Submitted to the Annals of Applied Statistics

SIGNAL-NOISE RATIO OF GENETIC ASSOCIATIONS AND STATISTICAL POWER OF SNP-SET TESTS

BY HONG ZHANG[1], MING LIU[2,*], JIASHUN JIN[3] AND ZHEYANG WU[2,†]

[1]*Merck Research Laboratories, Merck & Co., Inc., hong.zhang8@merck.com*

[2]*Department of Mathematical Sciences, Worcester Polytechnic Institute, *mliu5@wpi.edu; †zheyangwu@wpi.edu*

[3]*Department of Statistics, Carnegie Mellon University, jiashun@cmu.edu*

The SNP-set analysis is a powerful tool for dissecting the genetics of complex human diseases. There are three fundamental genetic association approaches to SNR-set analysis: the marginal model fitting approach, the joint model fitting approach, and the decorrelation approach. A problem of primary interest is how these approaches compare with each other. To address this problem, we develop a theoretical platform to compare the signal-to-noise ratio (SNR) of these approaches under the generalized linear model. We elaborate how causal genetic effects give rise to statistically detectable association signals, and show that when causal effects spread over blocks of strong linkage disequilibrium (LD), the SNR of the marginal model fitting is usually higher than that of the decorrelation approach, which in turn is higher than that of the unbiased joint model fitting approach. We also scrutinize dense effects and LDs by a bivariate model and extensive simulations using the 1000 Genome Project data. Last, we compare the statistical power of two generic types of SNP-set tests (summation-based and supremum-based) by simulations and an osteoporosis study using large data from UK Biobank. Our results help develop powerful tools for SNP-set analysis and understand the signal detection problem in the presence of colored noise.

1. Introduction. The SNP-set analysis is a powerful method in genome-wide association studies (GWAS) for dissecting the genetics of complex human diseases (Hoh, Wille and Ott, 2001; Wu et al., 2014; Kwak and Pan, 2016; Pasaniuc and Price, 2017; Guo and Wu, 2019). It combines genetic associations of multiple SNPs in a set to test the genetic association of the whole group. There are three fundamental genetic association approaches in GWAS (Marchini, Donnelly and Cardon, 2005; Wu and Zhao, 2009). The first approach follows the single-SNP analysis, which applies marginal regression or logit model to obtain SNP's Z-scores or p-values. It is commonly used due to computational simplicity. Meanwhile, its genetic effect estimation is biased because of correlations – non-causal SNPs in LD with causal SNPs often show statistical associations too. The second approach is to fit the joint regression or logit model to obtain the multiple SNP's Z-scores or p-values simultaneously. It is computationally more intensive but provides unbiased estimation. As a method jointly analyzing multiple SNPs together, it is often considered as being beneficial to revealing "joint associations" of multiple SNPs (Yang et al., 2012). The third approach is to decorrelate SNPs' Z-scores. It is often applied before combining the Z-scores for the global hypothesis test of the SNP-set. A classic example is Hotelling's t-squared statistic and other similar variance-component statistics (Hotelling, 1931; Luo et al., 2010). This approach is convenient for calculating the SNP-set's p-value because the test statistic's null distribution is often easier to derive when the Z-scores are transformed to independent and

Keywords and phrases: SNP-set analysis, causal genetic effect, linkage disequilibrium, signal-noise ratio, global hypothesis test, osteoporosis.

identical distributed (i.i.d.) components. The three genetic association approaches have been routinely applied in practice, sometimes with intuitive justifications from computational or genetic prospectives. However, it remains unclear how these approaches compare with each other. The problem is of primary interest in statistical genetics. Our paper aims to address this problem through a rigorous study that directly compares the three approaches' statistical signals emerged from causal genetic effects and LDs.

In a seemingly different setting, Hall and Jin (2010) studied the problem of detecting Rare-and-Weak signals in a Gaussian Means Models (GMM): given a high dimensional normal vector $\mathbf{T} = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the question is how to test whether $\boldsymbol{\mu} = \mathbf{0}$ or $\boldsymbol{\mu} \neq \mathbf{0}$. They assumed that only a few entries of $\boldsymbol{\mu}$ may be nonzero, and each nonzero entry is relatively small (so the signals are both rare and weak). The correlation matrix $\boldsymbol{\Sigma}$ is assumed available. They considered three data transformations, which we may call identical transformation, whitening (i.e., decorrelation) transformation, and innovated transformation, and studied the (approach-dependent) signal-to-noise ratios (SNRs) associated with each transformation. As noted in the literature, there are some interesting connections between signal detection and genetic association study. For example, He and Wu (2011) adapted the innovated transformation by Hall and Jin (2010) to the SNP-set analysis; see also Barnett, Mukherjee and Lin (2017). Motivated by such connections, we develop a theoretical framework by extending some ideas of Hall and Jin (2010) and use the framework to compare the three SNR-set approaches aforementioned.

To do so, we face several challenges. First, Hall and Jin (2010) only considered a very idealized model, and it is unclear how to extend the ideas there to our setting, where we must use a linear regression model or a generalized linear model (GLM) to characterize genetic architecture meaningfully. The extension is non-trivial. For example, the GMM assumes that signals and correlations are independent parameters. In contrast, under the GLM, the Z-scores' means and correlations are interconnected by causal effects and the correlations among the SNPs and other covariates. Moreover, the Z-scores under the GLM have different types to consider (e.g., various likelihood-based estimators). Second, the relationship between a signal detection approach for GMM and a genetic association approach is usually not apparent. Sometimes, it is hard to translate an approach for GMM to our setting correctly. For example, under the GMM, a signal detection approach based on the innovated transformation has been shown to increase the statistical power of the Higher Criticism (HC) test (i.e., the iHC test) (Hall and Jin, 2010). However, in an important study of the SNP-set analysis, the iHC was reported as "subject to considerable loss of power" (Barnett, Mukherjee and Lin, 2017). We explain that the reason why we have such contradicting conclusions on the innovated transformation is due to an inaccurate translation of the innovated transformation in GMM to the setting of SNP-set analysis. Third, further study is also needed to compare the SNRs in the genetic association approaches based on typical patterns of causal genetic effects and the LDs in the human genome.

Our contribution is three-fold. First, we extend the platform of Hall and Jin (2010) for GMM to the more practical GLM setting and develop a new framework by which we are able to answer the above genetic questions carefully. In particular, using this framework, we found in-depth relationships between genetic association approaches and the three transformations studied in Hall and Jin (2010). For example, we find that both the Z-scores from the marginal fitting and those from the joint fitting approach can be viewed as the direct results of innovated transformation approach in Hall and Jin (2010).

Second, we compare the SNRs of the three genetic association approaches aforementioned theoretically, and so are able to spell out their advantages and disadvantages. In particular, we prove that when causal effects are dispersed over blocks of strong LDs among SNPs, the SNRs of the marginal fitting approach are always higher than those of the decorrelation

approach, which are, in turn, higher than those of the joint fitting approach. We also address the case where the causal effects and correlations are both "dense" by an analytical study of a bivariate model as well as a simulation study based on the genotype data of the 1000 Genome Project (Siva, 2008). Bivariate model is a great setting to study complex statistical problems (Arias-Castro, Huang and Verzelen, 2020). Using it, we explain the principal mechanisms of signal boosting and cancellation depending on the relative magnitude and directionality of causal effects and LDs. The simulation study surveys how genetic effects and LD patterns in the human genome influence the SNRs. The numerical results support the theory well.

Last, we carefully studied the statistical power of two primary SNP-set tests: the summation-based combination (represented by Fisher's combination (Fisher, 1934) and the Sequence Kernel Association Test (SKAT) (Wu et al., 2011)) and the supremum-based combination (represented by the minimal p-value method and the HC Donoho and Jin (2004); Arias-Castro and Wang (2017)). Based on systematic simulations and the analysis of osteoporosis data from UK Biobank (Sudlow et al., 2015), the empirical results confirmed that the marginal fitting is often more powerful, although the joint fitting and the decorrelation could have complementary merits for detecting extra disease genes. Moreover, the supremum-based tests have an advantage when causal effects are strong in the common-variant analysis, whereas the summation-based tests are particularly beneficial for detecting weak and dense effects in the rare-variant analysis.

The remainder of the paper is organized as follows. Section 2 formulates the genetic architecture and the SNP-set test, and provides the definitions and properties of the SNR and related transformations. Section 3 presents theoretical studies on the SNRs of genetic association approaches. SNR simulations based on the 1000 Genome Project are given in Section 4. The statistical power study for the SNP-set tests is given in Section 5. Section 6 provides a GWAS of osteoporosis using UK Biobank data. Section 7 summarizes the study and briefly discusses the limitations and future research plans.

2. Entry-wise SNR and Transformations under GLM. Under the GMM, Hall and Jin (2010) considered an n-dimensional Gaussian vector $\mathbf{T} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$, where the vector of means $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)'$, the vector of noise variables $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} = (\Sigma_{ij})_{1\le i,j\le n}$ is the correlation matrix assumed as known. In contrast to the null $\boldsymbol{\mu} = \mathbf{0}$, we say we have a signal at entry j if $\mu_j \neq 0$, and the SNR is $|\mu_j|/\sqrt{\Sigma_{jj}} = |\mu_j|$ since $\Sigma_{jj} = 1$. Suppose $\mathbf{A}$ is an $n \times n$ matrix (which is non-random and may depend on $\boldsymbol{\Sigma}$) and consider the transformation $\mathbf{T} \mapsto \mathbf{AT}$. The SNR for entry j of $\mathbf{AT}$ is $|\tilde{\mu}_j|/\sqrt{\widetilde{\Sigma}_{jj}}$, where $\tilde{\boldsymbol{\mu}} = \mathbf{A}\boldsymbol{\mu}$ and $\widetilde{\boldsymbol{\Sigma}} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$. Hall and Jin (2010) showed that under some settings (e.g., the signals satisfy a so-called rare-weak model), we can find a proper matrix $\mathbf{A}$ such that for all j where $\mu_j \neq 0$ the post-transform SNR at entry j is always larger than the pre-transform SNR at entry j: $|\tilde{\mu}_j|/\sqrt{\widetilde{\Sigma}_{jj}} \ge |\mu_j|$. That is, the transformation of $\mathbf{T} \mapsto \mathbf{AT}$ simultaneously boosts the SNR at all signal entries.

The comparison among the genetic association approaches can be viewed (at least asymptotically) as a problem related to the above. In the SNP-set analysis, we first compute summarizing Z-scores, each for a SNP; then, we stack the Z-scores into a Z-vector as the input for testing the SNP-set's association. One crucial problem is how to compare the Z-scores obtained from different genetic association approaches. A nice idea from the discussion above is to connect these approaches by transformations and compare their entry-wise SNRs. However, we need to start from the GLM that defines a meaningful genetic architecture:

$$g(\mathrm{E}(\mathbf{Y}_k|\mathbf{X}_{k\cdot}, \mathbf{Z}_{k\cdot})) = \mathbf{X}'_{k\cdot}\boldsymbol{\beta} + \mathbf{Z}'_{k\cdot}\boldsymbol{\gamma}, \tag{1}$$

where Y_k quantifies the phenotypic trait of the kth subject, $k = 1, \ldots, N$, with the sample size N. $\mathbf{X}_{k\cdot} = (X_{k1}, \ldots, X_{kn})'$ is the genotype vector of n SNPs, $\mathbf{Z}_{k\cdot} = (Z_{k1}, \ldots, Z_{km})'$ is

the vector of m controlling covariates. This paper focuses on SNPs, including rare variants, but the model applies to other genetic markers as well. The controlling covariates could include the intercept, environmental factors, other genetic variants, or potential confounders to be controlled upon. The link function g characterizes the genotype-phenotype relationship depending on the distribution of Y_k. The constant coefficient vectors $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_n)'$ and $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_m)'$ are unknown. The nonzero elements of $\boldsymbol{\beta}$ are referred to as the causal genetic effects of the corresponding SNPs. The SNP-set analysis concerns whether at least one SNP is causal by testing the global hypotheses

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \text{ versus } H_1 : \boldsymbol{\beta} \neq \mathbf{0}. \tag{2}$$

Different genetic association approaches lead to different estimators $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, ..., \widehat{\beta}_n)'$. For a given estimator, the non-zero expectation $\mathrm{E}(\widehat{\beta}_i)$ of the ith SNP reflects its genetic association signal. The signal could come from its own causality $\beta_i \neq 0$ or its linkage to other causal SNP(s). The SNR definition involves two scenarios. In the scenario of finite sample size N, the SNR is defined as $\mathrm{SNR}(\widehat{\beta}_i) = |\mathrm{E}(\widehat{\beta}_i)|/\mathrm{SD}(\widehat{\beta}_i)$. In the scenario of the asymptotics with $\sqrt{N}(\widehat{\beta}_i - \theta_i) \xrightarrow{D} N(0, d_i^2)$ as $N \to \infty$ (here, $\xrightarrow{D}$ denotes the convergence in distribution), we define $\mathrm{SNR}(\widehat{\beta}_i) = |\theta_i|/d_i$. For the convenience of calculation by linear algebra, we unify the two scenarios by one notation in vector and matrix format:

$$\mathrm{SNR}(\widehat{\boldsymbol{\beta}}) = (\mathrm{diag}(\boldsymbol{\Sigma}))^{-\frac{1}{2}} |\boldsymbol{\theta}|, \tag{3}$$

where under the finite-sample scenario, the mean vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)' = \mathrm{E}(\widehat{\boldsymbol{\beta}})$ and the variance-covariance matrix $\boldsymbol{\Sigma} = (\Sigma_{ij})_{1 \leq i,j \leq n} = \mathrm{Var}(\widehat{\boldsymbol{\beta}})$. Under the asymptotics $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ are the asymptotic mean vector and the asymptotic variance-covariance matrix, respectively. Denote the diagonal matrix of $\boldsymbol{\Sigma}$ as $\mathrm{diag}(\boldsymbol{\Sigma}) = \mathrm{diag}\{\Sigma_{ii}; i = 1, ..., n\}$, and $(\mathrm{diag}(\boldsymbol{\Sigma}))^{-\frac{1}{2}} = \mathrm{diag}\{1/\sqrt{\Sigma_{ii}}; i = 1, ..., n\}$.

The Z-score statistics of $\widehat{\boldsymbol{\beta}}$ are denoted by $\mathbf{T} = (T_1, ..., T_n)'$, where $T_i = \widehat{\beta}_i/\mathrm{SD}(\widehat{\beta}_i)$ has unit variance. When $\mathrm{SD}(\widehat{\beta}_i)$ is unknown, we can estimate the standard deviation and define $T_i = \widehat{\beta}_i/\widehat{\mathrm{SD}}(\widehat{\beta}_i)$, or in vector format, $\mathbf{T} = \left(\mathrm{diag}(\widehat{\boldsymbol{\Sigma}})\right)^{-\frac{1}{2}} \widehat{\boldsymbol{\beta}}$. Obviously, $\mathrm{SNR}(\widehat{\boldsymbol{\beta}}) = \mathrm{SNR}(\mathbf{T})$ (assuming convergence in probability: $\widehat{\boldsymbol{\Sigma}} \xrightarrow{P} \boldsymbol{\Sigma}$ under the asymptotics). In general, the SNR vector is invariant to component-wise scaling by any nonzero constants.

We define transformations for $\widehat{\boldsymbol{\beta}}$ similar as Hall and Jin (2010). However, we consider two scenarios in line with realistic problems: 1) $\boldsymbol{\Sigma}$ is known; 2) it is unknown so that the transformations in real data analysis need to rely on its estimator $\widehat{\boldsymbol{\Sigma}}$. Let $\mathbf{U}$ be the inverse of the Cholesky factorization of $\boldsymbol{\Sigma}$, i.e., $\mathbf{U}$ is the lower triangular matrix such that $\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}' = \mathbf{I}$. The decorrelation transformation (DT) of $\widehat{\boldsymbol{\beta}}$ is defined as

$$\widehat{\boldsymbol{\beta}}^{\mathrm{DT}} = \mathbf{U}\widehat{\boldsymbol{\beta}}. \tag{4}$$

Let diagonal matrix $\mathbf{D} = \left(\mathrm{diag}\left(\boldsymbol{\Sigma}^{-1}\right)\right)^{-\frac{1}{2}}$. The innovated transformation (IT) of $\widehat{\boldsymbol{\beta}}$ is

$$\widehat{\boldsymbol{\beta}}^{\mathrm{IT}} = \mathbf{D}\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\beta}}. \tag{5}$$

Based on $\widehat{\boldsymbol{\Sigma}}$ and correspondingly $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{D}}$, we define the "estimated DT" (referred to as eDT) as $\widehat{\boldsymbol{\beta}}^{\mathrm{eDT}} = \widehat{\mathbf{U}}\widehat{\boldsymbol{\beta}}$. The "estimated IT" (eIT) is defined as $\widehat{\boldsymbol{\beta}}^{\mathrm{eIT}} = \widehat{\mathbf{D}}\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\beta}}$.

These transformations have several properties important for studying the SNR. First, it is trivial to see that both the DT and the IT rescale the estimators into unit variance since both $\mathrm{Var}(\widehat{\boldsymbol{\beta}}^{\mathrm{DT}}) = \mathbf{I}$ and $\mathrm{Var}(\widehat{\boldsymbol{\beta}}^{\mathrm{IT}}) = \mathbf{D}\boldsymbol{\Sigma}^{-1}\mathbf{D}$ have a unit diagonal. Second, the transformations for $\widehat{\boldsymbol{\beta}}$ and its Z-score vector $\mathbf{T}$ are equal, i.e., $\widehat{\boldsymbol{\beta}}^{\mathrm{DT}} = \mathbf{T}^{\mathrm{DT}}$ and $\widehat{\boldsymbol{\beta}}^{\mathrm{IT}} = \mathbf{T}^{\mathrm{IT}}$. In general, the DT

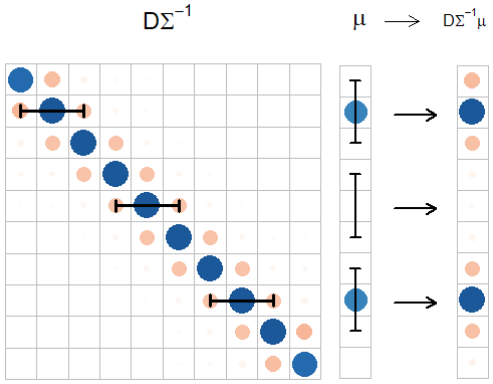


Fig 1: An illustrative case that the IT increases the SNRs. The dots represent nonzero elements of the matrix or vector (a larger size indicates a larger magnitude); empty cells represent negligible elements (i.e., zero or close-to-zero). The segments represent the band width of non-negligible correlations. Because diagonal elements of $\mathbf{D}\boldsymbol{\Sigma}^{-1}$ are larger than 1, if the signals are sparser than the correlations (i.e., the distances between original signals in $\boldsymbol{\mu}$ are wider than the correlation band in $\mathbf{D}\boldsymbol{\Sigma}^{-1}$), both the magnitude and the quantity of nonzero elements in $\mathbf{D}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ are increased over $\boldsymbol{\mu}$.

and the IT are invariant to any component-wise scaling by nonzero constants. Third, the DT of $\widehat{\boldsymbol{\beta}}$ equals the DT of $\widehat{\boldsymbol{\beta}}^{\text{IT}}$: $(\widehat{\boldsymbol{\beta}}^{\text{IT}})^{\text{DT}} = \widehat{\boldsymbol{\beta}}^{\text{DT}}$. Fourth, $\mathbf{T}$ and $\mathbf{T}^{\text{IT}}$ are mutually IT, i.e., $(\mathbf{T}^{\text{IT}})^{\text{IT}} = \mathbf{T}$. Furthermore, these equations remain true for the eDT and the eIT, and if $\widehat{\boldsymbol{\Sigma}} \xrightarrow{P} \boldsymbol{\Sigma}$, we have $\text{Var}(\mathbf{T}^{\text{eDT}}) \to \mathbf{I}$ and $\text{Var}(\mathbf{T}^{\text{eIT}}) \to \mathbf{D}\boldsymbol{\Sigma}^{-1}\mathbf{D}$. See Lemma 1.1 and the proof given in Section 1 of the Supplementary Material (Zhang et al., 2022).

In the following, we briefly discuss the essence for the DT and the IT to increase or decrease the SNR. That is, if $\boldsymbol{\Sigma}$ is a symmetric positive definite matrix with unit diagonal, then $1 \le U_{ii} \le \sqrt{(\boldsymbol{\Sigma}^{-1})_{ii}}$ for all $i = 1, \cdots, n$. The inequalities are strict for at least some i unless $\boldsymbol{\Sigma}$ is the identity matrix. See Lemma 1.2 and the proof in Section 1 of the Supplementary Material (Zhang et al., 2022). Based on these inequalities, we can show that the DT and the IT increase the SNR as long as the causal effects are "sparser" than their correlations (e.g., when the distances between causal SNPs are wider than LD blocks). Consider the one-effect case for example. Let $\boldsymbol{\mu} = \text{E}(\mathbf{T})$ with $\mu_j = c > 0$ at an arbitrary location j and $\mu_i = 0$ for all $i \neq j$. The SNR at the jth element is increased by the DT because $(\mathbf{U}\boldsymbol{\mu})_j = cU_{jj} \ge c$. The SNR after the IT is further increased because $(\mathbf{D}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})_j = c\sqrt{(\boldsymbol{\Sigma}^{-1})_{jj}} \ge cU_{jj}$. Furthermore, the stronger the correlations, the larger the increases are. On the other hand, if we start from $\mathbf{T}^{\text{IT}}$, by Lemma 1.1, its DT and IT will reversely transform $\mathbf{D}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ back to $\mathbf{U}\boldsymbol{\mu}$ and $\boldsymbol{\mu}$, respectively, and thus reduce SNRs. In this case, the nonzero elements in $\mathbf{D}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ does not have a sparse pattern as $\boldsymbol{\mu}$ does. That is, the DT and the IT could reduce signals under some dense signals. As for multiple causal effects, Figure 1 provides a schematic idea; rigorous studies are given in the next section.

3. Model fittings and SNR. This section connects the three genetic association approaches to the transformations and provides theoretical comparisons among their SNRs. For technical convenience, we first consider the linear model (LM) with known error variance under the finite sample scenario, then we study the GLM under asymptotics.

3.1. *Under the LM.* Consider the LM for modeling continuous traits:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{6}$$

where $\mathbf{X}_{N\times n}$ and $\mathbf{Z}_{N\times n}$ are given design matrices with the kth row vector being $\mathbf{X}'_{k\cdot}$ and $\mathbf{Z}'_{k\cdot}$ in (1), respectively. The error term $\boldsymbol{\epsilon} \sim N(\mathbf{0}_{N\times 1}, \sigma^2\mathbf{I}_{N\times N})$. We assume σ^2 is known and N is finite. Estimated σ^2 is to be discussed in the GLM setting under asymptotics.

For simple notations and meaningful interpretations, we can standardize genotype data $\mathbf{X}$ conditional on $\mathbf{Z}$. Specifically, let $\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ be the projection matrix onto the column space of $\mathbf{Z}$. Define diagonal matrix $\boldsymbol{\Omega} = \left(\text{diag}\left(\mathbf{X}'(\mathbf{I}-\mathbf{H})\mathbf{X}\right)\right)^{-\frac{1}{2}}$ with the diagonal elements $\Omega_{ii} = 1/\sqrt{\mathbf{X}'_i(\mathbf{I}-\mathbf{H})\mathbf{X}_i}$, where $\mathbf{X}_i$ is the ith column of $\mathbf{X}$. We can rewrite equation (6) as

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta}^* + \boldsymbol{\epsilon}^*, \tag{7}$$

where $\mathbf{Y}^* = (\mathbf{I}-\mathbf{H})\mathbf{Y}/\sigma$, $\mathbf{X}^* = (\mathbf{I}-\mathbf{H})\mathbf{X}\boldsymbol{\Omega}$, $\boldsymbol{\beta}^* = \boldsymbol{\Omega}^{-1}\boldsymbol{\beta}/\sigma$, and $\boldsymbol{\epsilon}^* = (\mathbf{I}-\mathbf{H})\boldsymbol{\epsilon}/\sigma$. The columns of $\mathbf{X}^*$ have unit norm, i.e., $\mathbf{X}^{*\prime}_i\mathbf{X}^*_i = 1$. The matrix

$$\mathbf{M} = \mathbf{X}^{*\prime}\mathbf{X}^* = \boldsymbol{\Omega}\mathbf{X}'(\mathbf{I}-\mathbf{H})\mathbf{X}\boldsymbol{\Omega} \tag{8}$$

is a correlation matrix because it is symmetric positive definite with the unit diagonal (assuming $\mathbf{X}'(\mathbf{I}-\mathbf{H})\mathbf{X}$ has full rank). $\mathbf{M}$ measures the correlations among genotype data conditional on $\mathbf{Z}$. For example, when $\mathbf{Z} = \mathbf{1}$ being the intercept vector, $\mathbf{M}$ is the matrix of Pearson's correlation coefficients among genotypes, a typical estimate of the LD r values among SNPs.

The interpretations of the causal genetic effects by $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$ are different. A nonzero β_i measures a subject's trait value change per unit change of genotype. It is the allelic effect when the genotype is typically defined as the copy number of minor allele. In contrast, β^*_i is a population-level genetic effect that involves both genetic and environmental variations:

$$\beta^*_i = \sqrt{\mathbf{X}'_i(\mathbf{I}-\mathbf{H})\mathbf{X}_i}\beta_i/\sigma, \quad i = 1, ..., n, \tag{9}$$

where $\mathbf{X}'_i(\mathbf{I}-\mathbf{H})\mathbf{X}_i$ measures the genetic variation conditional on $\mathbf{Z}$, and σ measures the variation of environmental factors. In the case of $\mathbf{Z} = \mathbf{1}$ and the genotype being the copy number of minor allele, $\mathbf{X}'_i(\mathbf{I}-\mathbf{H})\mathbf{X}_i/N = 2\widehat{q}_i(1-\widehat{q}_i)$, where $\widehat{q}_i$ is the estimated minor allele frequency (MAF) of the ith SNP. Therefore, a rare SNP with a small MAF has a weak effect β^*_i at the population level, unless its allelic effect β_i is high under the "Common Disease, Rare Variant" hypothesis (Schork et al., 2009). Moreover, $\boldsymbol{\beta}^*$ is directly related to the genetic heritability (i.e., the proportion of total trait variation due to genetic variation): $h = \frac{\sum_i^n \beta^{*2}_i/N}{1+\sum_i^n \beta^{*2}_i/N}$ when causal SNPs are independent (Wu et al., 2014).

Note that $\boldsymbol{\beta}^*$ is a component-wise scaling of $\boldsymbol{\beta}$, which implies two properties. First, their least-squares estimators (equivalent to the maximum likelihood estimators (MLE) under the LM) keep the same functional relationship: $\widehat{\boldsymbol{\beta}}^* = \boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\beta}}/\sigma$. Second, $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\boldsymbol{\beta}}$ have the same SNRs.

In the approach of studying genetic associations by the joint model fitting, we fit $\mathbf{Y}$ to the full data $\mathbf{X}^*$ (or equivalently, $(\mathbf{X}, \mathbf{Z})$). The estimators of the genetic effects are unbiased:

$$\widehat{\boldsymbol{\beta}}^*_{\text{J}} = \left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{Y}^* = \boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\beta}}_{\text{J}}/\sigma \sim N\left(\boldsymbol{\beta}^*, \mathbf{M}^{-1}\right). \tag{10}$$

Denote diagonal matrix $\boldsymbol{\Lambda} = \left(\text{diag}\left(\mathbf{M}^{-1}\right)\right)^{-\frac{1}{2}}$ with diagonal elements $\Lambda_{ii} = 1/\sqrt{(\mathbf{M}^{-1})_{ii}}$. Standardizing the estimated coefficients by $\boldsymbol{\Lambda}$ gives the Z-score statistics:

$$\mathbf{T}_{\text{J}} = \boldsymbol{\Lambda}\widehat{\boldsymbol{\beta}}^*_{\text{J}} \sim N\left(\boldsymbol{\Lambda}\boldsymbol{\beta}^*, \boldsymbol{\Lambda}\mathbf{M}^{-1}\boldsymbol{\Lambda}\right). \tag{11}$$

The SNRs are $|\text{E}(\mathbf{T}_{\text{J}})| = |\boldsymbol{\Lambda}\boldsymbol{\beta}^*|$. Note that the SNRs are smaller than the genetic effects, i.e., $|\boldsymbol{\Lambda}\boldsymbol{\beta}^*| \le |\boldsymbol{\beta}^*|$, because $0 \le \boldsymbol{\Lambda}_{ii} \le 1$ (by Lemma 1.2). The equal sign holds only if $\mathbf{M} = \mathbf{I}$, i.e., when genotypes are completely independent (conditional on $\mathbf{Z}$). That is, for given genetic effects, the LDs reduce the SNRs obtained from the joint fitting.

In the approach of studying genetic associations by the marginal model fitting between $\mathbf{Y}$ and $\mathbf{X}_i^*$ (or equivalently, $(\mathbf{X}_i, \mathbf{Z})$), $i = 1, ..., n$, the estimators are

$$\widehat{\boldsymbol{\beta}}_{\mathrm{M}}^* = \mathbf{T}_{\mathrm{M}} = \mathbf{X}^{*\prime}\mathbf{Y}^* = \boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\beta}}_{\mathrm{M}}/\sigma \sim N(\mathbf{M}\boldsymbol{\beta}^*, \mathbf{M}). \tag{12}$$

$\widehat{\boldsymbol{\beta}}_{\mathrm{M}}^*$ is readily the Z-scores $\mathbf{T}_{\mathrm{M}}$ because $\mathbf{M}$ already is a correlation matrix. Note that the correlations among input statistics in $\mathbf{T}_{\mathrm{M}}$ is consistent with the correlations among the genotypes, while the correlations of $\mathbf{T}_{\mathrm{J}}$ is the inverse. For example, if SNPs are positively correlated, $\mathbf{T}_{\mathrm{J}}$ has negative correlations. Furthermore, the means and the correlations of $\mathbf{T}_{\mathrm{J}}$ (similarly for $\mathbf{T}_{\mathrm{M}}$) are interconnected through data $\mathbf{X}$ and $\mathbf{Z}$ when $\boldsymbol{\beta} \neq \mathbf{0}$. It is different from the GMM, which assumes signals and correlations are independent parameters (Hall and Jin, 2010).

Deduction shows that $\mathbf{T}_{\mathrm{J}}$ in (11) and $\mathbf{T}_{\mathrm{M}}$ in (12) are mutually IT, i.e., $\mathbf{T}_{\mathrm{J}}^{\mathrm{IT}} = \mathbf{T}_{\mathrm{M}}$ and $\mathbf{T}_{\mathrm{M}}^{\mathrm{IT}} = \mathbf{T}_{\mathrm{J}}$. Furthermore, the decorrelation transformations for $\mathbf{T}_{\mathrm{J}}$ and $\mathbf{T}_{\mathrm{M}}$ give the same statistics. Let $\mathbf{U}$ be the inverse of the Cholesky factorization of $\mathbf{M}^{-1}$, i.e., $\mathbf{U}\mathbf{M}^{-1}\mathbf{U}' = \mathbf{I}$. We have

$$\mathbf{T}^{\mathrm{DT}} = \mathbf{T}_{\mathrm{J}}^{\mathrm{DT}} = \mathbf{T}_{\mathrm{M}}^{\mathrm{DT}} = (\mathbf{U}')^{-1}\mathbf{X}^{*\prime}\mathbf{Y}^* \sim N(\mathbf{U}\boldsymbol{\beta}^*, \mathbf{I}). \tag{13}$$

See Lemma 1.3 with deduction in Section 1 of the Supplementary Material (Zhang et al., 2022).

The following theorem provides sufficient conditions that guarantee an increasing order of the SNRs of $\mathbf{T}_{\mathrm{J}}$, $\mathbf{T}^{\mathrm{DT}}$ and $\mathbf{T}_{\mathrm{M}}$. Specifically, the SNRs of $\mathbf{T}_{\mathrm{J}}$ are component-wisely less than those of $\mathbf{T}^{\mathrm{DT}}$ and $\mathbf{T}_{\mathrm{M}}$, and the maximum SNR of $\mathbf{T}^{\mathrm{DT}}$ is less than some SNR of $\mathbf{T}_{\mathrm{M}}$. The proof is given in Section 1 of the Supplementary Material (Zhang et al., 2022).

THEOREM 3.1. *Consider the LM in (6) with coefficients $\beta_i \neq 0$ for $i \in \mathcal{M}^* = \{i_1, \ldots, i_K\} \subset \{1, \ldots, n\}$ and $\beta_i = 0$ otherwise. Consider either of the following conditions:*

1. *$K = 1$ (i.e., single causal effect).*
2. *$K > 1$ (i.e., multiple causal effects). Assume $\mathbf{X}_i'(\mathbf{I} - \mathbf{H})\mathbf{X}_j = 0$ for all i and j such that $|i - j| \geq b$, where $b = \min_{i_k, i_l \in \mathcal{M}^*} |i_k - i_l|$.*

For $\mathbf{T}_J$ in (11), $\mathbf{T}_M$ in (12), and $\mathbf{T}^{DT}$ in (13), we have that

$$|\mathrm{E}(T_{Ji})| \leq \min\{|\mathrm{E}(T_i^{DT})|, |\mathrm{E}(T_{Mi})|\}, \text{ for each } i = 1, \cdots, n, \text{ and}$$

$$\max_{i \in \{1,\ldots,n\}} |\mathrm{E}(T_i^{DT})| \leq |\mathrm{E}(T_{Mi_k})|, \text{ for some } i_k \in \mathcal{M}^*.$$

The theorem indicates that if there is only one causal SNP in the SNP-set to be tested, these signal relationships hold whenever correlations (primarily the LDs) exist. In the case of multiple causal SNPs, the result holds as long as they locate more separately than strong correlations (i.e., dispersed over blocks of strong LDs). In either case, other properties regarding the effects and correlations, such as their magnitudes and directions, do not change the order.

A good property of the joint fitting is that it gives an unbiased estimator of $\boldsymbol{\beta}$. However, SNP-set analysis aims to detect the existence of causal effects within the set, not to differentiate them. The joint fitting' SNRs are smaller because of the larger variances of $\mathbf{M}^{-1}$ in (10). Moreover, the joint fitting is computationally more intensive because of the need for matrix inversion. Therefore, under sparse causal effects, the joint fitting has a disadvantage in signal strength even if the SNP-set analysis is conceptually a joint analysis of multiple SNPs.

Theorem 3.1 clarifies how the iHC procedure should be interpreted under linear models. It is meant to apply the DT or the IT to $\mathbf{T}_{\mathrm{J}}$, instead of $\mathbf{T}_{\mathrm{M}}$ that was suggested in (Barnett, Mukherjee and Lin, 2017). The reason lies in the condition of effect sparsity for guaranteeing the IT to increase the SNR. If the causal genetic effects represented by nonzero β_i's are sparse, the signals of the joint fitting $\mathbf{T}_{\mathrm{J}}$ remain sparse. On the other hand, the signals of the marginal fitting are often no longer sparse because of the transformation $\mathrm{E}(\mathbf{T}_{\mathrm{M}}) = \mathbf{M}\boldsymbol{\beta}^*$. Therefore, for sparse causal effects, applying the IT to $\mathbf{T}_{\mathrm{M}}$ would reverse back to $\mathbf{T}_{\mathrm{J}}$ and thus reduce signals (recall that $\mathbf{T}_{\mathrm{J}}$ and $\mathbf{T}_{\mathrm{M}}$ are mutually IT by Lemma 1.3).

3.2. *Under the GLM.* This section extends the SNR framework to the GLM, which involves various types of Z-score statistics from different likelihood-based estimations. Correlations often need to be estimated. The normality of the Z-scores often requires asymptotics under $N \to \infty$ and regularity conditions carefully checked for mathematical rigor.

The GLM considers that the responses Y_k in (1) have the density function in the exponential family (McCullagh and Nelder, 1989):

$$f(y_k|\theta_k, \phi) = \exp\{(y_k\theta_k - b(\theta_k))/a_k(\phi) + c(y_k, \phi)\}, \tag{14}$$

where a_k, b and c are parameter functions determined by specific distributions. For example, for continuous traits, the normal distribution $Y_k \sim N(\mu_k, \sigma^2)$ corresponds to $a_k(\phi) = \sigma^2$, $b(\theta_k) = \theta_k^2/2$, and $\theta_k = \mu_k$. For binary traits, $Y_k \sim \text{Bernoulli}(p_k)$, which has $a_k(\phi) = 1$, $b(\theta_k) = \log(1 + e^{\theta_k})$, and $\theta_k = \log(p_k/(1 - p_k))$. In general, $\mathrm{E}(Y_k) = b'(\theta_k)$ and $\mathrm{Var}(Y_k) = a_k(\phi)b''(\theta_k)$ under regularity conditions such as the exchangeability of the differentiation and the integration of the log density function (Shao, 2010). Therefore, the covariates in (1) are linked to θ_k through $g(b'(\theta_k)) = \mathbf{X}'_{k\cdot}\boldsymbol{\beta} + \mathbf{Z}'_{k\cdot}\boldsymbol{\gamma}$. In this paper, we consider the commonly used canonical link, i.e., $g = (b')^{-1}$, such that $\theta_k = \mathbf{X}'_{k\cdot}\boldsymbol{\beta} + \mathbf{Z}'_{k\cdot}\boldsymbol{\gamma}$. Canonical link is used in both the LM and the logit model. ϕ is called the dispersion parameter, which is a nuisance parameter regarding the coefficient parameters $\boldsymbol{\alpha}' = (\boldsymbol{\beta}', \boldsymbol{\gamma}')$. Following the typical GLM inference, we assume $a_k(\phi) = \phi/a_k$ with known positive a_k, so that the MLE of $\boldsymbol{\alpha}$ does not depend on ϕ (Shao, 2010). Most commonly used GLM-modeling distributions satisfy this assumption, including the LM and the logit model.

For the joint model fitting, we consider the MLE and the one-step MLE of $\boldsymbol{\beta}$. The one-step MLE is the first iteration in computing the MLE using Fisher-scoring method (Shao, 2010). It can be considered an approximation of MLE, and it has a closed form that can be used to establish the transformational relationship with the score statistics to be discussed below.

Specifically, with n fixed and $N \to \infty$, as a classic result, the MLE $\widehat{\boldsymbol{\beta}}$ has an asymptotic normal distribution (see Lemma 1.4 in the Supplementary Material (Zhang et al., 2022) for the deduction under regularity conditions):

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N(\mathbf{0}, \phi\mathbf{V}(\boldsymbol{\alpha})). \tag{15}$$

The asymptotic variance-covariance matrix $\mathbf{V}(\boldsymbol{\alpha}) = \lim_{N\to\infty} N\mathbf{V}_N(\boldsymbol{\alpha})$ with

$$\mathbf{V}_N(\boldsymbol{\alpha}) = (\tilde{\mathbf{X}}'(\mathbf{I} - \tilde{\mathbf{H}})\tilde{\mathbf{X}})^{-1}, \tag{16}$$

where the regressors are weighted: $\tilde{\mathbf{X}} = \mathbf{W}^{\frac{1}{2}}\mathbf{X}$, $\tilde{\mathbf{Z}} = \mathbf{W}^{\frac{1}{2}}\mathbf{Z}$, and $\tilde{\mathbf{H}} = \tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'$, and the weighting diagonal matrix is $\mathbf{W}(\boldsymbol{\alpha}) = \mathrm{diag}\{a_k b''(\mathbf{X}'_{k\cdot}\boldsymbol{\beta} + \mathbf{Z}'_{k\cdot}\boldsymbol{\gamma}); k = 1, ..., N\}$.

Component-wise standardization of $\widehat{\boldsymbol{\beta}}$ gives the Z-score statistics:

$$\mathbf{T} = \sqrt{N}\widehat{\phi}^{-\frac{1}{2}}\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\beta}}, \tag{17}$$

where $\widehat{\phi}$ is the MLE of ϕ, the diagonal matrix $\widehat{\boldsymbol{\Lambda}} = (\mathrm{diag}(N\mathbf{V}_N(\widehat{\boldsymbol{\alpha}})))^{-\frac{1}{2}}$, and $\mathbf{V}_N(\widehat{\boldsymbol{\alpha}})$ is the matrix (16) evaluated at the MLE $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$. Let $\boldsymbol{\Lambda} = (\mathrm{diag}(\mathbf{V}(\boldsymbol{\alpha})))^{-\frac{1}{2}}$. The distribution is

$$\mathbf{T} - \sqrt{N}\phi^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{\beta} \xrightarrow{D} N(0, \boldsymbol{\Lambda}\mathbf{V}(\boldsymbol{\alpha})\boldsymbol{\Lambda}). \tag{18}$$

The one-step MLE is the first iteration in computing the MLEs by Fisher-scoring method, where the initial value is estimated under $H_0 : \boldsymbol{\beta} = \mathbf{0}$. Let $\widehat{\boldsymbol{\alpha}}^{(0)} = (\mathbf{0}', \widehat{\boldsymbol{\gamma}}^{(0)\prime})'$ under the restraint of H_0, where $\widehat{\boldsymbol{\gamma}}^{(0)}$ is the maximum quasi-likelihood estimator of $\boldsymbol{\gamma}$ using the controlling covariates $\mathbf{Z}$ only. Denote $\mathcal{U}(\boldsymbol{\alpha}) = \frac{\partial}{\partial\boldsymbol{\alpha}}\log L(\boldsymbol{\alpha}, \phi|\mathbf{y})$ the score function regarding $\boldsymbol{\alpha}$, and $\mathcal{I}(\boldsymbol{\alpha}) = \mathrm{Var}(\mathcal{U}(\boldsymbol{\alpha}))$ the Fisher information matrix. The one-step MLE of $\boldsymbol{\alpha}$ is $\widehat{\boldsymbol{\alpha}}^{(1)} = \binom{\widehat{\boldsymbol{\beta}}^{(1)}}{\widehat{\boldsymbol{\gamma}}^{(1)}} = \widehat{\boldsymbol{\alpha}}^{(0)} + \left(\mathcal{I}(\widehat{\boldsymbol{\alpha}}^{(0)})\right)^{-1}\mathcal{U}(\widehat{\boldsymbol{\alpha}}^{(0)})$. Accordingly, the one-step MLE of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}^{(1)} = \mathbf{V}_N(\widehat{\boldsymbol{\alpha}}^{(0)})\mathbf{X}'\mathbf{A}(\mathbf{Y} - \widehat{\boldsymbol{\mu}}^{(0)}), \tag{19}$$

where $\mathbf{V}_N(\widehat{\boldsymbol{\alpha}}^{(0)})$ is the matrix (16) evaluated at $\widehat{\boldsymbol{\alpha}}^{(0)}$, $\mathbf{A} = \text{diag}\{a_k; k = 1, ..., N\}$, and $\widehat{\boldsymbol{\mu}}_k^{(0)} = g^{-1}\left(\mathbf{Z}_{k\cdot}'\widehat{\boldsymbol{\gamma}}^{(0)}\right)$. For example, in the logit model, when $\mathbf{Z} = \mathbf{1}$ for the intercept, we have $\widehat{\boldsymbol{\mu}}_k^{(0)} = \bar{y}$, $\widehat{\boldsymbol{\gamma}}^{(0)} = \log\left(\bar{y}/(1-\bar{y})\right)$. Under proper conditions, especially $\|\left(\mathcal{I}(\boldsymbol{\alpha})\right)^{1/2}(\widehat{\boldsymbol{\alpha}}^{(0)} - \boldsymbol{\alpha})\| = O_p(1)$, we have

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}) \xrightarrow{D} N(0, \phi\mathbf{V}(\boldsymbol{\alpha})). \tag{20}$$

The Z-score statistics can be based on either $\widehat{\boldsymbol{\alpha}}^{(0)}$ or $\widehat{\boldsymbol{\alpha}}^{(1)}$:

$$\mathbf{T}_l^{(1)} = \sqrt{N}\widehat{\phi}^{-\frac{1}{2}}\widehat{\boldsymbol{\Lambda}}_l\widehat{\boldsymbol{\beta}}^{(1)}, \quad l = 0, 1, \tag{21}$$

where $\widehat{\boldsymbol{\Lambda}}_l = \left(\text{diag}\left(N\mathbf{V}_N(\widehat{\boldsymbol{\alpha}}^{(l)})\right)\right)^{-\frac{1}{2}}$, $l = 0$ or 1, is an estimator based on $\widehat{\boldsymbol{\alpha}}^{(0)}$ or $\widehat{\boldsymbol{\alpha}}^{(1)}$. In either case, $\mathbf{T}_l^{(1)}$ has the same asymptotic distribution as (18). These results are given by Lemma 1.5 in Section 1 of the Supplementary Material (Zhang et al., 2022).

Similarly, for the marginal model fitting, the marginal estimator of $\boldsymbol{\beta}$ can be obtained from the maximum marginal likelihood estimator (MMLE) (Fan et al., 2010) or the one-step MMLE. The MMLE relies on an iterative algorithm to maximize the marginal likelihood, similar to that for the MLE. The one-step MMLE of the ith coefficient β_i, $i = 1, ..., n$, follows equation (19) except using data $(\mathbf{X}_i, \mathbf{Z})$, where $\mathbf{X}_i$ is the ith column of $\mathbf{X}$. Let $\widehat{\boldsymbol{\Omega}}_0 = \left(\text{diag}\left([N\mathbf{V}_N(\widehat{\boldsymbol{\alpha}}^{(0)})]^{-1}\right)\right)^{-\frac{1}{2}}$. The vector format of all one-step MMLEs is

$$\widehat{\boldsymbol{\beta}}_{\text{M}}^{(1)} = \frac{1}{N}\widehat{\boldsymbol{\Omega}}_0^2\mathbf{X}'\mathbf{A}(\mathbf{Y} - \widehat{\boldsymbol{\mu}}^{(0)}). \tag{22}$$

The marginal score statistics broadly used in genetical studies (Barnett, Mukherjee and Lin, 2017) are actually the Z-scores of the one-step MMLEs:

$$\mathbf{T}_{\text{M}} = \frac{1}{\sqrt{N}}\widehat{\phi}^{-\frac{1}{2}}\widehat{\boldsymbol{\Omega}}_0\mathbf{X}'\mathbf{A}(\mathbf{Y} - \widehat{\boldsymbol{\mu}}^{(0)}) = \sqrt{N}\widehat{\phi}^{-\frac{1}{2}}\widehat{\boldsymbol{\Omega}}_0^{-1}\widehat{\boldsymbol{\beta}}_{\text{M}}^{(1)}. \tag{23}$$

Moreover, it can be shown that $\mathbf{T}_{\text{M}}$ is the eIT of $\widehat{\boldsymbol{\beta}}^{(1)}$ in (19) based on the estimated variance-covariance matrix $\widehat{\boldsymbol{\Sigma}} = \widehat{\text{Var}}(\widehat{\boldsymbol{\beta}}^{(1)}) = \widehat{\phi}\mathbf{V}_N(\widehat{\boldsymbol{\alpha}}^{(0)})$. Following that, we have

$$\mathbf{T}_{\text{M}} - \sqrt{N}\phi^{-\frac{1}{2}}\boldsymbol{\Omega}\mathbf{V}^{-1}(\boldsymbol{\alpha})\boldsymbol{\beta} \xrightarrow{D} N(0, \boldsymbol{\Omega}\mathbf{V}^{-1}(\boldsymbol{\alpha})\boldsymbol{\Omega}), \tag{24}$$

where $\boldsymbol{\Omega} = \left(\text{diag}\left(\mathbf{V}^{-1}(\boldsymbol{\alpha})\right)\right)^{-\frac{1}{2}}$. The above results are given by Lemma 1.6 in Section 1 of the Supplementary Material (Zhang et al., 2022).

We can also apply the eIT to the MLE $\widehat{\boldsymbol{\beta}}$ (or equivalently, its Z-score $\mathbf{T}$ in (17)), or adjust $\mathbf{T}_{\text{M}}$, to get the following two statistics,

$$\begin{aligned} \mathbf{T}^{\text{eIT}} &= \sqrt{N}\widehat{\phi}^{-\frac{1}{2}}\widehat{\boldsymbol{\Omega}}(N\mathbf{V}_N(\widehat{\boldsymbol{\alpha}}))^{-1}\widehat{\boldsymbol{\beta}}, \\ \mathbf{T}_{\text{M}}^{(1)} &= \frac{1}{\sqrt{N}}\widehat{\phi}^{-\frac{1}{2}}\widehat{\boldsymbol{\Omega}}_1\mathbf{X}'\mathbf{A}(\mathbf{Y} - \widehat{\boldsymbol{\mu}}^{(0)}), \end{aligned} \tag{25}$$

where $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Omega}}_1$ have the same formula as $\widehat{\boldsymbol{\Omega}}_0$ except using $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\alpha}}^{(1)}$ instead of $\widehat{\boldsymbol{\alpha}}^{(0)}$. $\mathbf{T}^{\text{eIT}}$ and $\mathbf{T}_{\text{M}}^{(1)}$ have the same asymptotic distribution as (24).

Regarding the decorrelation transformations, we can apply the eDT to the joint and marginal estimators too. Specifically, applying the eDT to the MLE in (17), the one-step MLE in (21), and the marginal score statistics in (23), respectively, we get

$$\begin{aligned} \mathbf{T}^{\text{D}} &= \left(\widehat{\phi}\mathbf{V}_N(\widehat{\boldsymbol{\alpha}})\right)^{-\frac{1}{2}}\widehat{\boldsymbol{\beta}}, \\ \mathbf{T}_l^{(1)\text{D}} &= \left(\widehat{\phi}\mathbf{V}_N(\widehat{\boldsymbol{\alpha}}^{(l)})\right)^{-\frac{1}{2}}\widehat{\boldsymbol{\beta}}^{(1)}, \quad l = 0, 1, \\ \mathbf{T}_{\text{M}}^{\text{D}} &= \left(\widehat{\boldsymbol{\Omega}}_0 N\mathbf{V}_N(\widehat{\boldsymbol{\alpha}}^{(0)})\widehat{\boldsymbol{\Omega}}_0\right)^{-\frac{1}{2}}\mathbf{T}_{\text{M}}. \end{aligned} \tag{26}$$

Under the same conditions that lead to the same asymptotic distribution for the MLE and the one-step MLE, these statistics have the same asymptotic distribution:

$$\mathbf{T}^{\mathrm{D}} - \sqrt{N}\,(\phi\mathbf{V}(\boldsymbol{\alpha}))^{-\frac{1}{2}}\,\boldsymbol{\beta} \xrightarrow{D} N(0,\mathbf{I}). \tag{27}$$

Some classic global hypothesis testing statistics are actually functions of the decorrelation-based Z-scores. In particular, through the L_2 norm (i.e., the sum of quadratic terms), the classic Wald and score statistics (Fahrmeir and Kaufmann, 1985) are, respectively,

$$W_N = \|\mathbf{T}^{\mathrm{D}}\|^2, \quad S_N = \|\mathbf{T}_0^{(1)\mathrm{D}}\|^2. \tag{28}$$

It has been shown that under proper conditions both statistics are asymptotically equivalent to the log likelihood ratio statistic (Fahrmeir, 1987)

$$\lambda_N = -2\log(L(\widehat{\boldsymbol{\alpha}}^{(0)})/L(\widehat{\boldsymbol{\alpha}})). \tag{29}$$

Note that under the LM, the MLE is obtained in one step. Both $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}^{(1)}$ reduce to (10), where $\phi = \sigma^2$, $\mathbf{W} = \mathbf{I}$, $\mathbf{A} = \mathbf{I}$, and $\widehat{\boldsymbol{\mu}}^{(0)} = \mathbf{HY}$. Similarly, $\mathbf{T}_{\mathrm{M}}$, $\mathbf{T}^{\mathrm{eIT}}$, and $\mathbf{T}_{\mathrm{M}}^{(1)}$ reduce to (12). σ^2 is replaced by its MLE when it is unknown.

Similar to Theorem 3.1, under the GLM and proper conditions, if the causal effects in $\boldsymbol{\beta}$ are sparser than the correlations among covariates, the SNRs of the marginal estimators are larger than the SNRs of the joint estimators, and the SNRs of the decorrelated estimators are in-between. The following theorem gives the rigorous statement.

THEOREM 3.2. *Consider the GLM defined by (1) and (14). Let* $\mathbf{V}(\boldsymbol{\alpha}) = \lim_{N\to\infty} N\mathbf{V}_N(\boldsymbol{\alpha})$ *with* $\mathbf{V}_N(\boldsymbol{\alpha})$ *in (16),* $\boldsymbol{\Lambda} = (\mathrm{diag}(\mathbf{V}(\boldsymbol{\alpha})))^{-\frac{1}{2}}$, *and* $\boldsymbol{\Omega} = \left(\mathrm{diag}\left(\mathbf{V}^{-1}(\boldsymbol{\alpha})\right)\right)^{-\frac{1}{2}}$. *Follow the conditions in Lemmas 1.4 – 1.6. The estimators and Z-score statistics involve three SNR vectors:*

- $\boldsymbol{\mu} = \phi^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{\beta}$ *for the joint-fitting* $\widehat{\boldsymbol{\beta}}$ *and* $\widehat{\boldsymbol{\beta}}^{(1)}$ *and their Z-scores in (17) and (21).*
- $\boldsymbol{\mu}_M = \phi^{-\frac{1}{2}}\boldsymbol{\Omega}\mathbf{V}^{-1}(\boldsymbol{\alpha})\boldsymbol{\beta}$ *for the marginal-fitting* $\widehat{\boldsymbol{\beta}}_M^{(1)}$ *and the Z-scores in (23) and (25).*
- $\boldsymbol{\mu}_D = \phi^{-\frac{1}{2}}(\mathbf{V}(\boldsymbol{\alpha}))^{-\frac{1}{2}}\boldsymbol{\beta}$ *for the decorrelation Z-scores in (26).*

Assume coefficients $\beta_i \neq 0$ *for* $i \in \mathcal{M}^* = \{i_1,\ldots,i_K\} \subset \{1,\ldots,n\}$ *and* $\beta_i = 0$ *otherwise. Consider either of the following two cases:*

1. $K = 1$ *(i.e., single causal effect).*
2. $K > 1$ *(i.e., multiple causal effects) and* $\mathbf{X}_i'(\mathbf{I} - \tilde{\mathbf{H}})\mathbf{X}_j/N \to 0$ *for all* i *and* j *such that* $|i-j| \geq b$, *where* $b = \min_{i_k, i_l \in \mathcal{M}^*} |i_k - i_l|$.

Then,

$$|\mu_i| \leq \min\{|\mu_{Di}|, |\mu_{Mi}|\}, \text{ for each } i = 1,...,n, \text{ and } \max_{i\in\{1,\ldots,n\}} |\mu_{Di}| \quad \leq |\mu_{Mi_k}|, \text{ for some } i_k \in \mathcal{M}^*.$$

3.3. *Under Bivariate Model.* The assumption of sparse causal effects may not always be satisfied. Here, we exhaust possible SNRs under the bivariate model of $n = 2$ SNPs. This case study helps understand in principle how the advantages of different genetic association approaches are influenced by the patterns of causal effects and the correlations (LDs).

Under the LM, we set the scaled effects in (7) as $\boldsymbol{\beta}^* = \binom{a}{b}$ and the correlation matrix in (8) as $\mathbf{M} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Under the GLM, they correspond to $\boldsymbol{\beta}^* = \phi^{-\frac{1}{2}}\boldsymbol{\Omega}^{-1}\boldsymbol{\beta}$ and $\mathbf{M} = \boldsymbol{\Omega}\mathbf{V}^{-1}(\boldsymbol{\alpha})\boldsymbol{\Omega}$ based on the analogous connection between equations (12) and (24). Therefore, the following SNR comparisons are valid for both models.

Direct calculation gives $\mathbf{M}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$ and $\mathbf{U} = \frac{1}{\sqrt{1-\rho^2}}\begin{pmatrix} \sqrt{1-\rho^2} & 0 \\ -\rho & 1 \end{pmatrix}$. Under the marginal fitting, the Z-score statistics in (12) are $\mathbf{T}_{\mathrm{M}} = \binom{T_{\mathrm{M}1}}{T_{\mathrm{M}2}} \sim N(\binom{a+\rho b}{\rho a+b}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$.

Based on $\mathbf{T}_{\mathrm{M}}$, we can also consider Stouffer's statistic, which is simply the standardized summation of the Z-scores (Stouffer et al., 1949): $T_{\mathrm{MS}} = \frac{T_{\mathrm{M1}}+T_{\mathrm{M2}}}{\sqrt{2(1+\rho)}} \sim N((a+b)\sqrt{\frac{1+\rho}{2}}, \quad 1)$.

For the joint model fitting, by (11) and the fact that $\mathbf{T}_{\mathrm{J}}$ is the IT of $\mathbf{T}_{\mathrm{M}}$, the statistics are $\mathbf{T}_{\mathrm{J}} = \binom{T_{\mathrm{J1}}}{T_{\mathrm{J2}}} \sim N(\sqrt{1-\rho^2}\binom{a}{b}, \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix})$. Based on $\mathbf{T}_{\mathrm{J}}$, Stouffer's statistic is $T_{\mathrm{JS}} = \frac{T_{\mathrm{J1}}+T_{\mathrm{J2}}}{\sqrt{2(1-\rho)}} \sim N((a+b)\sqrt{\frac{1+\rho}{2}}, \quad 1)$. Clearly, T_{JS} and T_{MS} have the same distribution.

The decorrelated Z-scores are $\mathbf{T}^{\mathrm{DT}} = \mathbf{U}\mathbf{T}_{\mathrm{M}} \sim N(\binom{a+\rho b}{b\sqrt{1-\rho^2}}, \mathbf{I})$. Note that $T_1^{\mathrm{DT}} = T_{\mathrm{M1}}$ and $T_2^{\mathrm{DT}} = T_{\mathrm{J2}}$. Accordingly, Stouffer's statistic is $T_{\mathrm{S}}^{\mathrm{DT}} = \frac{T_1^{\mathrm{DT}}+T_2^{\mathrm{DT}}}{\sqrt{2}} \sim N(\frac{a+b(\rho+\sqrt{1-\rho^2})}{\sqrt{2}}, \quad 1)$.

These six statistics involve nine SNR values. We can simplify the comparisons based on their formulas to understand the essential differences. First, due to the symmetry of the expressions we can assume $0 \le |a| \le b$ without loss of generality. Second, for $\mathbf{T}_{\mathrm{M}}$, $\mathbf{T}_{\mathrm{J}}$ and $\mathbf{T}^{\mathrm{DT}}$, we consider their maximal SNR (i.e., the second elements) because large SNRs help detect the effects' existence. Therefore, we simplify the comparisons to be among four SNRs:

$$
\begin{aligned}
|\mathrm{E}(T_{\mathrm{M2}})| &= \rho a + b; \\
|\mathrm{E}(T_{\mathrm{J2}})| = |\mathrm{E}(T_2^{\mathrm{DT}})| &= b\sqrt{1-\rho^2}; \\
|\mathrm{E}(T_{\mathrm{MS}})| = |\mathrm{E}(T_{\mathrm{JS}})| &= (a+b)\sqrt{\frac{1+\rho}{2}}; \\
|\mathrm{E}(T_{\mathrm{S}}^{\mathrm{DT}})| &= \frac{|a+b(\rho+\sqrt{1-\rho^2})|}{\sqrt{2}}.
\end{aligned} \tag{30}
$$

The direction and magnitude of a, b, and ρ together characterize the effect and correlation patterns and decide the SNRs. Figure 2 shows the pair-wise contrasts among these SNRs in (30) over the area of $a \in [-1,1]$ and $\rho \in [-1,1]$ at $b = 1$ (Figure S1 in Section 2 of the Supplementary Material (Zhang et al., 2022) gives the 3-D surfaces of these SNRs). The positive and negative differences are red- and green-colored, respectively. Overall, $|\mathrm{E}(T_{\mathrm{M2}})|$ is larger in a majority of the areas than the rest (as shown by the first row of the panels). It indicates a general advantage of the marginal fitting. For example, $|\mathrm{E}(T_{\mathrm{M2}})| \ge |\mathrm{E}(T_{\mathrm{J2}})|$ except in two relatively small regions, where a and ρ are both large but with opposite directions, causing effect cancellation under $|\mathrm{E}(T_{\mathrm{M2}})|$.

The analytical formulas in (30) help us understand the mechanism of how causal effects and the correlation determine the SNRs. For example, in case of sparse effect (i.e., $a = 0$), the marginal fitting's SNR, $|\mathrm{E}(T_{\mathrm{M2}})| = b$, is always the largest, a result consistent with Theorems 3.1 and 3.2. In case of dense effects (i.e., $0 < |a| \le b$), the relative advantages depend on a and ρ. First, under independence with $\rho = 0$, $|\mathrm{E}(T_{\mathrm{M2}})| = |\mathrm{E}(T_{\mathrm{J2}})| = b$ and $|\mathrm{E}(T_{\mathrm{MS}})| = |\mathrm{E}(T_{\mathrm{S}}^{\mathrm{DT}})| = (a+b)/\sqrt{2}$. These equations indicate that the marginal and joint fittings are equivalent (as expected), and Stouffer's combination is better when $a > (\sqrt{2}-1)b$. Second, in the equal-effect case (i.e., $a = b$, which is often assumed in theoretical studies (Hall and Jin, 2010)), it is straightforward to show that $|\mathrm{E}(T_{\mathrm{MS}})|$ by Stouffer's combination without decorrelation is always the largest for all ρ. At the same time, $|\mathrm{E}(T_{\mathrm{J2}})| = |\mathrm{E}(T_2^{\mathrm{DT}})|$ by the joint fitting or the decorrelation is always the smallest if $\rho > 0$, while $|\mathrm{E}(T_{\mathrm{M2}})|$ by the marginal fitting is always the smallest if $\rho < 0$. Third, in the case that a and b have the same direction (i.e., $a > 0$) and positive correlation (i.e., $\rho > 0$), $|\mathrm{E}(T_{\mathrm{J2}})|$ is always the smallest.

It is interesting to see how the correlation ρ and the effect directions (i.e., the signs of a and b) could strengthen or weaken signals. First, for the joint fitting, $|\mathrm{E}(T_{\mathrm{J2}})| = b\sqrt{1-\rho^2} < b$ for all a and ρ. That is, correlation always weakens signal under the joint fitting. Second,

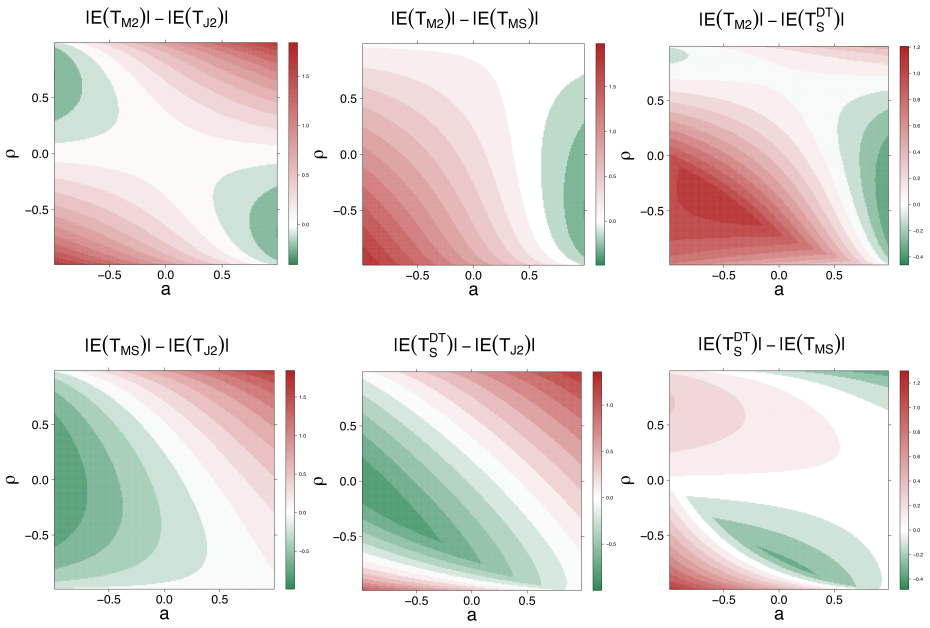


Fig 2: Contrasts of SNRs for marginal fitting, joint fitting, decorrelation, and summational statistics in (30). X-axis: $a \in [-1, 1]$; y-axis: $\rho \in [-1, 1]$; fixing $b = 1$. Red: Positive difference; Green: Negative difference.

$|\mathrm{E}(T_{\mathrm{M2}})|$ depends on the product $a\rho$. If $a\rho$ is in the same direction as b, the signal is strengthened; otherwise, the signal is weakened due to effect cancellation. Third, $|\mathrm{E}(T_{\mathrm{MS}})|$ in (30) is monotone in ρ. Therefore, the signal is weakened when $\rho < 0$, and is strengthened when $\rho > 0$. Also, the signal is always weakened due to cancellation whenever a and b have opposite signs. Moreover, if effects have different directions, $|\mathrm{E}(T_{\mathrm{MS}})|$ suffers more cancellation than $|\mathrm{E}(T_{\mathrm{M2}})|$. For example, when $a = -b$, we always have $|\mathrm{E}(T_{\mathrm{MS}})| = 0$, whereas $|\mathrm{E}(T_{\mathrm{M2}})|$ still is nonzero for all $\rho < 1$. Lastly, the influence of ρ to $|\mathrm{E}(T_{\mathrm{S}}^{\mathrm{DT}})|$ is more complicated. When a and b have the same direction, the largest signal is obtained at $\rho = 0.71$.

Based on the above comparisons we can conclude a few general rules. First, the marginal fitting has the advantage when effects are sparser than the correlation, or when effects and data correlations are in the same direction (e.g., a, b, and ρ have the same sign). Second, joint fitting could be the best in case of heavy effect-cancellations. For example, when a and b have the same direction but are negatively correlated, or when they have opposite directions but are positively correlated (e.g., $\rho = 0.5, a = -1, b = 1$ give $|\mathrm{E}(T_{\mathrm{J2}})| = 0.87 > |\mathrm{E}(T_{\mathrm{M2}})| = 0.5 > |\mathrm{E}(T_{\mathrm{S}}^{\mathrm{DT}})| = 0.26 > |\mathrm{E}(T_{\mathrm{MS}})| = 0$). Third, when correlation is negative and strong, i.e., $\rho \to -1$, $\mathrm{E}(T_{\mathrm{J2}}) = \mathrm{E}(T_{\mathrm{MS}}) = 0$. In this case, $|\mathrm{E}(T_{\mathrm{M2}})| = b - a$ and $|\mathrm{E}(T_{\mathrm{S}}^{\mathrm{DT}})| = (b - a)/\sqrt{2}$ are nonzero (unless $a = b$) and could even reach the largest value when effects a and b have opposite directions (i.e., $a < 0$).

4. SNRs Under Typical LDs of Human Genome. This section surveys how typical LDs among human SNPs influence the signal strengths by different genetic association approaches. SNP-sets were grouped by SNPs located in the regions of 20kbps length from the transcription start sites (TSSs) of genes. For each SNP-set, the genotype data $\mathbf{X}$ in (6) were retrieved from the 1000 Genomes Project Phase 3 (sample size $N = 2504$). The controlling covariates' data $\mathbf{Z}$ were generated by simulation, containing two independent variables $Z_1 \sim$

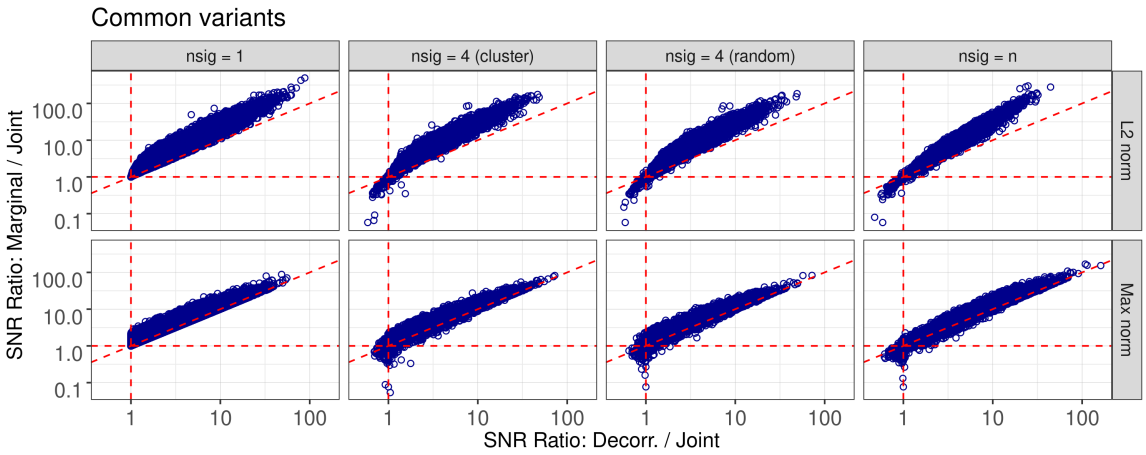


Fig 3: SNR comparisons based on common SNPs grouped by regions from gene TSSs. X-axis: the ratio of the SNR-norms from the decorrelation versus the joint fitting; Y-axis: ratio of the SNR-norms from the marginal fitting versus the joint fitting. Rows: L_2-norm and L_∞-norm of the SNR vectors; columns: four scenarios of causal effects.

Bernoulli(0.5) and $Z_2 \sim N(0,1)$. SNP-sets were discarded if their $\mathbf{M}$ matrices in (8) could not be stably inverted. We considered common SNPs (MAF ≥ 0.01) and rare SNPs (MAF < 0.01 but minor allele count (MAC) ≥ 10). In total, 770,014 common SNPs in 19,163 sets and 676,858 rare SNPs in 17,656 sets were studied; the median set size was 37 common (or 41 rare) SNPs.

For each SNP set of size n, causal effects were simulated by setting coefficients $\beta_i = 1$, or $\beta_i \sim$ Uniform(0, 2), $i \in \mathcal{M}^* = \{i_1, \ldots, i_K\} \subseteq \{1, \ldots, n\}$. Seven scenarios were considered: a) $K = 1$ and $\mathcal{M}^*$ is random. This scenario mimics sparse causal effects. b) $K = 4$ and $\mathcal{M}^*$ is random. The effects could be sparse or dense depending on n. c) $K = 4$ and $\mathcal{M}^*$ is clustered at two random locations (each with two adjacent causal SNPs). It mimics the locally clustered dense effects, which are of practical and theoretical interests (Ke, Fan and Wu, 2015). d) $K = n$, i.e., all SNPs were causal – a case of surely dense effects. The rest three scenarios e) – g) are correspondingly similar to b) – d) except the effect sizes are random $\beta_i \sim$ Uniform(0, 2). For the $K = 1$ scenario, the random magnitude of the single effect does not influence the ordering of the SNRs of joint fitting, marginal fitting, and decorrelation. The property has been guaranteed by Theorems 3.1 and 3.2, no matter whether this single effect is fixed or random.

We compare the SNR vectors of $\mathbf{T}_{\text{J}}$ in (11), $\mathbf{T}_{\text{M}}$ in (12), and $\mathbf{T}^{\text{DT}}$ in (13). The comparisons are analogously valid under the GLM. Two types of SNR-norms were applied to summarize the overall signal strengths of the SNR vectors: the Euclidean norm (L_2) and the maximum (L_∞-norm). Figure 3 shows the comparisons based on common SNPs in scenarios a) – d). Each dot represents a SNP set. The x-axis is the ratio of the SNR-norm from the decorrelation versus that from the joint fitting; the y-axis is the ratio between the marginal fitting and the joint fitting. A dot with $x > 1$ (or $y > 1$) indicates that the signal from the decorrelation (or the marginal fitting) is stronger than the signal from the joint fitting. A dot above the $y = x$ diagonal line indicates that the signal from the marginal fitting is stronger than that from the decorrelation (see Figure S3 for a direct comparison between the marginal fitting vs. the decorrelation).

A few interesting observations can be made. First, in the scenario of single causal SNP ($K = 1$) represented by the first column of the figure, all dots satisfy $y \geq x \geq 1$. It means

that for all SNP sets, the marginal fitting is uniformly better than the decorrelation, and the decorrelation is uniformly better than the joint fitting. This observation directly evidenced the conclusions of Theorems 3.1 and 3.2.

Second, in the other three scenarios with multiple causal SNPs, similar relationships still remain true for most SNP sets. The joint fitting outperformed the marginal fitting (or the decorrelation) in only about 5.4% (or 4.4%) sets. Most of such sets are relatively small: 83% of them have $n \leq 10$, which means that the causal effects are relatively dense – at least 40% SNPs were causal. Meanwhile, dense effects are a necessary but not sufficient condition for the advantage of the joint fitting. In the scenario that all SNPs are causal ($K = n$, represented by the last column of the figure), the joint fitting still gave weaker signals for most SNP sets. We found that joint fitting could provide stronger signals than the marginal fitting if the latter is suffered from signal cancellation, e.g., by negative LDs (an example of SNP set in gene *TUBG2* was examined; see Figure S2 in Section 2 of the Supplementary Material (Zhang et al., 2022)).

Third, the dots in the first row (L_2-norm) arise above the diagonal line slightly further than those in the second row (L_∞-norm). That is, the advantage of the marginal fitting over the decorrelation is more noticeable when summarizing the SNRs by L_2-norm (the sum of the squared SNRs) than by L_∞-norm (the maximum). Therefore, the marginal fitting could be more prominent to the summation-based SNP-set tests than to the supremum-based tests. This phenomenon was also supported by power comparisons in the next section.

The SNR comparisons for rare SNPs and scenarios e) – g) are summarized by Figures S4 and S5 in Section 2 of the Supplementary Material (Zhang et al., 2022). The patterns of relative advantages are similar. Meanwhile, the SNR-norm ratios have smaller values because the LDs among rare variants are weaker so that the three genetic association approaches are closer to each other.

5. Statistical Power of SNP-set Tests. This section applies systematic simulations to reveal how the genetic association approaches influence different SNP-set tests' power at various causal effects and LD patterns. We consider two primary types of SNP-set tests. First, the summation-based test statistics utilize additive functions to combine SNPs' Z-scores or p-values. The SKAT (Wu et al., 2011) and Fisher's combination test are two representatives. Second, the supremum-based test statistics apply supremum functions to the ordered SNP p-values $P_{(1)} \leq ... \leq P_{(n)}$ (Zhang and Wu, 2022). Two representatives are the minimal p-value test (minP) and the HC test. Furthermore, we use the LRT in (29) as a benchmark. Details of the testing procedures, including test statistics and p-value computation (by the empirical and analytical methods well-established in the literature as the best practice available to control the type I error rates), are described in Section 3 of the Supplementary Material (Zhang et al., 2022).

We simulated continuous and binary traits by the LM and the logit model, respectively:

$$Y_k = \mathbf{X}'_{k\cdot}\boldsymbol{\beta} + 0.5Z_{1k} + 0.75Z_{2k} + \epsilon_k, \text{ where } \epsilon_k \overset{i.i.d.}{\sim} N(0,1),$$
$$\text{logit}(\mathbb{P}(Y_k = 1)) = \mathbf{X}'_{k\cdot}\boldsymbol{\beta} - 1.25 + 0.5Z_{1k} + 0.1Z_{2k}, \quad k = 1, ..., N,$$

where $Z_{1k} \sim \text{Bernoulli}(0.5)$ and $Z_{2k} \sim N(0,1)$ are independent. We focused on three well-known osteoporosis genes: *LRP5*, *MEPE* and *SOX6*, which have typical gene sizes and various LD patterns. Their genotype data $\mathbf{X}$ were from the 1000 Genome Project. The numbers of common/rare SNPs are *LRP5*: 64/64; *MEPE*: 87/63; *SOX6*: 49/59. The LD matrices are shown by Figure S6 in Section 3 of the Supplementary Material (Zhang et al., 2022). The LDs among common SNPs are more substantial than those among rare SNPs. We assumed various causal effects: the numbers of causal SNPs $K \in \{1, 2, 3, 4, 6, 8, 10, 12\}$ with the ~~nonzero~~

causal SNPs' coefficients ~~$\beta_i = \beta \in \{0.1, 0.2, ..., 1\}$~~ $\beta_i = \beta = 0, 0.1, 0.2, ...$ (the zero effect is used to evidence that the type I error rates are well-controlled). The causal SNPs were either randomly allocated or grouped into two clusters at random locations; each cluster contained half of the causal SNPs, i.e., a scenario of locally clustered effects. The empirical statistical power was obtained based on 2500 simulations at the type I error rate $\alpha = 0.05$ (also empirically controlled) and $\alpha = 2.5 \times 10^{-6}$ (by analytical calculation methods).

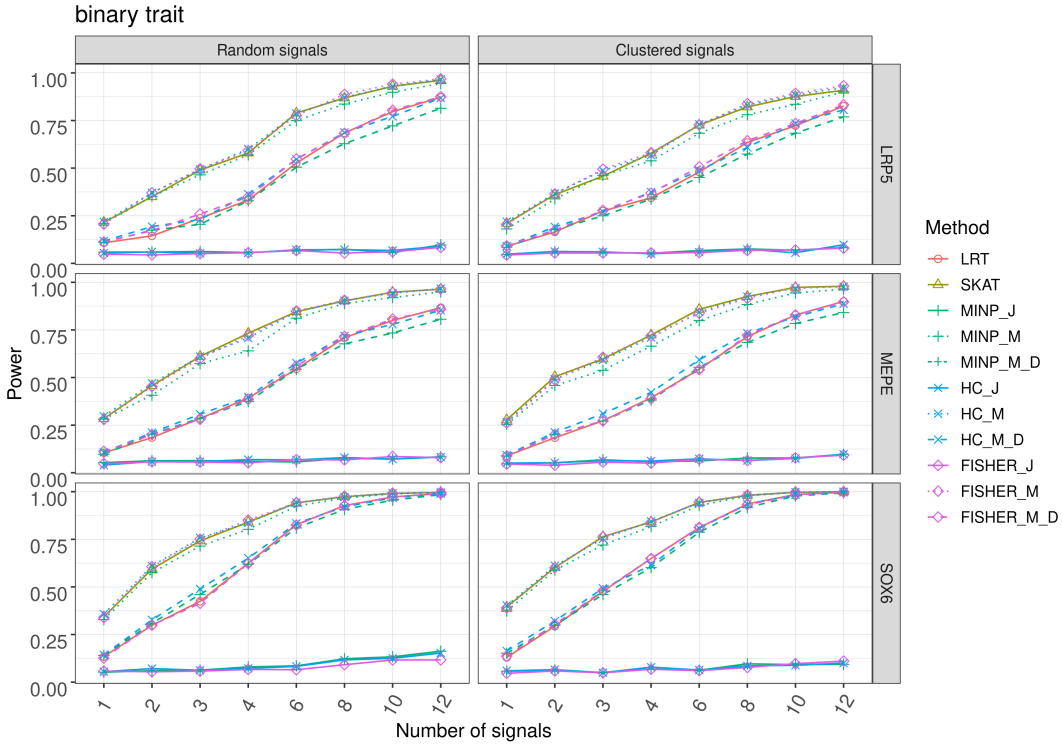
Depending on the LM or the logit model, SNP Z-scores by joint model fitting (denoted by _J in the following) were based on (11) or (17) (where σ was treated as unknown and estimated by the MLE). Similarly, Z-scores under the marginal fitting (_M) were based on (12) or (23). The Z-scores could be transformed by the estimated innovated transformation (_I) or by the estimated decorrelation transformation (_D). Figure 4 illustrates the power comparisons ($\alpha = 0.05$) for binary traits under various genetic effects β and causal SNP numbers K. The three genes are arranged in rows; the two signal patterns (random or clustered causal loci) are in columns. Results for binary traits under more settings and for continuous traits are given by Figures S7– ~~S10~~ S19 in Section 3 of the Supplementary Material (Zhang et al., 2022).

Comparison patterns are summarized in the following. Regarding the genetic association approaches, all three p-value combination tests (minP, HC, and Fisher) show that the marginal fitting (represented by the tests' dotted lines) leads to higher power than the decorrelation (dashed lines), which in turn has an advantage over the joint fitting. With the increase of effect size β or causal SNP number K, decorrelation-based tests catch up the power of the marginal-fitting based tests in a much faster pace than the joint-fitting based tests do.

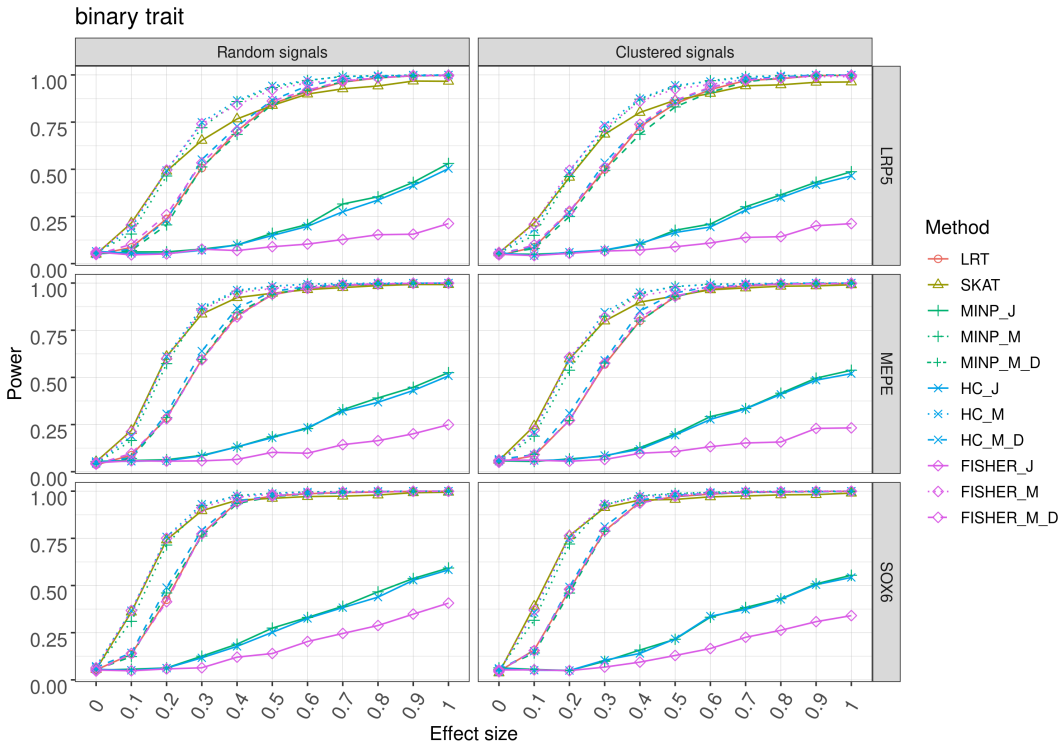
Regarding the types of SNP-set tests, the supremum-based minP and HC are similar; both show advantages when causal SNPs are fewer. Meanwhile, minP is slightly better when $K = 1$ and HC is sightly better when K is larger. This observation is consistent with the literature that HC is more robust to denser signals (Zhang, Jin and Wu, 2020; Zhang and Wu, 2022). As a summation-based test, Fisher's combination test has been shown suitable for detecting dense signals under independence (Littell and Folks, 1973; Zhang et al., 2020). With the correlated data of our sittings, Fisher shows a slight advantage over minP and HC when effects are dense and weak (e.g., $\beta = 0.1, 0.2$) under the marginal fitting or decorrelation. This result is consistent with a theoretical study of the HC (Donoho and Kipnis, 2021). However, minP and HC are often more powerful than Fisher under the joint fitting. Fisher has a larger power increase from the joint fitting to the marginal fitting than minP and HC. It echos the observation in Section 4 that the advantage of the marginal fitting is more noticeable when the signal strength is measured by the L_2-norm (i.e., the sum-squared SNRs) than by the L_∞-norm (i.e., the maximum). That is, the marginal fitting could be more prominent to the summation-based combination (Fisher) than the supremum-based tests (minP and HC).

The SKAT is defined based on the marginal fitting. When genetic effects are weak, it has similar power as other marginal-fitting-based tests (minP_M, HC_M, and Fisher_M). However, with the effect size increase, SKAT's power becomes relatively lower than these tests (sometimes even lower than the decorrelation-based tests). The LRT is reasonably close to the decorrelation-based minP/HC/Fisher; the observation is consistent with its asymptotic equivalence to the decorrelation-based Wald and score tests in (28) (Fahrmeir, 1987). The above comparison patterns remain stable whether the causal SNPs are randomly allocated or locally clustered (as shown by the two columns of the panels). They also well agree for continuous traits.

The comparisons for the rare SNPs are given by S11– S14 , S16, S17, and S19 in Section 3 of the Supplementary Material (Zhang et al., 2022). The comparison patterns are similar to those for common SNPs. However, there are several noticeable differences based on the features of rare SNPs. First, at the same $\boldsymbol{\beta}$ value (i.e., the allelic effect), the genetic association



(a) Genetic effect $\beta = 0.2$.



(b) Causal SNP number $K = 3$.

Fig 4: Power comparisons ($\alpha = 0.05$) under binary trait and common SNPs. Upper: Fix genetic effect $\beta = 0.2$ and vary causal SNP number K over x-axis; Lower: fix causal SNP number $K = 3$ and vary β. Three genes (LD patterns) are in rows; two signal patterns (random or clustered causal loci) are in columns. _J: the MLE (solid line); _M: the marginal score statistics (dotted line); _D: decorrelation (dashed line).

signals of rare variances are weaker at the population level due to smaller variations of rare SNPs, as was revealed by equation (9). Therefore, at the same $\boldsymbol{\beta}$ value, the power curves for rare SNPs are lower than those for common SNPs. Second, the LDs among rare SNPs are also weaker. Therefore, the differences among the genetic association approaches are less significant. Third, when causal effects are weak and dense, the advantages of SKAT and Fisher over minP and HC are more noticeable for rare SNPs than for common SNPs. A strong signal is more important for minP and HC to reach high power. The power comparisons at $\alpha = 2.5 \times 10^{-6}$ are in ~~Figures S15 (common SNPs) and S16 (rare SNPs)~~ Supplementary Figures S15 – S19 for binary and continuous traits under common and rare SNPs. The comparative observations are similar to those at $\alpha = 0.05$ in principle, although the advantage could slightly drop for the SKAT and rise for the LRT. We also applied power comparisons to verify the relationships among various Z-scores. The results well supported our theoretical study. See Figures S20 – S29 and discussions in Section 3 of the Supplementary Material (Zhang et al., 2022) for details.

6. GWAS Analysis of Osteoporosis. This section demonstrates the influences of genetic association approaches on the SNP-set tests by analyzing a large data of Osteoporosis from UK Biobank. The study participants are primarily Caucasians between 40 and 69 years old (Sudlow et al., 2015). We included osteoporosis cases with no pathological fractures according to the primary and secondary diagnosis fields of their hospital records. The SNP data is from genotyping by UK Biobank Axiom Array. As a typical quality control (QC), we excluded SNPs with missing rate $> 5\%$ or Hardy-Weinberg equilibrium testing p-value $< 10^{-6}$. We also removed subjects who had genotype missing rate $> 5\%$ or were estimated to be genetically related. After the QC, 14,469 cases (12,016 females and 2,453 males) and 409,205 controls (218,491 females and 190,714 males) were included in our analysis.

We grouped SNPs by genes and combined SNP p-values for testing genetic associations between genes and osteoporosis susceptibility. Specifically, for each gene, we applied a logit model that contained SNP genotypes and controlling covariates: age, sex, and the first five ancestry principal components (Abraham, Qiu and Inouye, 2017). Two-sided SNP p-values were calculated using the MLE-based Z-scores in (17), the marginal score statistics in (23), and the decorrelated Z-scores in (27). Each gene with more than one SNP was tested by the minP, the HC, and Fisher's combination test. We analyzed common SNPs (MAF ≥ 0.05), rare SNPs (MAF≤ 0.01), and combined SNPs (with MAC$\geq 2,000$ (Dey et al., 2017)), for which the number of involved SNPs (and genes) are 253,473 (16,526), 88,516 (16,563), and 467,948 (19,113), respectively. To put the results into context, we extensively searched literature and obtained a list of 679 genes with reported genetic associations with osteoporosis or bone mineral density (BMD) related traits (see Section 4 of the Supplementary Material (Zhang et al., 2022)).

6.1. *The analysis of common SNPs.* Figure 5 row 1 provides the Q-Q plots for all genes based on common SNPs. The genome-wide inflation is reasonably controlled. The top-hit genes at the genome-wide significance level of 0.05 are listed in Tables S1 – S3 in Section 4 of the Supplementary Material (Zhang et al., 2022). Most of them are in our known-gene list, indicating a reliable study.

Compared with the MLE by the joint fitting (denoted by J for simplicity), the marginal score statistics (M) and their decorrelated Z-scores (M_D) give smaller gene p-values overall. To see this point, first, in the Q-Q plots, the dots by M and M_D are higher than those by J. Second, M and M_D also lead to more gene hits than J does – the numbers of hits by M/M_D/J are 14/12/5 for the minP, 15/12/5 for the HC, and 11/13/4 for Fisher's combination test. Third, we studied 513 known genes mapped by common SNPs, considering that they

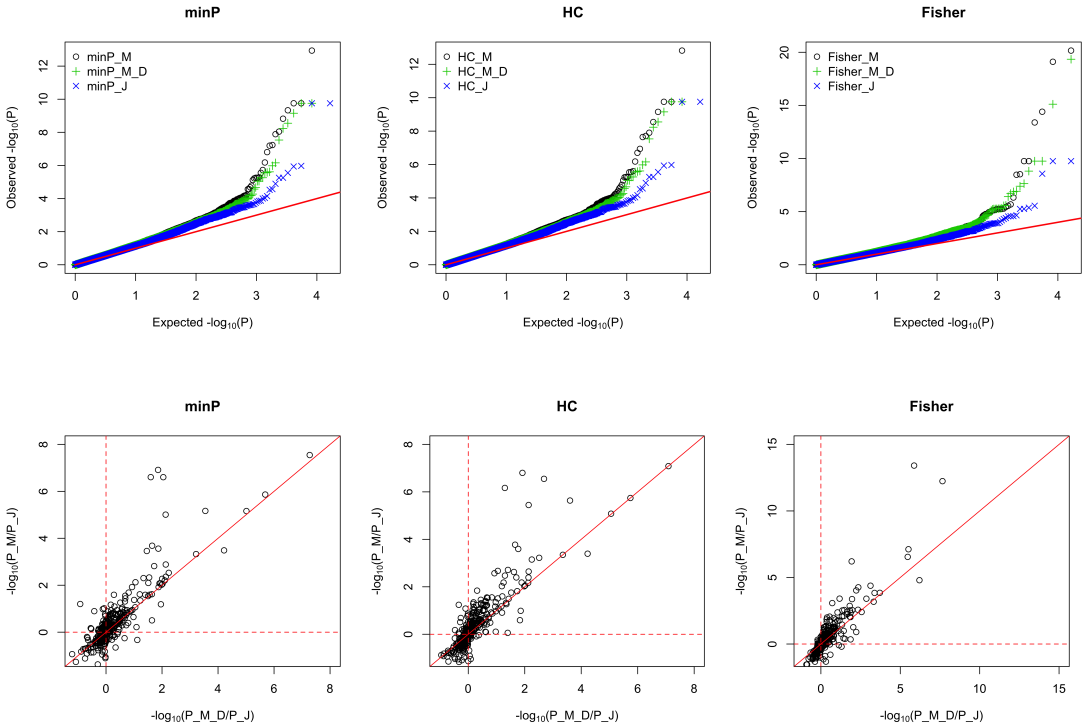


Fig 5: Gene-based SNP-set tests by common SNPs. Row 1: Q-Q plots for all genes. Three tests: the minP, the HC, and Fisher's combination. Three association approaches to get SNP p-values: the marginal score statistics (M), their decorrelated Z-scores (M_D), and the MLE-based Z-scores (J). Row 2: The $-\log_{10}$ ratio between gene p-values of the 513 known genes. The y-axis: gene p-value calculated by M versus that by J; x-axis: M_D versus J.

possess enriched causal genetic effects. The second row of Figure 5 summarizes the comparisons of these genes' p-values. A dot represents a gene. The y-axis is the $-\log_{10}$ of the ratio between the M-based p-value versus the J-based p-value; the x-axis is the log-ratio for M_D versus J. The plots show that if significant differences exist (represented by dots far away from the origin or the diagonal line), the marginal fitting is more likely to give smaller gene p-values.

Meanwhile, the MLE and the decorrelation procedure could still complement the marginal score statistics for detecting disease genes. For example, gene *HOXC9* is a known gene detected based on the MLE and the decorrelation procedure, but not by the marginal score statistics. Gene *ZBED4* is another known gene detected based on the decorrelation procedure, but not by the MLE or the marginal score statistics.

The top-hits contain two putative novel genes: *PRRC2A* and *UBA7*, which are likely relevant to osteoporosis. Gene *PRRC2A* has been shown to interact with genes *C1QBP* and *EIF3S6* (Lehner et al., 2004). *C1QBP* was connect to hip-BMD via enhancer SNP rs1806269 (Stelzer et al., 2016), and *EIF3S6* was connected to heel-BMD (Morris et al., 2019). Gene *UBA7*'s mRNA expression has been shown significant in hematopoietic stem cells and bone marrow (Stelzer et al., 2016). We suggest follow-up validations for these two genes.

Our results also demonstrate that the SNP-set tests often detect disease genes where single-SNP analysis could fail. For example, the minimum SNP p-values in gene *PRRC2A* ($3.76 \times$

10^{-5}, 1.5×10^{-3}, and 9.01×10^{-4} corresponding to M, J, and M_D, respectively) are all larger than 1.97×10^{-7}, the genome-wide significance threshold for the SNP p-values. Other examples include top-hit known genes *SUPT3H*, *HOXC9*, *ZBED4*, *MARK3*, and *TCIRG1* (see their minimum SNP p-values in Table S4 in Section 4 of the Supplementary Material (Zhang et al., 2022)).

6.2. *The analysis of rare and combined SNPs.* To analyze rare SNPs (MAF ≤ 0.01), we first overcame genome-wide inflation by using the saddle-point approximation to correct the SNP p-values obtained from the marginal score statistics (Dey et al., 2017). Correlation-based transformations are less influential for rare variants than common ones because the LDs among rare SNPs are weaker. Moreover, it is interesting to see that Fisher's combination test could detect more known genes than the minP and the HC (all three methods controlled genomic inflation excellently; see the Q-Q plot in ~~Figure S28~~ Supplementary Figure S31). Specifically, the minP and the HC detected only one gene *LRP5*, which is in our known gene list. Fisher's combination test detected this gene and other 13 genes, including known genes *SLC12A3*, *MYBPC3*, and *LMNA* (see Table S5 for the full list of top hits). The extra power of Fisher's combination test for rare-variant analysis echoes the power study results in Section 5.

The single-SNP analysis could not detect all top-hit genes except *LRP5*. Some of these putative novel genes have been reported in the literature as being relevant functionally or in model organisms. For example, gene *GCK* was shown associated with osteoarthritis through a close SNP rs3757837 (Evangelou et al., 2014). Its activity reduction was also related to osteoporosis in Drosophila (Mascolo et al., 2021). Gene *KCNH2* was shown associated with osteosarcoma and other bone diseases (Zeng et al., 2016). For gene *KCNQ1*, a study has shown that knockdown of *KCNQ1OT1* suppresses cell invasion and sensitizes osteosarcoma cells to cisplatin (Qi et al., 2019). Gene *RET* variants have been reported to be associated with a drug target for osteosarcoma (Kovac et al., 2021). Gene *INSR* is a receptor tyrosine kinase gene that mediates the pleiotropic actions of insulin. It has been considered connected with osteosarcoma because the insulin receptor on bone cells modulates the synthesis of collagen, and this role may be important in bone homeostasis (Pun, Lau and Ho, 1989). *INSR* has also been shown differentially expressed in human osteoarthritis chondrocytes (Rosa et al., 2011). A more detailed discussion is given in the Supplementary Material. We suggest follow-up validations for these genes.

We also carried out gene-set enrichment analyses based on the top genes' gene ontology (GO) terms and KEGG pathways. Table S6 lists four GO terms and two biochemical pathways related to osteoporosis, which are significantly enriched by top-hit genes. The GO terms are related to the bone-forming process (including "heart development", "regulation of secretion", and "tissue morphogenesis") and the bone regulation process (including "response to peptide") (Guo et al., 2019). The two enriched biochemical pathways are ovarian steroidogenesis and the mTOR signaling pathway. More details are given in Section 4 of the Supplementary Material (Zhang et al., 2022). Due to the limited space of the manuscript, we present the analysis results for the combined SNPs in Section 4 of the Supplementary Material (Zhang et al., 2022). The combined SNPs led to very consistent results with those of common SNPs.

7. Discussion. This work extended the SNR framework from the theoretical GMM setting to the applicational GLM setting. Based on that, we rigorously studied an important genetical problem on how causal genetic effects and LD patterns coherently influence the signal strengths of genetic association approaches and thus the power of the SNP-set test. The SNP-set test aims at determining the existence of causal SNPs rather than differentiating them. Therefore, even though the marginal model fitting based approach is biased in

estimating individual causal SNPs, it could provide higher signal strength for the SNP set. Computationally expensive joint model fitting (even if it is an unbiased estimation) and the decorrelation procedure (even if it is convenient for calculating p-values) are often less powerful, especially when the causal effects are dispersed over blocks of strong LDs. The study is meaningful to general applications; the results are valid for the global hypothesis testing problems that are based on the GLM with correlated covariates.

There are a few limitations of this work, which will be addressed in our future research. First, the current study on the influences of dense correlations is limited to the bivariate model or empirical methods. Analytical study for addressing more general correlations is interesting. Second, the current study assumes no gene-gene interactions, which could be an important factor in genetics. Further research could add such terms into the GLM and estimate the correlations among the cross-product terms. Third, the joint fitting and the decorrelation could still be critical complements to the marginal fitting for detecting novel disease genes. We are designing a testing procedure that automatically adapts to a proper transformation suitable for given correlation patterns and unknown causal effects. Fourth, although Cholesky decomposition is commonly used for decorrelation, other methods (such as eigenvalue decomposition) could lead to different SNRs (He and Wu, 2011). A further analytical study is interesting.

SUPPLEMENTARY MATERIAL

Supplement to "On signal-noise ratio of genetic effects and statistical power of SNP-set tests"

The Supplementary Material contains the lemmas, proofs of lemmas and theorems, extra numerical studies, and extra results of the GWAS of osteoporosis.

REFERENCES

ABRAHAM, G., QIU, Y. and INOUYE, M. (2017). FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33** 2776–2778.

ARIAS-CASTRO, E., HUANG, R. and VERZELEN, N. (2020). Detection of sparse positive dependence. *Electronic journal of statistics* **14** 702–730.

ARIAS-CASTRO, E. and WANG, M. (2017). Distribution-free tests for sparse heterogeneous mixtures. *Test* **26** 71–94.

BARNETT, I., MUKHERJEE, R. and LIN, X. (2017). The Generalized Higher Criticism for Testing SNP-set Effects in Genetic Association Studies. *Journal of the American Statistical Association* **112** 64–76.

DEY, R., SCHMIDT, E. M., ABECASIS, G. R. and LEE, S. (2017). A fast and accurate algorithm to test for binary phenotypes and its application to PheWAS. *The American Journal of Human Genetics* **101** 37–49.

DONOHO, D. L. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics* **32** 962-994.

DONOHO, D. L. and KIPNIS, A. (2021). The Impossibility Region for Detecting Sparse Mixtures using the Higher Criticism. *arXiv preprint arXiv:2103.03218*.

EVANGELOU, E., KERKHOF, H. J., STYRKARSDOTTIR, U., NTZANI, E. E., BOS, S. D., ESKO, T., EVANS, D. S., METRUSTRY, S., PANOUTSOPOULOU, K., RAMOS, Y. F. et al. (2014). A meta-analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. *Annals of the rheumatic diseases* **73** 2130–2136.

FAHRMEIR, L. (1987). Asymptotic testing theory for generalized linear models. *Statistics: A Journal of Theoretical and Applied Statistics* **18** 65–76.

FAHRMEIR, L. and KAUFMANN, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics* **13** 342–368.

FAN, J., SONG, R. et al. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *The Annals of Statistics* **38** 3567–3604.

FISHER, R. A. (1934). *Statistical Methods for Research Workers*, 5th ed ed. Oliver and Boyd, Edinburgh.

GUO, B. and WU, B. (2019). Powerful and efficient SNP-set association tests across multiple phenotypes using GWAS summary data. *Bioinformatics* **35** 1366–1372.

GUO, L., HAN, J., GUO, H., LV, D. and WANG, Y. (2019). Pathway and network analysis of genes related to osteoporosis. *Molecular medicine reports* **20** 985–994.

HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *The Annals of Statistics* **38** 1686-1732.

HE, S. and WU, Z. (2011). Gene-based Higher Criticism methods for large-scale exonic single-nucleotide polymorphism data. In *BMC proceedings* **5** S65. Springer.

HOH, J., WILLE, A. and OTT, J. (2001). Trimming, weighting, and grouping SNPs in human case-control association studies. *Genome Research* **11** 2115-2119.

HOTELLING, H. (1931). The generalization of Student's ratio. *The Annals of Mathematical Statistics* **2** 360-378.

KE, Z. T., FAN, J. and WU, Y. (2015). Homogeneity pursuit. *Journal of the American Statistical Association* **110** 175–194.

KOVAC, M., WOOLLEY, C., RIBI, S., BLATTMANN, C., ROTH, E., MORINI, M., KOVACOVA, M., AMELINE, B., KULOZIK, A., BIELACK, S. et al. (2021). Germline RET variants underlie a subset of paediatric osteosarcoma. *Journal of medical genetics* **58** 20–24.

KWAK, I.-Y. and PAN, W. (2016). Adaptive gene-and pathway-trait association testing with GWAS summary statistics. *Bioinformatics* **32** 1178–1184.

LEHNER, B., SEMPLE, J. I., BROWN, S. E., COUNSELL, D., CAMPBELL, R. D. and SANDERSON, C. M. (2004). Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region. *Genomics* **83** 153–167.

LITTELL, R. C. and FOLKS, J. L. (1973). Asymptotic optimality of Fisher's method of combining independent tests II. *Journal of the American Statistical Association* **68** 193–194.

LUO, L., PENG, G., ZHU, Y., DONG, H., AMOS, C. I. and XIONG, M. (2010). Genome-wide gene and pathway analysis. *European Journal of Human Genetics* **18** 1045-1053.

MARCHINI, J., DONNELLY, P. and CARDON, L. R. (2005). Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature genetics* **37** 413–417.

MASCOLO, E., LIGUORI, F., STUFERA MECARELLI, L., AMOROSO, N., MERIGLIANO, C., AMADIO, S., VOLONTÉ, C., CONTESTABILE, R., TRAMONTI, A. and VERNÌ, F. (2021). Functional Inactivation of Drosophila GCK Orthologs Causes Genomic Instability and Oxidative Stress in a Fly Model of MODY-2. *International Journal of Molecular Sciences* **22** 918.

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. CRC Press LLC, Florida.

MORRIS, J. A., KEMP, J. P., YOULTEN, S. E., LAURENT, L., LOGAN, J. G., CHAI, R. C., VULPESCU, N. A., FORGETTA, V., KLEINMAN, A., MOHANTY, S. T. et al. (2019). An atlas of genetic influences on osteoporosis in humans and mice. *Nature genetics* **51** 258–266.

PASANIUC, B. and PRICE, A. L. (2017). Dissecting the genetics of complex traits using summary association statistics. *Nature Reviews Genetics* **18** 117–127.

PUN, K., LAU, P. and HO, P. (1989). The characterization, regulation, and function of insulin receptors on osteoblast-like clonal osteosarcoma cell line. *Journal of Bone and Mineral Research* **4** 853–862.

QI, X., YU, X.-J., WANG, X.-M., SONG, T.-N., ZHANG, J., GUO, X.-Z., LI, G.-J. and SHAO, M. (2019). Knockdown of KCNQ1OT1 suppresses cell invasion and sensitizes osteosarcoma cells to CDDP by upregulating DNMT1-mediated Kcnq1 expression. *Molecular Therapy-Nucleic Acids* **17** 804–818.

ROSA, S., RUFINO, A., JUDAS, F., TENREIRO, C., LOPES, M. and MENDES, A. (2011). Expression and function of the insulin receptor in normal and osteoarthritic human chondrocytes: modulation of anabolic gene expression, glucose transport and GLUT-1 content by insulin. *Osteoarthritis and Cartilage* **19** 719–727.

SCHORK, N. J., MURRAY, S. S., FRAZER, K. A. and TOPOL, E. J. (2009). Common vs. rare allele hypotheses for complex diseases. *Current opinion in genetics & development* **19** 212–219.

SHAO, J. (2010). *Mathematical Statistics*, 2nd ed. Springer Verlag.

SIVA, N. (2008). 1000 Genomes project. *Nature biotechnology* **26** 256-256.

STELZER, G., ROSEN, N., PLASCHKES, I., ZIMMERMAN, S., TWIK, M., FISHILEVICH, S., STEIN, T. I., NUDEL, R., LIEDER, I., MAZOR, Y. et al. (2016). The GeneCards suite: from gene data mining to disease genome sequence analyses. *Current protocols in bioinformatics* **54** 1–30.

STOUFFER, S. A., SUCHMAN, E. A., DEVINNEY, L. C., STAR, S. A. and WILLIAMS, R. M. (1949). *The American Soldier: Adjustment during Army Life* **I**. Princeton University Press, New Jersey.

SUDLOW, C., GALLACHER, J., ALLEN, N., BERAL, V., BURTON, P., DANESH, J., DOWNEY, P., ELLIOTT, P., GREEN, J., LANDRAY, M. et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* **12** e1001779.

WU, Z. and ZHAO, H. (2009). Statistical power of model selection strategies for genome-wide association studies. *PLoS genetics* **5**.

WU, M. C., LEE, S., CAI, T., LI, Y., BOEHNKE, M. and LIN, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics* **89** 82-93.

WU, Z., SUN, Y., HE, S., CHO, J., ZHAO, H. and JIN, J. (2014). Detection boundary and Higher Criticism approach for sparse and weak genetic effects. *The Annals of Applied Statistics* **8** 824-851.

YANG, J., FERREIRA, T., MORRIS, A. P., MEDLAND, S. E., MADDEN, P. A., HEATH, A. C., MARTIN, N. G., MONTGOMERY, G. W., WEEDON, M. N., LOOS, R. J. et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature genetics* **44** 369.

ZENG, W., LIU, Q., CHEN, Z., WU, X., ZHONG, Y. and WU, J. (2016). Silencing of hERG1 gene inhibits proliferation and invasion, and induces apoptosis in human osteosarcoma cells by targeting the NF-κB pathway. *Journal of Cancer* **7** 746.

ZHANG, H., JIN, J. and WU, Z. (2020). Distributions and Power of Optimal Signal-Detection Statistics in Finite Case. *IEEE Transactions on Signal Processing* **68** 1021–1033.

ZHANG, H. and WU, Z. (2022). The general goodness-of-fit tests for correlated data. *Computational Statistics & Data Analysis* **167** 107379.

ZHANG, H., TONG, T., LANDERS, J. E. and WU, Z. (2020). TFisher: A powerful truncation and weighting procedure for combining p-values. *The Annals of Applied Statistics* **14** 178-201.

ZHANG, H., LIU, M., JIN, J. and WU, Z. (2022). Supplement to "On signal-noise ratio of causal genetic effects and statistical power of SNP-set tests". *Annals of Applied Statistics*.