**Responses to Reviewer Comments**

We thank both reviewers for their time and thoughtful evaluations of our manuscript. We are grateful to Reviewer 1 for their careful consideration of our revisions and for agreeing with the edits we previously introduced. We also sincerely thank Reviewer 2 for providing additional constructive comments, which have helped us further clarify and strengthen the manuscript. In this revision, we have carefully addressed each of Reviewer 2’s new concerns with additional analyses, methodological details, and clarifications. All changes are highlighted in yellow throughout the revised manuscript. We believe these updates comprehensively address the points raised and further improve the clarity and rigor of the work.

**Reviewers' Comments to Author:**

**Reviewer #1Comments to the Author**

**I acknowledge that KDPS prioritizes the retention of key individuals through phenotypic weights, and is particularly suitable for studies on rare phenotypes, but I still have the following issues for this work.**

**1. As a phenotype-aware method, KDPS may retain more cases but could also inadvertently preserve related individuals, potentially inflating false discovery rates compared to conservative, phenotype-naïve approaches. The manuscript did not assess whether KDPS adequately controls type I error in downstream association analyses, which is crucial for interpreting GWAS or similar results. Please note that, this comment was likely submitted in the last round, I do not know why it was ignored in the response.**

**Reviewer 2 – Response 1:** We thank the reviewer for raising this important comment again and sincerely apologize for the confusion in the earlier iteration. We have now clarified in the manuscript that KDPS ensures the removal of related individuals and does not retain kinship pairs after pruning. Specifically, we added the following sentence between lines 210–212:

“In both simulations and real-world applications, KDPS successfully pruned the complex relatedness networks, resulting in a final dataset of entirely unrelated individuals after kinship decoupling.”

Because KDPS operates on pre-calculated kinship matrices, the algorithm guarantees that, after pruning, the resulting dataset contains no individuals related above the user-defined kinship threshold. Thus, the retained cohort is free of relatedness according to the specified cutoff, mitigating concerns about inflated type I error rates in downstream analyses. We have updated the Results section to make this explicit, and we hope this addition addresses the reviewer’s concern.

**2. The heritability in the real-data application was consistently large; I suggest to evaluate the performance of KDPS regarding phenotypic heritability via suitable simulations.**

**Reviewer 2 – Response 2:** We thank the reviewer for raising this important point again and apologize that it was not sufficiently addressed in the earlier version of the manuscript. In direct response to this comment, we have now included a dedicated simulation study to evaluate the impact of phenotypic heritability on KDPS performance.

As described in the Methods (lines 164–181), we simulated phenotypes by first seeding cases at 10% prevalence and then propagating them to related individuals with probabilities scaled by both the kinship coefficient and a user-defined heritability indicator. This design allowed us to model a spectrum of traits ranging from non-heritable (indicator = 0, where relatives never inherit the phenotype) to strongly heritable (indicator ≥ 15, where nearly all close relatives inherit the phenotype).

In the Results (lines 264–273), we now report that subject retention ratios remained stable across all levels of heritability, showing that pruning is determined primarily by the relatedness network structure rather than phenotype assignment. In contrast, as expected, case retention ratios decreased as heritability increased, because higher heritability concentrates cases within related families that must be removed to achieve unrelatedness. These findings are summarized in the new Supplementary Figure S3.

Together, these additions clarify that KDPS performance is robust to varying levels of heritability with respect to subject retention, while the expected tradeoff in case retention reflects the necessary pruning of related cases. We hope this resolves the reviewer’s concern.

**3. The authors emphasized the high efficiency of computing time (such as processing 100,000 pairs of relationships < 15 minutes), but they did not mention the details of memory occupation, especially for biobank-scale genotypes and phenotypes.**

**Reviewer 2 – Response 3:** We appreciate the reviewer’s comment and agree that reporting memory usage is important for evaluating scalability. The memory footprint of KDPS is relatively modest, as the algorithm only requires kinship matrix files and phenotype files as inputs. To clarify this, we have added the following statement to the Results section (lines 212–215):

“Because KDPS only requires kinship matrix files and phenotype files as inputs, its memory footprint remained modest, never exceeding 4 GB of RAM. This makes the method practical and accessible on most consumer-grade computers and standard workstations.”

We believe this addition provides the necessary detail on memory efficiency to complement the discussion of usage of computational resources of KDPS.

**4. It will be better to offer more details for the real-data application to some phenotypes of the UK Biobank.**

We thank the reviewer for this helpful suggestion. In response, we have expanded the description of the real-data application in the Methods section (lines 183–194) to provide more details on the phenotypes and their extraction process. Specifically, we now state:

“Furthermore, KDPS was applied to four real-world phenotypes from the UK Biobank: schizophrenia (UKB ID 130874), acute myocardial infarction (UKB ID 131298), multiple sclerosis (UKB ID 131042), and alcohol drinking status (never drinkers, UKB ID 20117). Schizophrenia was defined using ICD-10 F20 diagnoses from hospital records (Fields 41202/41204), death registries, primary care, and self-report (Field 20002). Acute myocardial infarction was identified from hospital and death records (ICD-10 I21–I22; Fields 41202/41204, 40001/40002) and self-report (Field 20002). Multiple sclerosis was captured via ‘first occurrence’ fields (131042, 131043; ICD-10 G35) along with self-report and hospital data. Alcohol status was derived from questionnaire Field 20117 distinguishing never versus ever drinkers. These phenotypic traits were extracted and harmonized with respect to the pre-calculated pair-wise kinship coefficients for all UK Biobank individuals.”

We believe this additional detail improves clarity and provides the necessary context for how the phenotypes were defined and used in the KDPS application.