nature genetics

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41588-019-0481-0

In the format provided by the authors and unedited.

A global overview of pleiotropy and genetic architecture in complex traits

Kyoko Watanabe[1], Sven Stringer[1], Oleksandr Frei[2], Maša Umićević Mirkov[1], Christiaan de Leeuw[1], Tinca J. C. Polderman[1], Sophie van der Sluis[1,3], Ole A. Andreassen[2,4], Benjamin M. Neale[5,6,7] and Danielle Posthuma[1,3]*

[1]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University Amsterdam, Amsterdam, the Netherlands. [2]NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [3]Department of Clinical Genetics, Section of Complex Trait Genetics, Neuroscience Campus Amsterdam, VU Medical Center, Amsterdam, the Netherlands. [4]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [5]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [6]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. [7]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. *e-mail: d.posthuma@vu.nl

Table of Contents

1. Publicly available GWAS summary statistics

GWAS summary statistics were curated from multiple resources and were included only when the full set of SNPs were available. We excluded whole exome sequencing studies. This yielded 2,288 GWASs from 33 consortia and any other resources where summary statistics are available (last update 23rd October 2018). From dbGAP, we obtained 2,659 unique datasets (ftp://ftp.ncbi.nlm.nih.gov/dbgap/Analyses_Table_of_Contents.txt, accessed on July 2017) and extracted 896 GWAS summary statistics in which a matched publication was available and sample size for a specific trait was explicitly mentioned in the original study. We excluded non-GWAS studies (e.g. PAGE (Prenatal Assessment of Genomes and Exomes) studies) and GWASs with immune-chip, whole exome sequencing and replication cohorts (exact reasons of exclusion for each dataset is available in **Supplementary Table 25**). Together this resulted in a total of 3,555 sets of GWAS summary statistics. The complete list and detailed information for each GWAS with summary statistics is available in **Supplementary Table 3** (atlas ID 1-3184, 3785-4155).

2. UK Biobank release 2 dataset

2.1 UK Biobank study cohort

The current study includes data from the UK Biobank study (www.ukbiobank.ac.uk)[1]. The UK Biobank is a large population-based cohort that includes over 500,000 participants and aims to improve insight into a wide variety of health-related determinants and outcomes across the UK. All participants provided written informed consent; the UK Biobank received ethical approval from the National Research Ethics Service Committee North West-Haydock (reference 11/NW/0382), and all study procedure were in accordance with the World Medical Association for medical research. Between 2006 and 2010, approximately 9.2 million invitations to participate in the study were sent to all people aged 37-72 years old who were registered with the National Health Service (NHS) and were living within 25 miles from one of the 22 study research centers. In total, 503,325 participants were recruited in the study. Access to the UK Biobank data was obtained under the application number 16406. We analyzed imputed genotype data of UK Biobank second release (July 2017)[2], which includes 92,693,895 genetic variants in 487,422 individuals. Genotypes were imputed using a combination of two reference panels; a combination of the UK10K haplotype and 1000 Genomes reference panels, and Haplotype Reference Consortium (HRC) panel[3]. If variants were imputed in both panels, the HRC imputation was retained. Details about genotyping and

imputation are available in the original study[2]. We excluded variants imputed based on the combined UK10K/1000 Genomes panel according to recommendations from UK Biobank. To determine individuals of European ancestry, we projected ancestry principal components from the 1000 Genome[4] reference populations onto the called genotypes available in UK Biobank and grouped individuals into their closest ancestral population identified by the minimum Mahalanobis distance from the projected principal component scores [5]. We excluded subjects with a Mahalanobis distance >6 S.D. from their assigned population. Additional filtering of individuals was performed based on UKB-provided information on relatedness (subjects with most inferred relatives, 3rd degree or closer, were removed until no related subjects were present), discordant sex, sex aneuploidy, missing phenotype or covariate data and withdrawn consent.

Imputed variants were converted to hard calls filtering by an imputation INFO score at 0.9 and excluding multi-allelic SNPs, indels, SNPs without unique rsID and SNPs with minor allele frequency (MAF) <0.0001, resulting in a total of 10,847,151 remaining SNPs. To correct for population stratification, we computed European-specific principal components based on a set of 145,432 independent (r^2<0.1) SNPs with MAF>0.01 and INFO=1 using FlashPCA2[6] and first 20 PCs were used as covariates of the association analyses.

2.2 UK Biobank phenotype definition

From the 1,940 unique field IDs to which we had access, 755 had >50,000 subjects with non-missing values. We only used phenotype fields with first visit and first run (e.g. f.xxx.0.0) with exceptions for multi-coded phenotypes, which allowed to assign more than one code for a single subject (see below). They are assigned to field name using ukb_field.tsv obtained from http://biobank.ctsu.ox.ac.uk/crystal/download.cgi (last accessed 31st August 2017). Note that for newly available phenotypes for release 2, we annotated field names manually based on the UK biobank data showcase. From these phenotypes, we excluded baseline characteristics, phenotypes used as covariates, date and place phenotypes, status phenotypes (i.e. completion status, answered a specific question), ethnicity, genomic phenotypes and any other phenotypes that are not relevant for performing a GWAS. For each phenotype, we provided reason of exclusions in **Supplementary Table 1**. This resulted in 434 unique fields including 49 multi-coded phenotypes. 385 phenotypes were considered quantitative when the phenotype value was quantitative or ordinal. Phenotypes coded by yes/no were considered as binary with a few exceptions (**Supplementary Table 1**). Quantitative and binary phenotypes

with exceptional definitions are described in **Supplementary Table 1**. Multi-coded phenotypes (i.e., nominal phenotypes to which 1 (exclusive categories) or multiple (inclusive categories) answers were possible) are described in **Supplementary Table 2**.

For quantitative and binary phenotypes, subjects with phenotype codes -1 for "Do not know" or -3 for "Prefer to not answer" were excluded and the original phenotype code as described in the UK Biobank data showcase was used unless specified in Supplementary Text or **Supplementary Table 1** and **2**.

For 49 multi-coded phenotypes, we dichotomized each code to dummy binary phenotypes (cases for 1 and controls for 0) and included subjects with phenotype code -7 for "None of the above" as controls. Again, subjects with phenotype codes -1 for "Do not know" or -3 for "Prefer to not answer" were excluded.

For multi-coded phenotypes with exclusive categories (i.e., participants could select only 1 answer), we used information gathered during the first visit and first run (i.e., field codes ending with 0.0, i.e., f.xxx.0.0). These phenotypes were then dichotomized through dummy-coding (e.g., if 5 categories could be selected, 5 dummies were created in which each of the categories was contrasted against the other categories). For instance, the field f.1747.0.0 (hair color) has the following 5 categories; blonde, red, right brown, dark brown and black. In this case, the first dummy trait is blonde vs all others (red, right brown, dark brown and black) and the second is red vs all others (blonde, right brown, dark brown and black) and so on that results in 5 distinct GWAS.

For multi-coded phenotypes with inclusive categories (i.e., participants could select multiple answers, i.e., so-called multiple response questions), we used information of all runs gathered during a single visit (i.e., field codes ending with 0.x, i.e., f.xxx.0.x with exception for f.40006, f.40011 and f.40012 in which we used fields encoding f.xxx.x.0). These phenotypes were then re-coded such that every individual category was treated as a separate dichotomous question. For instance, participants could select multiple options for f.6139 (Gas or solid fuel cooking/heating; f.6139.0.0: "a gas hop or gas cooker", f.6139.0.1: "a gas fire that you used regularly in winter time", and f.6139.0.2: "an open solid fuel fire that you used regularly in winter time"). In this case, each of the three answer options was treated as a separate dichotomous question, with people endorsing the specific option coded as cases (1), and all others coded as controls (0).

The phenotypes gauging substance use (i.e., Smoking status and Alcohol Drinker status: fields f.20116.0.0 and f.20117.0.0, respectively) were originally coded as 0 (Never), 1 (Previous), and 2 (Current). For these phenotypes, we ran two separate case-control GWAS:

a) Never (0) versus Ever (1+2), using the data of all participants, and b) Previous (1) versus Current (2), using only the data of people who had ever used the specific substance.
ICD10 fields often included sub-traits (e.g., ICD10 field A01 has 5 sub-traits coded A01.0 - A01.4). These sub-traits were collapsed with the main category (i.e., endorsement of any of the sub-categories A01.0 - A01.4 was treated as endorsement of the main category A01).
After defining phenotypes, we recounted the total sample size used in the analyses by excluding subjects without genotype data or missing some covariates. To all available phenotypes, we applied the minimum sample size criterion of *i*) at least N=50,000 European subjects, and *ii*) in case of dichotomous phenotypes, both the number of cases and the number of controls exceeding N=10,000. All phenotypes that did not reach this criterion were excluded from further analysis. Note that the final total sample size encoded in the atlas database (http://atlas.ctglab.nl) might be less than 50,000 due to lack of genotype data or missing values in covariates.
GWAS was performed for up to 10,846,944 SNPs with MAF>0.0001 using PLINK 1.9[7], while correcting for array, age (f.54.0.0), sex (f.31.0.0), Townsend deprivation index (f.189.0.0), assessment centre (f.21003.0.0) and 20 PCs (computed based on European subjects). Linear or logistic models were used for quantitative or binary traits, respectively.

3. Pre-processing of GWAS summary statistics

Curated summary statistics were pre-processed to standardize the format. SNPs with p≤0 or >1, or non-numeric values such as "NA" were excluded. For summary statistics with non-hg19 genome coordinates, liftOver software was used to align to hg19. When only rsID was available in the summary statistics file without chromosome and position, genome coordinates were extracted from dbSNP 146. When rsID was missing, it was assigned based on dbSNP 146. When only the effect allele was reported, the other allele was extracted from dbSNP 146.

4. Phenotypic correlations across 458 traits

As we expect not all traits are phenotypically independent, especially the ones in the same trait domain, using the number of associated traits to define the level of pleiotropy is not fair since the number of traits per domain varies, and trait correlations within domains may also vary. To test how dependent the traits are, we extracted phenotypes of 458 UKB2 traits out of

558 selected traits for which we have individual level phenotype information. Then computed the phenotypic correlation (r_p) as the Pearson's correlation of standardized phenotype. We replaced non-significant correlation between pair of traits after Bonferroni correction As expected, our results showed that traits within the same domain have relatively higher phenotypic correlations (average $|r_p|$=0.10) compared to traits from different domains (average $|r_p|$=0.03). We next clustered traits based on the phenotypic correlations using hierarchical clustering by optimizing the number of clusters k while maximizing silhouette score. Non-significant correlations after Bonferroni correction were replaced with zero. We identified 161 clusters (**Supplementary Fig. 2a** and **b**). 45 traits did not cluster together with any other trait (i.e. clusters consisting only one trait) and 128 out of 161 clusters (79.5%) contain traits from the same domain (**Supplementary Fig. 2c**). These results show that, traits are less likely to be clustered with traits from different trait domains and generally tend to cluster within the same domain. This suggests that when assessing pleiotropy purely based on the number of associated traits, (i.e. regardless of trait domains), the observed stronger phenotypic correlations between traits within the same domain may induce pleiotropy. In addition, the 558 traits were not equally distributed across 24 trait domains. Therefore, we primarily defined the level of pleiotropy based on the number of associated domains, to avoid biasing the extent of pleiotropy due to within-domain phenotypic trait correlations. The number of associated traits was only used to distinguish trait-specific associations from potentially pleiotropic associations within domains.

5. Null distribution of trait-associated loci

To test whether the level of locus pleiotropy observed in 558 traits is higher than expected, we obtained two sets of null distribution of the trait-associated loci under a null hypothesis of no pleiotropy, i) across the entire genome and ii) across the genomic region overlapping with one of the trait-associated loci (i.e. 1707.0Mb).

To obtain a null distribution across the genome, we randomly distributed trait-associated loci on the genome for each of 470 traits, 100 times. The number of loci (and their size) per trait was kept constant across the permutations and each permutation consisted of 470 independent permutation of loci. To make sure that all loci from the same trait were LD independent, we used previously defined approximately independent LD blocks[8] obtained from http://bitbucket.org/nygcresearch/ldetect-data. The trait-associated loci from the same trait were assigned to distinct LD blocks at each permutation. Since LD blocks were often

larger than trait-associated loci, the center positions of trait-associated loci were randomly chosen within the assigned LD blocks.

To obtain a null distribution within the genomic region associated with any trait, we randomly assigned the trait-associated loci to one of the 3,362 grouped loci (defined based on the physical overlap) for each of 470 traits, 100 times. Again, the number of loci (and their size) per trait was kept constant across the permutations and each permutation consisted of 470 independent permutation of loci. To make sure that all loci from the same trait were LD independent and located within the 1707.0Mb of the genome, each locus were assigned to distinct grouped loci whose size is larger than the original trait-associated loci, per trait. When the assigned grouped locus is larger than the trait-associated locus, the center positions of trait-associated loci were randomly chosen within the assigned grouped locus.

For each permutation, we computed the total length of the genome covered by trait-associated loci and regions overlapping with loci from more than one trait.

The random permutation across the entire genome showed that 41,533 loci from 470 traits covers on average 87.4% of the genome (2442.7Mb) which is 1.43 times longer than we observed (p=1.2e-218, two-sided t-test). This suggests that trait-associated loci from one trait are more likely to be overlapping with loci from other traits. Indeed, the total length of genomic regions covered by trait-associated loci overlapping with at least one locus from other traits was on average 2383.3Mb consisting of 39,652 loci with density of 5.80 traits/bp based on random permutation, while the observed data showed 1593.3Mb consisting of 40,262 loci with density 8.64 traits/bp (p=1.2e-232).

Subsequently, the random permutation across 1707Mb showed that the total length of trait-associated loci overlapping with at least one locus from other traits was on average 1641.9Mb (96.2% of 1707.0Mb) consisting of 41458 loci. Although, the total length and the number of loci are higher than observed (1593.3Mb consisting of 40262 loci), the density was 8.45 traits/bp which is significantly lower than observed (p=9.2e-122).

Both results showed that trait-associated loci tend to be more densely distributed in a smaller region of the genome than expected and out observation of locus pleiotropy purely based on physical overlap is unlikely to be driven by chance.

6. Colocalization of the trait-associated loci

To distinguish pleiotropy due to LD overlap between associated loci from likely true pleiotropy (i.e. the same causal SNP is associated with multiple trait domains) of the loci, we

colocalized each trait-associated locus with all of the physically overlapping loci using coloc package in R[9] (**Methods**). The *coloc.abf* function assumes a single causal SNP for each trait and estimates the posterior probability of the following 5 scenarios for each testing region; H_0: neither trait has a genetic association, H_1: only trait 1 has a genetic association, H_2: only trait 2 has a genetic association, H_3: both trait 1 and 2 are associated but with different causal SNPs and H_4: both trait 1 and 2 are associated with the same single causal SNP. In this study, as we pre-define the trait-associated loci for each trait which already discard scenarios H_0 to H_2, we are only interested whether H_4 is most likely. We defined a pair of loci sharing a causal SNP when the posterior probability of sharing the same causal SNP was greater than 0.9.

We note that the Bayesian colocalization method we used in this study, implemented in R package *coloc*, assumes a single shared causal variant between two traits. This assumption will be violated in loci where there are multiple shared causal variants. Simulations performed by Giambartolomei *et al.*[9], show that the posterior probability for H_4 (i.e. both trait 1 and 2 are associated with the same, single causal SNP) reached >0.9 more than 50% of the times when two traits were influenced by multiple causal variants in one locus, and multiple causal variants were shared between the traits. This suggests that although we might miss some of the pairs of loci in which multiple causal variants are shared between two traits, the *coloc* method should capture the majority of the loci with *at least one* shared causal variant. Out of 41,533 loci, 40,262 loci were physically overlapping with at least one of the other loci. 88.4% (35,609 loci) of these were colocalized with at least one locus from a different trait, and 55.4% (22,319 loci) were colocalized with trait(s) from different trait domain(s) (**Supplementary Table 6**). We then re-grouped 35,609 loci based on the colocalization pattern, so that a grouped locus is a connected component of the graph consists of loci within a grouped locus based on physical overlap and pairs of colocalized loci are linked (note that it does not require all loci within a group to be colocalized with all other loci; see **Methods** for details). This resulted in 3,534 connected components. For each component, the number of traits and domains were counted (**Supplementary Fig. 4c** and **Supplementary Table 7**). Out of 2,091 grouped loci with more than one associated trait based on physical overlap, 1,453 loci contained more than one connected component (median 3 and average 5.2 independent components; **Supplementary Fig. 3**). For 1,626 multi-domain loci (defined by the physical overlap), the number of associated trait domains of the largest connected component per grouped locus showed on average 38.3% decrease compared to the number of domains in the grouped loci based on physical overlap (**Supplementary Fig. 4d**). In addition, we computed

the proportion of colocalized pairs for each connected component. When the proportion is 1, the connected component is a clique (fully connected graph, i.e. every locus is colocalized with every other). We found that the proportion of colocalized pairs tends to decrease along with the number of trait domains in the components (**Supplementary Fig. 4e**). This suggests that larger connected components likely include "hub" loci that are colocalized with a lot of loci from different traits and domains. However, 67 grouped loci contained associations with ≥5 domains and had a colocalized proportion >0.5, indicating high pleiotropy of these loci.

7. Gene length and LD for MAGMA gene analysis

We showed that more pleiotropic genes are significantly longer compared to genes that are not associated with any of the 558 traits. This is unlikely to be due to the fact that longer genes contain more SNPs and therefore would have a higher chance of including significant associations, since MAGMA controls for this by projecting a SNP matrix on principal components (PC) and pruning away PCs with small eigenvalues[10]. Therefore, the number of parameters (estimated number of LD independent SNPs) per gene is independent on the length of genes. Indeed, the number of parameters done not significantly differ between multi-domain, domain specific, trait specific and non-associated genes, except a slight increase between non-associated genes to multi-domain genes (**Supplementary Fig. 5c** and **Supplementary Table 9**).

8. Distribution of pleiotropic SNPs across chromosomes

To test if the pleiotropic SNPs are evenly distributed across chromosomes, we first pruned the 1,740,179 SNPs and extracted 822,052 independent SNPs at $r2<0.1$ with maximum distance 1Mb. We then performed the following 4 types of tests:

i) Enrichment of significant SNPs in each chromosome

For each chromosome, a Fisher's exact test (two-sided) was performed to test whether the number of significant SNPs in a certain chromosome deviated from that number based on all chromosomes.

ii) Enrichment of pleiotropic SNPs in each chromosome

For each chromosome, a Fisher's exact test (two-sided) was performed to test whether the proportion of pleiotropic SNPs (associated with more than one trait) on o particular chromosome is different from that proportion in all chromosomes.

iii) Enrichment of highly pleiotropic SNPs in each chromosome

Same as above, but comparing the proportion of highly pleiotropic SNPs (associated with more than one trait domain).

iv) Deviation of the level of pleiotropy of SNPs across chromosomes

For each chromosome, a Mann-Whitney U test (one-sided) was performed to test whether the number of associated domains of SNPs associated with at least one trait differed in a specific chromosome compared to other chromosomes. A statistically significant test result suggests a significantly higher level of pleiotropy of SNPs (in terms of the number of associated domains) in the tested chromosome compared to all others.

9. Pleiotropic gene-sets

The level of pleiotropy of a gene-set may depends on the definition of tested gene-sets, e.g. if only a limited number of genes are included in the current gene-sets, this might bias gene-set pleiotropy by the level of pleiotropy of genes available in the gene-sets. Therefore, we evaluated the distribution of pleiotropic genes across analyzed gene-sets. Out of 17,444 genes analyzed in this study (tested in all 558 traits), 16,401 genes (94.0%) were assigned to one of the 10,086 tested gene-sets. Genes that were not included in any of the gene-sets showed a significantly lower level of pleiotropy compared to genes involved in at least one of the gene-sets (p=2.2e-15, two-sided Mann-Whitney U test on the number of associated domains of genes). Therefore, a higher proportion of trait-specific gene-sets compared to trait-specific genes is less likely to be due to the exclusion of highly pleiotropic genes from the tested gene-sets.

10. Generic correlations across 558 traits

Although the proportion of trait pairs with a significant genetic correlation was generally higher within a domain than between domains, it is important to note that the trait domains are assembled based on external definitions which might partially reflect phenotypic

correlations, but not genetic similarity. To identify genetically similar clusters of traits, we performed hierarchical clustering on the r_g matrix and identified 62 clusters by optimizing the number of clusters k with maximizing silhouette score (non-significant r_g*'s* were replaced by zero, see **Methods** for details). One large cluster contained 258 traits (46.2% of 558 traits) that are clustered together mainly due to low r_g (or non-significance due to low power) across traits (**Supplementary Fig. 8**). Apart from that cluster, 18 traits were not clustered with any other traits and 7 clusters contained traits from the same domain, while the remaining 36 clusters contain traits from multiple domains (**Supplementary Fig. 8b**).

11. Distribution of MAF and effect sizes of lead SNPs

To test whether the observed proportion of rare lead SNPs (MAF<0.01) is higher or lower than expected, we compared it with the number of SNPs with MAF<0.01 in UKB2 reference panel, as >85% of traits are based on UKB2. In GWASs performed in this study, 10,846,943 SNPs were tested, of which 33.6% of SNPs had MAF<0.01. The proportion of rare lead SNPs (12.3%) is, therefore, less than expected (p<1e-323, two-sided Fisher's exact test). We note, however, that the lower proportion of rare lead SNPs may be biased by the different QC criteria across different studies, especially for non-UKB2 GWASs (i.e. it is possible that rare SNPs are excluded prior to GWASs).

Rare lead SNPs (MAF<0.01) showed the highest median of squared effect size (β^2) in the cognitive domain (β^2=0.24) followed by psychiatric (β^2=0.17) and mortality (β^2=0.15) domains (**Supplementary Table 20**). For common lead SNPS, most of the domains showed a median of β^2<0.001 except connective tissue (7.4e-3), neoplasms (2.2e-3) and dermatological (1.1e-3) domains. At a trait level, we again observed a wide variety in effect sizes, with the high median β^2 (>0.01) of common SNPs (MAF>=0.01) for lead SNPs of several psychiatric traits, such as 'Felt distant from other people in past month' and 'frequency of memory loss due to drinking alcohol in last year' (**Supplementary Table 20**). Genetic associations for the nutritional, environment and social interactions domains have a higher proportion of rare lead SNPs while the effect sizes of these are moderate compared to other domains (**Supplementary Fig. 10a**). On the other hand, the cognitive, psychiatric and neoplasms domains showed much smaller proportions of rare variants in genetic associations, yet with higher effect size. Within domains, traits also differed with respect to the frequency and effect sized of associated lead SNPs. For example, 'Frequency of tenseness / restlessness

in last 2 weeks' has a very small proportion of rare lead SNPs with high effect size, while 'Frequency of memory loss due to drinking alcohol in last year' has relatively higher proportion of rare lead SNPs with low effect size but higher effect size for common lead SNPs (**Supplementary Fig. 10b** and **Supplementary Table 21**).

12. Evaluation of performance of fine-mapping with reference panel

Benner *et al.* previously reported that the performance of fine-mapping using FINAMEP FINEMAP is affected by *i*) mismatch of reference panel from the GWAS, and *ii*) sample size of the reference panel used to estimate pair-wise LD correlation of SNPs[11].

In this study, 85.8% of analyzed GWASs (out of 558) are based on the UKB2 cohort (including 11 GWASs with meta-analyses with UKB2), and 8.2% of GWASs are based on UKB1 cohort which is subset of UKB2, and the remaining (6%) is non-UKB based. As a reference panel for FINEMAP, we randomly selected 100k individuals of European ancestry from UKB2 to compute a LD matrix. Therefore, a mismatch in reference panel is unlikely to cause any increase in error rates for 94% of the GWASs. For the 6% of remaining GWAS, are not based on the UKB cohort, yet for which UKB was also used as reference, the mismatch of reference panel might increase the error rate. Our alternative reference panel would be ~500 EUR subjects from the 1000 Genome Phase 3. However, as shown by Benner *et al.* a GWAS with 50k subjects requires at least 10k subjects for a reference panel to provide adequate performance[11]. Given that the average sample size was ~125k for the non UKB GWASs, 1000 Genome would clearly be too small to serve as a reference for LD estimation. Since we do not have access to the original cohort of each GWAS, the second-best option is to use UKB2 cohort which is large enough.

To address if 100k is sufficient for a reference panel for fine-mapping given our GWAS sample size, we conducted simulations to evaluate the performance of fine-mapping with different reference sample size. First, 100 non-overlapping loci were randomly selected from 41,041 loci associated with one of the 466 traits (that are fine-mapped in the current study, details are described in the next section). For each locus, SNPs within 50kb around the top SNP or locus boundary, whichever was larger, were selected. For each locus, we selected 5 causal SNPs as follow, the most significant (based on the P-value from the original GWAS) and 4 randomly selected SNPs from the locus. Using genotype data of 387,614 unrelated European subjects from UKB2, we simulated phenotypes as described previously[11] and estimated effect sizes of SNPs (and their standard error) based on linear regression using the

lm function in R. We created 5 sets of 5 causal SNPs for each locus, resulting in a total of 500 simulated data sets. We then performed fine-mapping using FINEMAP with reference panels of full subjects used to computed summary statistics (i.e. ~388k) and randomly selected subset of 100k subjects. A pair-wise LD matrix was created using LDstore software[11] (http://www.christianbenner.com/#). The maximum number of causal SNPs were set to 10. For each simulated data set, the area under the curve (AUC) was computed with ranking SNPs by posterior inclusion probability using the *auc* function implemented in the pROC package in R. To count true and false positives, we defined SNPs belonging to 95% credible sets (set of configs selected until the cumulative sum of the posterior probability reaches 0.95, see next section for more details) as causal defined by fine-mapping, and compared these with the 5 causal SNPs selected for each data set.

The median AUC was 0.79 for both reference panels with full and 100k subjects, and it did not significantly differ (**Supplementary Fig. 11**). The number of true positive and false positive SNPs also did not show significant difference between two sizes of reference panels (**Supplementary Fig. 11**). Therefore, we conclude that 100k is sufficiently large enough to achieve the similar performance as using exact same subjects for estimation of LD.

13. Fine-mapping of the trait-associated loci

Statistical fine-mapping utilizes evidence of the associations at each variant in the loci (effect sizes and LD structure) to assign posterior probability of each specific model at particular locus, which are then used to infer the posterior probabilities of each SNP being included in the model (posterior inclusion probability, PIP) and ascertain the minimum set of SNPs required to capture the likely causal variant. Fine-mapping was performed for a region of 50kb centered around the top SNP of the locus (unless the locus was larger than that window) using FINEMAP software[12]. The maximum number of SNPs in the causal configuration (k) was set to 10 and using randomly selected 100k unrelated European subjects from UKB2 as a reference panel. Loci overlapping with the MHC region (chr6:25Mb-36Mb) and loci with less than 10 SNPs were excluded. In total, 41,041 loci from 466 traits containing 76,732 lead SNPs were fine-mapped.

FINEMAP outputs a set of models (all possible combination of k causal SNPs in a locus) with posterior probability (PP) of being a causal model. A 95% credible set was defined by taking models from the highest PP until the cumulative sum of PP reached 0.95. Then 95% credible set SNPs were defined as unique SNPs included in the 95% credible set of models.

For each SNP, a posterior inclusion probability (PIP) was calculated as the sum of PPs of all models that contains that SNP. To select most likely causal SNPs, we further defined credible SNPs consists of SNPs with PIP>0.95.

The credible sets consist of 2,018,540 SNPs of which 783,067 are unique. 35,292 loci (86.0% of fine-mapped loci) contained the top SNPs in credible sets and 48,411 lead SNPs (63.1%) were included in credible sets. The number of causal SNPs were optimized at k=1 for the majority of loci (26,579 loci, 64.8%) while 982 loci (2.4%) showed k=10, which also suggests that there might be more than 10 as it was restricted to the maximum of 10. Each SNP in 95% credible sets also has posterior inclusion probability (PIP) which measures how likely the SNP is to be included in the causal configs. In total, 6,268 loci (15.3%) contained at least one credible SNPs with PIP> 0.95. This increased to 9,358 loci (22.8%), 16,182 loci (39.4%) and 34,750 (84.7%) by decreasing PIP to 0.8, 0.5 and 0.1, respectively. Loci that do not contain credible SNPs with high PIP are considered as unsolved mainly due to strong LD. Although coverage of true positive causal SNPs by fine-mapping using FINEMAP is high as shown by Benner *et al.*, false positive stayed considerably high[11]. We therefore focused on SNPs with PIP>0.95 in the current study which are referred as credible SNPs in the main text, however, we also provide results of functional enrichments for all SNPs in credible sets, SNPs with PIP>0.1, 0.5 and 0.8 (number of unique credible SNPs were 68,406, 21,702 and 15,176, respectively) in **Supplementary Table 22**.

We compared SNPs in the fine-mapped regions to all SNPs in the genome, and credible SNPs to SNPs in the fine-mapped regions. The enrichment pattern of SNPs in the fine-mapped regions was similar to SNPs in the trait-associated loci; i.e. significant enrichment of SNPs in intronic and flanking regions but the fold enrichment was much smaller (**Fig. 3d** and **Table 2**). This could be because the fine-mapped regions are often larger than the trait-associated loci by taking 50kb around the top SNPs of the trait-associated loci. In contrast, fold enrichment of exonic SNPs was slightly higher than trait-associated loci (**Table 2**). As we observed higher gene-density around the trait-associated loci, expanding the loci resulted in larger proportion of exonic regions. Both active chromatin state and eQTLs were also significantly enriched in SNPs in the fine-mapped regions, however, fold enrichment of eQTLs was notably less than trait-associated loci (**Fig. 3e** and **f** and **Table 2**). Similar to the lead SNPs, credible SNPs showed strong enrichment in exonic (E=3.47) and flanking regions (E=1.50), as well as intronic regions (E=1.14) compared to the SNPs in fine-mapped regions (**Table 2**). The credible SNPs were also significantly enriched in both active chromatin states (E=1.57) and eQTLs (E=4.67) where the enrichment of active chromatin states was

consistent with the lead SNPs while the enrichment of eQTLs was only seen in the credible SNPs but not in lead SNPs (**Fig. 3e** and **f** and **Table 2**). Enrichments of SNPs in exonic and flanking regions as well as in active chromatin and eQTLs were much higher in credible SNPs compared to lead SNPs, indicating that fine-mapping successfully re-assigned higher PIP for functional SNPs. By using different thresholds of PIP (i.e. PIP>0.8, 0.5 or 0.1), the results showed that enrichments of credible SNPs in exonic and flanking regions, active chromatin and eQTLs remained significant at any tested PIP threshold (i.e. PIP >0.8, 0.5 and 0.1; **Supplementary Table 22**). In addition, the proportion of those functional SNPs increased with PIP threshold (**Supplementary Table 22**), overall indicating that within a trait-associated locus, functional SNPs tend to have higher probability of being causal, as expected.

14. SNP heritability estimate based on LD score regression and SumHer

h^2_{SNp} is different from the broad sense heritability (H^2) which is generally based on twin or family studies and inferred from known genetic relationships. When h^2_{SNP} is lower than H^2, this 'still-missing heritability' may indicate that non-additive effects are responsible for some of the H^2 or that variants not included or captured in the GWAS such as rare variants or structural variants make a substantial contribution. In addition, phenotypic heterogeneity of the trait might also explain the low h^2_{SNP} compared to H^2[13]. Although h^2_{SNP} itself is not directly informative of the genetic architecture of a trait (in terms of the number of associated SNPs, allele frequencies or the effect sizes), it does provide insights into the relative contribution of additive SNPs effects across traits.

LDSC was introduced in 2015 and has been widely used since then to estimate h^2_{SNP} from summary statistics. LDSC assumes that the contribution of each SNP to the total h^2_{SNP} is constant[14]. Recently, the LDAK model has been introduced as an alternative, relying on the assumption that the contribution of each SNP to the total h^2_{SNP} is a function of its MAF and local LD[15]. SumHer implements the LDAK model, and like LDSC, it can be used to estimate h^2_{SNP} from summary statistics[16]. The models underlying LDSC and SumHer thus rely on different assumptions, and therefore estimated h^2_{SNP} values may differ between LDSC and SumHer. To facilitate a fair comparison between the two methods, we used the 1000 Genome Phase 3 European reference panel[4] and limited our analyses to HapMap3 SNPs (see **Methods**).

Overall, 74.7% (417/558) of traits showed higher estimates in SumHer than in LDSC (**Fig. 4a** and **Supplementary Table 23**), of which 268 traits showed more than a 2-fold increase. As

Speed *et al.* reported in their studies based on simulations, LDSC tends to under-estimate the h^2_{SNP} when a phenotype follows the LDAK model but at the same time SumHer tends to overestimate when a trait follows the LDSC model[15,16]. Notably, there were several traits with unexpectedly high h^2_{SNP} from SumHer, and for 8 traits, SumHer estimated $h^2_{SNP}>1$ which also had very high standard error (average 0.33). This is possible because the model is not designed to restrict the estimated of h^2_{SNP} between 0 and 1. In addition, as presented by Evans *et al.*, LDAK is highly sensitive to misspecification of the model assumptions; if causal variants of a trait are mainly common, LDAK model can highly bias the h^2_{SNP} estimate[17].

The main difference between LDSC and SumHer is in the weighting of SNPs. The contribution of *i*th SNP to the total SNP heritability in SumHer is defined as

$$h_i^2 = n_i \frac{q_i}{Q} h_{SNP}^2$$

with

$$q_i = \frac{[p_i(1-p_i)]^{1+\alpha}}{l_i^2}, \qquad Q = \sum_i q_i$$

where n_i is sample size, p_i is MAF and l_i^2 is LD score of the *i*th SNP. In LD score, α=-1 thus q=1 and Q is equal to the number of SNPs. From the study of Speed *et al.*[15], the recommended default is α=-0.25 which we also used in this study. To further investigate higher estimation of h^2_{SNP} by SumHer, we additionally performed SumHer with the following parameter settings and compared these with the estimates from the default SumHer model.

Model 1: $l_i^2 = l_i^2$ and α=-0.25 (default SumHer model)

Model 2: l_i^2= 1 and α=-0.25

Model 3: $l_i^2 = l_i^2$ and α=-1

Model 4: l_i^2= 1 and α=-1 (same as LDSC)

Note that model 4 is essentially the same model as LDSC, and model 2 is LDSC with MAF with and model 3 is LDSC with LD score weight. In each model, traits with Z score (estimated h^2/standard error) <2 were excluded from the analyses. We observed that estimates from model 2 is much less than model 1, which showed median of 66% decrease (**Supplementary Figure 12a**). This indicate large contribution of the LD score weight to the high estimate of SNP heritability. On the other hand, model 3 showed median of 15% increase from model 1 (**Supplementary Figure 12b**) and lastly, model 4 shoed median of 64% decrease from model 1 (**Supplementary Figure 12c**). Therefore, effects of MAF

weighting is much smaller than LD score weighting. In addition, the results suggest that LD score and MAF weights tend to bias SNP heritability in up and downwards, respectively, although it varies across the traits. Since the true underlying genetic model of a trait is unknown, the true value of h^2_{SNP} could be much higher than the LDSC estimates (which can be considered as a lower bound). However, the results clearly showed that unrealistically high estimates of SNP heritability in SumHer default model is mainly driven by LD score weight rather than MAF weight.

We also compared SumHer model 4 and LDSC estimates. Theoretically they are the same model, however, we observed that lower estimation of SumHer model 4 compared to LDSC with median 34% decrease (**Supplementary Figure 12d**). We note that possible reasons are difference of SNPs filtering (i.e. SNPs with chi-square>80 were filtered out from both LDSC and SumHer but additional SNPs which are in LD with one of them at r^2>0.1 were further excluded in SumHer), the reference panel, and implementation of the software. However, a thorough statistical evaluation of SumHer is beyond the scope of the current study.

Supplementary Figures

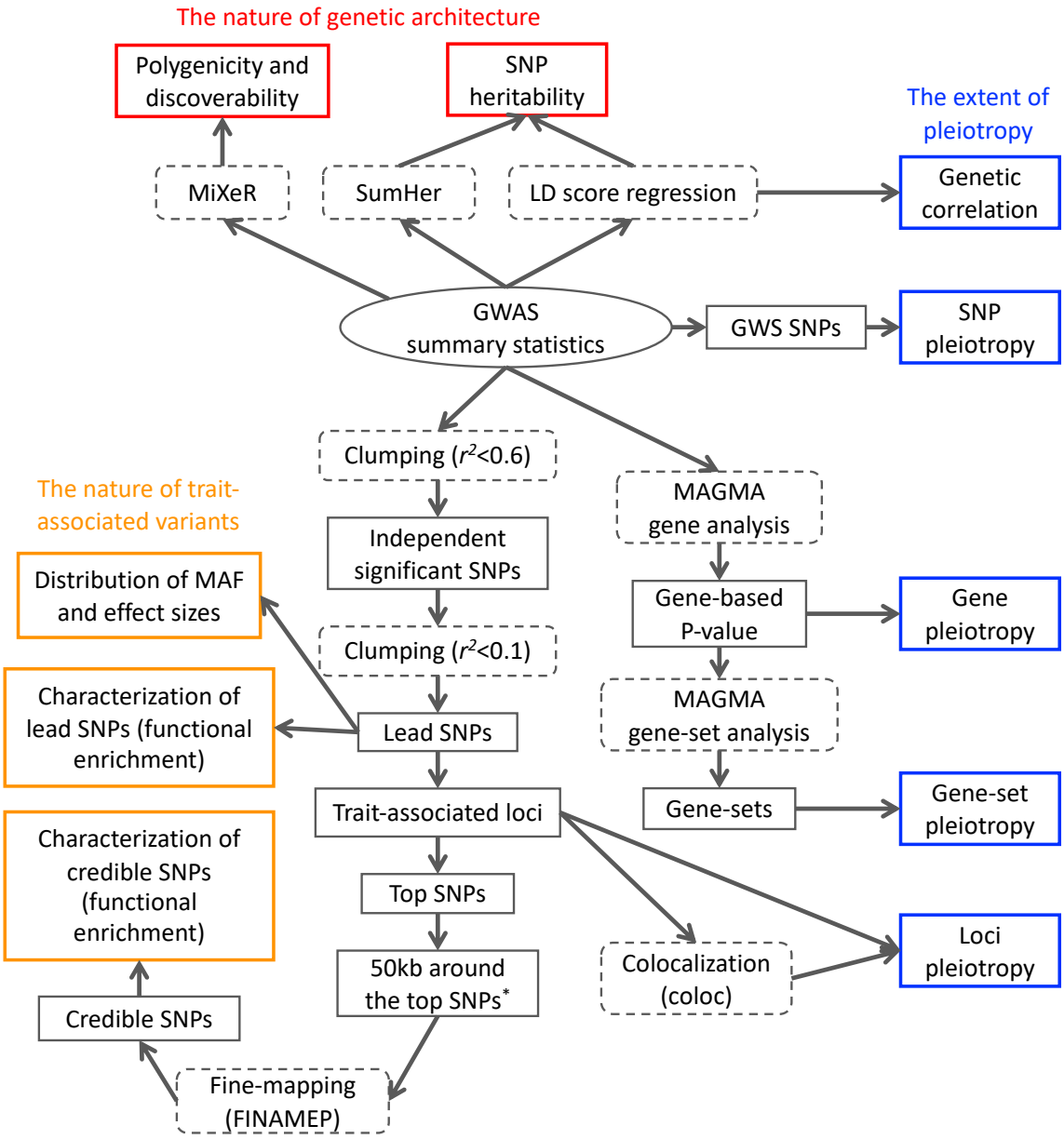


Fig. 1 Schematic overview of the analyses
Dashed grey boxes represent processes and/or software. Solid grey boxes represent results or outputs of the processes. Colored boxes are main analyses in this study, each corresponds a sub-heading in the main text. *Fine-mapped regions were defined by taking the largest range of 50kb around the top (most significant) SNPs and trait-associated loci.

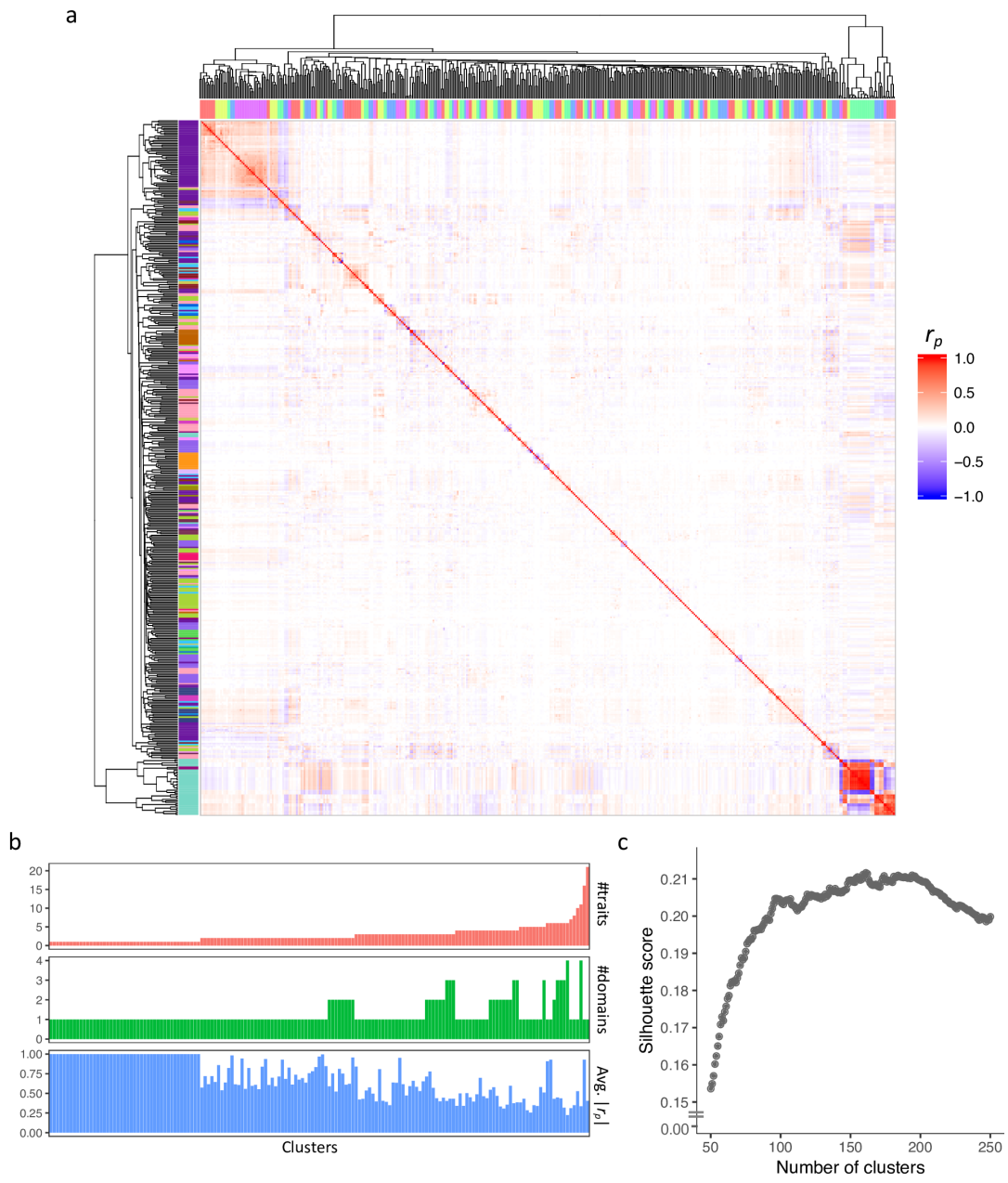


Fig. 2. Phenotypic correlation across 457 UKB2 phenotypes. a. Heatmap of pair-wise phenotypic correlation matrix. Both the columns and rows are hierarchically clustered. The colors at the top (x-axis) represent the clusters (consecutive traits with the same color are in the same cluster) and the colors at the left (y-axis) represent the domain of the trait. **b.** Number of traits, domains and average phenotypic correlation per cluster. **c**. Silhouette score with different number of clusters.

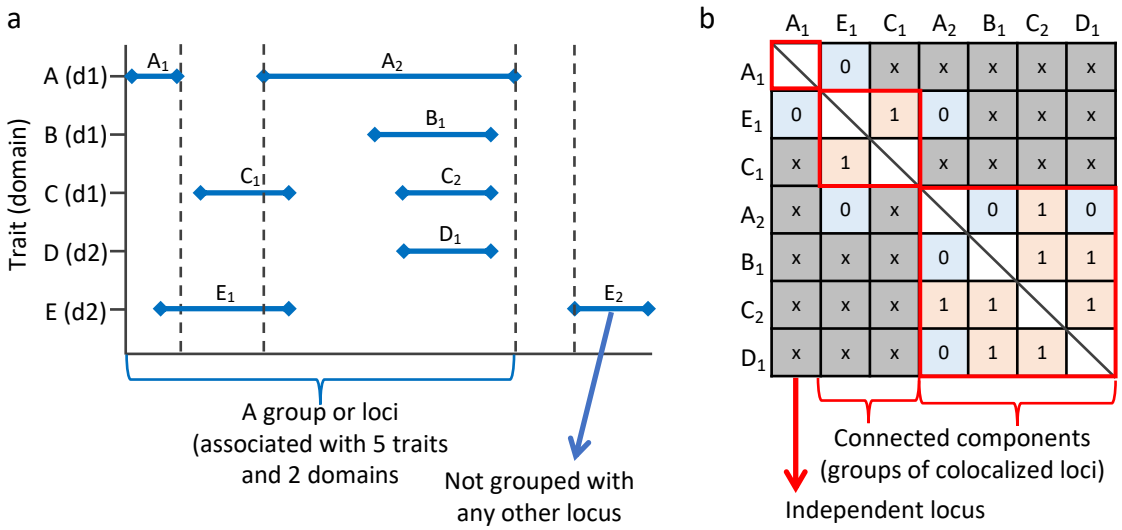


Fig. 3. Definition of grouped loci based on physical overlap and colocalization. a. A group of loci defined by physical overlap. X-axis is a chromosome coordinate, Y-axis is trait (domain in parentheses). Horizontal blue lines represent trait-associated loci. Seven loci except E_1 are, in this case, grouped into a single consecutive genomic region. This loci group is considered to be associated with 5 traits and 2 domains in the analyses. **b.** A matrix of colocalization pattern across loci within a physically overlapping group defined in **A**. Pair of loci is denoted as 'x' when there is no physical overlapped (colocalization was not performed), 0 when loci are not colocalized and 1 when loci are colocalized (sharing the same causal SNP). Red boxes represent connected component by considering the matrix as a undirected graph and cells with 1 as edges. In this case, there are 3 independent loci grouped based on colocalization and, for each group, the number of traits and domains are counted.

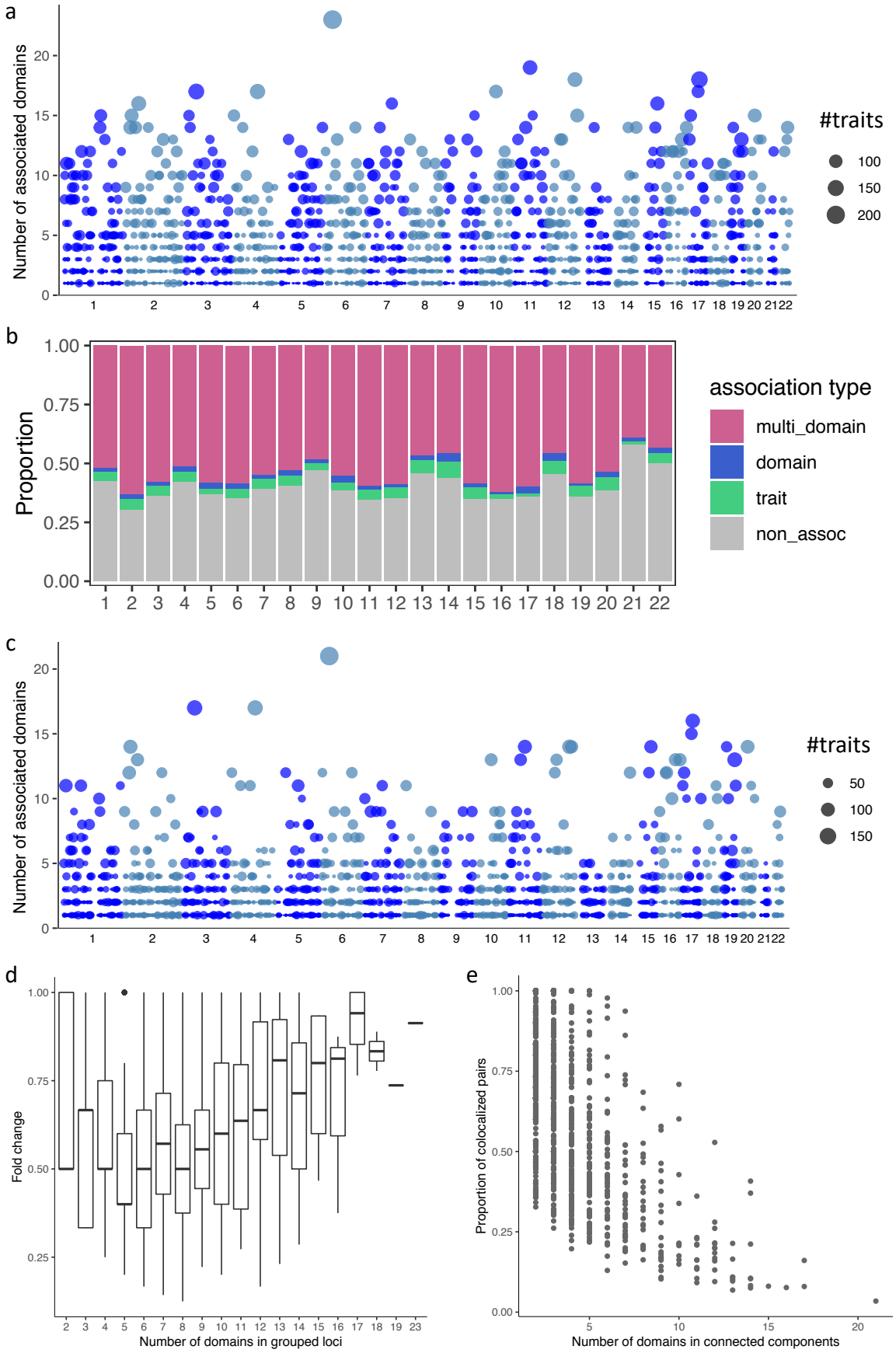


Fig. 4. Distribution of pleiotropic risk loci and colocalization. a. Manhattan like plot of pleiotropic trait-associated loci. Each data point represents a group of trait-associated loci

based on physical overlap. Full results are available in **Supplementary Table 4**. **b.** Proportion of the length of trait-associated loci with different association types per chromosome. multi_domain: associated with traits from >1 domain, domain: associated with >1 trait from a single domain, trait: associated with a single trait, non_assoc: not associated with any of 558 traits. **c**. Manhattan like plot for groups of colocalized loci (connected components of colocalized loci). Each data point represents the largest connected component of colocalized loci per a group of physically overlapping loci. **d.** Box plot of the proportional decrease of the number of domains in a connected component of colocalized loci compared to the grouped loci based on physical overlap. Outliers are plotted as dots. For each grouped locus, the largest connected component was selected. The x-axis is the number of domains in grouped loci based on physical overlap. **d.** The proportion of colocalized locus pairs per connected component of colocalized loci.

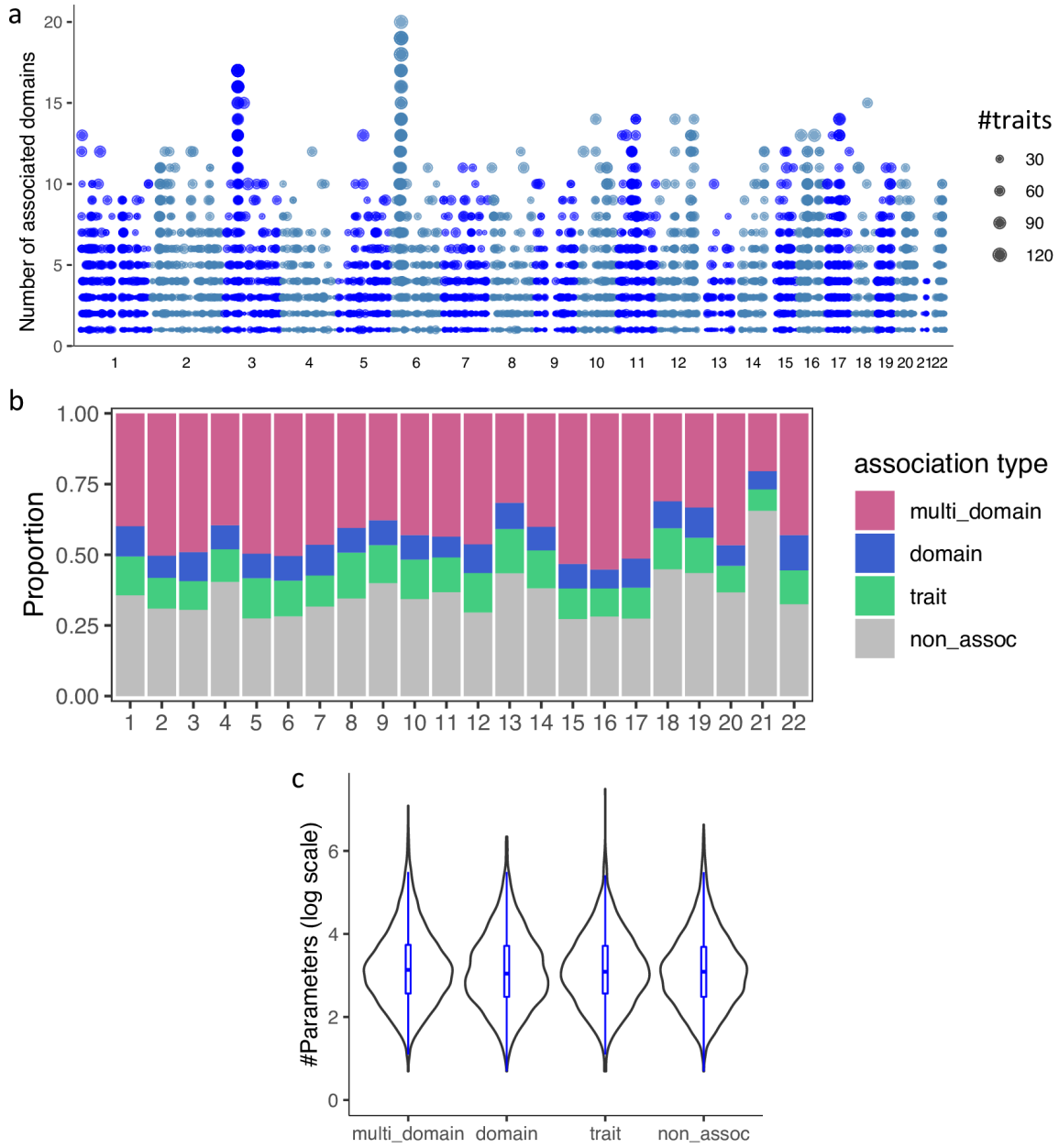


Fig. 5. Distribution of pleiotropic genes and features of genes and gene-sets. a. Manhattan like plot of pleiotropic genes, every dot represents a single gene. Full results are available in **Supplementary Table 7**. **b.** Proportion of genes with different association types per chromosome. multi_domain: associated with traits from >1 domain, domain: associated with >1 trait from a single domain, trait: associated with a single trait, non_assoc: not associated with any of 558 traits. **c.** Distribution of the number of parameters used in the MAGMA gene analysis with different association types.

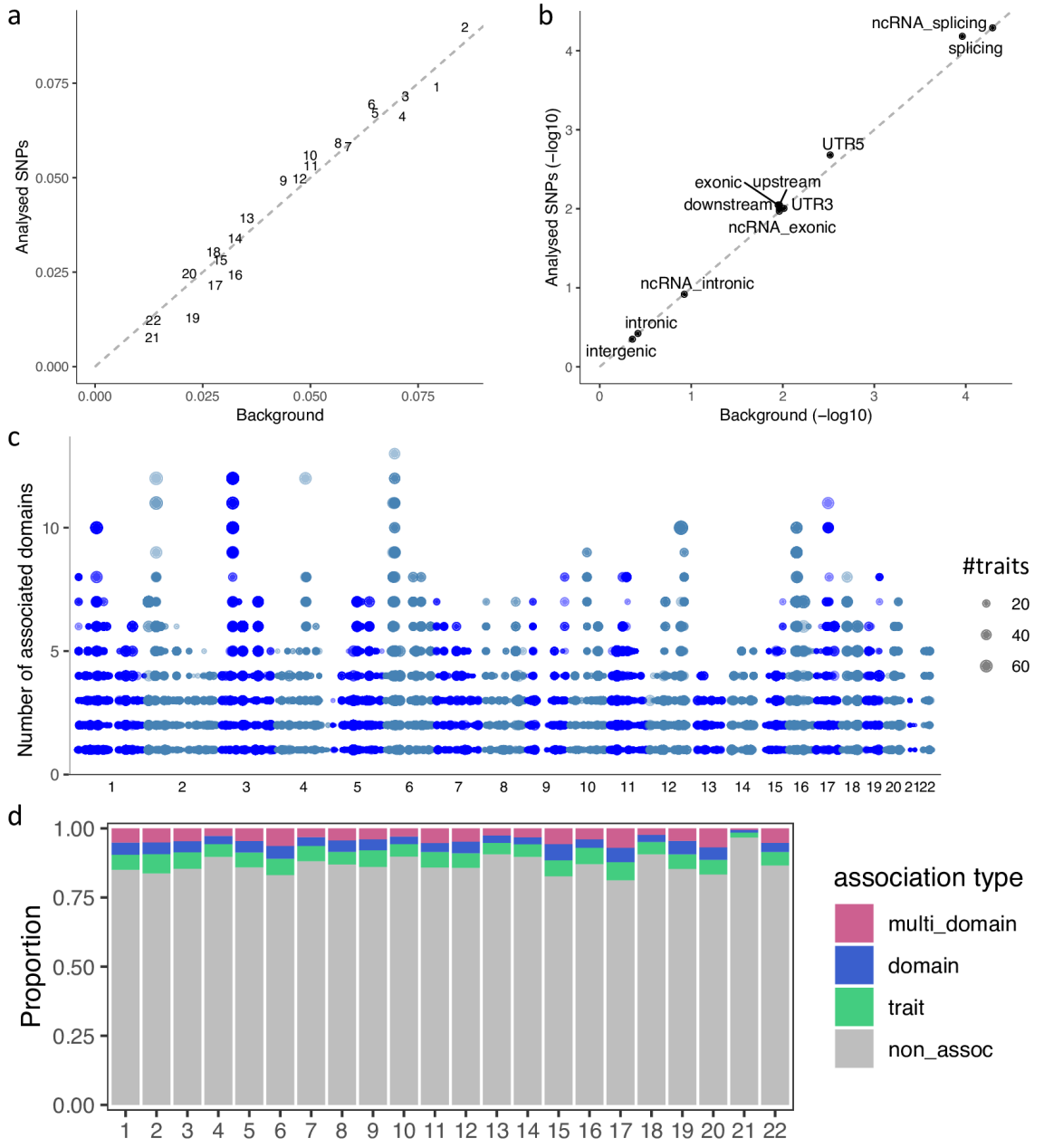


Fig. 6. Distribution of pleiotropic SNPs. a. Distribution of all SNPs in the genome (x-axis) and analyzed SNPs (y-axis) across chromosomes. **b.** Distribution of functional consequences of all SNPs in the genome (x-axis) and analyzed SNPs (y-axis). **c.** Manhattan like plot of pleiotropic SNPs. Each data point represents a single SNP and only SNPs associated with >1 trait are displayed. Full results are available in **Supplementary Table 12**. **d.** Proportion of SNPs with different association types per chromosome. multi_domain: associated with traits from >1 domain, domain: associated with >1 trait from a single domain, trait: associated with a single trait, non_assoc: not associated with any of 558 traits.

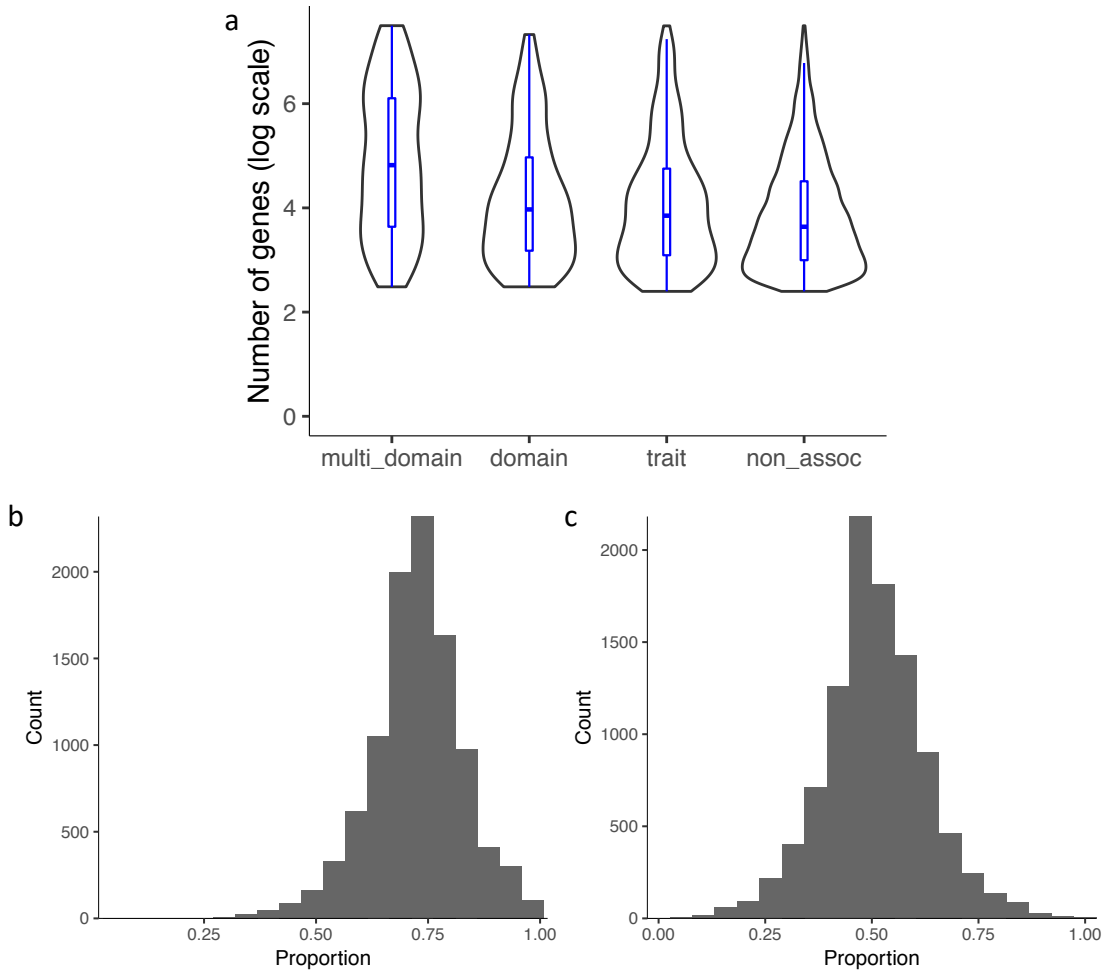


Fig. 7. Distribution of pleiotropic gene-sets. a. Distribution of the number of genes in the gene-sets with different association types. multi_domain: associated with traits from >1 domain, domain: associated with >1 trait from a single domain, trait: associated with a single trait, non_assoc: not associated with any of 558 traits. **b.** Histogram of the proportion of genes associated with at least one of the 558 traits in a gene-set. **c.** Histogram of the proportion of genes associated with traits from more than one domain in a gene-set.

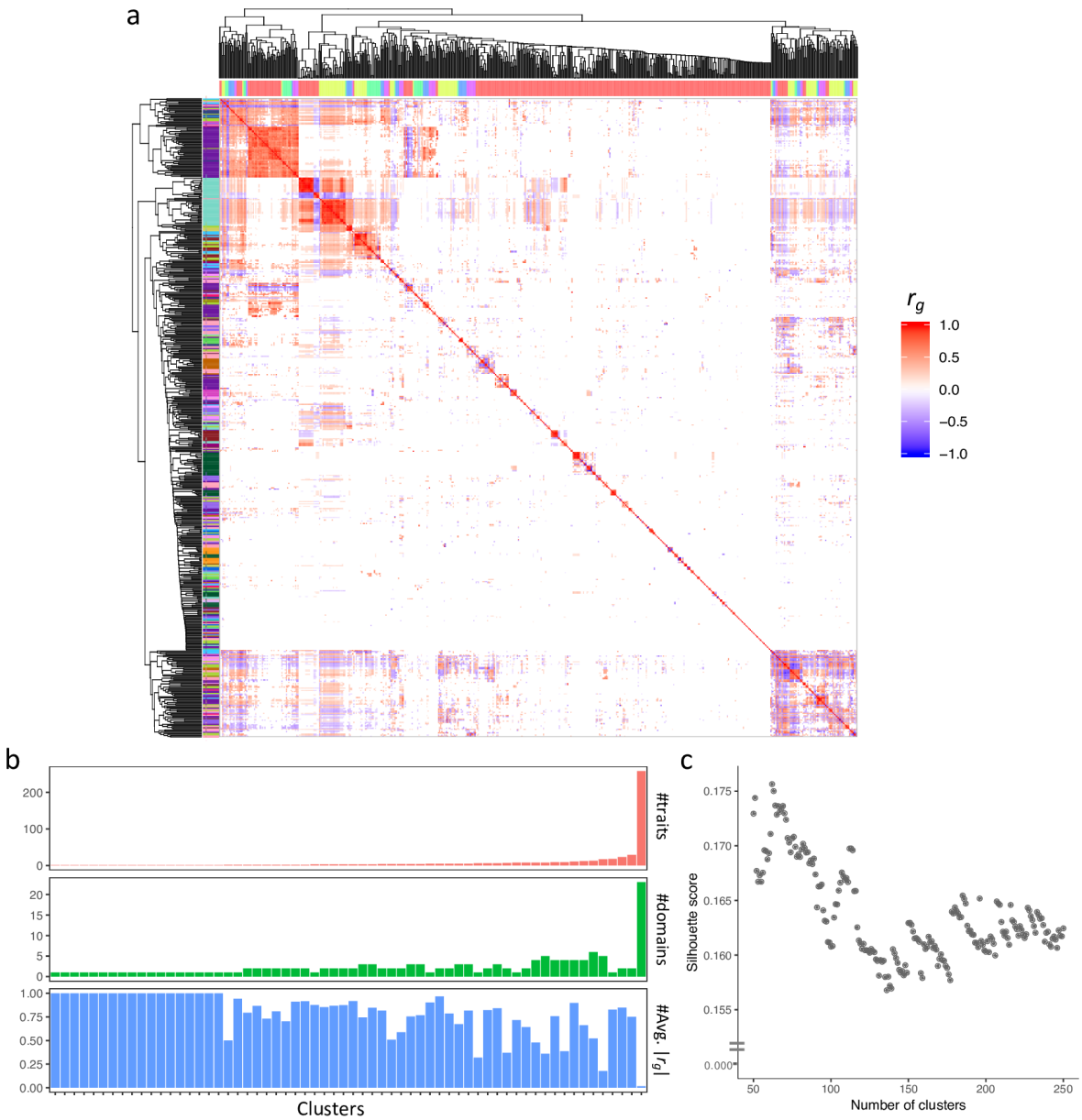
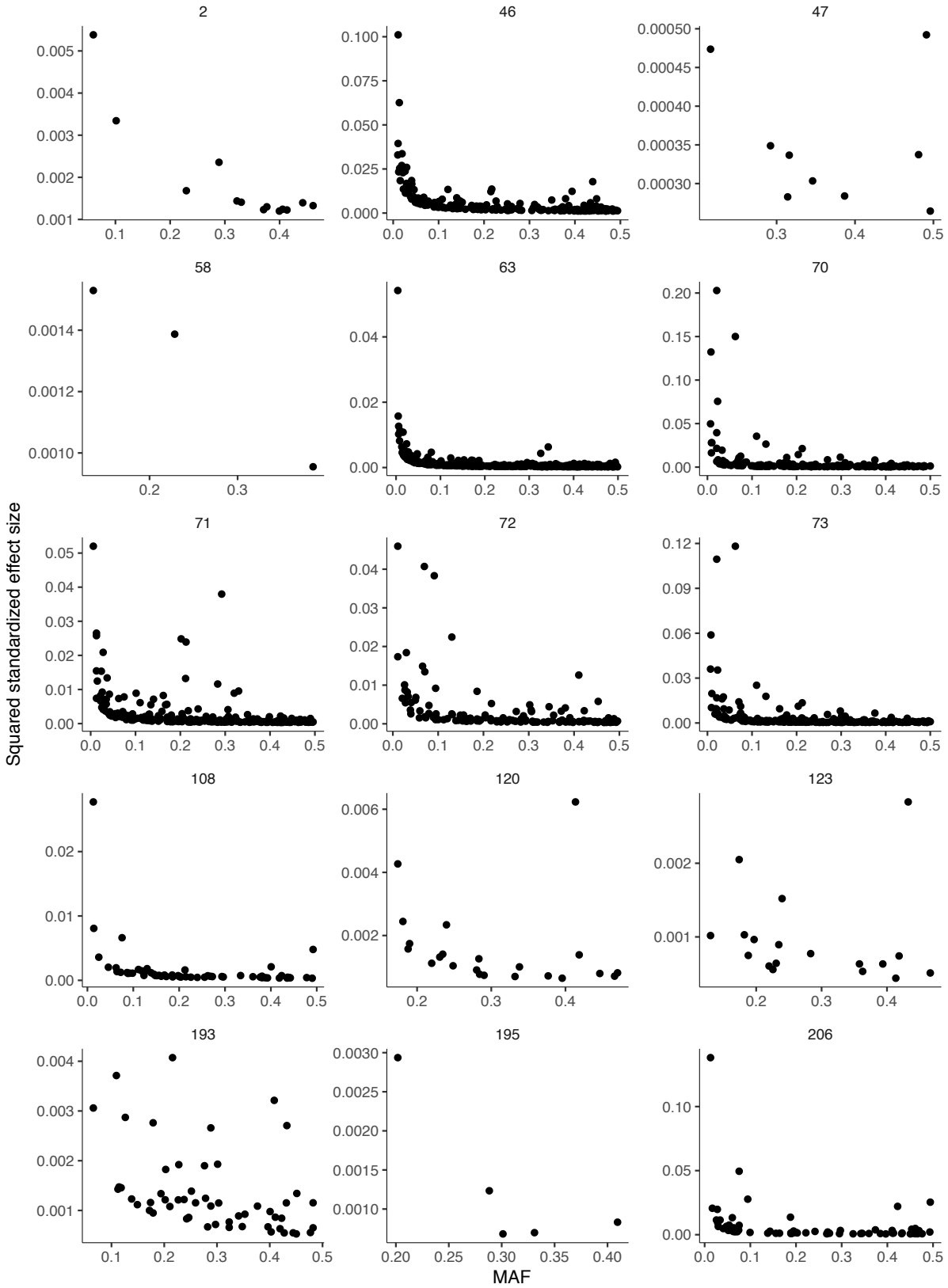
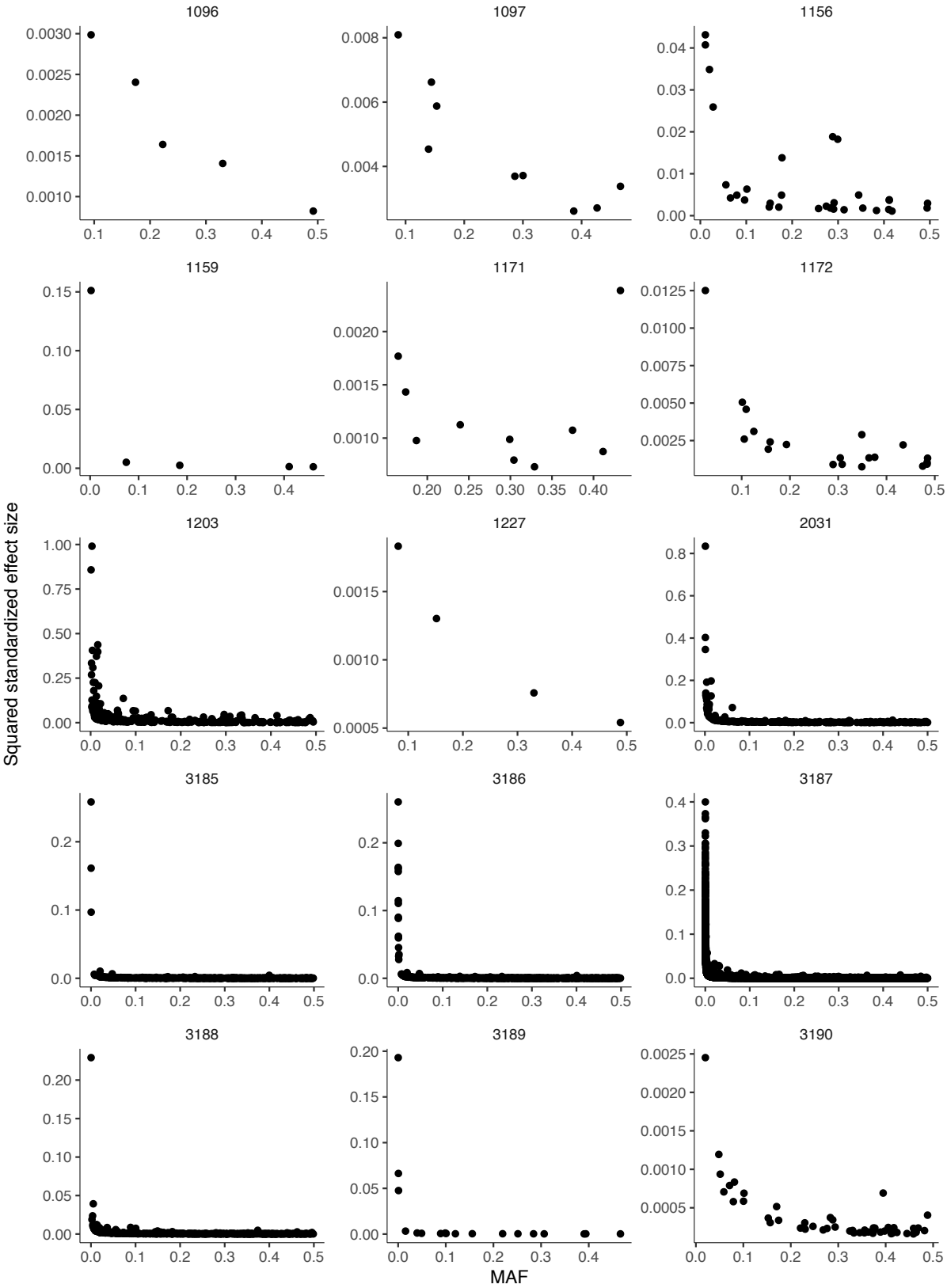
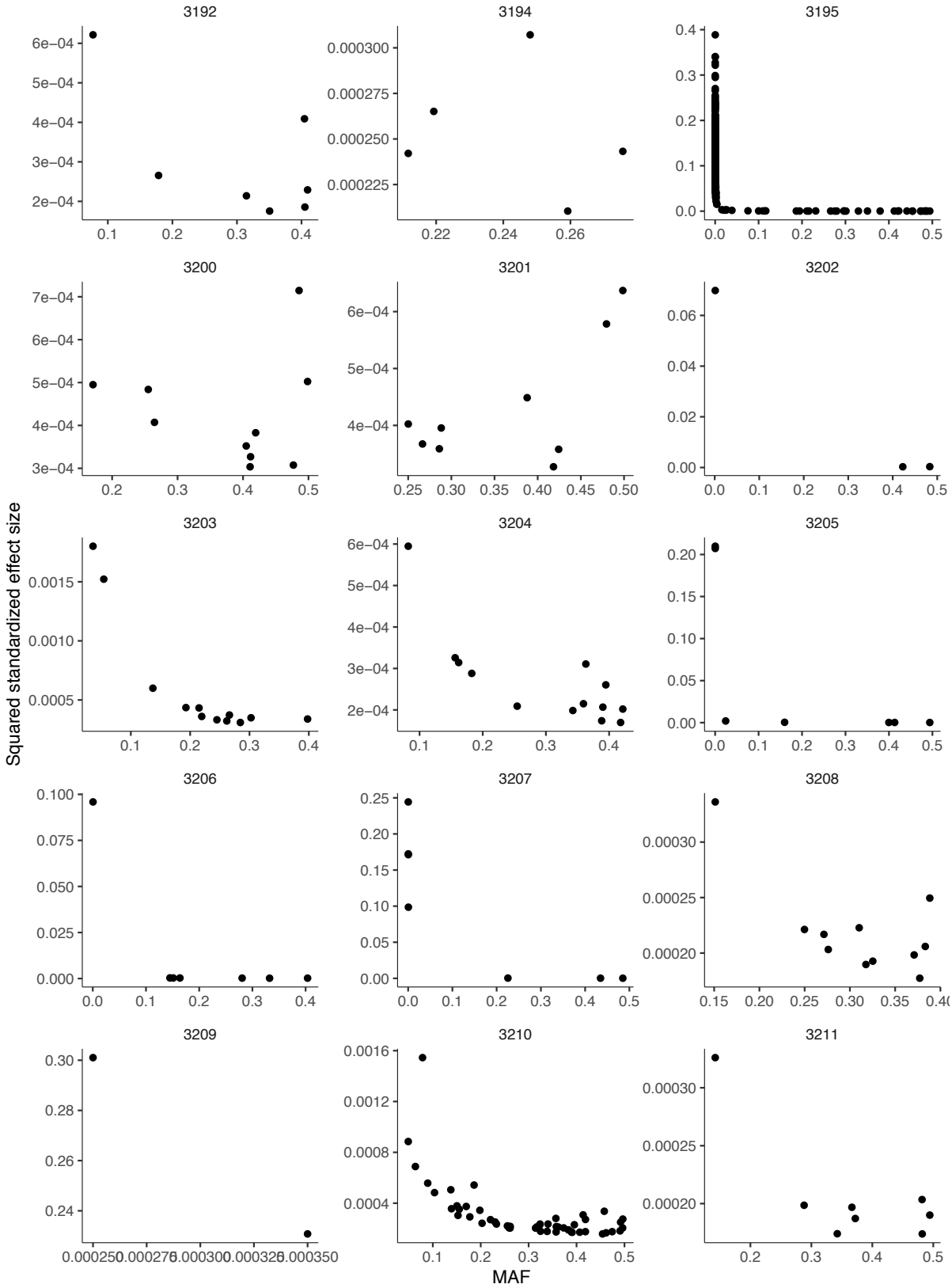
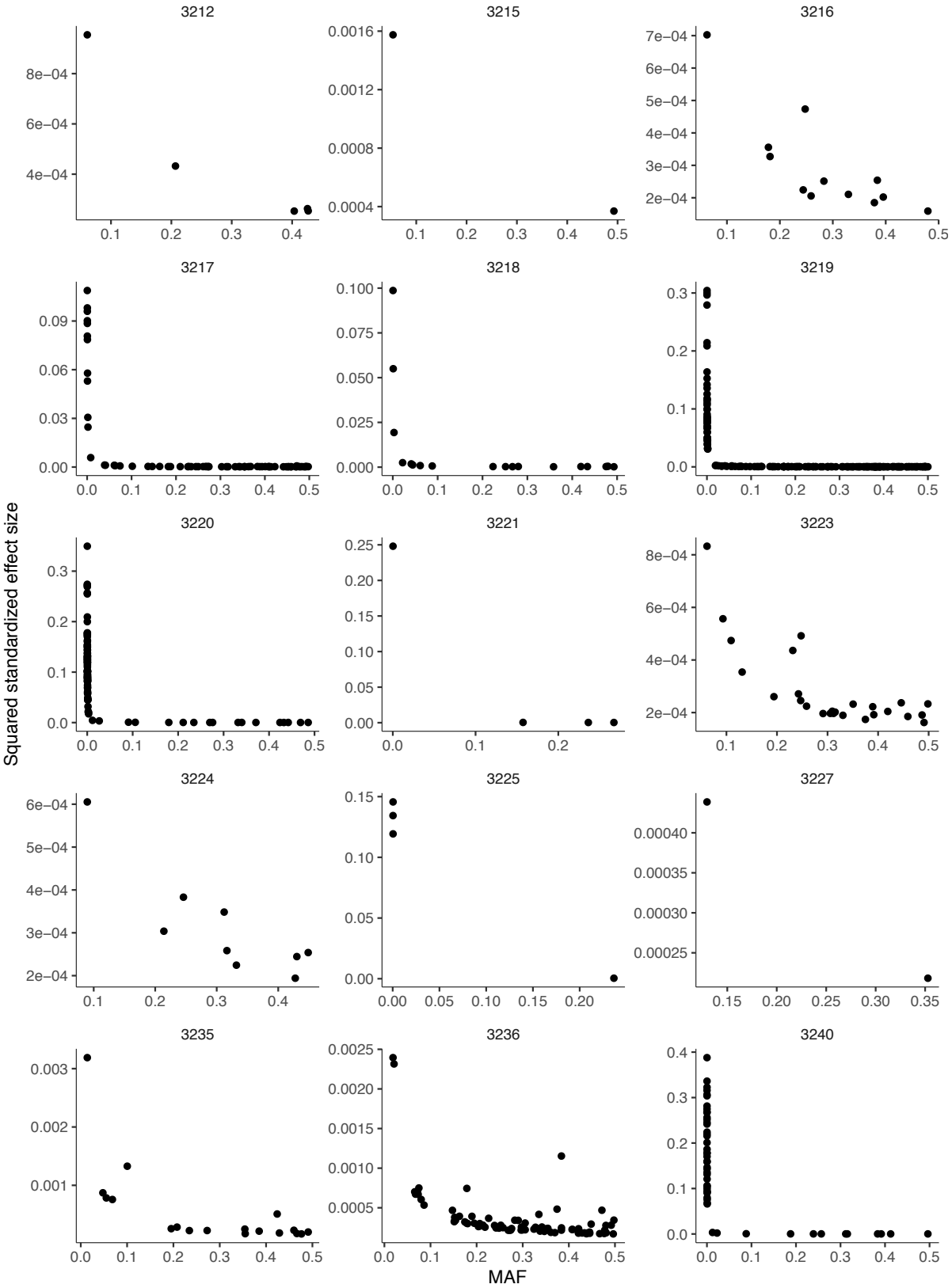


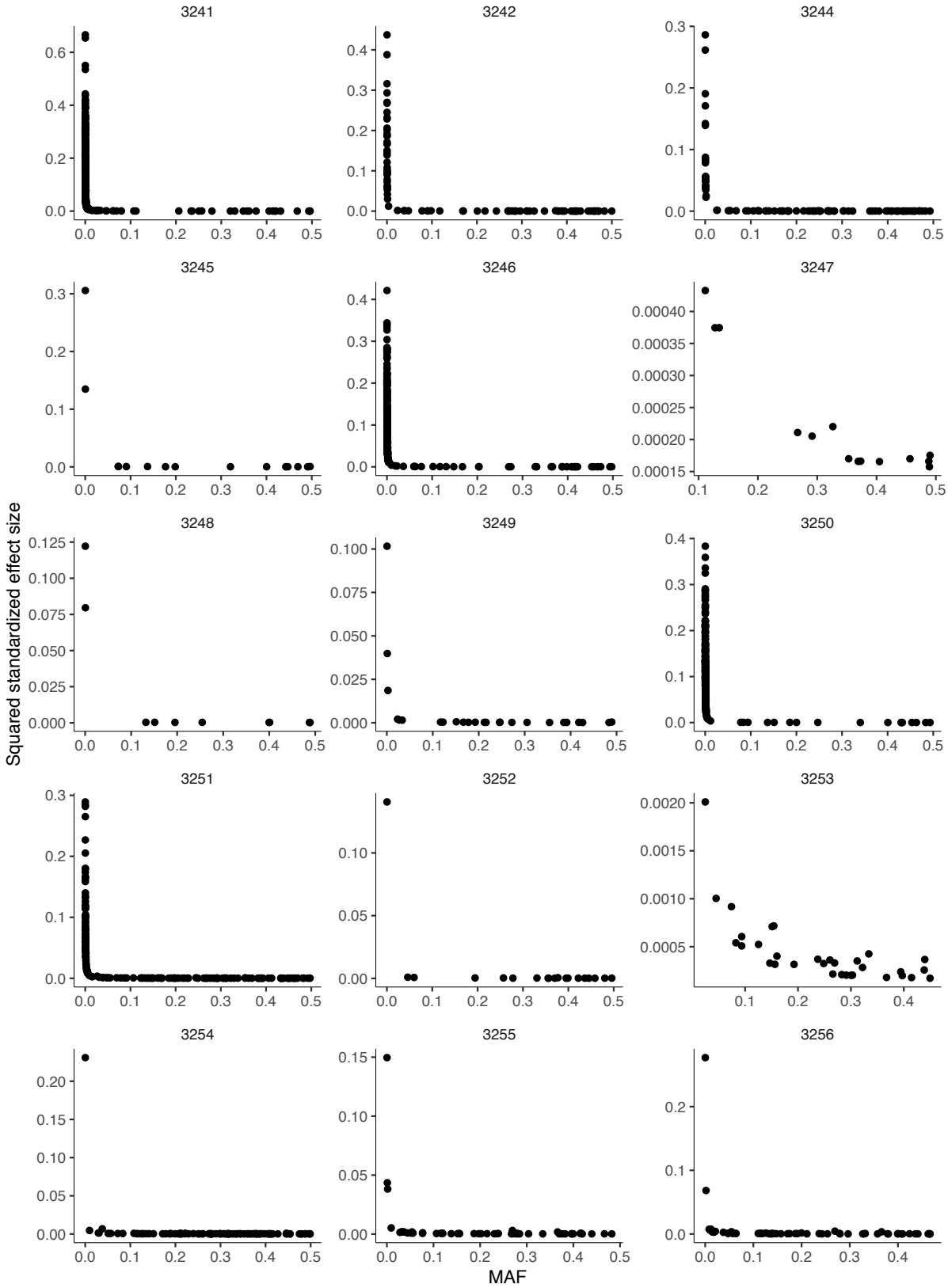
Fig. 8. Features of trait clusters based on genetic correlation. a. Heatmap of the genetic correlation (r_g) matrix. Both columns and rows are hierarchically clustered. The color bar at the top (x-axis) represents the clusters (consecutive traits with the same color are in the same cluster) and the color bar on the left (y-axis) represent domains of the traits. **b.** Number of traits, domains and average genetic correlation per cluster. **c.** Silhouette score with different number of clusters.

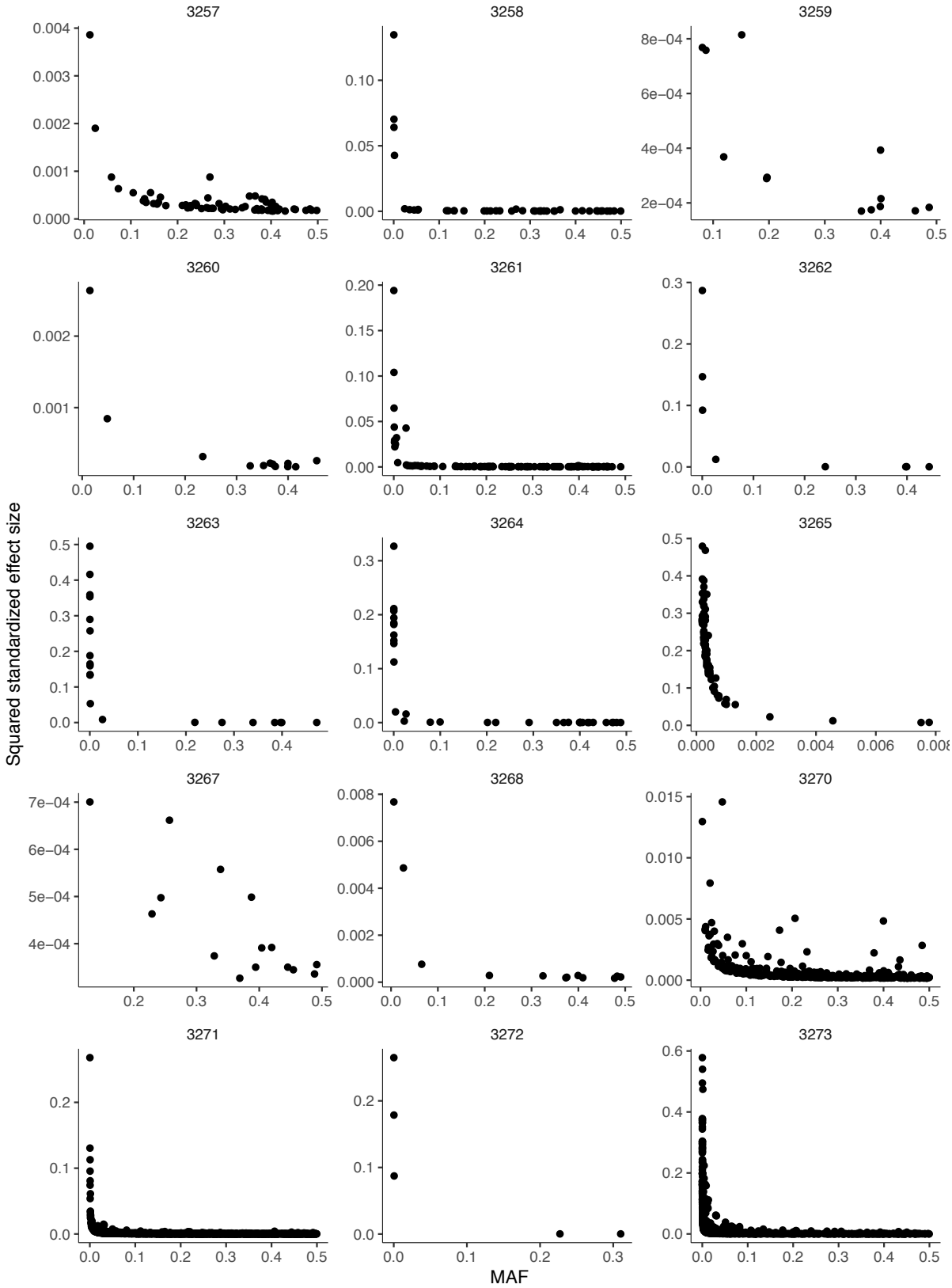

2
46
47
58
63
70
71
72
73
108
120
123
193
195
206
Squared standardized effect size
MAF

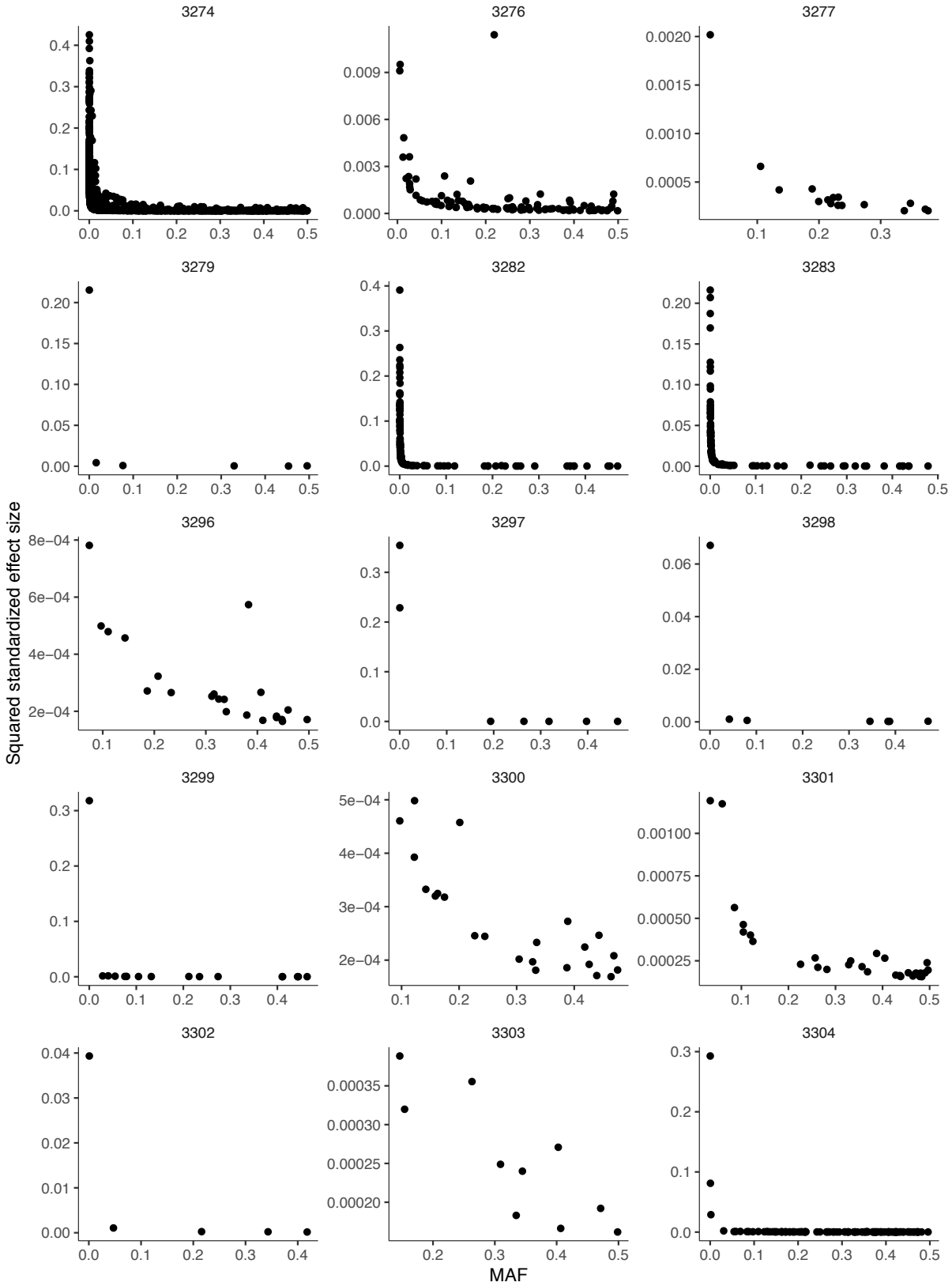

1096
1097
1156
1159
1171
1172
1203
1227
2031
3185
3186
3187
3188
3189
3190
Squared standardized effect size
MAF

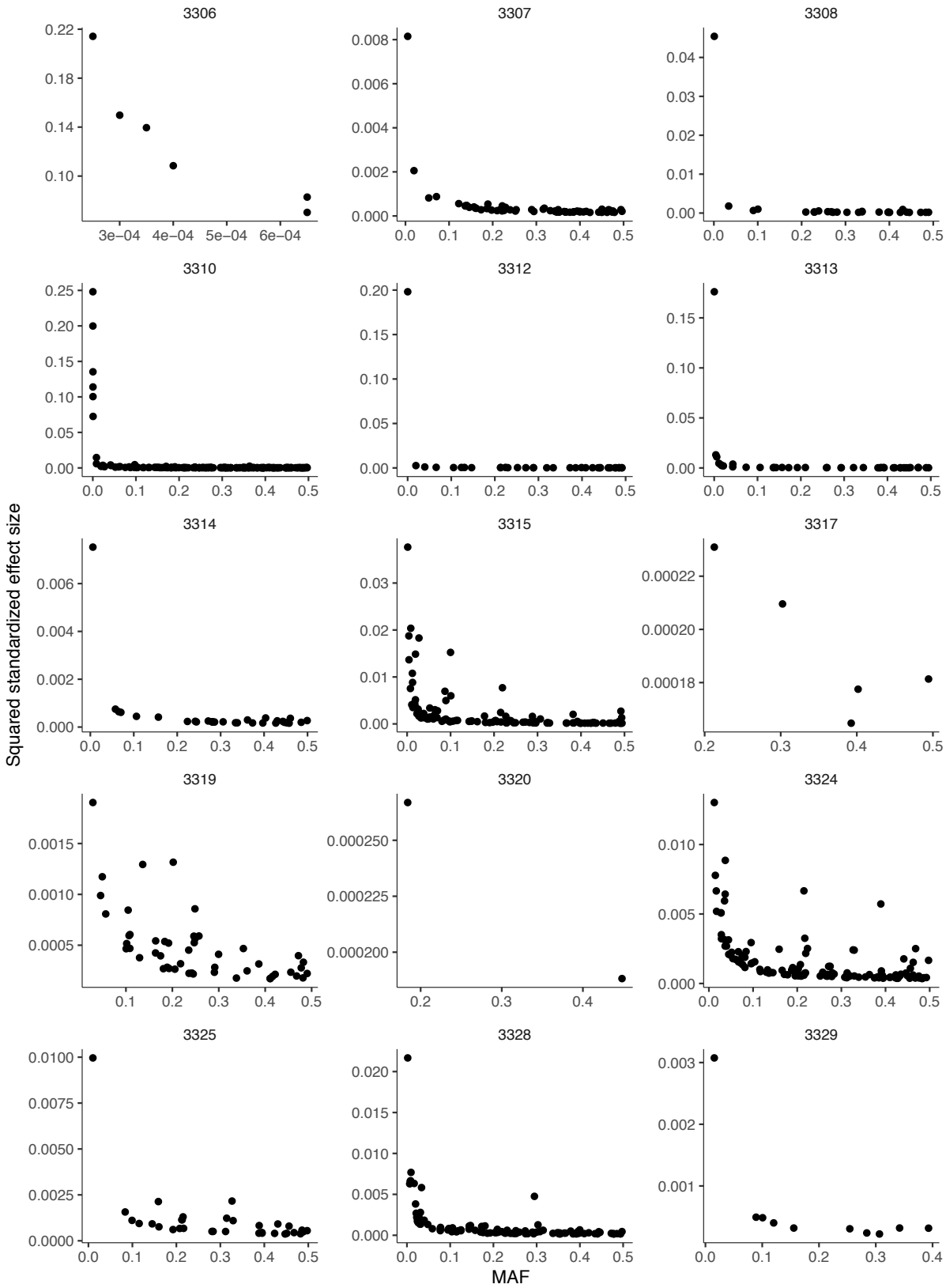

3192
3194
3195
3200
3201
3202
3203
3204
3205
3206
3207
3208
3209
3210
3211
Squared standardized effect size
MAF

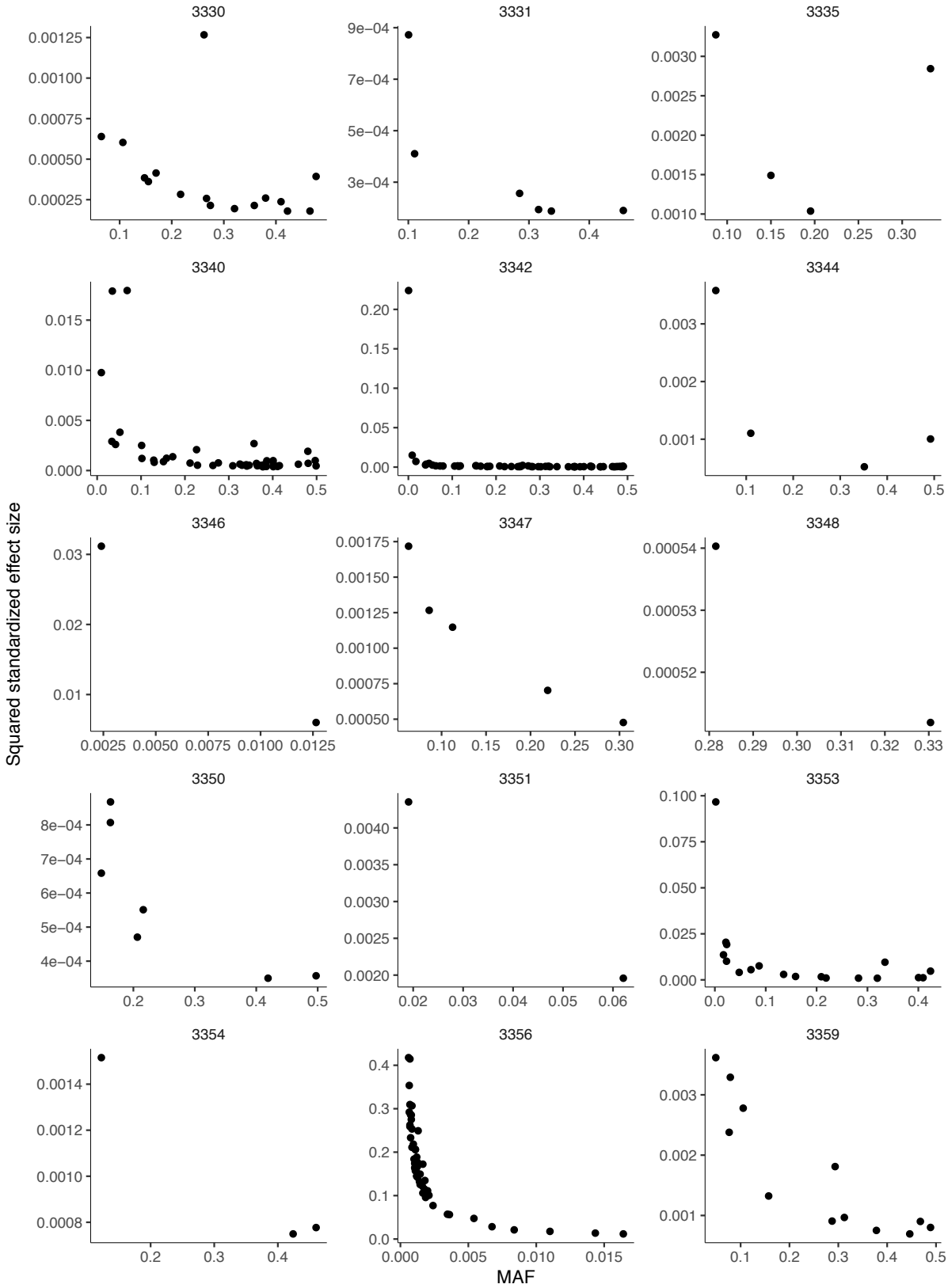

3212
3215
3216
3217
3218
3219
3220
3221
3223
3224
3225
3227
3235
3236
3240
Squared standardized effect size
MAF

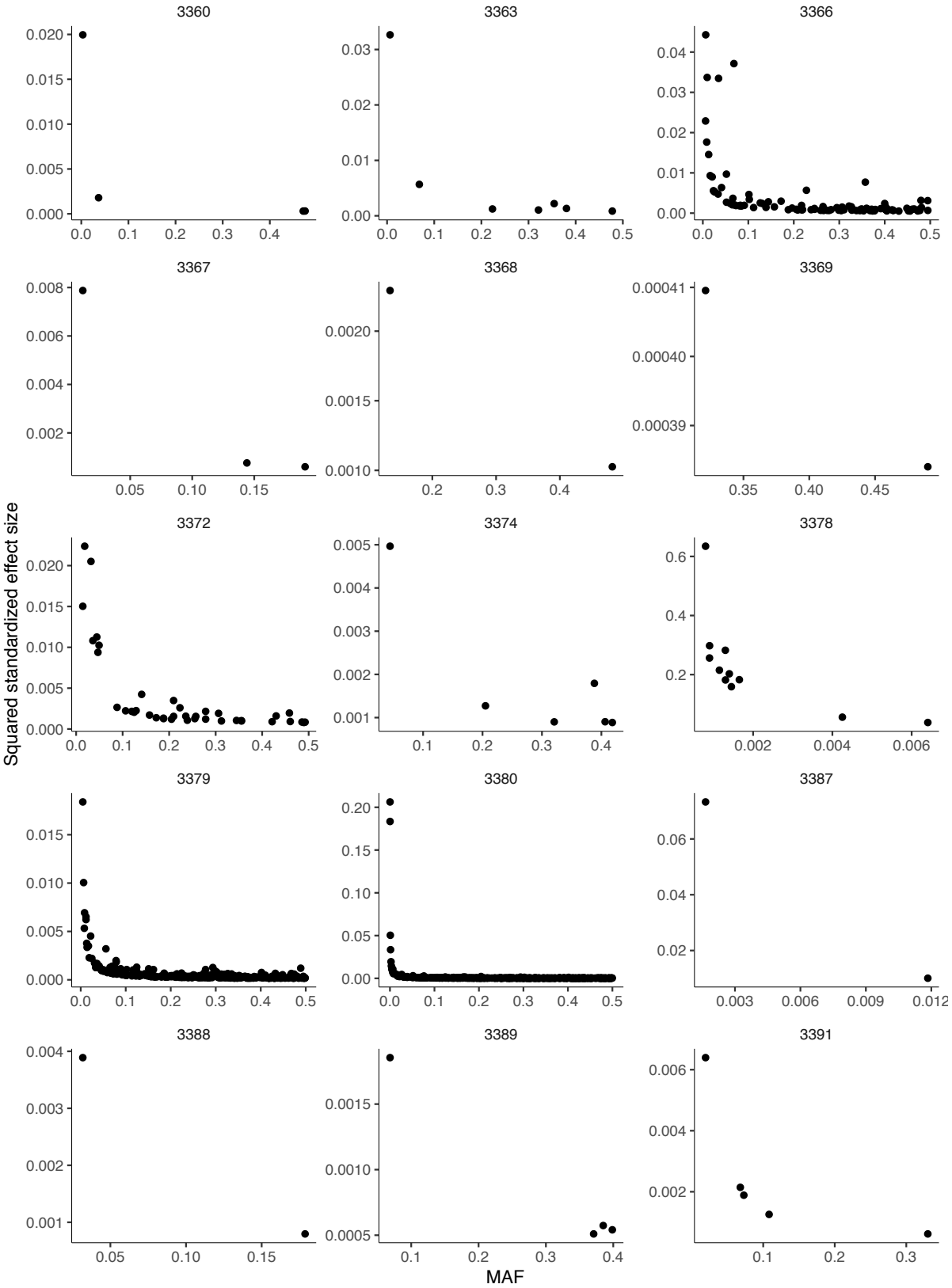

3241
3242
3244
3245
3246
3247
3248
3249
3250
3251
3252
3253
3254
3255
3256
Squared standardized effect size
MAF

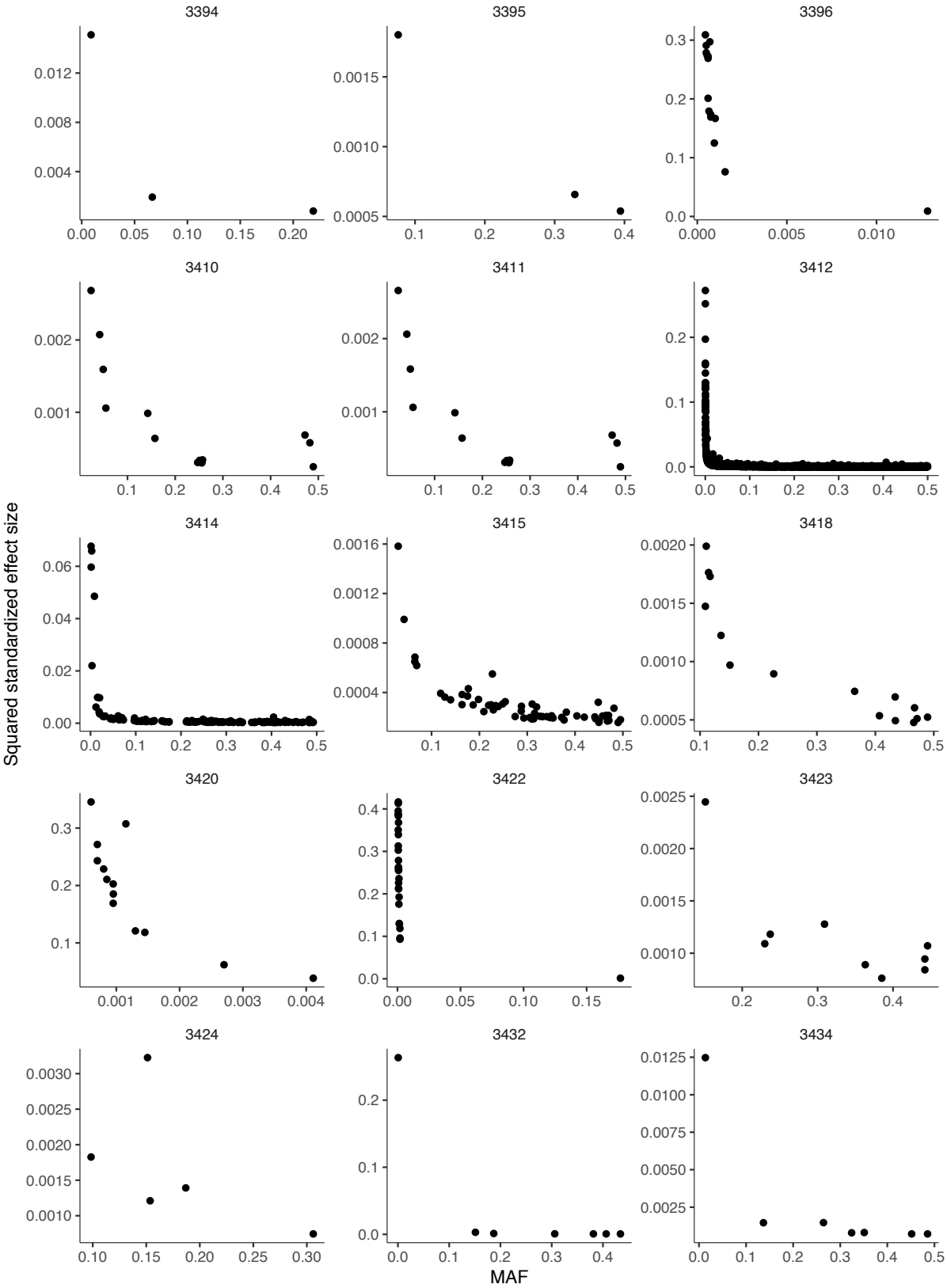

3257
3258
3259
3260
3261
3262
3263
3264
3265
3267
3268
3270
3271
3272
3273
Squared standardized effect size
MAF

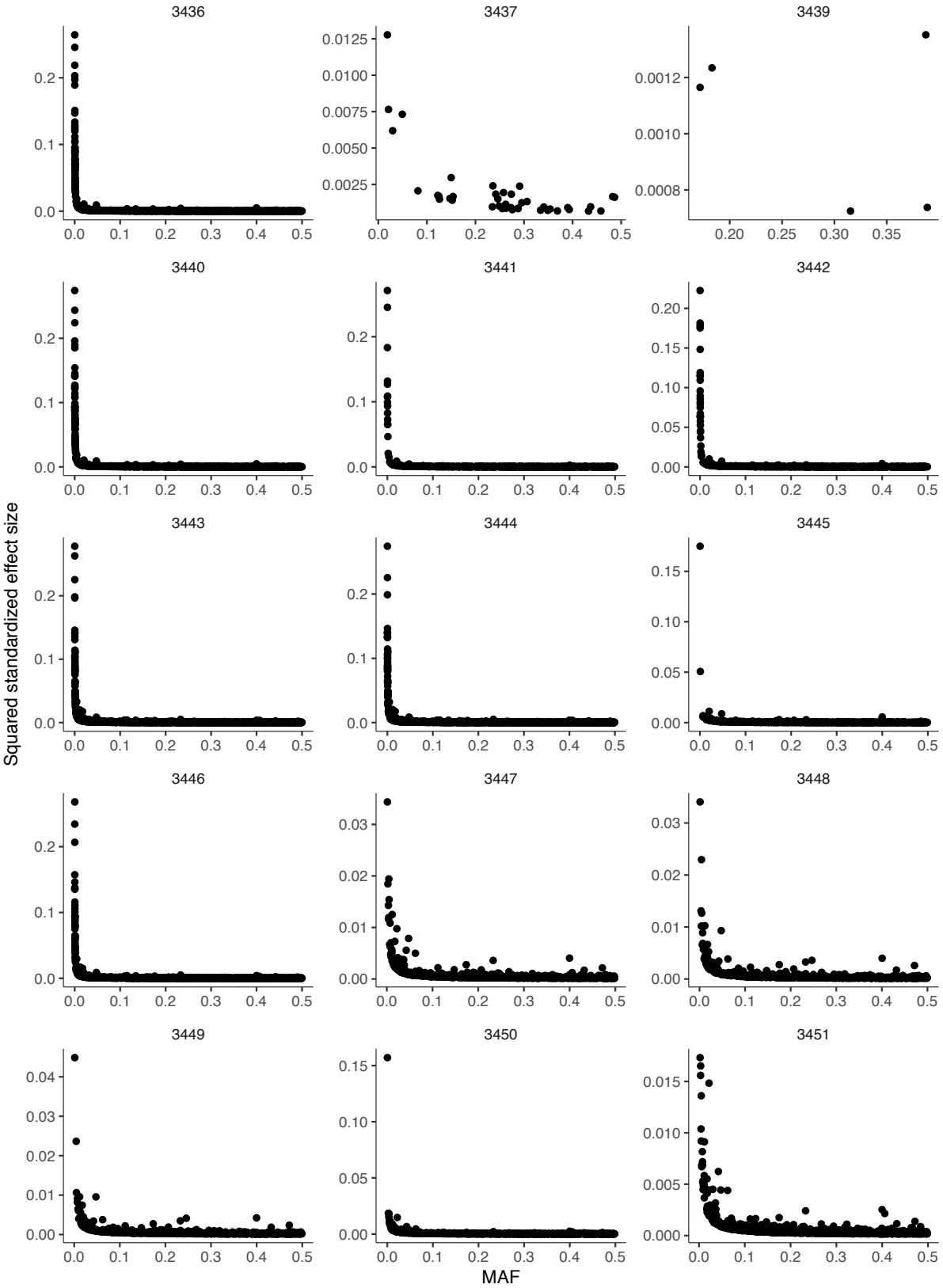

3274
3276
3277
3279
3282
3283
3296
3297
3298
3299
3300
3301
3302
3303
3304
Squared standardized effect size
MAF

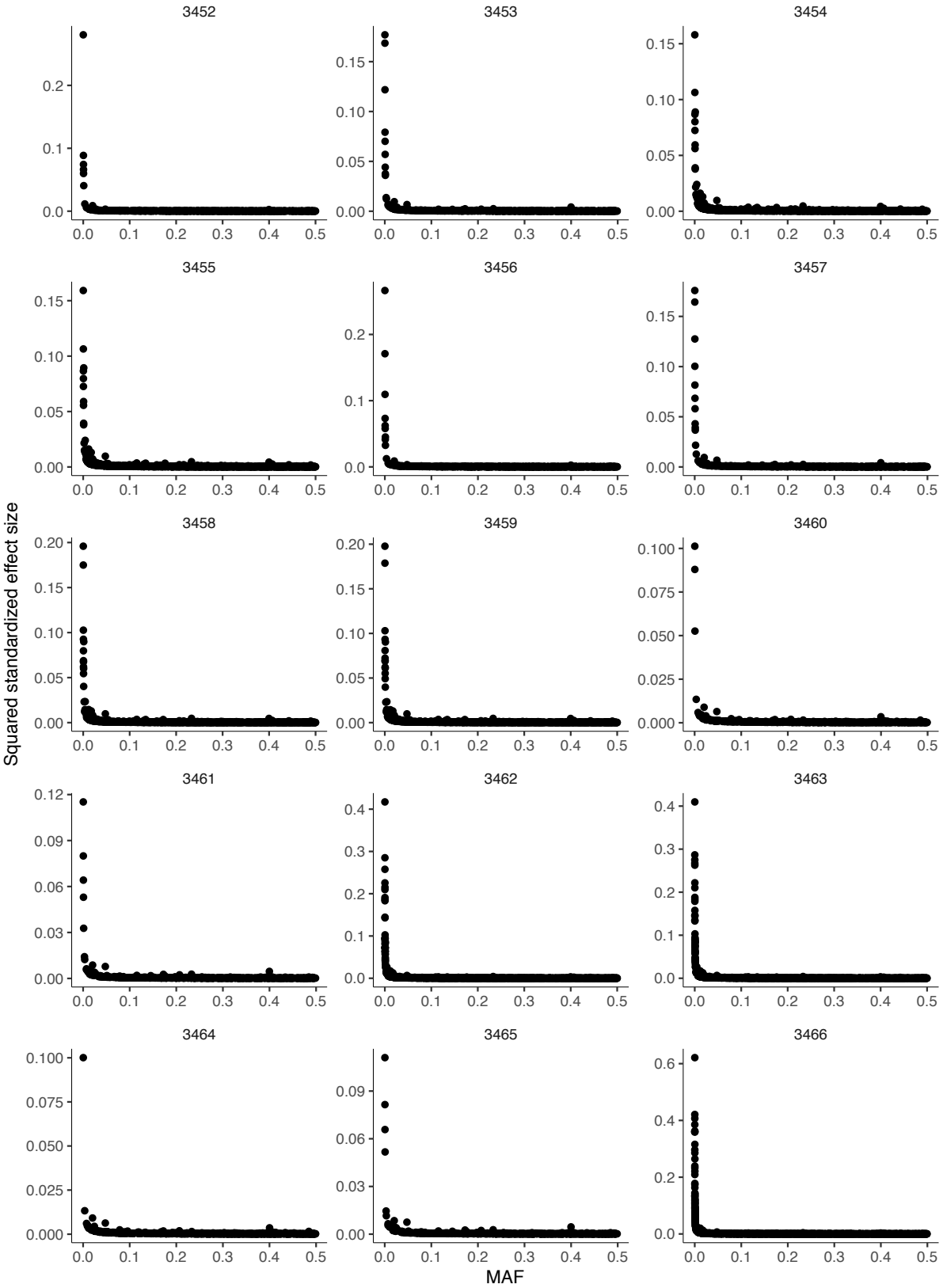

3306
3307
3308
3310
3312
3313
3314
3315
3317
3319
3320
3324
3325
3328
3329
Squared standardized effect size
MAF

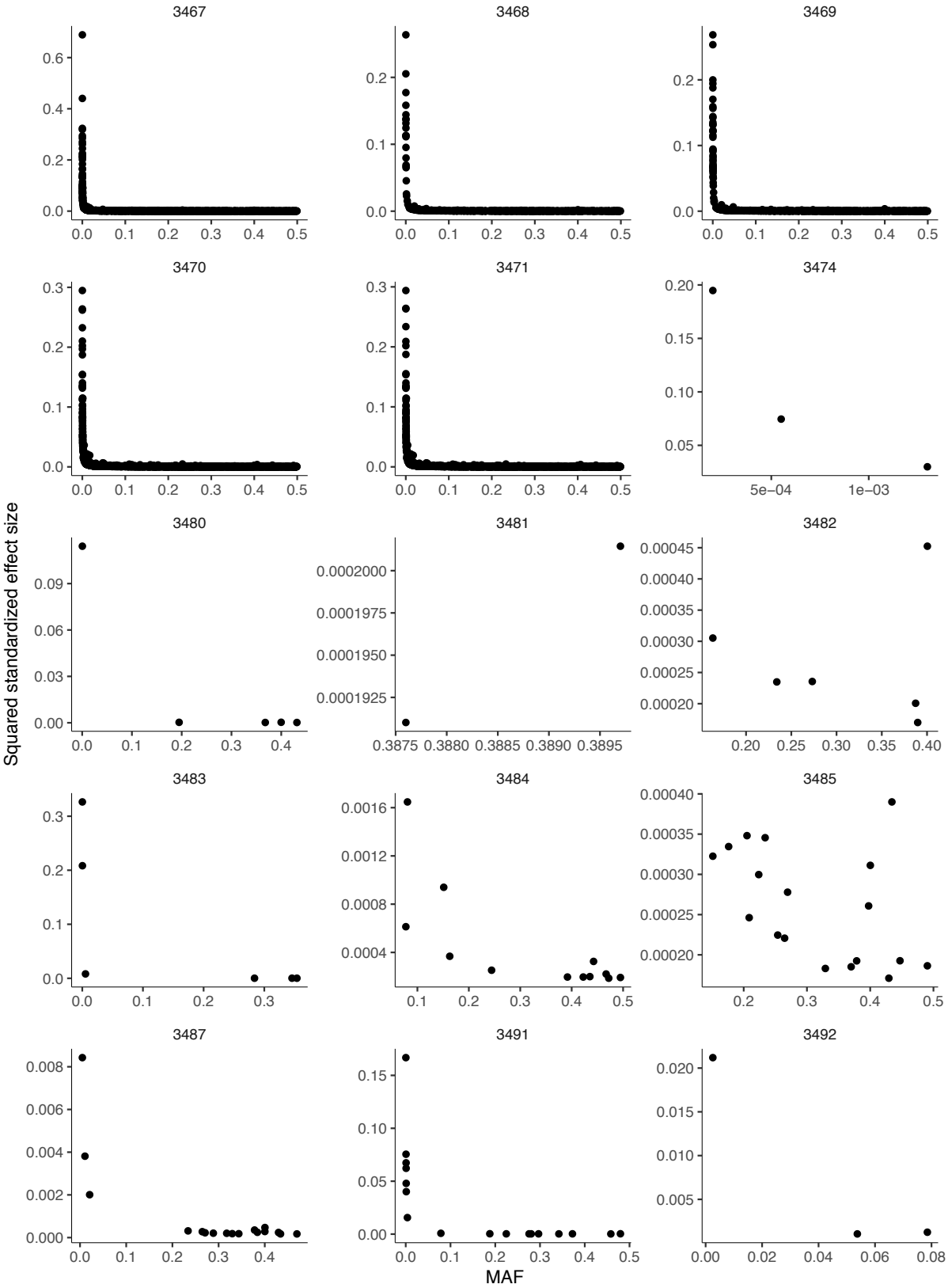

3330
3331
3335
3340
3342
3344
3346
3347
3348
3350
3351
3353
3354
3356
3359
Squared standardized effect size
MAF

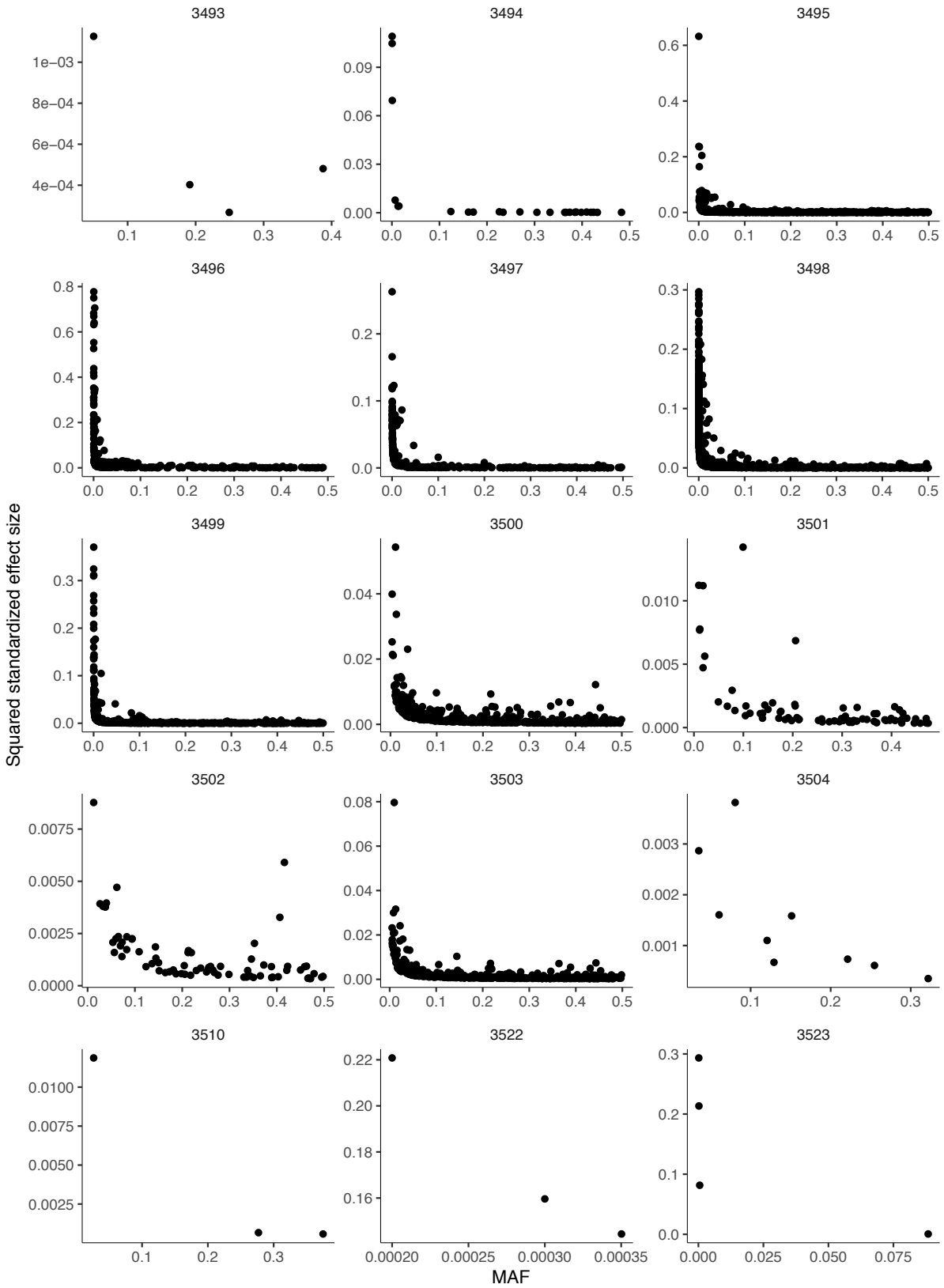

3360
3363
3366
3367
3368
3369
3372
3374
3378
3379
3380
3387
3388
3389
3391
Squared standardized effect size
MAF

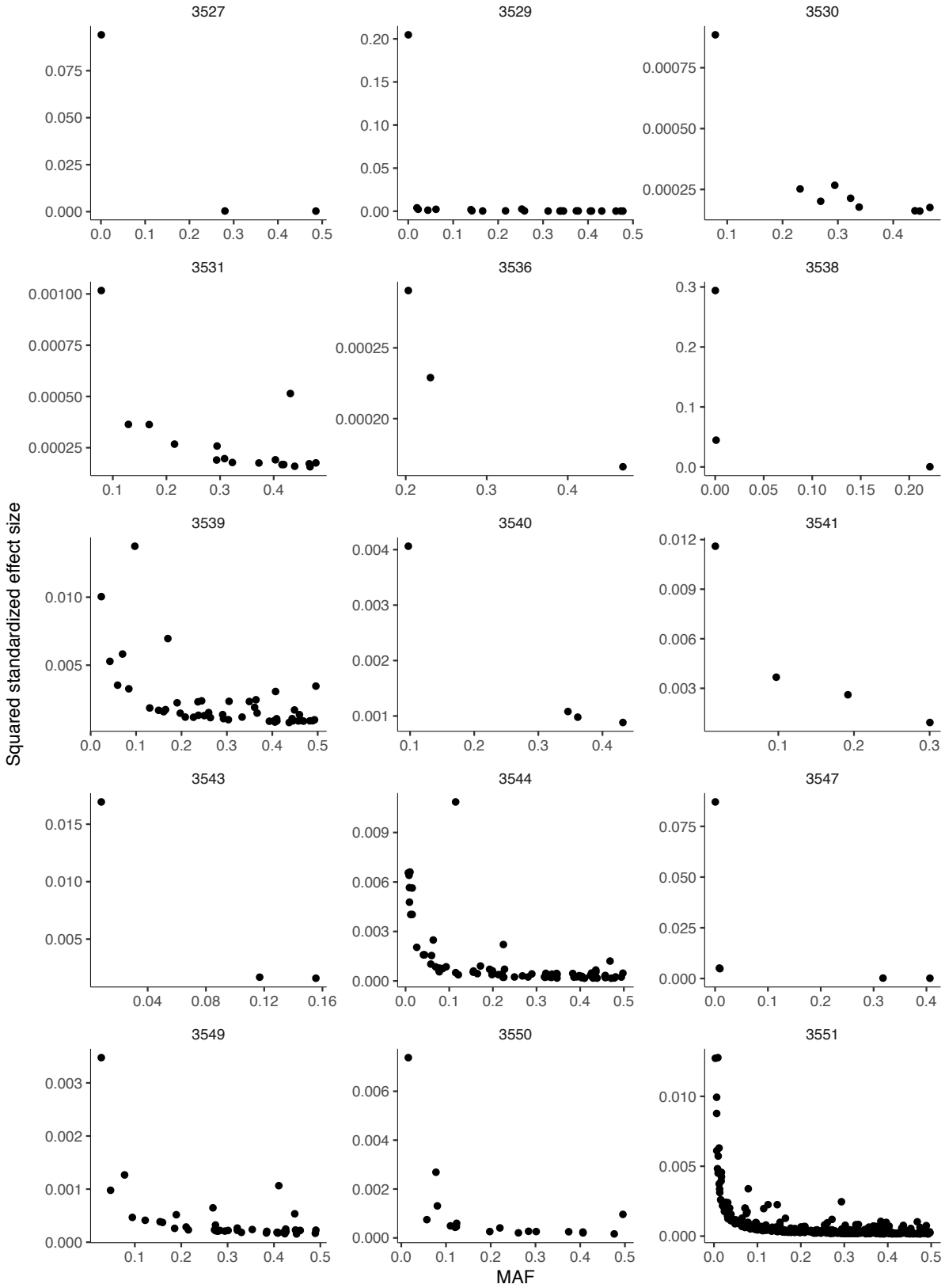

3394
3395
3396
3410
3411
3412
3414
3415
3418
3420
3422
3423
3424
3432
3434
Squared standardized effect size
MAF

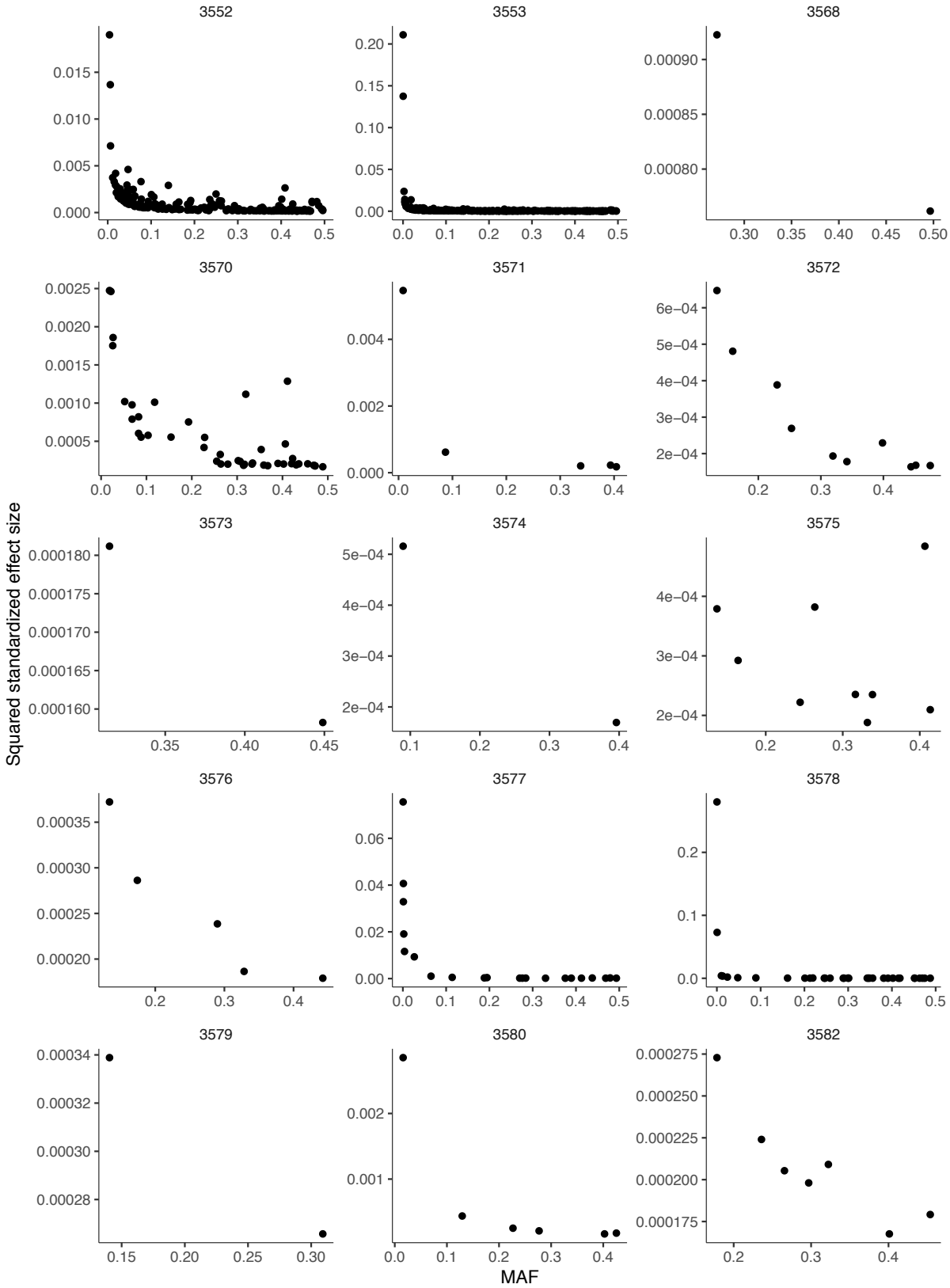

3436
3437
3439
3440
3441
3442
3443
3444
3445
3446
3447
3448
3449
3450
3451
Squared standardized effect size
MAF

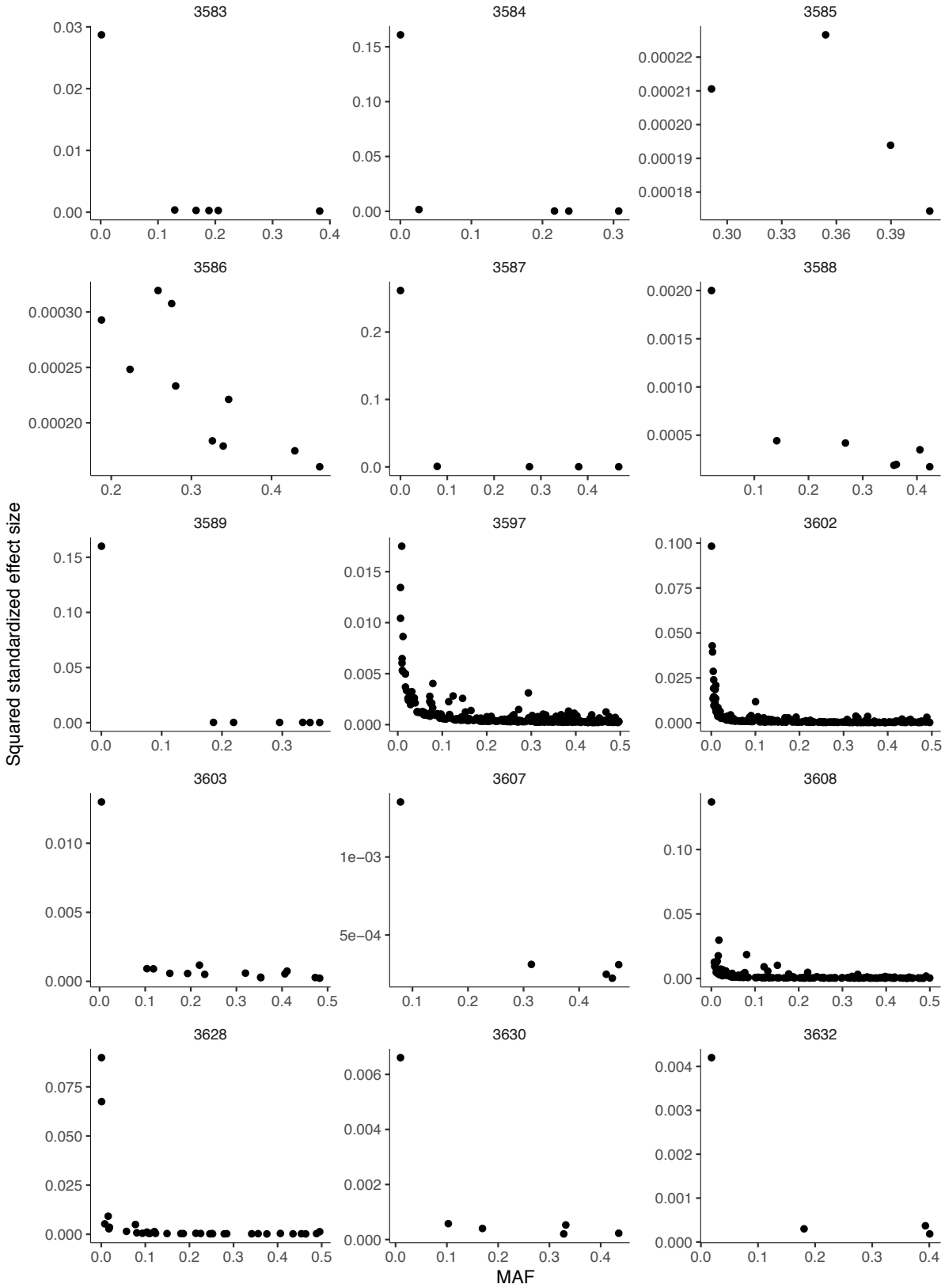

3452
3453
3454
3455
3456
3457
3458
3459
3460
3461
3462
3463
3464
3465
3466
Squared standardized effect size
MAF

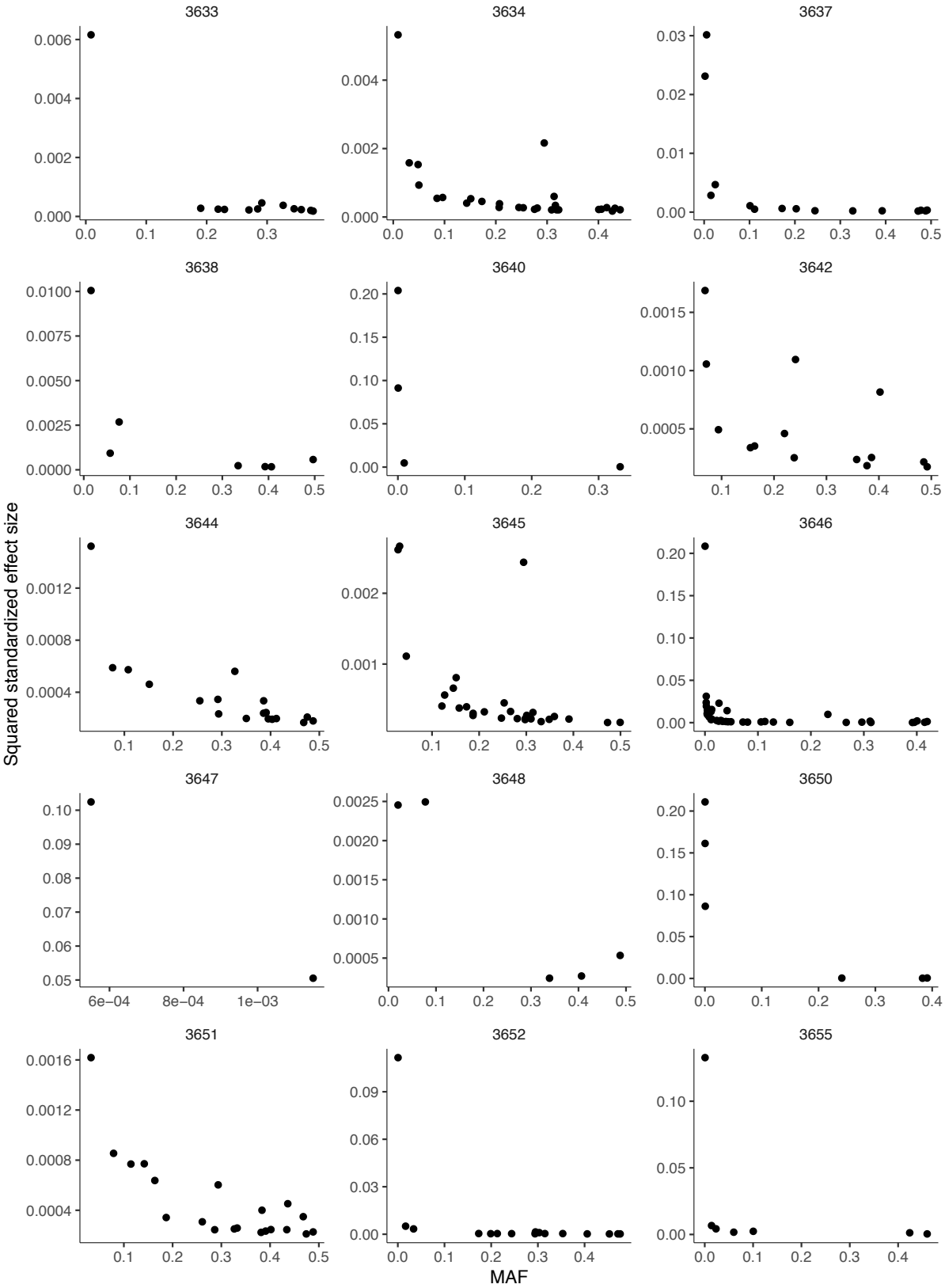

3467
3468
3469
3470
3471
3474
3480
3481
3482
3483
3484
3485
3487
3491
3492
Squared standardized effect size
MAF

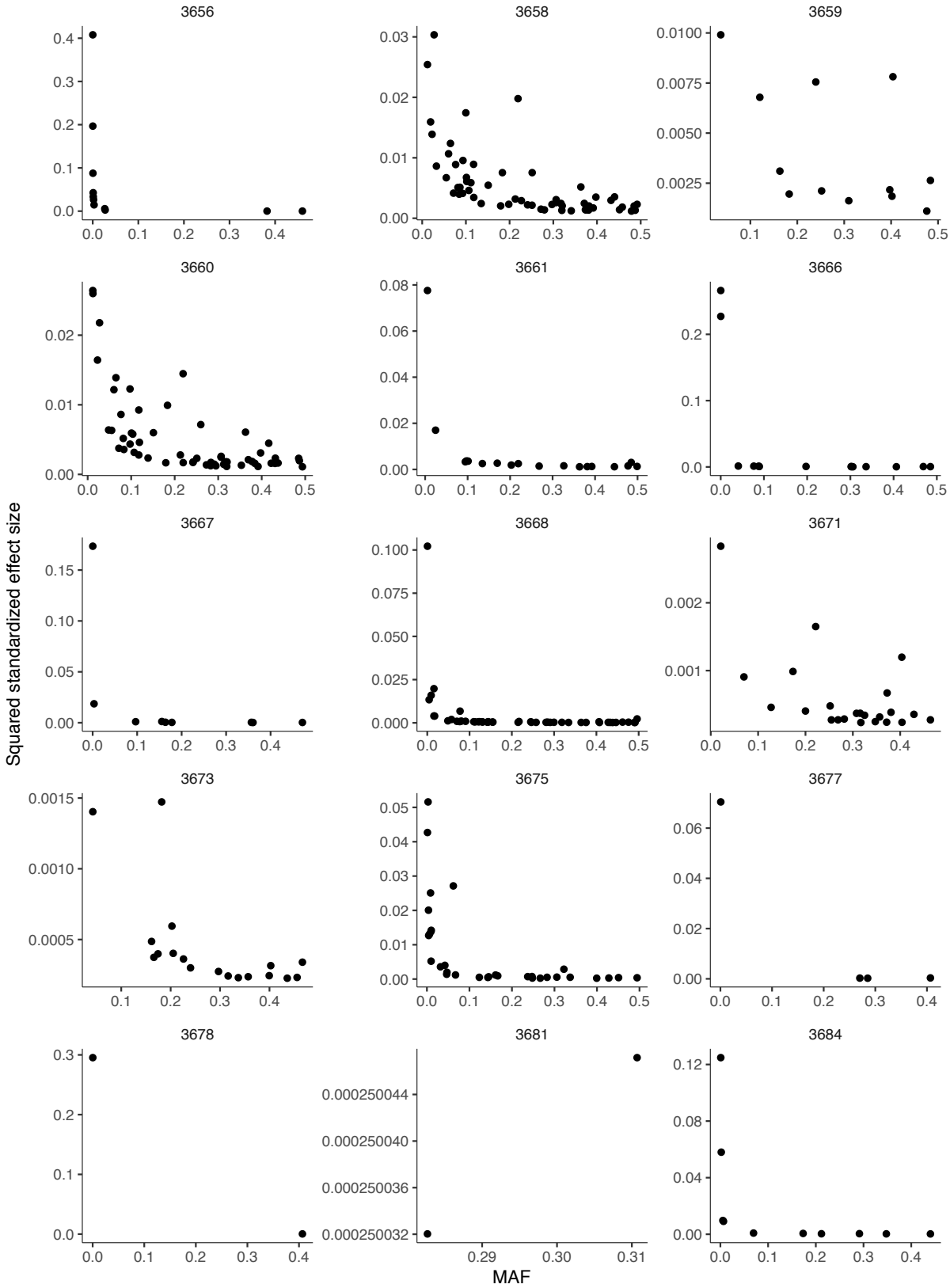

3493
3494
3495
3496
3497
3498
3499
3500
3501
3502
3503
3504
3510
3522
3523
Squared standardized effect size
MAF

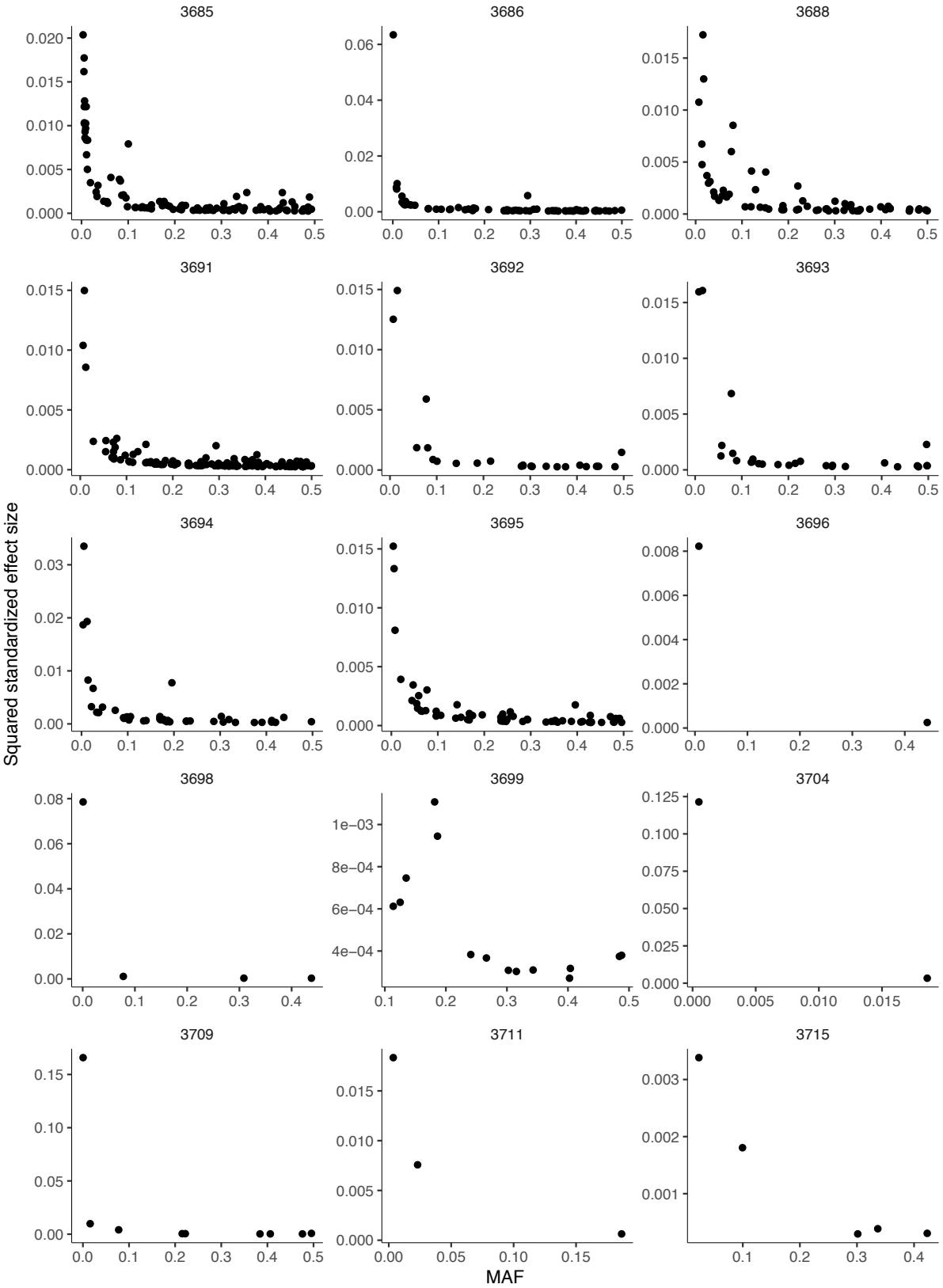

3527
3529
3530
3531
3536
3538
3539
3540
3541
3543
3544
3547
3549
3550
3551
Squared standardized effect size
MAF

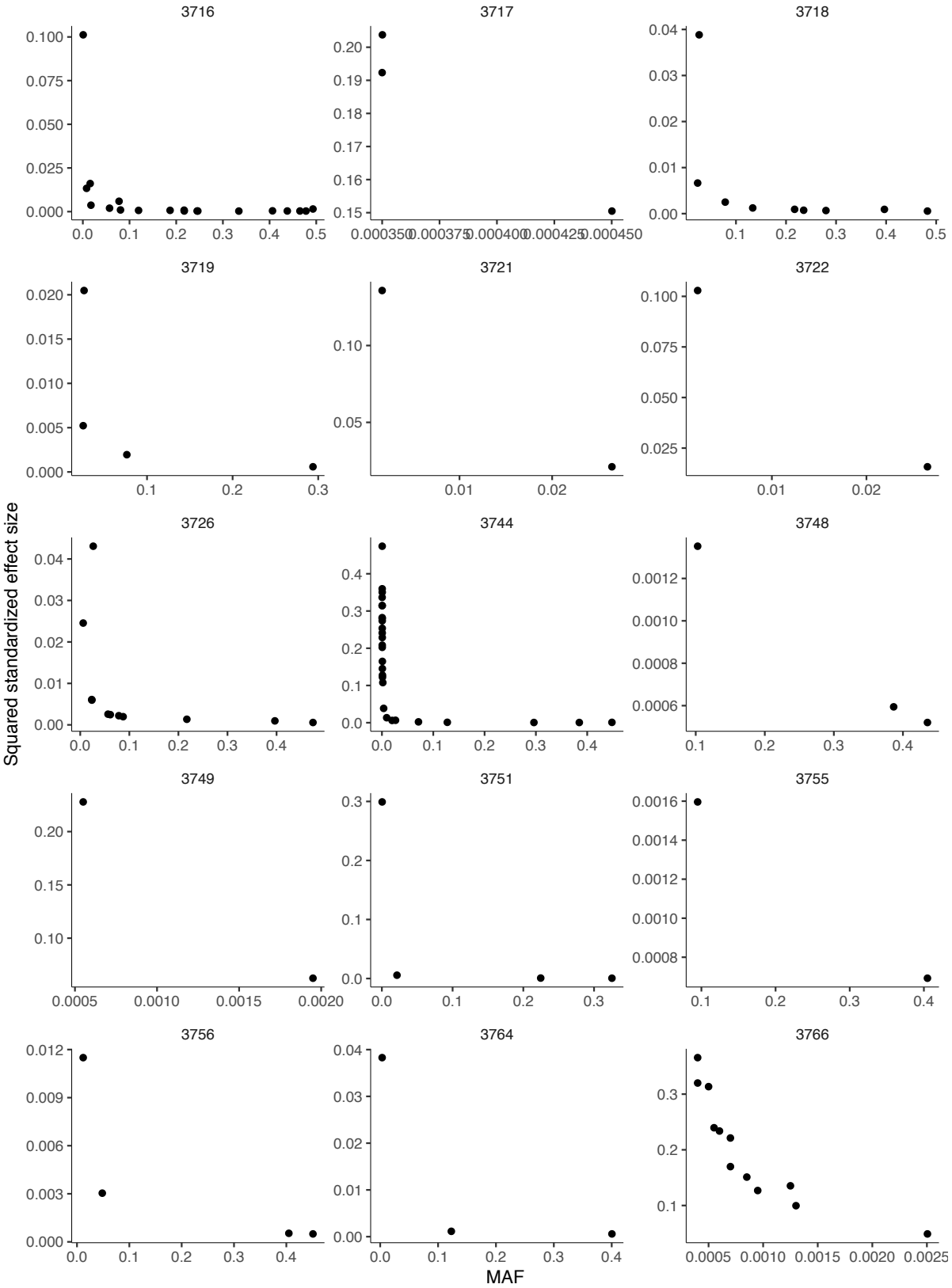

3552
3553
3568
3570
3571
3572
3573
3574
3575
3576
3577
3578
3579
3580
3582
Squared standardized effect size
MAF

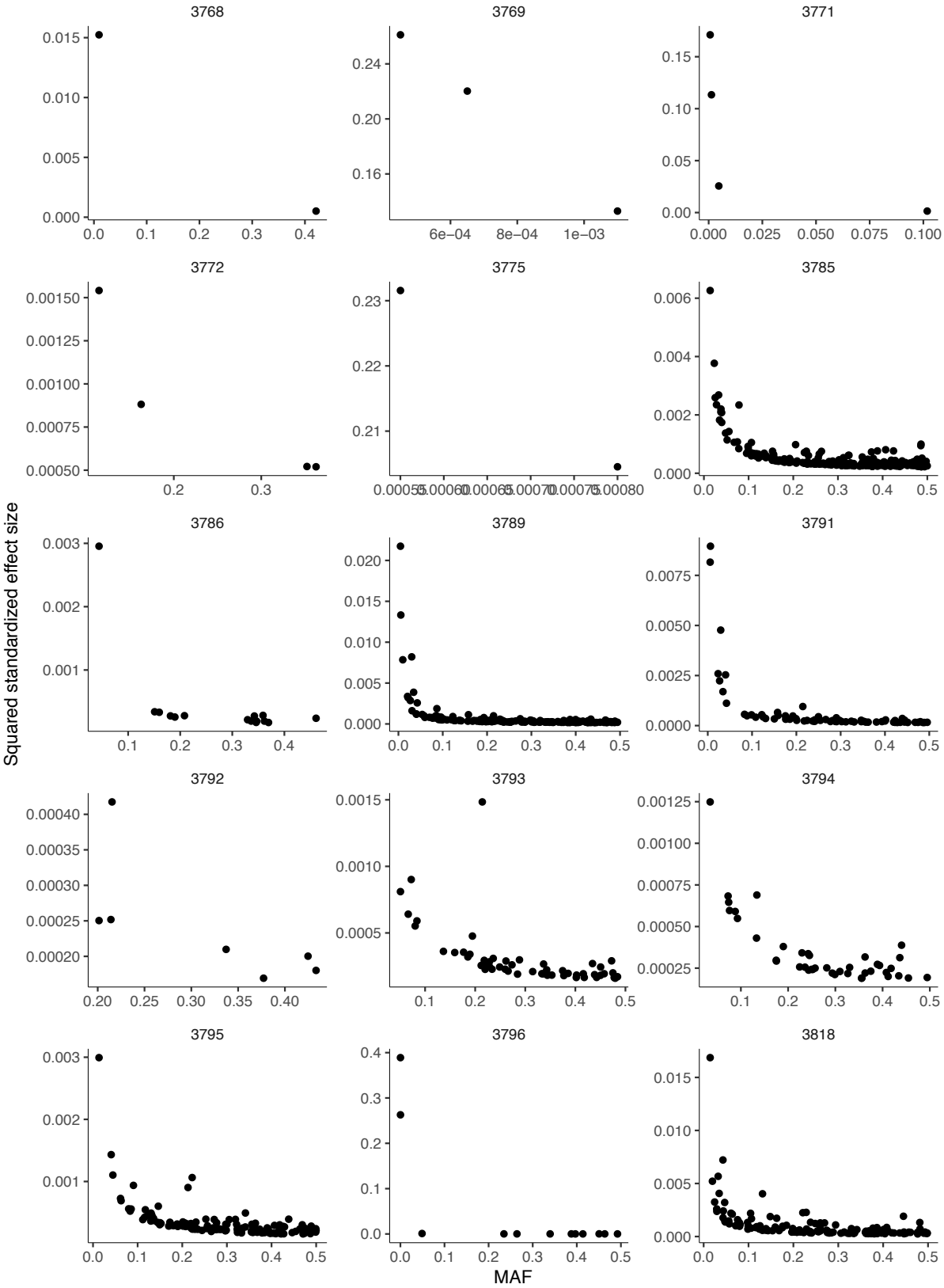

3583
3584
3585
3586
3587
3588
3589
3597
3602
3603
3607
3608
3628
3630
3632
Squared standardized effect size
MAF

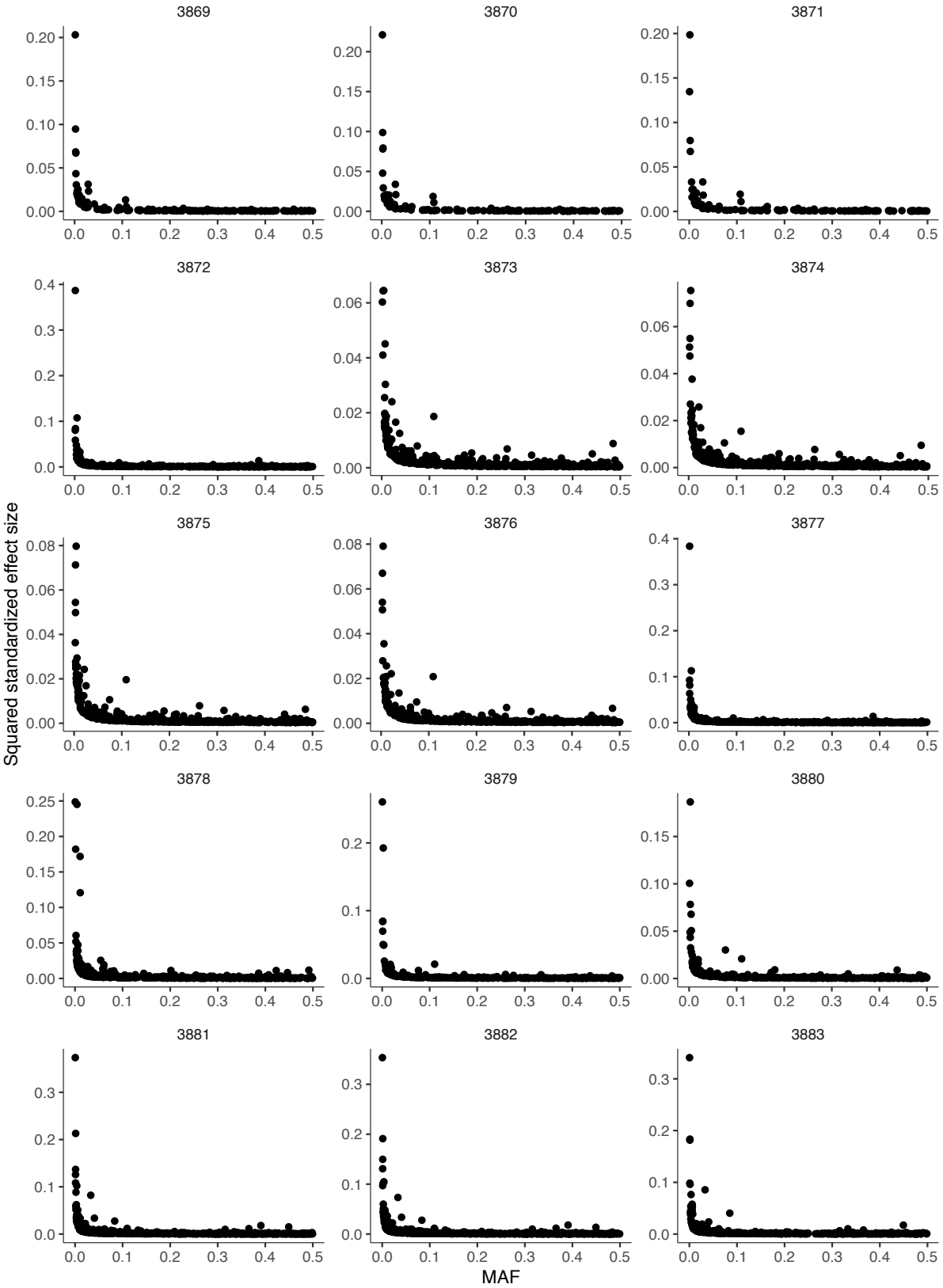

3633
3634
3637
3638
3640
3642
3644
3645
3646
3647
3648
3650
3651
3652
3655
Squared standardized effect size
MAF

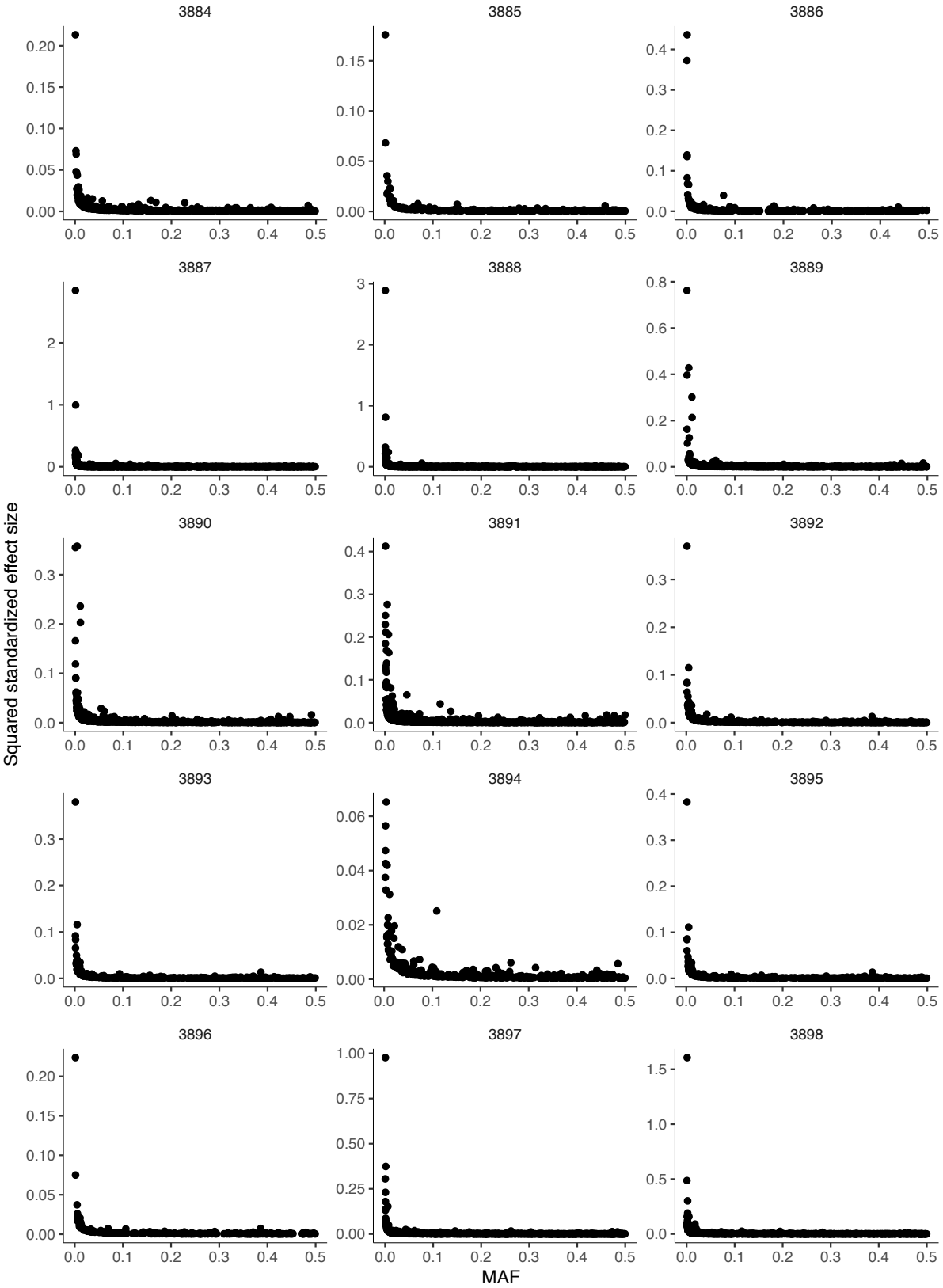

3656
3658
3659
3660
3661
3666
3667
3668
3671
3673
3675
3677
3678
3681
3684
Squared standardized effect size
MAF

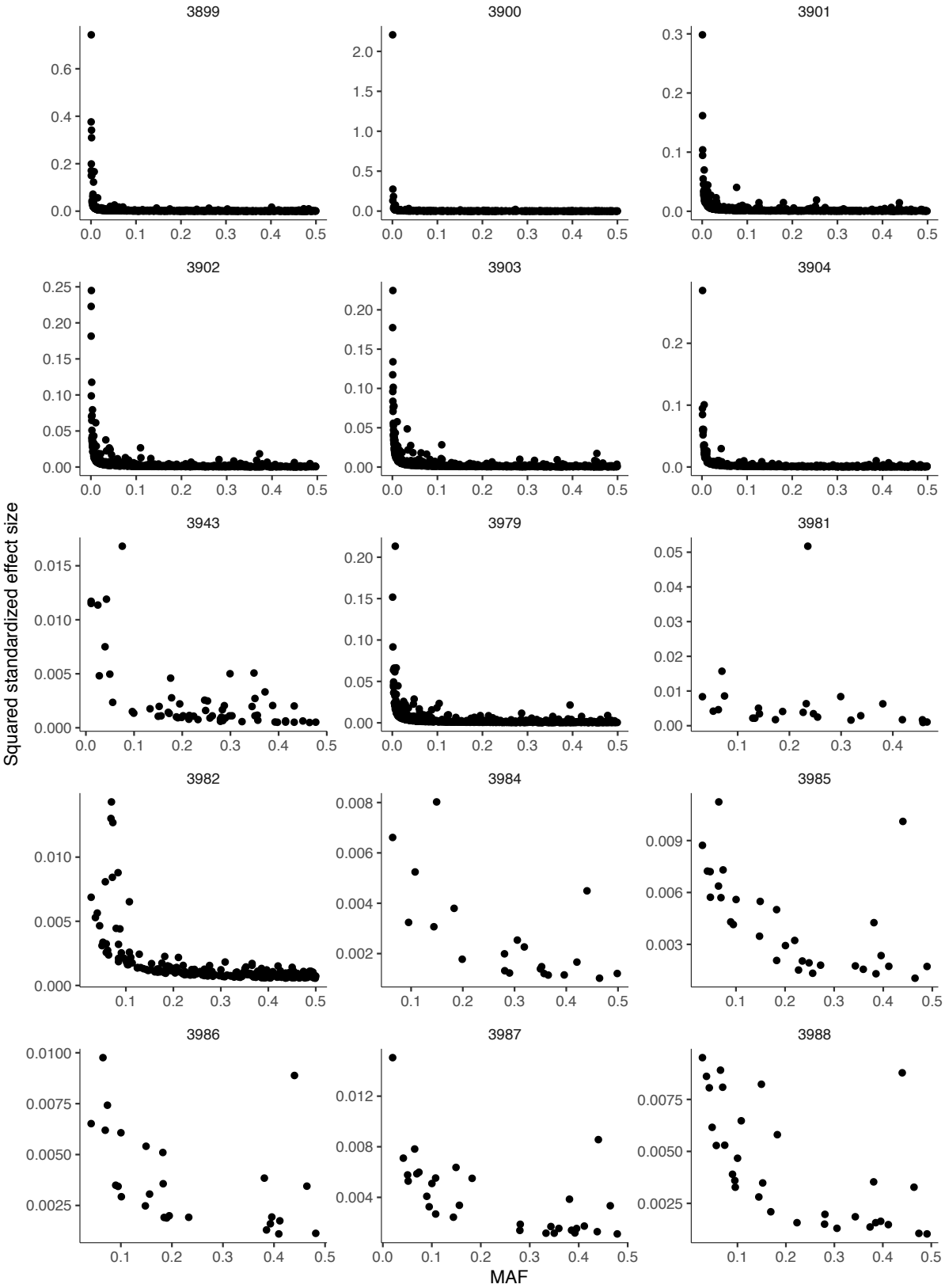

3685
3686
3688
3691
3692
3693
3694
3695
3696
3698
3699
3704
3709
3711
3715
Squared standardized effect size
MAF

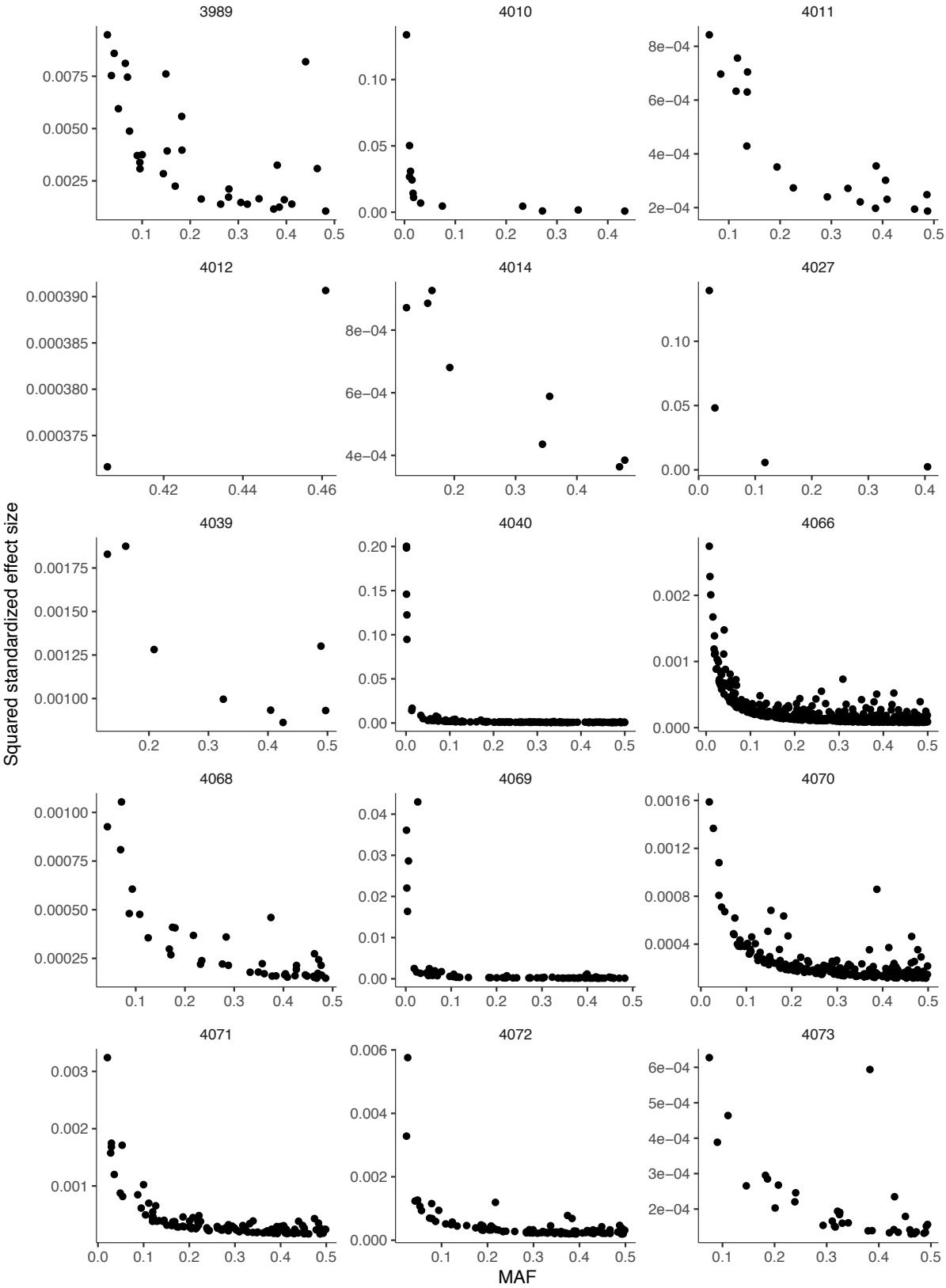

3716
3717
3718
3719
3721
3722
3726
3744
3748
3749
3751
3755
3756
3764
3766
Squared standardized effect size
MAF


3768
3769
3771
3772
3775
3785
3786
3789
3791
3792
3793
3794
3795
3796
3818
Squared standardized effect size
MAF


3869
3870
3871
3872
3873
3874
3875
3876
3877
3878
3879
3880
3881
3882
3883
Squared standardized effect size
MAF


3884
3885
3886
3887
3888
3889
3890
3891
3892
3893
3894
3895
3896
3897
3898
Squared standardized effect size
MAF


3899
3900
3901
3902
3903
3904
3943
3979
3981
3982
3984
3985
3986
3987
3988
Squared standardized effect size
MAF


3989
4010
4011
4012
4014
4027
4039
4040
4066
4068
4069
4070
4071
4072
4073
Squared standardized effect size
MAF

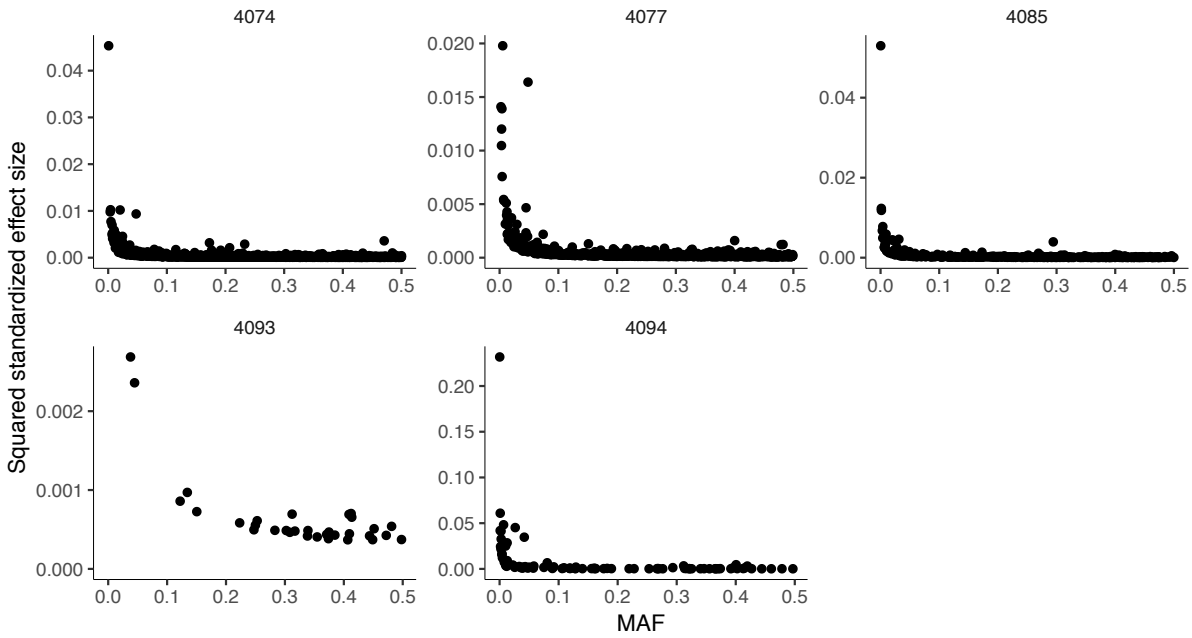


Fig. 9. Distribution of effect size and MAF of lead SNPs for each trait. Each of 410 traits with >1 lead SNPs are displayed. The number above the plot is the ID of the trait in the database matching with **Supplementary Table 3**.

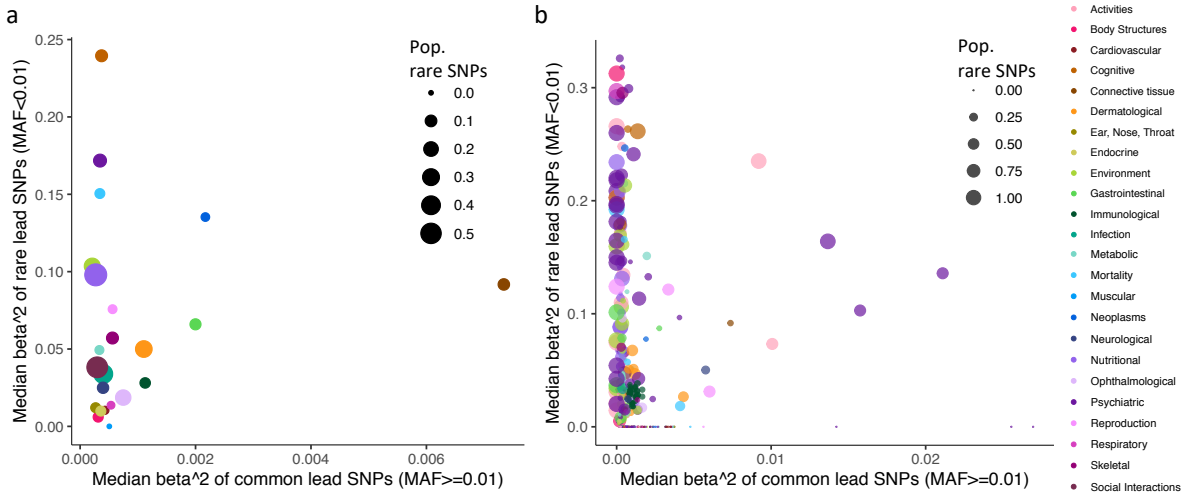


Fig. 10. Effect size and MAF of lead SNPs. a. Median of squared standardized effect size of unique lead SNPs per domain. **b.** Median of squared standardized effect size of unique lead SNPs per trait colored by domain.

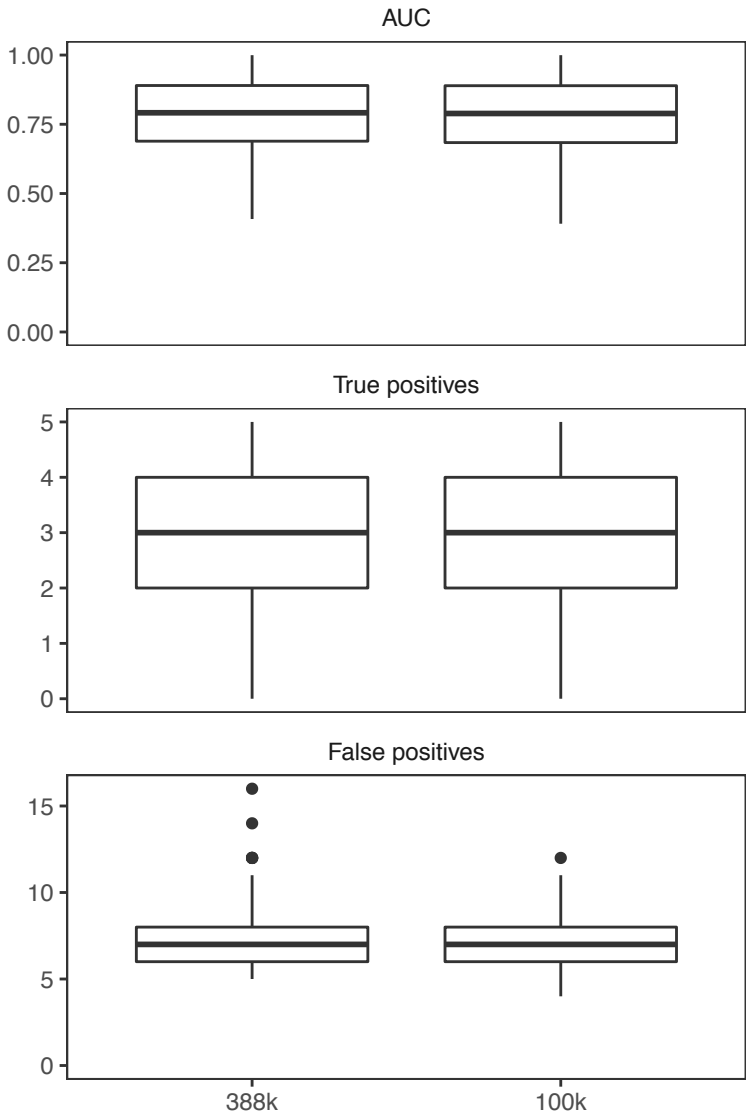


Fig. 11. Performance of fine-mapping with different size of reference panel with FINEMAP. The box plots of area under the curve (AUC, top), number of true positive (middle) and false positive (bottom) based on 500 simulations. Y-axis is the size of reference panel.

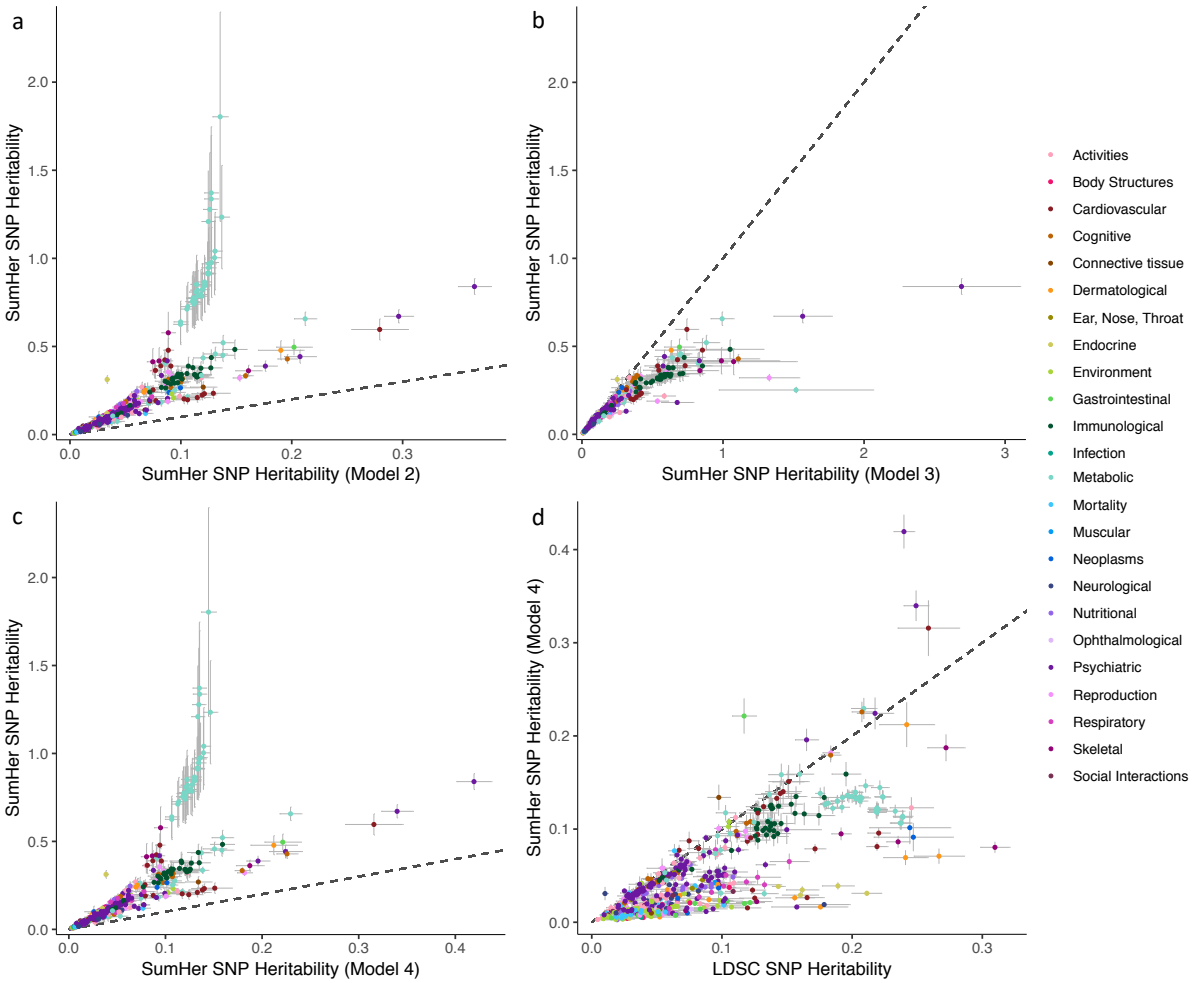


Fig. 12. Comparison of SNP heritability estimates based on SumHer with different parameters and LDSC. a. Comparison of SNP heritability estimated by SumHer with default parameter (with MAF and LD weights; y-axis) and without LD weight (with MAF weight; x-axis). **b**. Comparison of SNP heritability estimated by SumHer with default parameter (with MAF and LD weights; y-axis) and without MAF weight (with LD weight; x-axis). **c**. Comparison of SNP heritability estimated by SumHer with default parameter (with MAF and LD weights; y-axis) and without LD nor MAF weight (x-axis). d. Comparison of SNP heritability estimated by SumHer without LD nor MAF weight (y-axis) and LDSC (x-axis). Horizontal and vertical lines represent standard error of estimates for corresponding axis. Traits are colored by their domain. The full results are available in **Supplementary Table 23**.

References

1. 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,** 1–10 (2015).
2. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562,** 203–209 (2018).
3. Mccarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48,** 1279–1283 (2016).
4. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
5. Webb, B. T. *et al.* Molecular genetic influences on normative and problematic alcohol use in a population-based sample of college students. *Front. Genet.* **8,** 30 (2017).
6. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2 : principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33,** 2776–2778 (2017).
7. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
8. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32,** 283–285 (2016).
9. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10,** e1004383 (2014).
10. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11,** e1004219 (2015).
11. Benner, C. *et al.* Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *Am. J. Hum. Genet.* **101,** 539–551 (2017).
12. Benner, C. *et al.* FINEMAP : efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32,** 1493–1501 (2016).
13. Wray, N. R. & Maier, R. Genetic basis of complex genetic disease: the contribution of disease heterogeneity to missing heritability. *Curr. Epidemiol. Reports* **1,** 220–227 (2014).
14. Bulik-sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).
15. Speed, D. *et al.* Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49,** 986–992 (2017).
16. Speed, D. & Balding, D. J. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat. Genet.* **51,** 277–284 (2019).
17. Evans, L. M. *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).