | 51.03 | ▲ 60.94 | ▲ 63.82 | ▲ 64.49 | 51.03 | ▲ 60.94 | ▲ 63.82 | ▲ 64.49 | 65.98[e] |
| 5th_Additive | 40-90 (982) | 50-100 (103) | 75-150 (7) | 91.42 | 8.58 | 7.7 | 7.7 | 7.7 | 7.7 | 7.23 | 7.23 | 7.23 | 7.23 | 22.95[e] | 27.95[e] | 22.95[e] | 22.95[e] | 22.95[e] | 27.95[e] | 22.95[e] | 22.95[e] | na |
| 5th_Dominance | 40-90 (982) | 75-150 (103) | 75-150 (7) | 61.67 | 38.33 | 36.61 | ▲ 38.24 | ▲ 38.24 | ▲ 38.24 | 35.04 | ▼ 23.62 | ▼ 23.62 | ▼ 23.62 | 36.59 | ▲ 38.22 | ▲ 38.22 | ▲ 38.22 | 36.59 | ▲ 38.22 | ▲ 38.22 | ▲ 38.22 | 38.65[e] |
| 5th_Overdominance | 40-90 (982) | 75-150 (103) | 50-100 (7) | 55.12 | 44.88 | 34.61 | ▲ 41.87 | ▲ 44.1 | ▲ 44.82 | 33.09 | ▼ 26.46 | ▼ 28.64 | 46.06[e] | 34.61 | ▲ 41.84 | ▲ 44.08 | ▲ 44.8 | 34.61 | ▲ 41.84 | ▲ 44.08 | ▲ 44.8 | 46.34[e] |
| 5th_Heterosis-like | 40-90 (982) | 75-150 (103) | 40-90 (7) | 56 | 44 | 34.25 | ▲ 41.23 | ▲ 43.33 | ▲ 43.93 | 32.73 | ▼ 25.95 | ▼ 28 | 45.16[e] | 34.24 | ▲ 41.21 | ▲ 43.3 | ▲ 43.91 | 34.24 | ▲ 41.21 | ▲ 43.3 | ▲ 43.91 | 45.46[e] |
| 6th_Additive | 40-95 (982) | 45-100 (103) | 50-150 (7) | 97.56 | 2.44 | 2.3 | 2.3 | 2.3 | 2.3 | 2.15 | 2.15 | 2.15 | 2.15 | 22.59[e] | 22.59[e] | 22.59[e] | 22.59[e] | 22.59[e] | 22.59[e] | 22.59[e] | 22.59[e] | na |
| 6th_Dominance | 40-95 (982) | 50-150 (103) | 50-150 (7) | 76.6 | 23.4 | 20.67 | ▲ 23.03 | ▲ 23.31 | ▲ 23.31 | 19.57 | ▼ 13.02 | ▼ 13.39 | ▼ 13.39 | 20.65 | ▲ 23.05 | ▲ 23.29 | ▲ 23.29 | 20.65 | ▲ 23.05 | ▲ 23.29 | ▲ 23.29 | 24.16[e] |
| 6th_Overdominance | 40-95 (982) | 50-150 (103) | 45-100 (7) | 76.64 | 23.36 | 19.5 | ▲ 22.56 | ▲ 23.24 | ▲ 23.24 | 18.45 | ▼ 12.71 | ▼ 13.35 | ▼ 13.35 | 19.44 | ▲ 22.55 | ▲ 23.22 | ▲ 23.22 | 19.49 | ▲ 22.55 | ▲ 23.22 | ▲ 23.22 | 24.29[e] |
| 6th_Heterosis-like | 40-95 (982) | 50-150 (103) | 40-95 (7) | 76.57 | 23.43 | 16.19 | ▲ 20.62 | ▲ 22.57 | ▲ 23.21 | 15.28 | ▼ 11.49 | 60.84[e] | 24.11[e] | 16.18 | ▲ 20.61 | ▲ 22.56 | ▲ 23.19 | 16.18 | ▲ 20.61 | ▲ 22.56 | ▲ 23.19 | 24.31[e] |

Green triangles indicate improved estimation when compared to the analysis on the original genotypes (a). Red triangles indicate underestimations when compared to the analysis on the original genotypes (a). $\sigma_e^2$ & $\sigma_g^2$ represent the the enviromental and genetic variances, respectively. $\sigma_g^2$ is equivalent to the broad-sense heritability. [a]On original genotypes to capture additive effects; [b]On adjusted genotypes for dominance effects to capture additive and dominance effects; [c]On adjusted genotypes for dominance and overdominance effects to capture additive, dominance, and overdominace effects; [d]On adjusted genotypes for dominance,overdominance, and heterosis-like effects to capture additive, dominance, overdominace, and heterosis-like effects; [e]Overestimated.

## Discussion

Functional, regional, and MAF stratified SNPs have been applied to estimate SNP heritability (7, 11, 12, 14, 16, 39). Nevertheless, these approaches provide sub-genome estimates of genetic variances rather than phenotype-specific, *i.e.*, causal SNP specific, estimates. These approaches do not exclude SNPs having upper-tail association *p*-values. Therefore, all these SNP heritability estimation strategies are prone to illogical negative contributions by non-causal SNPs having upper-tail association *p*-values. Taken together, the results in this study indicate that as the inclusion of extra SNPs, other than the user-defined significant GWAS hits, is needed for a proper estimation of the heritability, the SNP-pruning for large *p*-values is also essential to avoid underestimations, *i.e.*, biologically meaningless negative contributions. To meet these challenges, SAFE-$h^2$ performs a dynamic "**S**NP **A**llocation **F**or **E**stimating the narrow-sense **H**eritability". In this process, SAFE-$h^2$ exploits the SNP heritability estimating tools EMMAX (19), LDAK (8), and GCTA-GREML (20) to estimate SNP heritability for *p*-value-guided SNP clusters. SAFE-$h^2$ prepares a list of *p*-value-guided SNP clusters and outputs a profiling graph of heritability estimates (Fig. 4a,b,c). This facilitates the incorporation of SNP hits while their association *p*-values are ascending until a *p*-value threshold at which the SNP heritability reaches its experimental or data-specific maximum. The heritability profiling by SAFE-$h^2$ also found the EMMAX algorithm as the most stable model for proper estimation of SNP heritability (Fig. 1, Extended Data Fig. 2, Table 2). Low statistical power is a consequence of inadequate genotypic coverage in experiments performing GWAS. To bypass this challenge, SAFE-$h^2$ restricts the contributions of false-positive SNP hits to SNP heritability estimations. SAFE-$h^2$ performs this task by comparing heritability profiles of random phenotypes with the heritability profile of phenotype-of-interest (Fig. 2, 4e). SAFE-$h^2$ can also extract intra-locus non-additive allelic effects, and hence lead to better estimates of SNP heritability. An example of SAFE-$h^2$ output graphs for joint analysis of additive and non-additive allelic effects is shown in Fig. 4d. In contrast to the GCTA-GREMLd model, SAFE-$h^2$ employes an allele adjustment algorithm and obtains identical results as compared to the original genotypic data at the presence of additive-only allelic effects (Table 2). SAFE-$h^2$ achieved the best performance using EMMAX model to capture additive and non-additive allelic effects of the predefined single biallelic-locus and did not lead to overestimation errors. However, GCTA-GREMLd model resulted in overestimation errors, although at small levels, but in all the measurements from the predefined single biallelic-locus analyses (Table 2). GCTA-GREMLd was also prone to SNP heritability underestimations on the real phenotypes as its estimates for additive and non-additive effects were smaller than its own estimates for additive-only effect (Fig. 3j, Extended Data Fig. 3). In summary, SAFE-$h^2$ facilitates phenotype-specific rather than genome or sub-genome specific estimates of SNP heritability with the ability to minimize the likelihood of false-positive contributions. SAFE-$h^2$ further refines the estimations by joint analysis of intra-locus non-additive and additive allelic effects. SAFE-$h^2$ composes, by allele adjustments, genotypic datasets that can also facilitate the linear association tests to address intra-locus allelic effects of additive and non-additive jointly.

## Materials and Methods

### *Datasets*

The simulated homozygote genotypes were arranged in 29 chromosomes with 300kb distance between SNPs to meet linkage equilibrium. For every genotype, reference and alternative alleles were assigned randomly at different chromosomal positions. Finally, VCF files for 1,000 and 10,000 genotypes were assembled with 200k SNPs having MAF from 1% to 10%. In these datasets, one of the 200k SNP positions was designed as non-random to be correlated with the predefined artificial non-random phenotypes. For having multiple predefined associated SNPs, we designed 11, 23, or 37 extra predefined SNPs to be integrated into the same genomic datasets. Further proof-of-concept analyses were accomplished on the human chromosome 22 from HapMap phase 1 genomic data (21). This included 1,092 genotypes and 177,675 SNPs after dataset filtering (MAF ≥1%). As for the simulated genotypes, one simulated non-random SNP position was introduced among the HapMap variants to be associated with the predefined artificial non-random phenotypes. The human chr.22 SNP-set was also extended by simulation to 10,000 individuals. For this, the original SNP dataset was shuffled three times among the chromosomal positions to get a new genotype set. This was repeated to

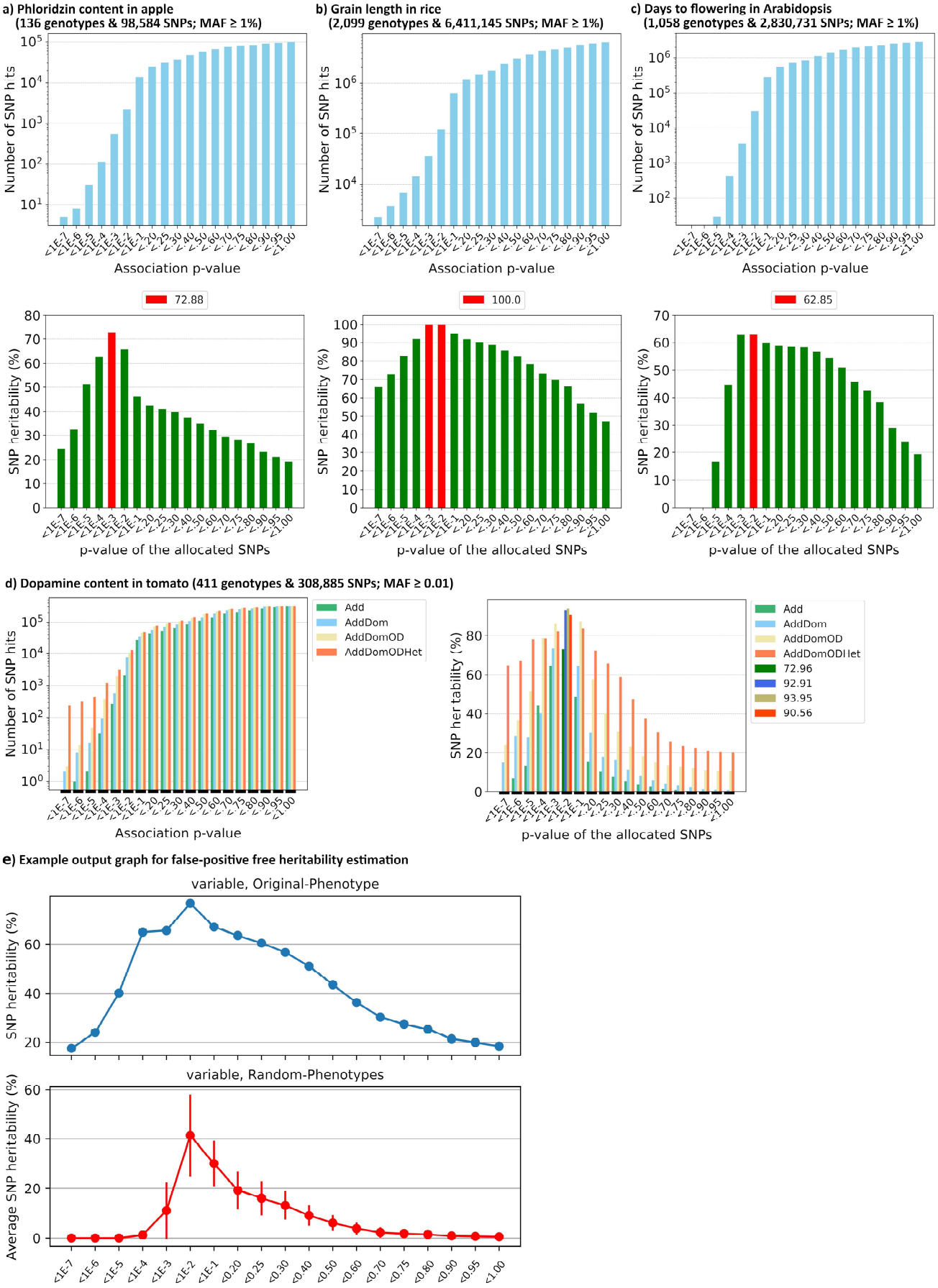

a) Phloridzin content in apple
(136 genotypes & 98,584 SNPs; MAF ≥ 1%)
b) Grain length in rice
(2,099 genotypes & 6,411,145 SNPs; MAF ≥ 1%)
c) Days to flowering in Arabidopsis
(1,058 genotypes & 2,830,731 SNPs; MAF ≥ 1%)
Number of SNP hits
Association p-value
72.88
100.0
62.85
SNP heritability (%)
p-value of the allocated SNPs
d) Dopamine content in tomato (411 genotypes & 308,885 SNPs; MAF ≥ 0.01)
Add
AddDom
AddDomOD
AddDomODHet
72.96
92.91
93.95
90.56
e) Example output graph for false-positive free heritability estimation
variable, Original-Phenotype
variable, Random-Phenotypes
Average SNP heritability (%)

**Figure 4.**

**Fig. 4. SAFE-$h^2$ outputs the bar graphs of both the *p*-value-guided SNP clusters and the SNP heritability profiles.** The SNP heritability estimates are highlighted by red color in (a), (b), (c), and by darker colors in (d). Add: heritability based on additive effects, AddDom: heritability based on additive and dominance effects, AddDomOD: heritability based on additive, dominance and overdominance effects, AddDomODHet: heritability based on additive, dominance, overdominance, and heterosis-like effects, MAF: minor allele frequency. e) Bars represent the standard deviations among six random phenotypes.

have 9 new datasets. The original and new datasets were then horizontally assembled to build a dataset of 10,000 genotypes represented by 177,675 SNPs. The final simulated dataset had 177,445 variants after filtering for MAF of ≥1%. The artificial phenotypes were produced as random or non-random integers assigned randomly among the genotypes or among the genotypic groups of the predefined associated SNPs. Comprehensive analyses were also conducted using *Arabidopsis* and rice public datasets. For *Arabidopsis*, 14 different phenotypes were obtained from the AraPheno database (22). These phenotypes were from a panel of 328 to 1,058 genotypes. The imputed genomic variants for these panels (23, 24) contained 2,621,362 to 2,840,072 post-filtering SNPs. The public rice data was obtained from the rice SNP-Seek database (25). The analyses were performed using 10 phenotypes that were previously recorded in 1,422 to 2,102 rice genotypes. The rice genomic data was also publicly available in rice SNP-Seek database (25–27). The genotypes contained from 6,131,325 to 6,414,800 post-filtering SNPs. Additional datasets from both animal and plant species were analyzed to examine the negative contributions to heritability as well as the incorporation of intra-locus allelic effects of non-additive into the SNP heritability estimates. These included clutch size in 314 wild avians (29) genotyped by 40,766 SNPs (MAF ≥0.01), 203 lean broiler lines versus 272 fat broiler lines (30) genotyped by 47,402 SNPs (MAF ≥0.01), four fruit and flowering related phenotypes in 96 tomato genotypes (31) represented by 16,782 SNPs (MAF ≥0.01), the contents of 14 secondary metabolites in 411 tomato genotypes (37) represented by 308,885 SNPs (MAF ≥0.01), the contents of 7 secondary metabolites in 136 apple genotypes (32) represented by 98,584 SNPs (MAF ≥0.01), apple fruit weight, sugar and dicarboxylic acid contents in 368 genotypes (33) represented by 3,055,912 SNPs (MAF ≥0.01), six growth related phenotypes in 752–755 goat genotypes (34) represented by 40,639 SNPs (MAF ≥0.01), the abundance of 12 microbiome sub-communities in 827 genotypes of foxtail millet represented by 161,562 SNPs (38), and a merged dataset for susceptibility to diabetes mellitus in 1,022 genotypes represented by 826 SNPs (MAF ≥0.01) that derived from two independent studies in cats (35, 36).

## *Data analysis*

Plink (40) (www.cog-genomics.org/plink/1.9/ ) and EMMAX (19) (http://csg.sph.umich.edu//kang/emmax/download/index.html) were used for the initial processing of the genomic data and the association analyses. MAF was set to ≥1% in all the analyses. A complete record was available on the latitudes of the geographical origins for the *Arabidopsis* genotypes when analyzing the stem length and the leaf mineral contents. This geographical information was applied as quantitative covariates in the analyses. SAFE-$h^2$ was developed to perform a dynamic SNP allocation for estimation of narrow-sense heritability. A comprehensive manual is provided together with the software. Briefly, SAFE-$h^2$ starts with the primary scanning of input files and executes the user-defined combination of heritability estimation models on the *p*-value guided clusters of the SNP hits. It determines an experiment and dataset-specific *p*-value threshold; SNPs with *p*-values larger than the threshold are distributed randomly among genotypes in relation to the phenotypic variation and impose negative contributions to heritability estimation levels. We call these SNPs as non-causal. They lead to higher genetic similarities and thereby, genetic variance and SNP heritability underestimations. In fact, this *p*-value threshold is the best criterion to differentiate the causal and non-causal SNPs under accepted experimental conditions, *i.e.*, the genomic and genotypic coverage. SAFE-$h^2$ requires a *p*-value file for SNPs, the Plink b-files ("bim", "bed", and "fam"), and optional covariate files specific for every estimation model. SAFE-$h^2$ exploits the four different models of EMMAX (19), LDAK-GCTA, LDAK-Thin (8), and GCTA-GREML (20) to estimate the SNP heritability. GCTA-GREML inbred algorithm can also be applied for inbred genotypes or for populations with considerable levels of inbreeding. To operate, SAFE-$h^2$ requires the binary files of the mentioned tools (see the SAFE-$h^2$ manual with full description). In return, SAFE-$h^2$ provides the output data files and related graphs for both the clustered SNP hits and the SNP heritability profiles. It also features the *p*-value threshold and the estimated SNP heritability at that threshold on the graphs. Depending on experiments, different levels of false-positive rates are always expected in GWAS. To bypass this challenge, SAFE-$h^2$ uses an innovative approach to exclude

the false-positive SNP hits. SAFE-$h^2$ determines a safe *p*-value interval with its upper-end to be used as another *p*-value threshold for heritability estimations that have minimum contributions from false-positive SNP hits. In this process, the user simulates six independent random phenotypes within the range of the original phenotype-of-interest for all the available genotypes in the dataset. To do the analysis, SAFE-$h^2$ requires a "bim" file, a "bed" file, fam files for the real (one "fam" file) and random phenotypes (six "fam" files), SNP *p*-value files for the real (one file) and random phenotypes (six files), and optional covariate files specific for estimation models. By using these input files, SAFE-$h^2$ creates the SNP heritability profiles for all the generated random phenotypes as well as the real phenotype-of-interest using an identical genomic dataset. There is also a possibility to choose the estimation models of interest. Analyzing random phenotypes subsequently determines a safe *p*-value interval having minimum heritability level at the lower-tail of heritability profile. The upper-end of this interval can act as a *p*-value threshold to omit the false-positive SNP hits from the estimation models and keep their contribution to heritability at its minimum. Finally, SAFE-$h^2$ generates an aligned graph of heritability profiles for random and original phenotypes. The SNP heritability of the original phenotype can accordingly be defined at upper-end of the determined *p*-value interval to keep the false-positive contributions at its minimum level.

SAFE-$h^2$ also offers allelic adjustments for linear approximation on intra-locus allelic effects including dominance, overdominance, and heterosis-like effects. This facilitates capturing of non-additive allelic effects. SAFE-$h^2$-preADOH builds two VCF files for the dominance adjustments by converting the heterozygote nucleotide positions into either of the homozygotes. SAFE-$h^2$-preADOH also builds two VCF files for the overdominance adjustments by converting heterozygote nucleotide positions into either of homozygotes and, homozygotes into their heterozygote counterparts. For the heterosis-like effects, SAFE-$h^2$-preADOH builds one more VCF file by converting one of homozygotes into the other homozygote in every nucleotide position. Here, the user can perform an independent association test or alternatively can command SAFE-$h^2$-preADOH to execute the Plink's glm function on the five adjusted and one original genotype datasets. To construct the combinatory genotype datasets, SAFE-$h^2$-preADOH applies the association *p*-values to select the best linear-fit (with smallest association *p*-value) among the original and the five adjusted versions for every SNP position. At the end, SAFE-$h^2$-preADOH provides three genotypic datasets that can be used to estimate the additive-dominance effects, additive-dominance-overdominace effects, and additive-dominance-overdominace-heterosis effects. Here, every SNP-ID gets a specific tag of [Add] for having additive effects, [Dom] or [dom] for having dominance effects, [OD] or [od] for having overdominance effects, or Het for having heterosis-like effects on the phenotype of interest. Finally, SAFE-$h^2$-ADOH performs the heritability profiling as explained for SAFE-$h^2$ and on the three adjusted genotype datasets as well as the original genotype dataset.

## Supplementary Information

**Extended Data Fig. 1: The SNP heritability profile of random quantitative phenotypes computed on the simulated genomic variants.** Initial GWAS was performed using EMMAX to determine the association levels for the SNPs. The four different approaches of EMMAX, LDAK-GCTA, LDAK-Thin, and GCTA-GREML (by "inbred" algorithm) were then applied to estimate the SNP heritability for the clusters of SNP hits captured using the association *p*-values. N represents the sample sizes, *i.e.*, number of genotypes. MAF stands for the minor allele frequency. The average SNP heritability estimates were obtained for 10 independent random phenotypes within identical ranges. The error bars represent 99% confidence intervals.

**Extended Data Fig. 2: The SNP heritability profiles of artificial quantitative phenotypes were computed on simulated genomic variants including multiple SNPs being associated with the artificial phenotypes at different rates.** Initial GWAS was performed using EMMAX to determine the association levels for the SNPs. The four different approaches of EMMAX, LDAK-GCTA, LDAK-Thin, and GCTA-GREML (by "inbred" algorithm) were then applied to estimate the SNP heritability for the clusters of SNP hits captured using the association *p*-values. N represents the sample sizes, *i.e.*, number of genotypes. MAF stands for the minor allele frequency. Ensuring additive-only effects, the non-random phenotypes were defined among the genotypic groups of the introduced non-random SNP positions, *i.e.*, Ref*i*/Ref*i* and Alt*i*/Alt*i* for the simulated inbred genotypes. The intra-genotype assignment of the phenotypes, while being within their specific ranges, was random. The inter-genotype variances of the phenotypes are shown as $\sigma_g^2$. $\sigma_e^2$ represents the intra-genotype variations.

**Extended Data Fig. 3: SAFE-$h^2$ estimates of SNP heritability under additive-only model, combined additive and non-additive models, and GREMLd model for different phenotypes in plant species.** The analyses of tomato and foxtail millet genotypes using GCTA-GREML model were performed based on the "inbred" algorithm of GCTA. The average inbreeding coefficients (Fhat) of tomato and foxtail millet genotypes were > 0.8. Add: heritability based on additive effects, AddDom: heritability based on additive and dominance effects, AddDomOD: heritability based on additive, dominance and overdominance effects, AddDomODHet: heritability based on additive, dominance, overdominance, and heterosis-like effects, GREMLd: heritability estimated by GREMLd model, MAF: minor allele frequency, OUT: operational taxonomic unit.

**Extended Data Table 1: Heritability profiling by SAFE-$h^2$ reveals negative contributions by SNPs with large association *p*-values.**

**Code availability:** SAFE-$h^2$ is available and attached to this publication as suppl. File 1. It is also available at figshare (DOI: 10.6084/m9.figshare.24025728, Link: https://figshare.com/s/7193d332894f11deb1a5). It contains the compiled SAFE-$h^2$ scripts with related installation and usage manual as well as example datasets.

**Author contributions:** B.D conceived the study. B.D performed the analyses and compiled the SAFE-$h^2$ script. B.D and M.N wrote the manuscript. B.D and M.N supervised the study.

**Competing interests:** The authors declare no competing interests.

## References


1. A. Tenesa, C. S. Haley, The heritability of human disease: estimation, uses and abuses. *Nat Rev Genet* **14**, 139–149 (2013).

2. O. Zuk, *et al.*, Searching for missing heritability: Designing rare variant association studies. *Proceedings of the National Academy of Sciences* **111**, E455–E464 (2014).

3. P. M. Visscher, W. G. Hill, N. R. Wray, Heritability in the genomics era — concepts and misconceptions. *Nat Rev Genet* **9**, 255–266 (2008).

4. M. N. Weedon, *et al.*, Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* **40**, 575–583 (2008).

5. G. Lettre, *et al.*, Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* **40**, 584–591 (2008).

6. D. F. Gudbjartsson, *et al.*, Many sequence variants affecting diversity of adult human height. *Nat Genet* **40**, 609–615 (2008).

7. J. Yang, *et al.*, Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565–569 (2010).

8. D. Speed, G. Hemani, M. R. Johnson, D. J. Balding, Improved Heritability Estimation from Genome-wide SNPs. *Am J Hum Genet* **91**, 1011–1021 (2012).

9. M. Tang, T. Wang, X. Zhang, A review of SNP heritability estimation methods. *Briefings in Bioinformatics* **23**, bbac067 (2022).

10. J. Yang, J. Zeng, M. E. Goddard, N. R. Wray, P. M. Visscher, Concepts, estimation and interpretation of SNP-based heritability. *Nat Genet* **49**, 1304–1310 (2017).

11. L. M. Evans, *et al.*, Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat Genet* **50**, 737–745 (2018).

12. S. H. Lee, *et al.*, Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* **44**, 247–250 (2012).

13. E. A. Stahl, *et al.*, Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* **51**, 793–803 (2019).

14. D. Speed, N. Cai, M. R. Johnson, S. Nejentsev, D. J. Balding, Reevaluation of SNP heritability in complex human traits. *Nat Genet* **49**, 986–992 (2017).

15. T. A. Manolio, *et al.*, Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

16. D. Speed, A. Kaphle, D. J. Balding, SNP-based heritability and selection analyses: Improved models and new results. *BioEssays* **44**, 2100170 (2022).

17. B. M. L. Baselmans, L. Yengo, W. van Rheenen, N. R. Wray, Risk in Relatives, Heritability, SNP-Based Heritability, and Genetic Correlations in Psychiatric Disorders: A Review. *Biological Psychiatry* **89**, 11–19 (2021).

18. C.-J. S. Barry, *et al.*, How to estimate heritability: a guide for genetic epidemiologists. *International Journal of Epidemiology* **52**, 624–632 (2023).

19. H. M. Kang, *et al.*, Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* **42**, 348–354 (2010).

20. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics* **88**, 76–82 (2011).

21. G. A. McVean, *et al.*, An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).

22. Ü. Seren, *et al.*, AraPheno: a public database for Arabidopsis thaliana phenotypes. *Nucleic Acids Research* **45**, D1054–D1059 (2017).

23. C. Alonso-Blanco, *et al.*, 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell* **166**, 481–491 (2016).

24. B. Arouisse, A. Korte, F. van Eeuwijk, W. Kruijer, Imputation of 3 million SNPs in the Arabidopsis regional mapping population. *The Plant Journal* **102**, 872–882 (2020).

25. L. Mansueto, *et al.*, Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic Acids Research* **45**, D1075–D1081 (2017).

26. S. R. McCouch, *et al.*, Open access resources for genome-wide association mapping in rice. *Nat Commun* **7**, 10532 (2016).

27. The 3,000 rice genomes project. *GigaScience* **3**, 2047-217X-3–7 (2014).

28. Z. Zhu, *et al.*, Dominance Genetic Variation Contributes Little to the Missing Heritability for Human Complex Traits. *The American Journal of Human Genetics* **96**, 377–385 (2015).

29. A. Husby, *et al.*, Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life-history trait. *Proc Biol Sci* **282**, 20150156 (2015).

30. D. Dou, *et al.*, Genome-wide association studies for growth traits in broilers. *BMC Genomic Data* **23**, 1 (2022).

31. E. Yamamoto, *et al.*, A simulation-based breeding design that uses whole-genome prediction in tomato. *Sci Rep* **6**, 19454 (2016).

32. K. A. McClure, *et al.*, Genome-wide association studies in apple reveal loci of large effect controlling apple polyphenols. *Hortic Res* **6**, 1–12 (2019).

33. L. Liao, *et al.*, Unraveling a genetic roadmap for improved taste in the domesticated apple. *Molecular Plant* **14**, 1454–1471 (2021).

34. M. Moaeen-ud-Din, R. Danish Muner, M. S. Khan, Genome wide association study identifies novel candidate genes for growth and body conformation traits in goats. *Sci Rep* **12**, 9891 (2022).

35. Y. Forcada, M. Boursnell, B. Catchpole, D. B. Church, A genome-wide association study identifies novel candidate genes for susceptibility to diabetes mellitus in non-obese cats. *PLOS ONE* **16**, e0259939 (2021).

36. I. Hernandez, *et al.*, Complex Feline Disease Mapping Using a Dense Genotyping Array. *Frontiers in Veterinary Science* **9** (2022).

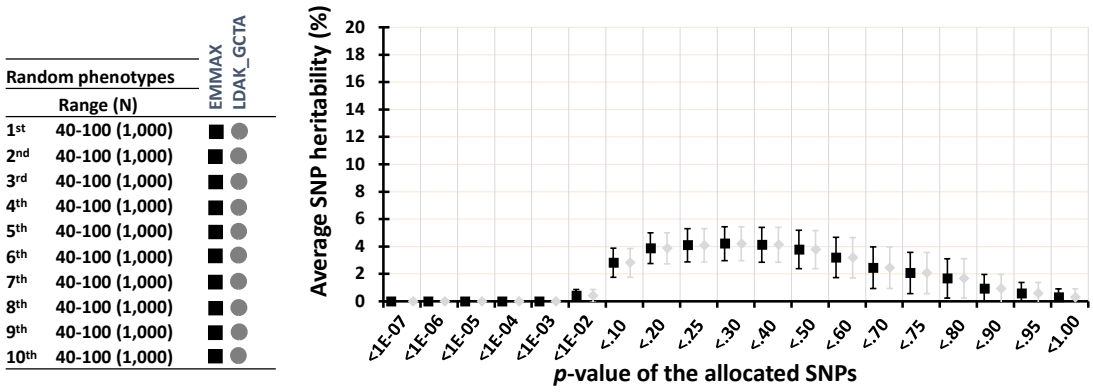
37. G. Zhu, *et al.*, Rewiring of the Fruit Metabolome in Tomato Breeding. *Cell* **172**, 249-261.e12 (2018).

38. Y. Wang, *et al.*, GWAS, MWAS and mGWAS provide insights into precision agriculture based on genotype-dependent microbial effects in foxtail millet. *Nat Commun* **13**, 5913 (2022).

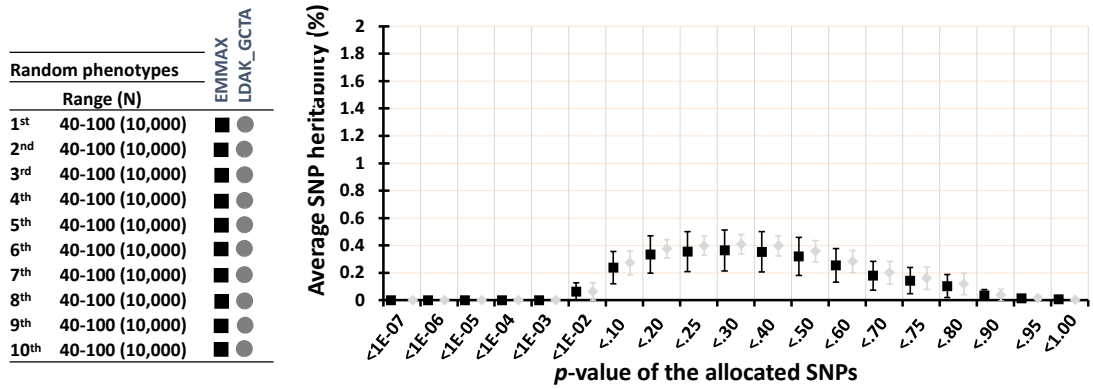
39. L. Yengo, *et al.*, A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).

40. C. C. Chang, *et al.*, Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742-015-0047–8 (2015).

**a) Analysis of 10 artificial random phenotypes on a panel of 1,000 simulated inbred genotypes (100 SNPs; MAF = 1.6%–10%)**




**b) Analysis of 10 artificial random phenotypes on a panel of 10,000 simulated inbred genotypes (100 SNPs; MAF = 1.6%–10%)**

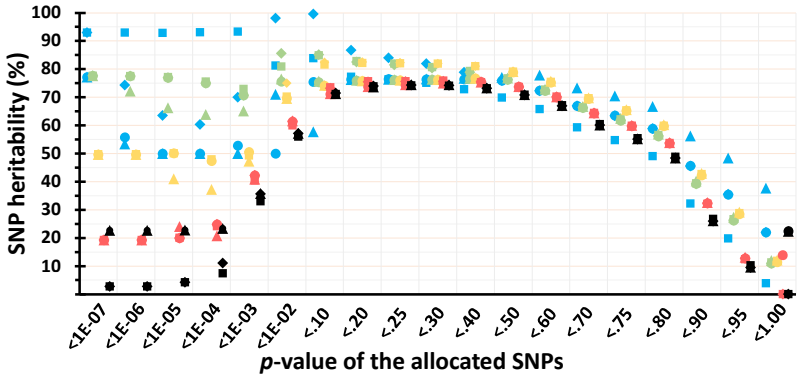



**Extended Data Fig. 1: The SNP heritability profile of random quantitative phenotypes computed on the simulated genomic variants.** Initial GWAS was performed using EMMAX to determine the association levels for the SNPs. The four different approaches of EMMAX, LDAK-GCTA, LDAK-Thin, and GCTA-GREML (by "inbred" algorithm) were then applied to estimate the SNP heritability for the clusters of SNP hits captured using the association *p*-values. N represents the sample sizes, *i.e.*, number of genotypes. MAF stands for the minor allele frequency. The average SNP heritability estimates were obtained for 10 independent random phenotypes within identical ranges. The error bars represent 99% confidence intervals.

**a) Analysis of 5 artificial non-random phenotypes on a panel of 1,000 simulated inbred genotypes (200,011 SNPs; MAF = 1.6%—10%)**

| Phenotypes | The introduced non-random SNP positions | | | | |
|---|---|---|---|---|---|
| | Ref*i*/Ref*i* Genotype | Alt*i*/Alt*i* Genotype | | | EMMAX / LDAK_GCTA / LDAK_Thin / GCTA_GREML |
| | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | |
| 1st | 40-50 (900) | 80-100 (100) | 6.995 | 93.005 | |
| 2nd | 40-60 (900) | 75-100 (100) | 22.389 | 77.611 | |
| 3rd | 40-70 (900) | 70-100 (100) | 50.267 | 49.733 | |
| 4th | 40-80 (900) | 60-100 (100) | 80.690 | 19.310 | |
| 5th | 40-90 (900) | 50-100 (100) | 97.170 | 2.830 | |

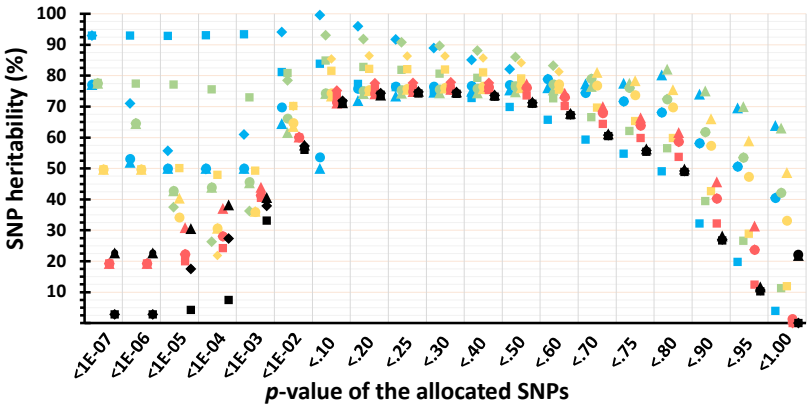
*i*: 1st-12th associated SNP.




**b) Analysis of 5 artificial non-random phenotypes on a panel of 1,000 simulated inbred genotypes (200,023 SNPs; MAF = 1.6%—10%)**

| Phenotypes | The introduced non-random SNP positions | | | | |
|---|---|---|---|---|---|
| | Ref*i*/Ref*i* Genotype | Alt*i*/Alt*i* Genotype | | | EMMAX / LDAK_GCTA / LDAK_Thin / GCTA_GREML |
| | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | |
| 1st | 40-50 (900) | 80-100 (100) | 6.995 | 93.005 | |
| 2nd | 40-60 (900) | 75-100 (100) | 22.389 | 77.611 | |
| 3rd | 40-70 (900) | 70-100 (100) | 50.267 | 49.733 | |
| 4th | 40-80 (900) | 60-100 (100) | 80.690 | 19.310 | |
| 5th | 40-90 (900) | 50-100 (100) | 97.170 | 2.830 | |

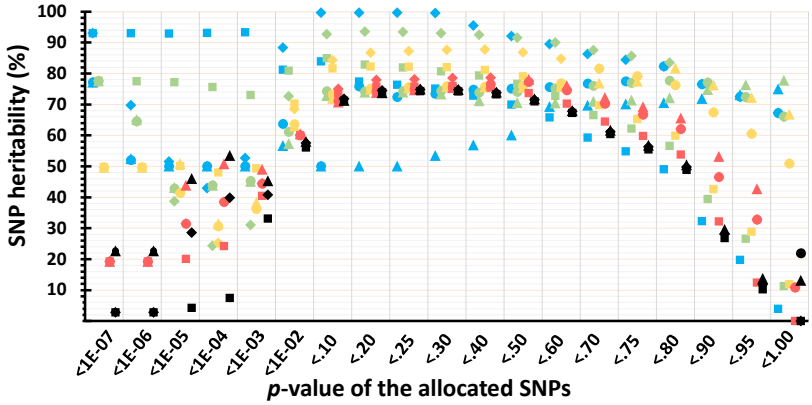
*i*: 1st-24th associated SNP.



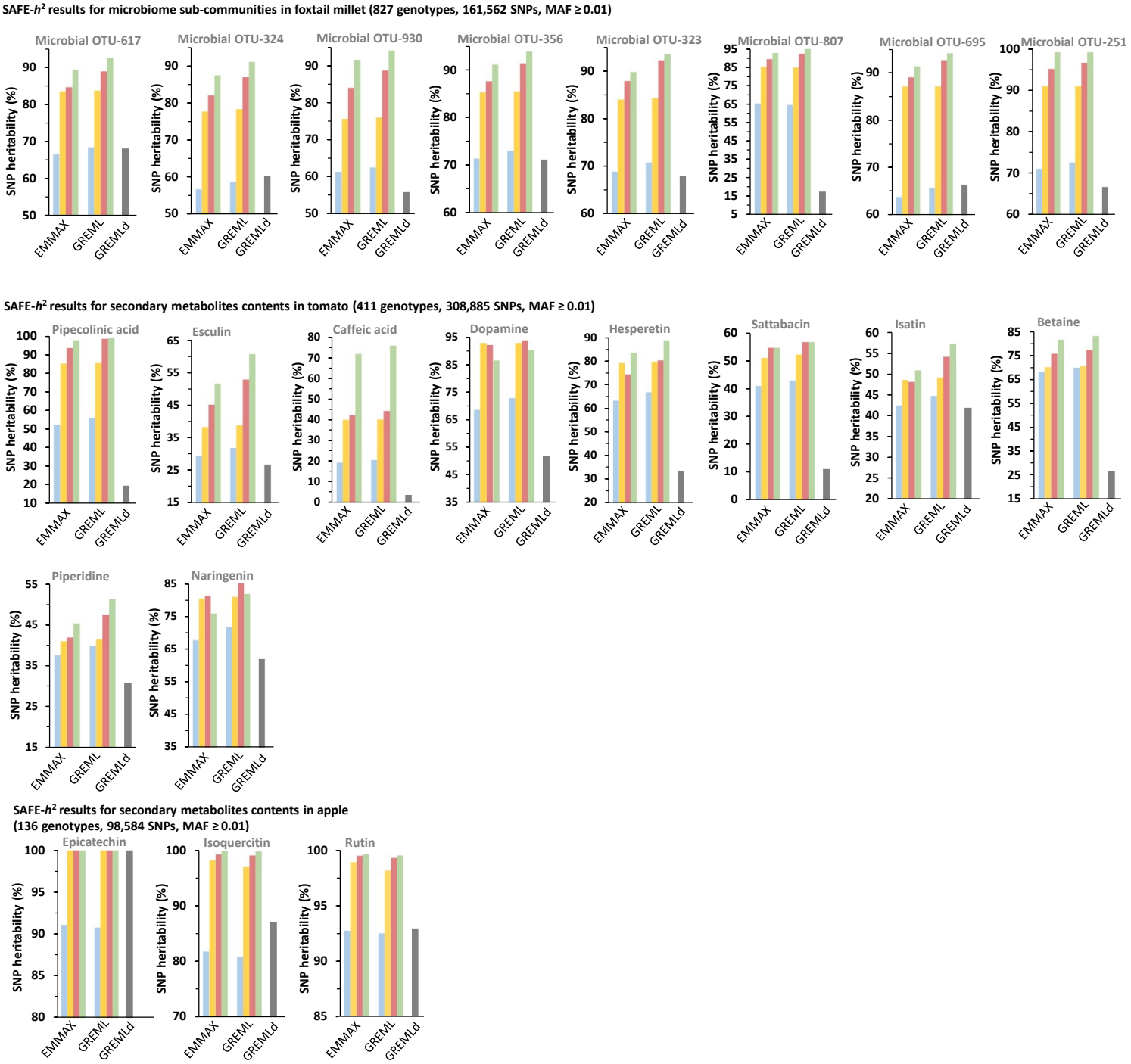

**c) Analysis of 5 artificial non-random phenotypes on a panel of 1,000 simulated inbred genotypes (200,047 SNPs; MAF = 1.6%—10%)**

| Phenotypes | The introduced non-random SNP positions | | | | |
|---|---|---|---|---|---|
| | Ref*i*/Ref*i* Genotype | Alt*i*/Alt*i* Genotype | | | EMMAX / LDAK_GCTA / LDAK_Thin / GCTA_GREML |
| | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | |
| 1st | 40-50 (900) | 80-100 (100) | 6.995 | 93.005 | |
| 2nd | 40-60 (900) | 75-100 (100) | 22.389 | 77.611 | |
| 3rd | 40-70 (900) | 70-100 (100) | 50.267 | 49.733 | |
| 4th | 40-80 (900) | 60-100 (100) | 80.690 | 19.310 | |
| 5th | 40-90 (900) | 50-100 (100) | 97.170 | 2.830 | |

*i*: 1st-48th associated SNP.




**Extended Data Fig. 2: The SNP heritability profiles of artificial quantitative phenotypes were computed on simulated genomic variants including multiple SNPs being associated with the artificial phenotypes at different rates.** Initial GWAS was performed using EMMAX to determine the association levels for the SNPs. The four different approaches of EMMAX, LDAK-GCTA, LDAK-Thin, and GCTA-GREML (by "inbred" algorithm) were then applied to estimate the SNP heritability for the clusters of SNP hits captured using the association *p*-values. N represents the sample sizes, *i.e.*, number of genotypes. MAF stands for the minor allele frequency. Ensuring additive-only effects, the non-random phenotypes were defined among the genotypic groups of the introduced non-random SNP positions, *i.e.*, Ref*i*/Ref*i* and Alt*i*/Alt*i* for the simulated inbred genotypes. The intra-genotype assignment of the phenotypes, while being within their specific ranges, was random. The inter-genotype variances of the phenotypes are shown as $\sigma_g^2$. $\sigma_e^2$ represents the intra-genotype variations.




**Extended Data Fig. 3: SAFE-$h^2$ estimates of SNP heritability under additive-only model, combined additive and non-additive models, and GREMLd model for different phenotypes in plant species.** The analyses of tomato and foxtail millet genotypes using GCTA-GREML model were performed based on the "inbred" algorithm of GCTA. The average inbreeding coefficients (Fhat) of tomato and foxtail millet genotypes were > 0.8. Add: heritability based on additive effects, AddDom: heritability based on additive and dominance effects, AddDomOD: heritability based on additive, dominance and overdominance effects, AddDomODHet: heritability based on additive, dominance, overdominance, and heterosis-like effects, GREMLd: heritability estimated by GREMLd model, MAF: minor allele frequency, OUT: operational taxonomic unit.

**Extended Data Table 1. Heritability profiling by SAFE-$h^2$ reveals negative contributions by SNPs with large association *p*-values**

| | Phenotypes | Number of Genotypes/Variants | [SAFE-$h^2_{SNP}$] - [$h^2_{SNP}$ based on all SNPs] EMMAX | GCTA-GREML[a] |
|---|---|---|---|---|
| **Tomato** | Cracked fruit | 96 / 16,782 | 3.95% | 2.37% |
| | Day-to-flowering | 96 / 16,782 | 0.00% | 0.00% |
| | Small fruit | 96 / 16,782 | 17.96% | 18.66% |
| | Iregular shape | 96 / 16,782 | 2.71% | 2.52% |
| | Caffeic acid | 411 / 308,885 | 12.16% | 12.80% |
| | Catechin | 411 / 308,885 | 22.25% | 21.92% |
| | Dopamine | 411 / 308,885 | 68.64% | 72.95% |
| | Esculin | 411 / 308,885 | 19.53% | 21.04% |
| | Hesperetin | 411 / 308,885 | 61.11% | 66.61% |
| | Naringenin | 411 / 308,885 | 51.78% | 54.29% |
| | Piperidine | 411 / 308,885 | 8.53% | 8.64% |
| | Viscumamide | 411 / 308,885 | 2.83% | 2.82% |
| | Betaine | 411 / 308,885 | 49.64% | 50.07% |
| | Isatin | 411 / 308,885 | 1.42% | 1.28% |
| | Pipecolinic acid | 411 / 308,885 | 52.12% | 56.12% |
| | Riboflavin | 411 / 308,885 | 48.99% | 50.73% |
| | Sattabacin | 411 / 308,885 | 26.69% | 27.50% |
| | Quercetin | 411 / 308,885 | 62.37% | 67.00% |
| **Apple** | Quercetin | 136 / 98,584 | 30.78% | 32.19% |
| | Catechin | 136 / 98,584 | 13.03% | 11.99% |
| | Avicularin | 136 / 98,584 | 72.53% | 71.99% |
| | Phloridzin | 136 / 98,584 | 53.78% | 52.99% |
| | Rutin | 136 / 98,584 | 90.21% | 90.07% |
| | Isoquercetin | 136 / 98,584 | 46.78% | 46.85% |
| | Epicatechin | 136 / 98,584 | 60.93% | 61.52% |
| | Fruit-weight | 368 / 3,055,912 | 2.10% | 2.21% |
| | Sucrose | 361 / 2,927,501 | 20.20% | 20.60% |
| | Fructose | 448 / 2,682,418 | 15.52% | 15.95% |
| | Tartrate | 461 / 2,767,199 | 19.39% | 10.73% |
| **Foxtail millet** | OUT-327 | 827 / 161,562 | 29.10% | 28.84% |
| | OUT-403 | 827 / 161,562 | 44.80% | 44.31% |
| | OUT-487 | 827 / 161,562 | 15.17% | 14.61% |
| | OUT-564 | 827 / 161,562 | 19.23% | 19.09% |
| | OUT-323 | 827 / 161,562 | 49.12% | 49.73% |
| | OUT-356 | 827 / 161,562 | 41.99% | 41.79% |
| | OUT-930 | 827 / 161,562 | 35.71% | 35.13% |
| | OUT-324 | 827 / 161,562 | 29.64% | 29.98% |
| | OUT-617 | 827 / 161,562 | 26.30% | 25.78% |
| | OUT-807 | 827 / 161,562 | 53.93% | 52.17% |
| | OUT-695 | 827 / 161,562 | 39.36% | 39.63% |
| | OUT-251 | 827 / 161,562 | 49.04% | 49.31% |
| **Cats** | Dibetes mellitus | 1,022 / 826 | 12.07% | 11.39% |
| **Goats** | Height | 752 / 40,639 | 2.32% | 2.37% |
| | Body-length | 752 / 40,639 | 1.76% | 1.85% |
| | Chest length | 752 / 40,639 | 0.23% | 0.23% |
| | Heart-girth | 752 / 40,639 | 2.60% | 2.75% |
| | Pubic bone-length | 752 / 40,639 | 0.00% | 0.03% |
| | Weight | 752 / 40,639 | 4.86% | 5.03% |
| **Wild Avians** | Clutch-size | 314 / 40,766 | 52.92% | 57.16% |
| **Broiler lines** | Lean vs Fat | 475 / 47,402 | 0.00% | --- |
| **Average** | | | 29.27% | 29.91% |

[a]The "inbred" algorithm of GCTA was used to analyze the tomato and foxtail millet genotypes. The average inbreeding coefficients (Fhat) for these gentypes were > 0.8.