nature genetics

TECHNICAL REPORT

https://doi.org/10.1038/s41588-018-0225-6

Identifying loci affecting trait variability and detecting interactions in genome-wide association studies

Alexander I. Young[1,2*], Fabian L. Wauthier[1,3] and Peter Donnelly[1,3*]

Identification of genetic variants with effects on trait variability can provide insights into the biological mechanisms that control variation and can identify potential interactions. We propose a two-degree-of-freedom test for jointly testing mean and variance effects to identify such variants. We implement the test in a linear mixed model, for which we provide an efficient algorithm and software. To focus on biologically interesting settings, we develop a test for dispersion effects, that is, variance effects not driven solely by mean effects when the trait distribution is non-normal. We apply our approach to body mass index in the subsample of the UK Biobank population with British ancestry (n ~408,000) and show that our approach can increase the power to detect associated loci. We identify and replicate novel associations with significant variance effects that cannot be explained by the non-normality of body mass index, and we provide suggestive evidence for a connection between leptin levels and body mass index variability.

Nearly all genome-wide association study (GWAS) associations have been discovered by testing the simplest additive model[1]. A long-standing controversy exists about the importance of departures from the additive model in human genetics[2]. In addition, the extent and nature of interactions between genetic variants and environmental factors remains poorly characterized. The advent of large population-based cohorts, such as the UK Biobank[3], provides large samples of genotyped individuals along with rich lifestyle and environment information. With the right statistical methodology, this will accelerate characterization of the interaction between genes and environment.

The problem of searching for interaction effects is harder than for additive effects, in part because the number of possible interaction models grows superlinearly with the number of possibly interacting variables[4]. The variance of a quantitative phenotype differs with the genotype of loci involved in interactions[5,6]. Therefore, one can reduce the search space of possible interaction models by screening genome-wide loci for variance effects.

Interactions involving genetic variants are one cause of a more general phenomenon: genetic effects on phenotypic variability[7–9]. Control of phenotypic variability, both within and between individual organisms, is a fundamental property of biological systems[10]. It is likely that organisms have evolved to suppress certain kinds of variability resulting from developmental processes and environmental stimuli[7]. The understanding of genetic effects on phenotypic variability is poor compared to the understanding of genetic effects on mean trait values[7], in part because methods for investigating effects on variability are not as well developed.

Most published methods for detecting loci affecting phenotypic variability concentrate on testing for a variance effect alone[11,12], even though such loci are also likely to affect the mean of the phenotype. Methods that jointly test for mean and variance effects[13,14] are not as well developed as those for additive association testing, and their implementation in a linear mixed-model framework relies on algorithms that scale cubically with sample size[14,15], making them impractical to apply to the large sample sizes needed to detect variance effects on complex human traits.

Here we introduce a two-degree-of-freedom test for jointly testing mean and variance effects on quantitative traits. If the trait distribution is non-normal, then additive effects at a locus will induce variance effects that are unlikely to be of interest. We show how to account for this and thereby detect variance effects not driven by mean effects. We implement the test in a mixed-model framework and develop an algorithm for fitting this model, whose computations scale linearly with sample size. To illustrate application of the method, we analyze body mass index (BMI) in the subsample population of the UK Biobank with predominantly British ancestry (n ~408,000).

Results

Variance models. We introduce a variance effect that is analogous to the additive effect on the mean. Because the variance is always positive, we cannot use a linear model, which is unbounded; instead, we use a log-linear model. Let $\sigma_g^2 = \mathrm{Var}(Y \mid G = g)$ be the variance conditional on the allele count at a single nucleotide polymorphism (SNP), G being $g \in \{0, 1, 2\}$. This model has the form $\log(\sigma_g^2) = \mu_v + \alpha_v g$ for some constant μ_v and a log-linear variance effect α_v. We note that when variance effects are small, log-linear variance effects should be approximately equivalent to the linear variance effects modeled in a previous meta-analysis of variance effects[12]. We fit this in a model that also allows for an additive effect on the mean:

$$\mathrm{M_{AV}}: \; Y|G = g \sim N(\mu + \alpha g, \; \exp(\mu_v + \alpha_v g))$$

which we call the additive variance (AV) model.

To model more complicated relationships between genotype and phenotypic variance, we introduce a general variance effect, δ_v, which models nonlinear changes in log-variance with genotype as

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [2]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. [3]Department of Statistics, University of Oxford, Oxford, UK. *e-mail: alexander.young@bdi.ox.ac.uk; donnelly@well.ox.ac.uk

$$\log(\sigma_g^2) = \mu_v + \alpha_v g + \delta_v (g-2f)^2$$

where f is the allele frequency. In the Supplementary Note, we show how these variance models fit into a hierarchy of models for the effect of genotype on phenotype; we also show the connection between these models and the mutual information between genotype and phenotype.

Variance tests. There is a natural one-degree-of-freedom test for a log-linear variance effect, as well as a general two-degree-of-freedom test for both a log-linear and general variance effect. Cao et al.[13] previously suggested using the two-degree-of-freedom variance test, which is very similar to the Bartlett test.

To assess the power of the two approaches, we simulated traits affected by a genetic variant (frequency = 0.5) with both log-linear and general variance effects (Methods). For the sample size considered (n = 100,000), the one-degree-of-freedom test was more powerful unless the variance of the heterozygote deviated by ≥ 2% from a log-linear variance model (Supplementary Table 1). We also show in the Supplementary Note that the functional form of the effect on the log-variance for an interacting locus is linear up to a correction on the order of $O(\alpha_v^2)$.

Therefore, we propose a test to discover loci with mean and variance effects that compares the likelihood of the AV model, M_{AV}, to the likelihood of the null model. We call this the AV test (Fig. 1). Because of the reduced degrees of freedom, this test should have improved power over the test proposed by Cao et al.[13] unless the variance of the heterozygote deviates substantially from a log-linear model.

We note that dominance (a nonlinear relationship between phenotypic means and allele counts) could potentially induce spurious variance effect estimates from a model that assumes a linear relationship between allele counts and phenotypic means. We recommend checking for this for variants displaying evidence of a variance effect under the AV model by fitting a model that includes a dominance effect (Supplementary Note).

Non-normality and variance effects. So far, models and test statistics have been derived under an assumption of normality. When the phenotypic distribution is non-normal, the null distribution of the variance test statistics is improperly calibrated[16]. Inverse normal transformation of the phenotype can be used to ensure that the null distribution of the variance test statistic is properly calibrated. However, there is a relationship between the mean and the variance of any non-normal distribution. This implies that, when trait values follow a non-normal distribution, any genetic variant that affects the mean of the trait will also affect the variance of the trait. This phenomenon can be seen in genome-wide estimated mean and variance effects on BMI taken from summary statistics provided by the GIANT (Genetic Investigation of ANthropometric Traits) consortium[12,17] (see URLs; Supplementary Fig. 3). This kind of variance effect is unlikely to be of direct interest and is not indicative of an interaction effect.

Variance effects driven by a mean–variance relationship can, in theory, be removed by a variance-stabilizing transformation if the mean–variance relationship is known. However, variance effects due to a mean–variance relationship are not, in general, removed by inverse normal transformation.

We introduce 'dispersion effects', which are effects on phenotypic variance independent of the general mean–variance relationship of the phenotypic distribution, implying that they are insensitive to phenotypic transformation (Methods and Supplementary Note). We show that, when additive effects are small, there is an approximately linear relationship between additive effect (α_l), log-linear variance effect (α_{vl}), and dispersion effect (d_l) at each locus l (Supplementary Note). Thus, $\alpha_{vl} \approx r_{av}\alpha_l + d_l$, where r_{av} parameterizes the expected log-linear variance effect due to the general mean–variance relationship and can be estimated from genome-wide regression of log-linear variance effect estimates on additive effect estimates (Methods). Given r_{av}, the dispersion effects are estimated by $\hat{d}_l = \hat{\alpha}_{vl} - r_{av}\hat{\alpha}_l$.

The heteroskedastic linear mixed model (HLMM). Linear mixed models can be used to model the effects of many genome-wide SNPs on the mean of a trait[18,19]. The effects are modeled as random effects, typically drawn from a normal distribution. Depending on how the SNPs are selected, linear mixed models can increase power and/or control for population structure and relatedness in GWAS[19]. The HLMM is a generalization of the standard mixed model that also models the effect of a set of variables on the residual variance of a trait. While such models have been proposed before[15], existing algorithms for them scale with the cube of the sample size[15], making them impractical for large samples. We give an algorithm for fitting the model that scales linearly when the number of random effects is fixed (Methods).

While the HLMM offers a potential gain in both power and robustness, it comes at a computational cost. In some settings, it will be appropriate to simply fit the heteroskedastic linear model, which is very fast. We provide software for fitting both the mixed and linear versions of the heteroskedastic model (see Code availability).

Simulation of inference on a non-normal phenotype. To demonstrate inference on a non-normal phenotype, we simulated 10 non-normal phenotypes (n = 150,000) using a gamma model (Methods). In the gamma model, every SNP with an additive effect on the untransformed phenotype also has a linear effect on phenotypic variance. For each phenotype, out of 22,000 simulated SNPs with varying frequencies, 1,000 were chosen to have additive effects on the phenotype, with 10% of the phenotypic variance explained by the combined additive effects. One of the 1,000 SNPs with additive effects was chosen to have a log-linear dispersion effect of 0.05 as well as an additive effect in the same direction as the dispersion effect. An inverse normal transformation was performed on the phenotypes before analysis.

Inference of additive and AV models was performed by maximum likelihood (Supplementary Note); dispersion effects were inferred (Methods). The Kolmogorov–Smirnov test was used to detect deviations from theoretical null distributions. Test statistics were combined across the ten independent phenotypes for the Kolmogorov–Smirnov tests. For the null hypothesis SNPs, we found no evidence for deviation from the theoretical null distributions for any of the test statistics examined: additive (n = 210,000, D = 0.0015, P = 0.23); log-linear variance (n = 210,000, D = 0.0017, P = 0.58); AV (n = 210,000, $D = 8\times10^{-4}$, P = 1.00); and dispersion (n = 210,000, D = 0.0019, P = 0.60).

For the 999 SNPs with additive effects on the non-normal phenotype, but no dispersion effect, the log-linear variance test statistic showed strong evidence for deviation from the null hypothesis (n = 9,990, D = 0.11, $P < 2.2\times10^{-16}$), with a mean log-likelihood ratio 56% larger than would be expected under the null hypothesis. However, there was no evidence that the dispersion effect test statistic deviated from the null hypothesis (n = 9,990, D = 0.0078, P = 0.58). In contrast, for the locus with a dispersion effect, the average P value for a dispersion effect was 6.8×10^{-11}. This demonstrates that our method can powerfully distinguish between loci that show variance effects because of a general mean–variance relationship and loci that have dispersion effects.

To simulate inference with the HLMM, we modeled random effects for the top 1,000 SNPs ranked according to the additive test statistics for each simulation. For the null hypothesis SNPs, we found no evidence for deviation from the theoretical null distributions for the additive test statistic (n = 210,000, D = 0.0019,

$P = 0.43$). However, we found evidence for slight deflation of the log-linear variance (mean = 0.984, $P = 3.0 \times 10^{-5}$ from the Kolmogorov–Smirnov test) and dispersion test statistic (mean = 0.990, $P = 0.008$). While there may be a very slight deflation of the variance test statistic for the null hypothesis SNPs, this does not imply there is reduced power for causal SNPs.

Compared to the non-mixed model, the additive chi-squared statistics were 9.5% higher on average for causal SNPs without dispersion effects; the log-linear variance statistics were 20.3% lower. For SNPs with a true dispersion effect, the dispersion chi-squared test statistics were 14.6% higher on average compared to the non-mixed-model analysis. This shows that modeling the random effects of SNPs can increase the power to detect loci with additive and dispersion effects.

Application to BMI in the UK Biobank. We analyzed the subsample of the UK Biobank population with predominantly white British ancestry as defined by the UK Biobank quality control process[20]. We applied an inverse normal transformation to BMI and adjusted for age, sex, and 40 principal components (Methods). After sample quality control (Methods), there were 408,250 individuals with both genotype and BMI data. We split the sample into two: an 'unrelated' subsample ($n = 276{,}415$), with no pairs related at the third-degree level or higher; and the complementary 'related' subsample ($n = 131{,}835$), which contains all third-degree or closer relative pairs. We did this to demonstrate two different uses of the mixed model: (1) to increase power in a homogeneous, unrelated sample; and (2) to control for relatedness. We selected 500 SNPs to model the random effects for the unrelated sample and 1,000 SNPs for the related sample (Methods). We then combined the additive and log-linear variance effects from the two subsamples to give single estimates of additive and log-linear variance effects for each SNP (Supplementary Table 2).

To visualize the genome-wide results of fitting the AV model, we introduce the 'Manhattan Sunset' plot (Fig. 2), which displays the additive and log-linear variance test statistics stacked on top of each other, highlighting any loci with evidence of variance effects. We note that many SNPs in the major histocompatibility complex (MHC) region on chromosome 6 appear to have both additive and variance effects on BMI (Fig. 2). Because of its complexity, we did not pursue further analysis of this region here.

We determined that a locus was significant if its P value for the AV test was below the accepted conventional level for genome-wide significance, namely 5×10^{-8} (Methods). Out of 328 loci that are genome-wide significant under either the additive or AV test (Supplementary Table 3), 48 loci have a smaller P value from the AV test than from the additive test, including the *FTO* locus. However, most loci have a smaller P value for the additive test; 56 loci that are genome-wide significant under the additive test are not genome-wide significant under the AV test.

While there may be a slight loss of power for most associated loci, for loci with substantial dispersion effects, a gain in power is likely. Fifteen loci were genome-wide significant under the AV test, although not under the additive test; some of these showed evidence for dispersion effects, as we outline next.

Evidence for dispersion effects. Untransformed BMI is non-normal, so variance effects may be due to a general mean–variance relationship that has not been removed by inverse normal transformation. To move beyond this, we estimated the dispersion effects for all test SNPs genome-wide (Methods). We inferred that, for inverse normally transformed BMI, $r_{av} \approx 0.135$ (Fig. 3) (Methods). While the genome-wide additive and log-linear variance test statistics show similar patterns of deviation from the null hypothesis, the dispersion test statistics follow a different pattern (Fig. 4). This is further evidence that the component of the variance test statistics driven by the mean effects has been largely removed from the dispersion test statistics. The systematic excess, relative to the null hypothesis, of the large test statistics in the right-hand half of the dispersion quantile-quantile plot (Fig. 4c) shows that there are likely to be many SNPs with real dispersion effects.

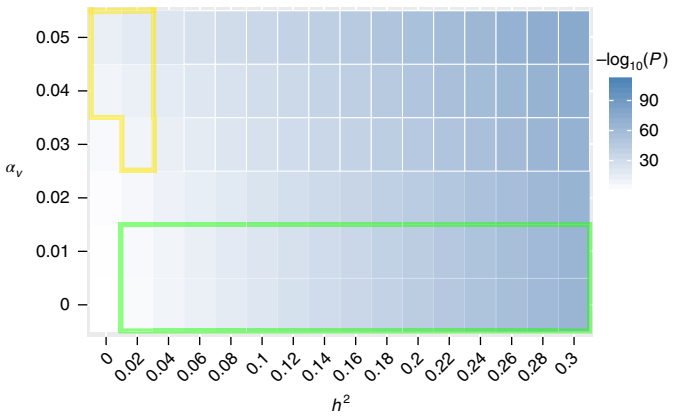

Fig. 1 | Association signal of AV test for simulated phenotypes with different parameters. The expected $-\log_{10}(P)$ of the AV test for different additive and log-linear variance effects of the test locus is indicated by the shading. Phenotypes were simulated for 100,000 unrelated individuals (Methods). The test locus had a frequency of 0.5; the strength of the additive effect is parameterized by the amount of variance explained, h^2. The log-linear variance effect is indicated on the y axis and corresponds approximately to the proportional change in phenotypic variance per allele. We have highlighted two regions of parameter space: the area inside the green lines is where the association signal is stronger under the additive test than under the AV test; the area inside the yellow lines is where the AV test is genome-wide significant ($P < 5 \times 10^{-8}$) but the additive test is not.

One locus reached genome-wide significance for a dispersion effect, around SNP rs900400 ($P = 1.2 \times 10^{-8}$, Fig. 5). Furthermore, some other loci with strong evidence under the AV model display evidence of dispersion effects: we list those loci with $P < 10^{-3}$ for a dispersion effect and $P < 5 \times 10^{-8}$ under the AV test in Table 1.

There are three loci in this list (around rs1538749, rs1801282, and rs900400) not previously associated with BMI at genome-wide significance levels, although rs1538749 and rs900400 have been associated with other obesity-related phenotypes at genome-wide significance levels (including hip circumference and waist-to-hip ratio for rs1538749 and waist circumference and waist-to-hip ratio for rs900400)[21]. Of the other loci in the table, there are two known loci with strong effects on BMI (*GIPR* and *FTO*), a synonymous codon variant (rs2303223) in *ZNF668* in strong linkage disequilibrium (LD; $r^2 = 0.993$) with a previously identified missense variant (rs749670) in *ZNF646*[21], and a variant at the *TCF7L2* locus (rs10787472), which contains variants with known strong effects on type 2 diabetes risk.

Using results from GIANT meta-analyses[12,17], we examined whether additive and variance effects could be replicated. We replicated the variance effect for rs900400 ($P = 1.1 \times 10^{-3}$) (Table 1). The estimated additive effect for rs900400 is so weak that it is consistent with zero in both our data ($P = 0.12$) and the GIANT meta-analysis ($P = 0.18$). The C allele of SNP rs900400 has been associated with lower birth weight[22,23], and the locus provides the strongest genome-wide association for this phenotype. By using self-reported birth weight data from the UK Biobank, we showed that the variance effect of rs900400 was not mediated through its effect on birth weight (Methods). The rs900400 SNP also shows significant genome-wide association with expression of *TIPARP* in adipocytes[24] and with age at menarche[25].

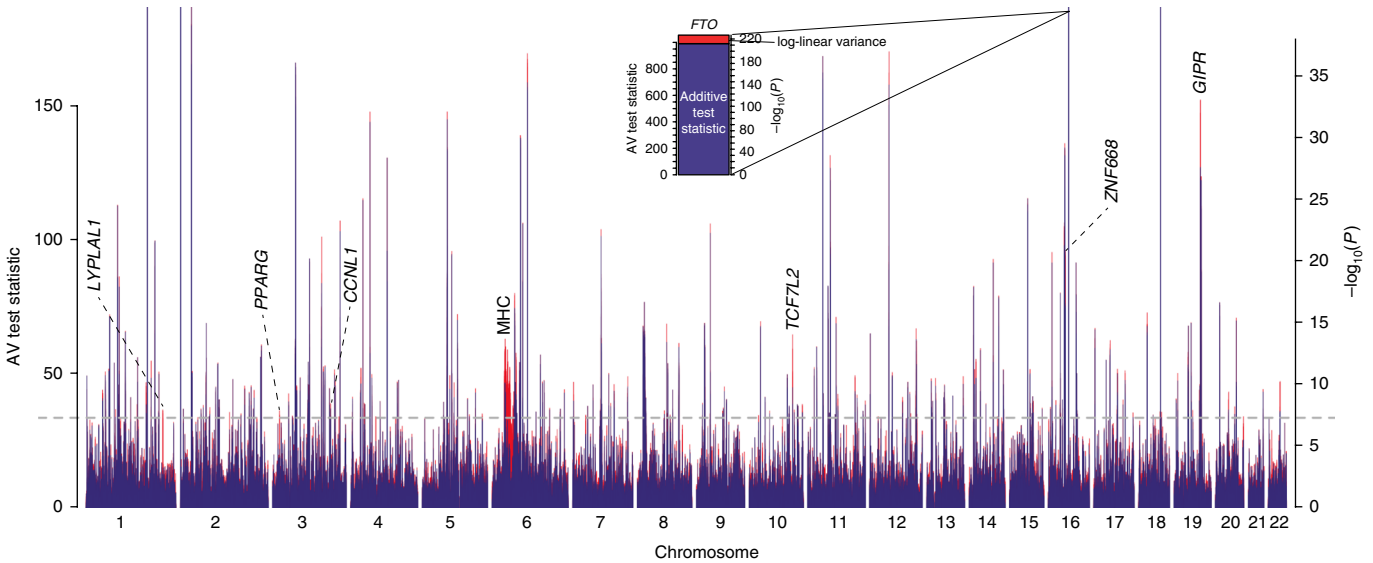

Fig. 2 | Manhattan Sunset plot visualizing the genome-wide additive and log-linear variance test statistics for BMI. At each SNP, the additive chi-squared test statistic is plotted as a blue bar, and the chi-squared test statistic for a log-linear variance effect is added on top of this in red, the combined height of which gives the AV test statistics, the $-\log_{10}(P)$ of which is marked on the right-hand *y* axis. Test statistics were derived from combining effect estimates from the related and unrelated subsamples of the UK Biobank population (Methods). The names of the nearest protein-coding genes are indicated for loci that passed genome-wide-significance ($P<5\times10^{-8}$) for the AV test and displayed evidence of a dispersion effect ($P<1\times10^{-3}$) (Table 1). In addition, the MHC region is indicated.

For rs1801282, the additive effect we observed was replicated in the GIANT meta-analysis ($P=4.2\times10^{-7}$). The estimated variance effect is consistent between our analysis (0.022) and the GIANT meta-analysis (0.015), but it does not quite reach statistical significance in the GIANT meta-analysis ($P=0.084$). The SNP rs1801282 is a missense variant in the gene *PPARG*. *PPARG* is involved in fatty acid storage and glucose metabolism, and is a target for a class of type 2 diabetes drugs (thiazolidinediones). The minor G allele has been repeatedly associated with reduced type 2 diabetes risk[26]. We found that the minor G allele is also associated with increased BMI and increased variability in BMI. The additive effects on type 2 diabetes and BMI are similar to the action of thiazolidinediones, which treat type 2 diabetes but have been found to cause fat gain[27]. We did not find evidence that the variance effect is mediated through an interaction with type 2 diabetes status ($P=0.58$) (Methods).

For rs1538749, the additive effect was consistent but not statistically significant ($P=0.097$) in the GIANT meta-analysis. The variance effect could not be replicated, with an effect estimate close to zero ($P=0.84$).

We found that nearly half of the previously identified variance effect of the *FTO* risk allele[12] (rs1421085; 0.027 per risk allele) can be explained by its additive effect, with a much smaller estimated dispersion effect (0.017 per risk allele).

Putative gene–environment interactions. By using a previously published approach to test jointly for interactions between lifestyle and/or environmental variables and variation at the *FTO* locus[28], we tested for interactions between the SNPs in Table 1 and diet, physical activity, sleep duration, TV watching, frequency of alcohol consumption, and socioeconomic status (Townsend deprivation index) (Methods). The results of fitting these interaction models are shown in Supplementary Table 4. Here, we report the interactions with $P<0.005$, except for SNP rs1421085 (*FTO*), on which we have previously reported[28].

The only SNP other than rs1421085 (*FTO*) with strong evidence for a specific gene–environment interaction was rs900400. We found evidence for an interaction with physical activity ($P=5.0\times10^{-5}$). There was also suggestive evidence for an interaction with diet ($P=6.2\times10^{-3}$). The results suggest that the variability-increasing allele may enhance the effect of physical activity and diet variation on BMI. Modeling the interactions reduced the estimated variance effect of rs900400 from 0.0200 to 0.0185, indicating that the modeled interactions do not explain all of the variance effect of rs900400. In contrast, the modeled interactions reduced the variance effect of rs1421085 (*FTO*) from 0.027 to 0.018, indicating that the interactions explain a substantial fraction of the variance effect of the *FTO* locus.

Leptin and BMI dispersion effects. The T allele of SNP rs900400 has also been associated with higher levels of circulating leptin after adjusting for BMI[29]. To investigate whether there might be a more general connection between leptin levels and BMI variability, we next asked whether SNPs associated with changes in leptin levels tended to be associated with BMI variability. Specifically, we took the top 100 lead SNPs from independent loci with evidence for effects on leptin levels adjusted for BMI[29] and compared their estimated effects on leptin adjusted for BMI with their estimated BMI dispersion effects. We observed a small but significant linear relationship between the effect of the SNP on leptin adjusted for BMI and on BMI dispersion. We estimate that a SNP that increases leptin levels by 1 s.d. per allele would have an expected dispersion effect on BMI of 0.032 ($P=0.026$, Methods and Supplementary Fig. 4). Following a Mendelian randomization argument, these results suggest that there may be a connection between leptin and BMI variability.

Discussion

We have introduced a new framework for association testing that extends widely used additive tests by testing for a linear change in phenotypic mean (exactly as in additive tests) and a log-linear change in phenotypic variance. A new visualization tool, the Manhattan Sunset plot, provides a genome-wide picture of evidence for both additive and log-linear variance effects.

Table 1 | Loci with evidence for dispersion effects

SNP	Position on chromosome	Gene	Alleles	EAF	Additive			Log-linear variance			AV		Dispersion			GIANT additive			GIANT variance		
					Est.	s.e.	*P*	Est.	s.e.	*P*	*P*	$\frac{\chi^2_V}{\chi^2_{AV}}$	Est.	s.e.	*P*	Est.	s.e.	*P*	Est.	s.e.	*P*
rs1538749[a]	1:219,649,865	*LPYLAL1*	A/C	0.45	−0.0083	0.0022	3.9×10^{-4}	−0.016	0.0032	9.7×10^{-7}	1.1×10^{-8}	0.656	−0.015	0.0032	4.8×10^{-6}	−0.0063	0.0038	9.7×10^{-2}	−0.0012	0.0058	8.4×10^{-1}
rs1801282	3:12,393,125	*PPARG*	G/C	0.12	0.0152	0.0033	2.3×10^{-5}	0.022	0.0048	1.8×10^{-5}	1.3×10^{-8}	0.506	0.020	0.0048	9.8×10^{-5}	0.0231	0.0046	4.2×10^{-7}	0.0150	0.0085	8.4×10^{-2}
rs900400	3:156,798,775	*CCNL1*	T/C	0.61	0.0037	0.0022	1.2×10^{-1}	0.020	0.0032	5.5×10^{-9}	1.2×10^{-8}	0.933	0.019	0.0032	1.2×10^{-8}	0.0054	0.0040	1.8×10^{-1}	0.0200	0.0061	1.1×10^{-3}
rs10787472	10:114,781,297	*TCF7L2*	C/A	0.47	−0.0156	0.0022	2.1×10^{-11}	−0.015	0.0032	9.0×10^{-6}	9.3×10^{-15}	0.305	−0.013	0.0032	1.3×10^{-4}	−0.0155	0.0031	4.1×10^{-7}	−0.0160	0.0057	5.0×10^{-3}
rs2303223	16:31,075,175	*ZNF668*	A/G	0.37	−0.0209	0.0022	2.4×10^{-18}	−0.015	0.0032	9.5×10^{-6}	1.5×10^{-21}	0.204	−0.012	0.0033	3.0×10^{-4}	−0.0184	0.0031	3.7×10^{-9}	−0.0120	0.0058	3.7×10^{-2}
rs1421085	16:53,800,954	*FTO*	C/T	0.40	0.0746	0.0022	2.0×10^{-217}	0.027	0.0032	1.1×10^{-15}	8.8×10^{-230}	0.061	0.017	0.0032	4.3×10^{-7}	0.0813	0.0031	8.8×10^{-151}	0.0260	0.0064	5.9×10^{-5}
rs10423928	19:46,182,304	*GIPR*	T/A	0.81	0.0332	0.0027	1.9×10^{-29}	0.021	0.0040	5.4×10^{-7}	9.4×10^{-34}	0.165	0.017	0.0040	7.2×10^{-5}	0.0299	0.0052	8.9×10^{-9}	0.0150	0.0077	6.0×10^{-2}

[a]For rs1538749, we used a proxy, rs11118304, with $r^2 \approx 1$, to obtain the relevant summary statistics from the GIANT consortium. The statistics for loci were genome-wide significant ($P < 5 \times 10^{-8}$) under a joint test for additive and log-linear variance effects; $P < 10^{-3}$ for a dispersion effect[12]. Additive and log-linear variance effect estimates were derived from combining estimates from the related and unrelated subsamples of the UK Biobank (see Methods). Gene names are the nearest protein-coding gene. The effects are given per copy of the first allele (per copy of the C allele if the alleles are given as, for example, C/T) for inverse normally transformed BMI. The ratio of the log-linear variance test statistics to the AV test statistics is given under the $\frac{\chi^2_V}{\chi^2_{AV}}$ column. The additive effect and *P* value from the GIANT consortium meta-analysis of BMI[11] and the variance effect from the GIANT meta-analysis of variance effects on BMI[12] are given for external validation. Est., estimated effect; EAF, estimated allele frequency.

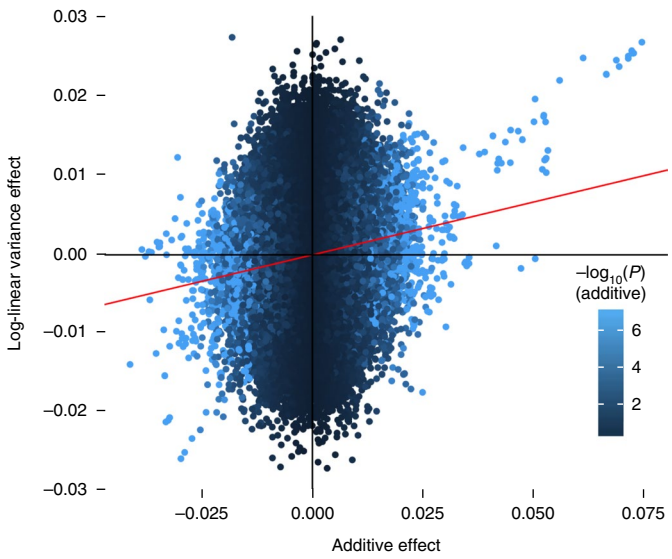

Fig. 3 | Relationship between additive and log-linear variance effects. Estimated additive (*x* axis) and log-linear variance (*y* axis) effects on BMI are plotted for all genome-wide SNPs, shaded in proportion to the $-\log_{10}(P)$ for an additive effect, up to a maximum of $-\log_{10}(5 \times 10^{-8})$, the conventional boundary for genome-wide significance. Due to the mean–variance relationship of the untransformed BMI, any locus with an additive effect is expected to have a log-linear variance effect, even after inverse normal transformation. The red line shows the expected log-linear variance effect given a particular additive effect on inverse normally transformed BMI, where we have inferred this relationship empirically (Methods). Note that SNPs on chromosome 6 were not included to exclude SNPs from the MHC region.

When applied to GWAS and related studies, this approach (1) can be more powerful for discovery of associated loci, and (2) can specifically identify loci with effects on phenotypic variability. We have demonstrated both of these potential advantages through simulations and application to BMI in the UK Biobank.

Mixed-model approaches have several advantages[19]. The test we propose can be performed in the HLMM; we provide an algorithm that scales linearly with sample size for a fixed number of random effects. We have demonstrated in simulations that our method has increased power over other tests that jointly test for mean and variance effects because of (1) a reduction in the degrees of freedom of the variance test statistics and (2) a reduction in the signal-to-noise ratio from modeling the random effects of SNPs with evidence for association.

When multiple haplotypes with different phenotypic effects are present in the region of a SNP, this can generate association between genotype and phenotypic variance at the SNP[8]. While detailed investigation of this phenomenon is beyond the scope of this article, it might explain some of the additive and variance effects of SNPs in the MHC region. However, this phenomenon is unlikely to explain the variance effect of rs900400, since it occurs between two close recombination hotspots, or the variance effects of the other SNPs in Table 1, since the SNPs are not in LD ($r^2 < 0.01$) with other SNPs with strong evidence for additive effects (Methods).

Variance effects that are explained by a general mean–variance relationship are unlikely to reflect interactions or other biologically meaningful phenomena. We developed a method that can identify when a SNP has a variance effect beyond that which can be explained by a general mean–variance relationship, which we term a 'dispersion effect'. As one application, we showed that around half of the reported effect of the *FTO* locus on BMI variability[12] is

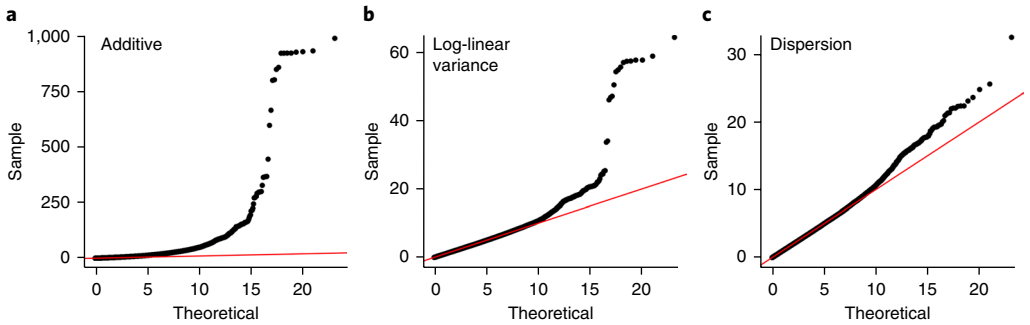

Fig. 4 | Quantile-quantile plots for test statistics. a–c, We compared the sample quantiles of the additive (**a**), log-linear variance (**b**), and dispersion test statistics (**c**) to the theoretical quantiles of the null distribution. Under the null distribution, the test statistics asymptotically follow a chi-squared distribution on 1 d.f. The red diagonal line indicates the expected sample quantiles under the null distribution. Additive test statistics were adjusted for inflation using LD score regression[30]; log-linear variance and dispersion test statistics were adjusted for inflation using genomic control (Methods). Note that, because of widespread association in the MHC region (Fig. 2), we removed SNPs from chromosome 6 from these plots.

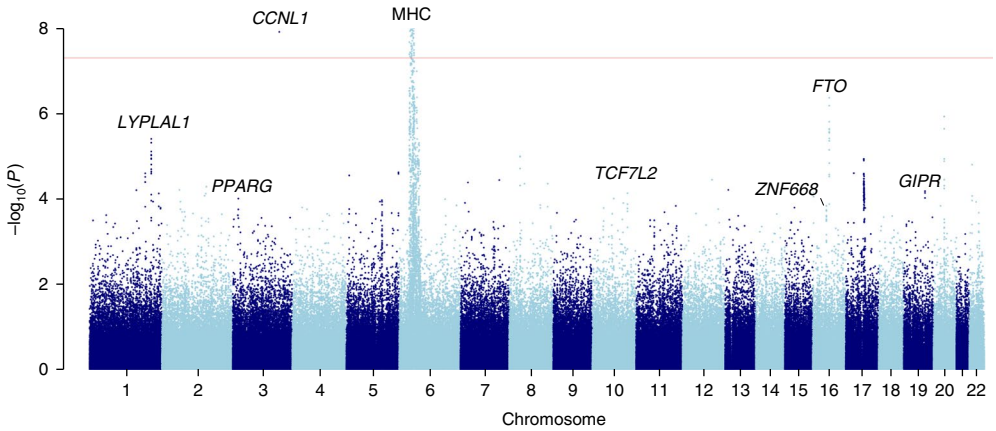

Fig. 5 | Manhattan Sunset plot for dispersion effects. The *P* value for a dispersion effect is plotted for each tested SNP, with the chromosome indicated on the *x* axis. The names of the nearest protein-coding genes are indicated for loci that are genome-wide significant ($P < 5 \times 10^{-8}$) under the AV test and have $P < 10^{-3}$ for a dispersion effect (Table 1). Note that SNP rs900400 (*CCNL1*) occurs between two close recombination hotspots, with no other genotyped SNPs in strong LD with rs900400. In addition, the MHC is highlighted.

explained by a general mean–variance relationship; thus, it may not be biologically interesting.

While only rs900400 showed genome-wide significant evidence for a dispersion effect, six other SNPs that were genome-wide significant under the AV test showed some evidence for a dispersion effect (Table 1). More generally, genome-wide test statistics showed an excess of large dispersion effect estimates (Fig. 4c), implying that these effects are prevalent and may be worthy of further investigation. It has been previously observed that rs900400 affects leptin levels[29]; we found suggestive evidence for a more general connection between leptin and BMI variability.

The statistical testing framework we have developed can be applied to many other traits, enabling the discovery of genetic factors that affect trait variability. Beyond an intrinsic biological interest in factors determining trait variability, this approach can guide the search for gene–gene and gene–environment interactions. Its application in large population biobanks will enable well-powered studies that increase our understanding of the interactions influencing phenotypic variation in human populations.

URLs. GIANT consortium, https://portals.broadinstitute.org/collaboration/giant/index.php?title=GIANT_consortium&oldid=251; UK Biobank Project, http://www.ukbiobank.ac.uk/; LD scores: https://data.broadinstitute.org/alkesgroup/LDSCORE/.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0225-6.

Received: 15 April 2017; Accepted: 3 August 2018;
Published online: 15 October 2018

References

1. Price, A. L., Spencer, C. C. A. & Donnelly, P. Progress and promise in understanding the genetic basis of common diseases. *Proc. Biol. Sci.* **282**, 20151684 (2015).
2. Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**, e1000008 (2008).
3. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS. Med.* **12**, e1001779 (2015).
4. Marchini, J., Donnelly, P. & Cardon, L. R. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat. Genet.* **37**, 413–417 (2005).
5. Paré, G., Cook, N. R., Ridker, P. M. & Chasman, D. I. On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. *PLoS Genet.* **6**, e1000981 (2010).
6. Struchalin, M. V., Dehghan, A., Witteman, J. C., van Duijn, C. & Aulchenko, Y. S. Variance heterogeneity analysis for detection of potentially interacting genetic loci: method and its limitations. *BMC Genet.* **11**, 92 (2010).

7. Hill, W. G. & Mulder, H. A. Genetic analysis of environmental variation. *Genet. Res. (Camb).* **92**, 381–395 (2010).
8. Forsberg, S. K. G. et al. The multi-allelic genetic architecture of a variance-heterogeneity locus for molybdenum concentration in leaves acts as a source of unexplained additive genetic variance. *PLoS Genet.* **11**, e1005648 (2015).
9. Ivarsdottir, E. V et al. Effect of sequence variants on variance in glucose levels predicts type 2 diabetes risk and accounts for heritability. *Nat. Genet.* 1398–1402 (2017).
10. Kitano, H. Biological robustness. *Nat. Rev. Genet.* **5**, 826–837 (2004).
11. Rönnegård, L. & Valdar, W. Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. *BMC Genet.* **13**, 63 (2012).
12. Yang, J. et al. FTO genotype is associated with phenotypic variability of body mass index. *Nature* **490**, 267–272 (2012).
13. Cao, Y., Wei, P., Bailey, M., Kauwe, J. S. K. & Maxwell, T. J. A versatile omnibus test for detecting mean and variance heterogeneity. *Genet. Epidemiol.* **38**, 51–59 (2014).
14. Cao, Y., Maxwell, T. J. & Wei, P. A family-based joint test for mean and variance heterogeneity for quantitative traits. *Ann. Hum. Genet.* **79**, 46–56 (2015).
15. Rönnegård, L., Felleki, M., Fikse, F., Mulder, H. A. & Strandberg, E. Genetic heterogeneity of residual variance: estimation of variance components using double hierarchical generalized linear models. *Genet. Sel. Evol.* **42**, 8 (2010).
16. Box, G. E. P. Non-normality and tests on variances. *Biometrika* **40**, 318–335 (1953).
17. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
18. Zhang, Z. et al. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360 (2010).
19. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106 (2014).
20. Bycroft, C. et al. Genome-wide genetic data on ~500,000 UK Biobank participants. Preprint at https://www.biorxiv.org/content/early/2017/07/20/166298 (2017).
21. Turcot, V. et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat. Genet.* **50**, 26–41 (2018).
22. Horikoshi, M. et al. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat. Genet.* **45**, 76–82 (2013).
23. Freathy, R. M. et al. Variants in *ADCY5* and near *CCNL1* are associated with fetal growth and birth weight. *Nat. Genet.* **42**, 430–435 (2010).
24. Hivert, M. F. et al. Genetic determinants of adiponectin regulation revealed by pregnancy. *Obesity (Silver Spring)* **25**, 935–944 (2017).
25. Perry, J. R. B. et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
26. Zeggini, E. et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341 (2007).
27. Larsen, T. M., Toubro, S. & Astrup, A. PPARgamma agonists in the treatment of type II diabetes: is increased fatness commensurate with long-term efficacy? *Int. J. Obes. Relat. Metab. Disord.* **27**, 147–161 (2003).
28. Young, A. I., Wauthier, F. & Donnelly, P. Multiple novel gene-by-environment interactions modify the effect of FTO variants on body mass index. *Nat. Commun.* **7**, 12724 (2016).
29. Kilpeläinen, T. O. et al. Genome-wide meta-analysis uncovers novel loci influencing circulating leptin levels. *Nat. Commun.* **7**, 10494 (2016).
30. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

Acknowledgements

This work was supported by Wellcome Trust grant 095552/Z/11/Z to P.D. and grants 090532/Z/09/Z and 20314/Z/16/Z as core support for the Wellcome Trust Centre for Human Genetics. A.Y. was supported by a Wellcome Trust Doctoral Studentship (099670/Z/12/Z) and by the Li Ka Shing Foundation.

Author contributions

A.Y. developed the method, led its application to the UK Biobank data, and wrote the paper. F.L.W. was involved in the development and application of the method. P.D. supervised the research and wrote the paper. All work undertaken by F.L.W. was done while F.L.W. was at University of Oxford.

Competing interests

P.D. is a founder and director of Genomics plc, and a partner of Peptide Groove LLP. The remaining authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0225-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.I.Y. or P.D.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Estimation of dispersion effects. Variation at a locus l has a log-linear dispersion effect d_l on a (not necessarily normal) phenotype Y if $\mathrm{E}[Y|G_l=g]=\mu_{lg}$ and

$$\mathrm{Var}(Y \mid G_l=g)=h(\mu_{lg})\exp(d_l(g-2f_l))$$

where f_l is the allele frequency, $2f_l$ is the mean genotype value, and h is the function that gives the mean–variance relationship for the untransformed phenotype distribution. We parameterize the model in terms of mean normalized genotypes, $(g-2f_l)$, so that the h function is the same for all loci. When the effects on the phenotypic mean are small and additive, there is an approximately linear relationship between the additive effect (α_l), the log-linear variance effect (α_{vl}), and the dispersion effect (d_l) at each locus l (Supplementary Note); that is, $\alpha_{vl}\approx r_{\mathrm{av}}\alpha_l+d_l$. It is possible to estimate r_{av} by regression of the estimates of α_{vl}, $\hat{\alpha}_{vl}$, on the estimates of α_l, $\hat{\alpha}_l$, across genome-wide loci (for all l). Standard regression estimates are downwardly biased because of noise in the estimation of α_l, but this bias can be corrected (Methods). Thus, we show (Supplementary Note):

$$r_{\mathrm{av}}\approx\frac{\mathrm{Cov}_l(\hat{\alpha}_{vl},\hat{\alpha}_l)}{\mathrm{Var}_l(\hat{\alpha}_l)}\left(1+\frac{\mathrm{E}_l[\mathrm{Var}(\hat{\alpha}_l)]}{\mathrm{Var}_l(\hat{\alpha}_l)-\mathrm{E}_l[\mathrm{Var}(\hat{\alpha}_l)]}\right)$$

where the subscripts on Cov_l and Var_l indicate that expectations are to be taken over the loci, whereas $\mathrm{Var}(\hat{\alpha}_l)$ represents the within-locus sampling variance, which is known from the asymptotic distribution of the maximum likelihood estimator. The first factor, $\mathrm{Cov}_l(\hat{\alpha}_{vl},\ \hat{\alpha}_l)/\mathrm{Var}_l(\hat{\alpha}_l)$, is the standard regression coefficient of $\hat{\alpha}_{vl}$ on $\hat{\alpha}_l$ across the loci, while the second term is an adjustment factor for the bias in the standard regression coefficient due to noise in the estimation of α_l.

We use robust regression to estimate $\mathrm{Cov}_l(\hat{\alpha}_{vl},\hat{\alpha}_l)/\mathrm{Var}_l(\hat{\alpha}_l)$ to prevent being overly influenced by outliers. Specifically, we use M-estimation[31] with prior weights determined by $1/\mathrm{Var}(\hat{\alpha}_l)$. To estimate the adjustment factor, we use the sample estimates of $\mathrm{Var}_l(\hat{\alpha}_l)$ and $\mathrm{E}_l[\mathrm{Var}(\hat{\alpha}_l)]$. We do not claim that this method is statistically optimal to estimate r_{av}; however, it worked well in simulations, suggesting the method is sufficient for large samples from a polygenic and heritable trait.

When many loci genome-wide are examined for a heritable and polygenic trait, error in estimating r_{av} will be negligible and can be ignored. Thereby, we estimate dispersion effects by $\hat{d}_l=\hat{\alpha}_{vl}-r_{\mathrm{av}}\hat{\alpha}_l$, which has a normal asymptotic sampling distribution (Supplementary Note):

$$\hat{d}_l\sim N\left(d_l,\ \mathrm{Var}(\hat{\alpha}_{vl})+r_{\mathrm{av}}^2\mathrm{Var}(\hat{\alpha}_l)\right).$$

The HLMM. The HLMM is specified as

$$Y\sim N\left(X\alpha,h^2GG^T+\exp(\mathrm{diag}(V\beta))\right),$$

where G is a normalized $N\times L$ matrix of genotypes, and h^2 corresponds to the variance explained by the additive associations of the l genotypes in G. This differs from the standard linear mixed model because of the parameterized heteroskedasticity of the residual error term, $\exp(\mathrm{diag}(V\beta))$. By modeling the effect of a test SNP on both the mean of the trait (putting it in X) and the residual variance of the trait (putting it in V), the additive and log-linear variance effects of the SNP can be estimated, enabling the two-degree-of-freedom AV test to be performed. Principal components and other covariates can also be included as mean covariates (in X) and variance covariates (in V) (Supplementary Note).

Existing computational methods for fitting linear mixed models based on spectral decomposition[32], which are appropriate when $N\gg L$, do not work when the residual variance is genotype-dependent. We have developed an algorithm for efficient inference of the parameters of the HLMM when the number of SNPs in the random effect is small compared to the sample size. The covariance matrix for the HLMM is

$$\Sigma=h^2GG^T+D,\qquad D=\exp(\mathrm{diag}(V\beta))$$

To compute the likelihood and gradient efficiently, we use low-rank update formulae for the inverse and determinant of the covariance matrix, expressed as

$$\Sigma^{-1}=D^{-1}-h^2D^{-1}G\Lambda^{-1}G^TD^{-1};\ \ \Lambda=I+h^2G^TD^{-1}G.$$

Similarly, the log-determinant of $\sum$ can be expressed as $\log|\Sigma|=\log|\Lambda|+\log|D|$. This allows us to rewrite the log-likelihood of the model as the log-likelihood of a diagonal system plus a low-rank correction. Let l be the log-likelihood of the model. Then

$$2l=\ -n\log(2\pi)-\log\ |D|-(y-X\alpha)^TD^{-1}(y-X\alpha)-\log\ |\Lambda|$$
$$+h^2\left[G^TD^{-1}(y-X\alpha)\right]^T\Lambda^{-1}\left[G^TD^{-1}(y-X\alpha)\right]$$

which can be computed in $O(NL^2+L^3)$ operations. The derivatives of the likelihood can be computed in the same complexity class; we give expressions for these and further details of the algorithm implementation in the Supplementary Note. We note the similarity to computational approaches previously used in general linear mixed models[33]. The NL^2 term dominates when $N\gg L$, and this method has effectively the same scaling with sample size as methods based on spectral decomposition for additive association testing in a mixed model[31]. However, our method requires an $O(NL^2)$ operation at every iteration, making it more computationally demanding than spectral decomposition, which requires only one $O(NL^2)$ operation. We benchmarked our algorithm for $N=100{,}000$ and $L=500$ on a 4.2 GHz Intel Core i7 processor. Fitting one model took around 1.3 CPU minutes, implying that an analysis of 400,000–800,000 SNPs would take between 9 and 17 hours on a server with 1,000 cores.

Power simulations. To test the power of the AV test relative to the additive test when detecting a test locus G, we simulated the phenotypes for 100,000 individuals according to the model $Y|G=g\sim N(\alpha g, e^{\alpha_v g})$, where the genotype was simulated from a binomial(2,0.5) distribution independently for each individual. The additive effect, α, varied such that the variance explained by the additive effect varied from 0 to 0.3%, and α_v varied from 0 to 0.05. For each pair (α, α_v), 1,000 independent phenotypes were simulated. We fitted models by maximum likelihood.

We also investigated how the association signals changed with sample sizes ranging from 10,000 to 100,000 for a fixed additive effect of $\alpha=0.02$ (corresponding to an explained variance of 0.02%), with α_v varying from 0 to 0.05 in increments of 0.005.

To compare the power of different variance effect tests, we simulated phenotypes for 100,000 individuals according to the model $Y|G=g\sim N(\alpha g, e^{\alpha_v g+\delta_v(g-1)^2})$, where δ_v represents the general variance effect and sets the deviation in log-variance of the heterozygote from a log-linear model. The genotype was simulated from a Binomial(2,0.5) distribution independently for 100,000 individuals. The additive effect, α, was set so that the variance explained was 0.02%, δ_v varied from 0 to 0.05, and α_v varied from 0 to 0.05. For each pair (δ_v, α_v), 1,000 independent phenotypes were simulated.

Simulation of non-normal phenotypes. We simulated phenotypes for 150,000 individuals from simulated 'genomes' of 22,000 independent SNPs. The frequency of each SNP was randomly sampled from a uniform distribution on [0.05,0.5], and each individual's genotype was drawn independently from a binomial distribution with the probability parameter equal to the frequency. We randomly selected 1,000 of 22,000 SNPs to have a causal effect on the phenotype. One of the 1,000 SNPs was selected to have a dispersion effect, which we called G_1. We generated an additive genetic component by sampling normally distributed effect sizes for the 999 causal SNPs other than G_1. The SNP with the dispersion effect, G_1, was given a positive additive effect scaled to explain 0.1% of the variance of the overall additive genetic component, which explained 10% of the phenotypic variance and includes the additive effects of all 1,000 causal SNPs including G_1. The overall additive genetic component was scaled to have a mean of 0 and a variance of 1.

To simulate the non-normal phenotypes, we used a gamma model. If Y has a gamma distribution with a shape parameter (k) and scale parameter (θ), then $Y\sim\mathrm{Gamma}(k,\theta)$. We parameterized each individual's distribution to give a certain heritability for the trait and a dispersion effect of d_1 to the SNP G_1. Each individual's phenotype, Y_i, was drawn from a gamma distribution with parameters chosen so that $\mathrm{E}[Y_i]=\mu+A_i$, where A_i is the value of the additive genetic component of individual i. Thus, for some $s>0$,

$$\mathrm{Var}[Y_i]=s(\mu+A_i)\exp(d_1(g_{i1}-f_1))=s\mathrm{E}[Y_i]\exp(d_1(g_{i1}-f_1))$$

where g_{i1} is the genotype of individual i for G_1. This shows the linear relationship between the conditional means and variances of a gamma-distributed phenotype. This implies that, according to the Law of Total Variance, $\mathrm{Var}(Y)\approx\mu s+1$. This implies that the heritability is $h^2\approx(\mu s+1)^{-1}$. For a given μ and h^2, one can solve for s. We chose $h^2=0.1$ and $\mu=5$ for our simulations, implying that $s=1.8$. To achieve this, each individual's phenotype is drawn from a gamma distribution, $Y_i\sim\mathrm{Gamma}(k_i,\theta_i)$, with

$$k_i=\exp(-d_1(g_{i1}-f_1))(\mu+A_i)/s$$

$$\theta_i=s\ \exp(d_1(g_{i1}-f_1))$$

The phenotypes simulated in this way were inverse normally transformed before fitting the models.

UK Biobank data. We used genotype data from the UK Biobank Project (see URLs). Quality control is described in the UK Biobank genotyping quality control document[20]. We used the sample of 409,703 individuals identified by the UK Biobank as having predominantly white British ancestry. We then excluded individuals from the analysis that had been flagged by the UK Biobank as having a

putative sex chromosome aneuploidy, excess relatives, or excess heterozygosity. See also the Life Sciences Reporting Summary.

Relatedness was determined by the UK Biobank[20]. We split the sample into two groups: one without any pair related at the third-degree level or higher; and its complement. The final 'related' sample had 131,835 individuals and the final 'unrelated' sample had 276,415 individuals.

Analysis of BMI. An inverse normal transformation was applied to BMI. For mean and variance covariates, we used age (Data-Field 21022), sex (Data-Field 31), age^2, age^3, age × sex, age^2 × sex, age^3 × sex, genotyping array, and the 40 principal components provided by the UK Biobank[20]. Throughout our analysis of BMI, we only considered SNPs with a minor allele frequency > 5% and missingness < 5%.

To choose the SNPs for the random effects in each sample, we ranked the SNPs by their negative log P values for additive effects from fitting the model without the random effects. The advantages of selecting SNPs in this way have been elucidated[34], although there may be disadvantages compared to using a full-rank random effect when there is family relatedness[19].

For the unrelated subsample, we selected 500 SNPs from the ranked list in descending order, only adding a SNP if it had an $r^2 < 0.1$ with all previously added loci. For the related sample, we first fitted the model without random effects and ranked SNPs by the strength of the evidence for an additive effect on BMI. We selected 1,000 SNPs from this list in descending order, only adding a SNP if it had a minor allele frequency > 0.4 and had an $r^2 < 0.01$ with all previously added loci (Methods). We used a more stringent r^2 threshold for the related sample to ensure SNPs were independent; thus, together they gave a better estimate of relatedness than if we had included loci in LD. We selected a greater number of common, independent loci for the related sample to better control for the relatedness present in the sample. To fit the mixed model, we fitted the model including the random effects for the selected loci other than the loci on the same chromosome as the test SNP.

For computational efficiency, we first fitted a null model for each chromosome with all of the mean and variance covariates, obtaining $\widehat{\alpha}_0$ and $\widehat{\beta}_0$ as the maximum likelihood estimates of α and β. We then performed the transformation $Y \rightarrow \mathrm{diag}(\exp(-0.5V\widehat{\beta}_0))(Y - X\widehat{\alpha}_0)$ to remove the influence of known covariates on the mean and the residual variance.

We combined additive effects across the two subsamples in a fixed-effects meta-analysis. We did the same for the log-linear variance effects. We calculated the Wald statistic for additive and log-linear variance effects by taking the square of the ratio of the estimated effects and their standard errors. The Wald statistics is asymptotically equivalent to a log-likelihood ratio. For the additive test statistics, we used LD score regression[30] to infer an inflation factor (see URLs for the source of the LD scores). For the log-linear variance test statistics, we calculated the sample median across all test SNPs and took the ratio of the sample median to 0.456 as an inflation factor[35]. We performed two-degree-of-freedom tests for each SNP as the sum of the inflation-adjusted chi-squared test statistics for additive and log-linear variance effects.

To estimate r_{av}, we used the procedure outlined earlier for all SNPs excluding those on chromosome 6. We excluded SNPs on chromosome 6 to exclude SNPs in the MHC region, many of which show evidence for additive and log-linear variance effects. We then calculated the Wald statistics for dispersion effects and took the ratio of the sample median to 0.456 as an inflation factor[35].

We pruned the list of SNPs that were genome-wide significant under either the additive or AV test to ensure that our results did not include SNPs in LD with SNPs with stronger additive associations. For each chromosome, we took the SNP with the strongest additive association first; we then added SNPs in order of additive association strength, excluding those with $r^2 > 0.01$ with a SNP already chosen. We did this to ensure that any locus we identified as having a dispersion effect was independent from other loci with strong additive effects.

If a SNP exhibits a dominance effect, it can lead to spurious inference of a log-linear variance and/or dispersion effect under the AV model. By fitting a model that also included dominance effects, we found no evidence that the estimated log-linear variance effects of the SNPs in Table 1 were driven by dominance effects (Supplementary Table 5).

Gene–environment interaction analysis. We used a previously published model for testing for interactions between rs1421085 (*FTO*) and various lifestyle and environmental factors[28]. In this model, the SNP and its interactions with multiple lifestyle and environmental factors are fitted jointly. Furthermore, because the UK Biobank contains many correlated measures of diet and physical activity, these variables were collapsed into diet and activity scores. In brief, each activity and diet variable was weighted by its strength and direction of association with BMI to create a score that well predicted BMI from measures in a particular category (diet or physical activity). We used the same weights for construction of these scores as in a previous analysis of rs1421085[28].

The lifestyle and environmental variables used are as described in a previous analysis of rs1421085 (*FTO*)[28]. We mean-imputed missing observations of the lifestyle and environmental variables. For mean and variance covariates, the test SNP, age, sex, age^2, age^3, age × sex, age^2 × sex, age^3 × sex, genotyping array, and the 40 principal components provided by UK Biobank[20] were used. In addition to these covariates, the lifestyle and environment variables and their interaction with the test SNP were also used as mean covariates. We fitted the models in the combined sample comprised of both the related and unrelated subsamples. To help control for relatedness and confounding variables, we modeled the random effects for the 1,000 SNPs selected for the analysis of the related sample, excluding the SNPs on the same chromosome as the test SNP.

Diabetes and rs1801282. To test for an interaction between rs1801282 and diabetes status, we used Data-Field 2443 (diabetes diagnosed by doctor). We removed those who answered 'Do not know' and 'Prefer not to answer.' For mean and variance covariates, rs1801282 genotype, age, sex, age^2, age^3, age × sex, age^2 × sex, age^3 × sex, genotyping array, and the 40 principal components provided by the UK Biobank[20] were used. In addition to these covariates, 'diabetes diagnosed by doctor' and its interaction with rs1801282 were also used as mean covariates.

Birth weight and rs900400. If rs900400 affects BMI variability through birth weight, then fitting birth weight and rs900400 jointly should reduce the estimated variance effect of rs900400. To test this, we used the self-reported birth weight data (Data-Field 20022) available for 230,477 genotyped participants without missing calls for rs900400. For mean and variance covariates, we used rs900400 genotype, birth weight, age, sex, age^2, age^3, age × sex, age^2 × sex, age^3 × sex, genotyping array, and the 40 principal components provided by the UK Biobank[20]. The estimated log-linear variance effect of rs900400 in this model was 0.0186 (s.e. = 0.0043). After dropping birth weight from the model, the estimated log-linear variance effect of rs900400 in the same sample was 0.0181. The increase in the variance effect when fitting rs900400 jointly with birth weight may be due to the fact that higher birth weight is associated with decreased variability of BMI ($P = 9.0 \times 10^{-9}$, two-sided Z-test), whereas the birth weight increasing allele of rs900400 is associated with increased BMI variability. This implies that the birth weight effect of rs900400 could be masking its effect on BMI variability, independent of birth weight. This is in direct contradiction to the hypothesis that the effect of rs900400 on BMI variability is due to rs900400's effect on birth weight.

Variance effects of leptin SNPs. We took summary statistics from a GWAS of circulating leptin levels adjusted for BMI[29]. (Because higher BMI is associated with higher leptin levels, this analysis looked for SNPs that affect leptin levels through pathways other than BMI.) We took the intersection of the SNPs from the leptin study with the SNPs on the UK Biobank array. We constructed a list of 100 SNPs at approximately independent loci with evidence for association with leptin. We added SNPs to our list in order of negative log P value for a leptin effect. To ensure the SNPs in our list were approximately independent of each other, we only added a SNP to the list if it had an $r^2 < 0.1$ with all the SNPs already in the list. To estimate the expected dispersion effect per unit increase in leptin effect, we used the same kind of procedure we used to estimate the mean–variance relationship parameter, r_{av} (see earlier). We used robust regression (M-estimation) with weights equal to the inverse of the sample variance of the dispersion effect, and we adjusted the regression coefficient for downward bias due to noise in the estimation of the leptin effects.

As a check, we performed an analogous analysis for the top 100 lead SNPs from independent loci for leptin adjusted for BMI, and regressed their additive effects for BMI on their effects for leptin adjusted for BMI. As expected (because leptin levels have already been adjusted for BMI), we saw no significant effect ($P = 0.47$).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability. The 'hlmm' code used is freely available under an MIT license at https://github.com/AlexTISYoung/hlmm.

Data availability

The primary data analyzed in this study come from the UK Biobank. Applications for access can be made on the UK Biobank website.

References

31. Marazzi, A. *Algorithms, Routines, and S-Functions for Robust Statistics*. (Chapman and Hall/CRC, New York, 1993).
32. Lippert, C. et al. FaST linear mixed models for genome-wide association studies. *Nat. Methods* **8**, 833–835 (2011).
33. Wolfinger, R., Tobiast, R., Sall, J., Tobias, R. & Sall, J. Computing Gaussian likelihoods and their derivatives for general linear mixed models. *SIAM J. Sci. Comput.* **15**, 1294–1310 (1994).
34. Lippert, C. et al. The benefits of selecting phenotype-specific variants for applications of mixed models in genomics. *Sci. Rep.* **3**, 1815 (2013).
35. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).

nature research

Corresponding Author: Peter Donnelly

Date: 17/04/2018

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   We used individuals of white, British ancestry in the UK Biobank, for whom we also had data on body mass index (BMI).

2. Data exclusions

   Describe any data exclusions.

   We excluded individuals flagged as problematic by the UK Biobank quality control documentation: excess heterozygosity, putative sex chromosome aneuploidy, or excess relatives.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   We replicated the variance effects of SNPs rs900400 and rs1801282 in data from an external meta-analysis (GIANT).

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   Not applicable to a genome-wide association study.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Not applicable to a genome-wide association study.

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☒	☐	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☒	☐	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☒	☐	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	HLMM (Heteroskedastic linear mixed model) software. Available under MIT license at https://github.com/AlexTISYoung/hlmm

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	Not applicable to a genome-wide association study

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Not applicable to a genome-wide association study

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Not applicable to a genome-wide association study
b. Describe the method of cell line authentication used.	Not applicable to a genome-wide association study
c. Report whether the cell lines were tested for mycoplasma contamination.	Not applicable to a genome-wide association study
d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	Not applicable to a genome-wide association study

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	Not applicable to a genome-wide association study

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	General descriptions have been provided by the UK Biobank