**Responses to Reviewer Comments**

We thank both reviewers for their time and thoughtful evaluations of our revised manuscript. We are delighted that Reviewer 1’s concerns have been fully addressed. In the updated manuscript, we have carefully addressed each of Reviewer 2’s new concerns with additional analyses, methodological details, and clarifications, detailed below. All changes are highlighted in yellow throughout the revised manuscript.

**Reviewers' Comments to Author:**

**Reviewer #2 Comments 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 apologize for the lack of clarity. We have revised the manuscript to clarify and detail that KDPS generates an unrelated set of subjects post decoupling. KDPS operates on pre-calculated kinship matrices, the algorithm generates a dataset that contains no individuals related above the user-defined kinship threshold. Specifically, we added the following sentence Results section lines 240–242) to make this explicit:

“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.”

**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. We now include a dedicated simulation study to evaluate the impact of phenotypic heritability on KDPS performance. 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 necessity of pruning related cases.

As described in the Methods (lines 194–211), 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-de4fined 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 263–272), 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.

**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 thank the reviewer for pointing out this oversight. Detailing the computational 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 211–214):

“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.”

**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. We have expanded the description of the real-data application in the Methods section (lines 182–193) 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.”