**Responses to Reviewer Comments**

We thank the reviewers for their time and thoughtful comments. We have carefully considered and responded to the reviewer’s comments. We have provided detailed responses to the reviewer comments below, with references to various parts of the manuscript that have been modified. We believe these changes have addressed the concerns raised by the reviewers and improved the manuscript overall. Note: all new inserted text in noted via yellow highlighting.

**Reviewers' Comments to Author:**

**Reviewer #1Comments to the Author**

**This study introduced a novel tool, Kinship Decouple and Phenotype Selection (KDPS), designed to address statistical biases arising from genetic relatedness in genomic studies. Notably, KDPS innovated by incorporating phenotype prioritization during the removal of related individuals, aiming to maximize the retention of subjects with target phenotypes (e.g., rare diseases or specific exposures) and thereby enhance statistical power. Overall, this research contributes valuable knowledge to the field, but addressing these concerns will strengthen its impact.**

**1.The manuscript states: “...sequentially eliminating subjects with the lowest phenotypic weight who are related to more than m−f subjects, where m is the number of related pairs of the subject in the cohort who is related to the most people...” (Page 4, Lines 51–55). However, it remains unclear whether m refers to the initial maximum connectivity or is dynamically updated after each iteration. If m is calculated once from the initial network, the algorithm may fail to adapt to evolving network structures during iterative pruning. If m is dynamically updated (e.g., recalculated after each iteration), the frequency of updates and computational overhead must be explicitly described.**

**Reviewer 1 – Response 1:** We thank the reviewer for this important question. We have revised the manuscript to sufficiently articulate the necessary details. As suggested, we have clarified that the parameter \*m\* is recalculated dynamically after each iteration, enabling the algorithm to adaptively respond to changes in the evolving network topology. We have revised the manuscript text (Materials and methods, lines: 141-146) accordingly to improve clarity:

“After each iteration, the algorithm recalculates m to reflect the current maximum number of relationships in the updated network, ensuring that the pruning criterion m − f dynamically adapts to the evolving structure of the cohort. This iterative update continues until no subject exceeds the relatedness threshold and helps maintain optimal pruning sensitivity as the network is refined.”

**2.While the study mentions that the fuzziness score (f) is user-configurable and tested values f = 0,1,2,5,10, it did not provide practical guidelines for selecting f based on network density (e.g., sparse vs. dense kinship networks) or phenotype distribution (e.g., rare vs. common phenotypes).**

**Reviewer 1 – Response 2:** We appreciate the reviewer’s valuable suggestion and have now expanded our explanation of the fuzziness score. We have included practical guidelines for its use. Specifically, we now provide examples of typical use cases informed by the distribution of relatedness observed in population-scale datasets. We added Supplementary Figures. Figure S1 to introduce and provide an overview of fuzziness score concept and Figure S2 illustrates empirical statistics of relatedness from a representative biobank-scale cohort. Moreover, we now include a discussion and guideline for the use fuzziness score in the main text (Methods section, lines 127-132 and Results section, lines 190-196):

“Given a higher fuzziness score, the algorithm downplays the importances of complex network topology, de-prioritize the removal of individuals related to many other subjects, and prioritize the removal of individuals with lower phenotypic weights. With lower fuzziness score, the algorithm adheres more closely to the network topology and prioritize the removal of individuals that are related to more subjects to minimize the number of subjects that need to be removed.”

“In biobank scale cohorts such as the UK Biobank, most kinship relatedness is expected to be pair-wise relationships instead of complex relatedness networks (Supplementary Figure S2). Accordingly, KDPS can resolve most relatedness scenarios with a fuzziness score of 0, which offers a practical default for balancing decoupling and phenotype retention. In cases involving complex relatedness networks or ultra-rare phenotypes, a higher fuzziness score (e.g., 3 - 5) may be warranted to prioritize phenotype over topology.”

**3.The manuscript mentions that KDPS supports composite scores to handle scenarios with “multiple phenotypes and exposures of interest” but provides no validation in simulated or real-world datasets. This omission undermines the reliability of this feature. To ensure robustness, validation of composite scores should be added.**

**Reviewer 1 – Response 3:** We thank the reviewer for highlighting this important point. To address this concern, we have added a simulation evaluating the performance of KDPS using a composite phenotype defined using two independent binary traits. The simulation demonstrates KDPS’s ability to prioritize individuals who meet multiple criteria by leveraging composite weights. These results are detailed in the Results (lines 211-217) and Discussion (lines 255-261) sections and summarized in Supplementary Table S2:

“In more complex scenarios where multiple phenotypes of interest are involved, KDPS also demonstrated the capability to maximize targeted subject retainment based on a composite weight. A simulation involving two independent binary phenotypes (~20% prevalence each) showed that applying composite weights, prioritizing subjects with both traits, resulted in a 42% (19 to 27) increase in the number of retained individuals with both conditions compared to equal-weight pruning (Supplementary Table S2).”

“Moreover, the use of composite weights enables highly flexible prioritization strategies, allowing users to specify phenotype combinations such as categorical values of a particular type (e.g., case status) in addition to a numeric variable within a defined range (e.g., BMI between 18–25). This capacity broadens the applicability of KDPS to complex study designs, enabling tailored subject retention across diverse phenotype-driven analytical objectives.”

**Reviewer #2 Comments to the Author**

**This study (BIB-25-0577) introduces the Kinship Decouple and Phenotype Selection (KDPS) tool, designed to improve subject selection in genetic and epidemiological research by incorporating phenotype prioritization. While the tool demonstrates promising capabilities, several aspects require further clarification and investigation:**

**1. Fuzziness Score Selection: The performance of KDPS is influenced by the fuzziness score, but the manuscript does not provide empirical guidance on how to select an optimal value for this score in practice.**

**Reviewer 2 – Response 1:** We thank the reviewer for their thoughtful observation. We have added detailed explanation of the fuzziness score concept and provided practical guidelines for its application as detailed in response to “**Reviewer 1 – Response 2**” above. In brief, these include addition of Supplementary Figures S1 and S2 and revisions to the manuscript.

**2. Simplistic Simulation Framework:**

**The simulation process lacks detail, particularly regarding the generation of simulated phenotypes. Key parameters—such as the distribution of genetic effects and heritability—are not described, raising concerns about the robustness of the benchmarking results.**

**Reviewer 2 – Response 2:** We thank the reviewer for this important feedback. To address this, we have expanded the description of the simulation framework in both the Methods (lines 154-160) and Results (lines 203-211) sections. We now include greater detail on how the kinship matrix was derived (by anonymizing and adapting the real UK Biobank kinship structure) and how phenotypes were simulated. Furthermore, Table 1 has been revised to detail the parameters used in phenotype simulations. Simulation results evaluating KDPS on binary, categorical, and continuous traits have also been added:

“For simulations, a complex relatedness network (~100,000 relationship pairs) was considered based on the UK Biobank kinship matrix. The matrix was anonymized by removing subject identifiers and randomly permuting individual labels, preserving the underlying topology and relationship structure while ensuring de-identification. Simulated phenotype data included three configurations: a binary trait, a categorical trait with three levels, and a quantitative trait drawn from a mean-centered, normally distributed range. Detailed simulation parameters can be found in Table 1.”

“In the evaluation of the performance of KDPS on multi-class categorical phenotypes, compared to phenotype-naïve pruning, phenotype-aware KDPS increased the retention of disease-relevant individuals by ~79% for disease 1 and ~56% for disease 2, demonstrating substantial gains in preserving prioritized classes (Supplementary Table s1). When KDPS was applied on a continuous phenotype using a simulated normally distributed quantitative trait, compared to phenotype-naïve pruning, phenotype-aware KDPS resulted in modest upward shifts across the distribution of retained subjects. The minimum phenotype value increased by 2%, and the mean by 0.09% (Supplementary Table s1).”

**3. Impact of Phenotype Heritability:**

**The manuscript does not address how the heritability of a phenotype influences KDPS's performance. Since heritability affects the detectability of genetic associations, this is a critical factor to evaluate.**

**Reviewer 2 – Response 3:** We appreciate the reviewer’s thoughtful suggestion. In the real-world application of KDPS, we have specifically included phenotypes with a broad range of heritability estimates: schizophrenia (~80%), multiple sclerosis (~30%), acute myocardial infarction (~40–50%), and alcohol consumption (~20–30%). These choices were intended to demonstrate KDPS's generalizability across varying genetic architectures. We have clarified this design in the manuscript to highlight that KDPS performs consistently across phenotypes with different levels of heritability (Discussion section, lines 249-255 and 272-275):

“In the real-world examples, KDPS was applied to phenotypes with diverse genetic architectures and heritabilities, including schizophrenia (heritability ~80%)(Sullivan, Daly and O’Donovan 2012), multiple sclerosis (~30%)(International Multiple Sclerosis Genetics Consortium 2019), acute myocardial infarction (~40–50%)(Marenberg et al. 1994, Inouye et al. 2018), and alcohol drinking status (~20–30%)(Verhulst, Neale and Kendler 2015, Clarke et al. 2017), all of which showed strong phenotype retention performance.”

“While we considered a diverse set of real and simulated phenotypes, this set is not exhaustive, and we expect the generalizable framework of KDPS should apply broadly to phenotypes with varying population prevalence and genetic contribution.”

**4. Generalizability to Underrepresented Populations:**

**The UK Biobank (UKB) and many genetic studies suffer from significant underrepresentation of non-European populations. It remains unclear whether KDPS performs equally well in diverse ancestry groups, which is essential for ensuring broad applicability.**

**Reviewer 2 – Response 4:** We thank the reviewer for this insightful point. While KDPS is designed to be agnostic to the ancestry of the input data, its performance depends on the accuracy of the underlying kinship or relatedness matrix. We agree that ancestry diversity and admixture may pose challenges for IBD estimation, which in turn may impact KDPS performance. Therefore, we have added a note of caution in the Discussion, advising users to use appropriate admixture-aware methods when generating kinship matrices in ancestrally diverse or admixed populations (lines 275-280):

“Additionally, it is also important to consider population structure in the generation of the relatedness matrix. In ancestrally diverse or admixed populations, standard IBD or kinship estimation methods may be inaccurate or biased due to the confounding effect of genetic admixture (Dou et al. 2017). Users are advised to select appropriate methods that account for ancestry when generating the subject relatedness matrix (Thornton et al. 2012, Conomos et al. 2016).”