We thank all reviewers for very helpful comments. This letter addresses the major questions raised by the reviewers.

Specific questions by Reviewer 1. Please see the response below for "distribution assumptions" and "global null and group of coefficients". We will correct our references and typos in the table. Thanks.

Specific questions by Reviewer 2. (1) We agree with the reviewer that the low signal-to-noise ratio (SNR) setting is very interesting and hurts us in high dimensions. Indeed, when the SNR is below a certain threshold, *no* testing procedure can succeed; therefore our goal is to completely characterize this detection boundary and establish a valid test as soon as the SNR exceeds this detection boundary. Our upper and lower bounds established in Section 3 are steps towards such exact characterizations. (2) We refer the reviewer to our response below for "more discussions of our simulations". On the design of the simulations, in the first part of the simulation section, we have varied the following factors: • signal-to-noise ratio • distribution of X • distribution of noise by varying (i) the Frobenius norm of Σ; (ii) the eigen-structure of Σ; (iii) the noise distribution for different scenarios. We also examine the effect of dimension in the second part of the simulation section. We shall elaborate more in our revised version to make these more clear.

Other comments: (1) While the optimality of our proposed test is validated in various settings, the optimal procedure in the most general form still remains open. We agree it is very interesting and non-trivial to see whether a modified F-test can be adapted for optimal testing in the low-SNR setting. (2) We share the reviewer's intuition that Lemma 1 should hold as in the small p case. However, it is non-trivial to demonstrate this result in the proportional regime where p grows together with n. In addition, considering the local alternatives for small $\beta^T\Sigma\beta$, only makes the testing problems more challenging. (3) The ARE compares the samples required for two tests to achieve the same power and it is defined as n_2/n_1. Therefore, $\mathrm{ARE} < 1$ means the first test is more effective. Sorry there was a typo in our manuscript and we will correct it in the revised version. Thanks for catching this. Indeed, there is randomness associated with S_k in the testing procedure. To conquer this issue, we provided high-probability guarantees in Section 2 and 3. Please also see the response below for "distribution assumptions".

Specific questions by Reviewer 3. Please see below for "global null and group of coefficients" and "more discussions of our simulations". We will elaborate more in our revised version when the eigen-spetra are not as nicely behaved. For Table 2, the Type I error rates are reflected in the column $\|\beta\|_2 = 0$; we shall make it more clear in the revised version.

Specific questions by Reviewer 4. Please see the response below for "global null and group of coefficients" and "novelty of the method and theory". We will add the comparisons of the running time in our revision. Thanks.

Global null and group of coefficients. To better demonstrate our main idea, our focus so far is mainly on testing the global null, however, our results and techniques can be substantially extended to testing other hypotheses. Built upon an improved argument of the high-dimensional F-test (see [36]), our framework can be *provably* adapted to testing whether $H_0 : G\boldsymbol{\beta} = r_0$ or $H_1 : G\boldsymbol{\beta} \neq r_0$ for $G \in \mathbb{R}^{q\times p}$ and $r_0 \in \mathbb{R}^q$ with $q \leq p$. For example, to test the joint significance of a group of coefficients, our test combines a sketching step (over the complement set of features) with the classical F-test. Fig 1 demonstrates its efficacy in the same setting as in our Section 4.

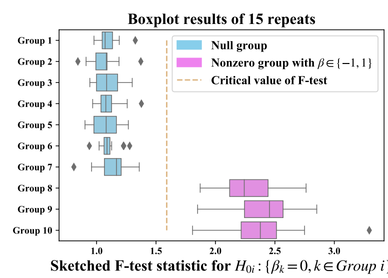

Fig 1: Sketched F-test for group testing. Here $n = 200, p = 500$ and $q = 50$.

Distribution assumptions. In our settings, each row of the design matrix X and noise vector Z follow Gaussian distributions and are independent of each other. We shall make these more explicit in our theorem statements. We will also make it clear that Theorem 1 and 3 hold beyond Gaussian settings, and a set of weaker distributional assumptions are stated in Appendix A. Moreover, we remark that the *Bayesian-type assumption (A)* in Proposition 1 is made only for illustrative purposes and is not used outside of line 192-215. Sorry for the confusion. In fact, the specific assumption (A) is invoked to provide intuition for the ARE expression in (8). We will also make it clear that we apply the *frequentist* approach throughout the entire paper, and all the theorems hold for any specific (β, Σ, σ) under mild conditions.

More discussions of our simulations. In most of our simulations, we consider cases where the eigenvalues of Σ enjoy a decaying structure, in which setting, the designs are of intrinsically lower dimensions. For such decaying structures, the U-statistics type tests (e.g. CGZ, ZC) by design are not suitable, therefore are not as competitive. These simulation results support our theoretical findings in Theorem 2 and 3. We emphasize that the optimality of our procedure relies on the intrinsic low-dimensional structure (in Definition 2/Appendix B); when there is no such structure, it is impossible to do feature-dimension reduction without losing information, and the optimal test for the global null remains open.

Novelty of the method and theory. As far as we know, the proposed procedure is the first attempt to analyze in details how sketching techniques work for testing regression coefficients. By a novel definition of the intrinsic dimension, we provide a systematic approach to determine the optimal sketching dimensions. While our theoretical results are built upon random matrix theory and the minimax decision framework, some technical lemmas introduced new techniques and are of independent interests.