Phenotype-aware decoupling of related subjects

Wanjun Gu1,2, Jiachen Xi1, Steven Cao1, Rany M. Salem1

1Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla, CA, 92093.

2University of California, San Francisco, San Francisco, CA 92093.

**Word count**: XYZ

**Tables**: 0, **Figures**: 1

**Corresponding author and to whom reprint requests are to be addressed:**

Rany M. Salem, PhD, MPH.

Herbert Wertheim School of Public Health and Longevity Science,

University of California, San Diego

9500 Gilman Drive #0725, La Jolla CA, 92093

Tel: 858-246-0433

Fax: 858- 534-4642

Email: rsalem@health.ucsd.edu

**Abstract**

Relatedness within genomic cohorts is a potential source of bias for many genetic analyses. Existing tools to break relatedness are phenotype naïve, indiscriminately remove subjects to break relationship, risking the loss of valuable data, especially in studies targeting uncommon and rare phenotypes. To address this limitation, we developed the Kinship Decouple and Phenotype Selection (KDPS) tool, with a novel algorithm designed to enhance the precision of subject selection in genetic and epidemiological research by incorporating phenotype prioritization. KDPS separates related individuals by considering kinship or identity by descent (IBS) scores, while simultaneously prioritizing subjects based on phenotypes of interest. This approach enables the retention of valuable subjects for analysis, even in the face of necessary exclusions due to relatedness. Furthermore, KDPS accommodates a wide range of phenotypes, including binary, ordinal, and quantitative types, and allows for customization to either prioritize specific phenotypes or maximize sample size. In simulations based on the UK Biobank dataset and real-world datasets from dbGaP, KDPS demonstrated significant improvements in retaining subjects with desired phenotypes and computational efficiency compared to previously published software. The ability of this method to process biobank-scale studies within practical timeframes marks a considerable advancement over existing techniques. KDPS offers tailored solutions for complex analytical challenges and broad applicability in genetics and epidemiology research. To our knowledge, KDPS is the first tool to perform phenotype-aware decoupling, paving the way for more precise genetic and epidemiological analyses.

**Introduction**

Geneticists utilize a broad suite of sophisticated methodologies to decode the complex tapestry of genotype-phenotype relationship, including the use of genome-wide association studies (Uffelmann et al. 2021), pleiotropy analyses (Glazner and Thompson 2012, Hu et al. 2014, Feng, Leckman and Zhang 2004), quantitative trait locus (QTL) mapping (Zeng 1994), and transmission disequilibrium tests (TDT) (Spielman, McGinnis and Ewens 1993). These tools are invaluable for unraveling genetic connections that account for family structure and relatedness, allowing for a more precise understanding of genetic inheritance patterns and their influence on diseases and traits. However, the landscape of genetic methodologies is diverse, and despite the advances that accommodate relatedness, a significant number of genetic analyses: selection scans (Akbari et al. 2018, Booker, Jackson and Keightley 2017), admixture mapping (Ali-Khan and Daar 2010), gene by environment interaction analysis (Ottman 1996, Virolainen et al. 2023), and analyses across stratum (Winkler et al. 2017), require studies of unrelated individuals. ~~I~~nclusion of related subjects in these scenarios violates model assumptions of independence and results in inflation of significance of test statistics.

Over the past two decades, population-based genetic studies have become a pivotal methodology for elucidating the relationships between genetic variants and complex human traits. The expansion of dataset sizes introduces complications such as cryptic relatedness (Voight and Pritchard 2005), especially in expansive initiatives like the UK Biobank (Bycroft et al. 2018) or the Trans-Omics for Precision Medicine (TOPMed) Program (Taliun et al. 2021). Cryptic relatedness entails unrecognized genetic connections among participants, presenting considerable obstacles in these large-scale endeavors. Consequently, it is imperative to refine the selection of study participants by assessing their genetic relatedness, albeit at the cost of excluding individuals with relevant phenotypes. This strategy, while reducing the inclusivity of certain phenotypic representations, is crucial to ensure the retention of a robust case sample size. The rationale behind this lies in the need to investigate not only prevalent conditions such as diabetes (Ong et al. 2023) or obesity (Hruby and Hu 2015) but also the gene-environment interactions associated with these traits. The lack of sufficient cases often results in underpowered analyses in gene-environment interaction studies and many other analytic techniques. Furthermore, the imperative to optimize the number of case subjects becomes even more critical when examining the genetics of rarer phenotypes, where every individual case contributes significantly to the understanding of the genetic architecture. For conditions with lower incidence rates, such as colorectal cancer (Rawla, Sunkara and Barsouk 2019), neuroblastoma (Yan et al. 2020), psychiatric disorders like schizophrenia (Charlson et al. 2018) and autism (Talantseva et al. 2023), as well as autoimmune diseases like Lupus (Tian et al. 2023), algorithms designed to maximize case sample size while excluding related individuals can drastically benefit the studies uncovering the genetics of the traits and complex trait correlations.

There are several tools and approaches to manage related subjects in a study, including KING (Manichaikul et al. 2010), PLINK2 (Chang et al. 2015), Friends and family (de Jager et al. 2017), SampleSeq2 (Edwards and Li 2012) and FastIndep (Abraham and Diaz 2014). However, none of which take phenotypes into account. The currently most used decoupling method executed in PLINK2 employs a greedy algorithm that indiscriminately removes individuals most related to others, resorting to random selection among equally related subjects. Naïve non-selective pruning is particularly problematic in scenarios in which the phenotype of interest is rare or uncommon within the cohort. Eliminating subjects with valuable phenotypic traits, which may compromise the sample size and the power analysis. Recognizing this gap, we introduce the Kinship Decouple and Phenotype Selection (KDPS) method. KDPS separates related individuals by considering kinship or identity by descent (IBS) scores and can simultaneously prioritize retention of subjects based on phenotypes of interest. This innovative approach aims to maximize the number of subjects with desired phenotypes, ensuring a robust sample for subsequent analysis even after the necessary exclusion of related individuals.

**Materials and methods**

Loading relatedness and phenotype data

KDPS requires an kinship relatedness (Manichaikul et al. 2010) or identity by descent (IBS) score (Su et al. 2012) file and phenotypic data file. This initial step is crucial for ensuring that only subjects present in both the kinship and phenotype datasets are included in further analysis (Figure 1A). KDPS accepts two categories of phenotypes: categorical, with two or more categories, or numerical. Users are required to specify the primary phenotypes of interest. For categorical phenotypes, an ordered list indicating priority is necessary. Whereas for numerical phenotypes, users must designate whether subjects with higher or lower values are to be prioritized. In scenarios where multiple phenotypes are under consideration, users can generate a combined score, which facilitate prioritization based on a combination of traits (e.g., sex and body height), thereby allowing for nuanced selection within the study sample. Additionally, two users parameters are required: relatedness cutoff values (Kinship or IBS) and a fuzziness score can be set, allowing for a degree of tolerance towards the variability in the number of connections each subject has within the network. A fuzziness score (f) instructs the program to assign subjects who are related with m individuals and subjects who are related to m – f (f < m) individuals with the same kinship weight and prioritize the pruning of their relatedness network based on their corresponding phenotype weights. This score plays a key role in determining which subjects are systematically excluded based on their phenotype prioritizations, e.g., prioritizing subjects with phenotype of interest versus maximizing set of unrelated subjects.

Pruning complex relatedness network

The relatedness pruning process initiates by identifying sets of subjects that are related to each other. First subjects that are not related with any other subjects in the study are removed from the kinship matrix to minimize data size. Next, pairs of subjects that are each related only to each other (related pairs) are identified. These pairs are then segregated according to the predefined phenotype prioritization criteria, significantly streamlining the relatedness matrix (Figure 1A). The algorithm next proceeds stepwise increasing the relatedness group size to tackle more complex relatedness networks, employing a dual strategy based on the fuzziness score. With a fuzziness score of zero, the program adopts a simplified approach by prioritizing the removal of super-subjects. Super-subjects are individuals linked to multiple subjects within the network, whereas these linked subjects are not related with any other subjects in the study. This targeted pruning is particularly effective in cohort studies, where such super-subjects can constitute a considerable portion of the network. Conversely, with a fuzziness score above zero, a greedy algorithm is employed, 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, and f represents the fuzziness score. This iterative removal continues until only related pairs remain, which are then subjected to a final work-up step. The culmination of this process is a curated list of subjects to be excluded, thereby refining the cohort to a set of unrelated individuals.

Benchmarking the method performance using simulations

We evaluated the performance of the KDPS method across simulation scenarios reflective of real-world cohorts. Utilizing the kinship structure derived from the UK Biobank dataset (Bycroft et al. 2018) and a corresponding set of categorical and numerical phenotypes, a complex related network (n ~100,000) was simulated. The performance of KDPS was then assessed under a diverse array of parameters, as detailed in Table 1. Furthermore, we have also applied KDPS on real-world phenotypes in the UK Biobank dataset. We tested removing related subjects regarding four binary phenotypes of interest: schizophrenia, acute myocardial infarction, multiple sclerosis and alcohol drinking (ever consumed alcohol). A fuzziness score of 0 was used and results were compared between the approach where phenotypic information was considered during the subject prioritization process (phenotype-aware) and the approach where phenotypic information was not considered (phenotype-naïve).

**Results**

Simulation test results

The computational time of the KDPS method increases logarithmically in relation to increasing fuzziness score, adhering to an approximate O(log(n)) complexity. Specifically, with a dataset comprising 50,000 subjects with over 10,000 relationships, the execution time of KDPS spans approximately 1.5 minutes at a fuzziness score of zero, increasing to over 10 minutes when the fuzziness score is set to 10 (Figure 1B). Additionally, the processing time required by KDPS exhibits a logarithmic dependency on the aggregate level of relatedness among subjects, as demonstrated in simulations with total sample size of XYZ and increasing number of relatedness, where fuzziness score was set at zero (Figure 1C). Notably, in simulations with 100,000 relationships, analogous to the interconnections within the UK Biobank cohort, KDPS completes its run within approximate 15 minutes while using a single thread and less than 4 Gb of memory. Real world computation times may diverge, influenced by the intricacy of the relatedness network and additional factors such as the fuzziness score.

In addition to demanding more computational resources, the choice of fuzziness score also subtly influences the retention ratio, defined as the percentage of subjects remaining after kinship decoupling compared to the original total. For instance, in a simulation involving 50,000 subjects, an increase in fuzziness score from zero to ten marginally reduces the retention ratio from 0.51 to 0.49 (Figure 1D). Moreover, we compared performance of phenotype selection and retentions using KDPS and standard phenotype-agnostic approaches (e.g., PLINK2), particularly in augmenting the prevalence of subjects possessing the phenotype of interest post-kinship decoupling (Figure 1E). In simulations where the baseline prevalence of the phenotype of interest was set at 0.2, KDPS significantly enhances this retention of subjects with phenotype of interest to approximately 30%, contrary to the phenotype-naïve approach that retained 20% of samples with phenotype of interest post decoupling.

Results for the real-world datasets

To comprehensively evaluate the efficacy of the KDPS within practical applications, we conducted a series of tests with multiple phenotypes retrieved from the UK Biobank cohort study. The outcomes, detailed in Table 2, highlight the capability of KDPS to preserve a significantly higher number of subjects possessing phenotypes of interest when juxtaposed against conventional phenotype-agnostic methodologies. When prioritizing case subjects, KDPS introduced a 11.8% increase in the number of case subjects with schizophrenia, 11.1% increase in subjects with acute myocardial infarction, 12.1% increase in subjects with multiple sclerosis and 8.7% increase in subjects who have self-reported to have never consumed alcohol. In the test using real-world datasets, KDPS also demonstrated remarkable efficiency, successfully completing the decoupling processes for all tested phenotypes in the UK Biobank within 35 minutes.

**Discussion**

In this report, we introduce KDPS, a novel tool and algorithm to address the challenges of phenotype-aware kinship decoupling in genetic and epidemiological research. KDPS substantially improves the utility and precision of existing methods by accounting for phenotype data in subject selection. This approach not only enhances the relevance of selected individuals by aligning them with specific phenotypes of interest but also extends its utility across a spectrum of analytical and practical research scenarios. The utility of KDPS is apparent in scenarios in which analysis method cannot accommodate relatedness and when statistical power by maximizing disease sample count is needed.

In scenarios where multiple phenotypes are under consideration, users can generate a combined score, which facilitate prioritization based on a combination of traits (e.g., sex and body height), offering tailored selection schema for complex design and analytical challenges.

Strengths of KDPS including efficiently of algorithm to process biobank-scale studies within a practical timeframe. KDPS can accommodate a broad set of phenotypes for prioritization, including numeric (binary, ordinal and quantitative measures) and categorical phenotype definitions. Simulation and real-world applied analyses illuminate KDPS's computational efficiency and its capacity to substantially conserve subjects with desired traits, presenting a notable advancement over preceding methodologies. This efficiency is vital for enabling researchers to undertake kinship decoupling tasks.

There are important considerations and limitations of KDPS. One potential challenge arises when dealing with datasets substantially larger than UK Biobank (~500K subjects) and/or extensive and complex relatedness. The increase sample size, amount and complexity of relatedness may result in significantly extended processing times. Since phenotype and unrelated selection are typically only performed once per study, this may not represent a significant barrier to use. Moreover, future improvements with novel algorithms or reimplementing KDPS in a lower-level programming language, such as C++, which could offer enhanced performance efficiencies. Furthermore, KDPS, similar to other subject selection methodologies, is susceptible to introduction of collider bias (Tönnies, Kahl and Kuss 2022). Collider bias occurs when the selection of subjects based on certain criteria inadvertently skews the analysis, potentially leading to misleading associations between study variables. This bias is a critical consideration in genetic and epidemiological studies, where the integrity of findings is paramount. To mitigate the impact of collider bias, researchers are advised to employ strategies such as conducting sensitivity analyses, *e.g.*, performing analyses in randomly selected subjects in parallel with KDPS-selected cohorts.

In conclusion, KDPS is a fast, computationally efficient, and powerful tool for phenotype-aware kinship decoupling, offering substantial improvements in both the inclusion of relevant subjects and computational efficiency. The integration of KDPS paves the way for phenotype-informed selection of unrelated subjects, offering broad applicability in genetics and epidemiology research.

**Data and code availability**

The latest release of KDPS and documentation can be found at <https://github.com/Broccolito/kdps>. The KDPS R package can be installed via GitHub release. Scripts for the benchmarking are available at <https://github.com/Broccolito/kdps_dev>.

**Acknowledgements**

The authors extend their sincere gratitude to Dr. Erin Richard and Mijia Ma from the University of California, San Diego, for their invaluable support and insightful comments regarding both the methodology and the manuscript. This work was supported in part by awards from the Foundation of the National Institute of Health and National Institute of Health grants R00HL122515 and DK135868 to R.M.S.

**Disclosures:** RMS reports a service contract with Travere. The authors have declared that no other conflict of interest exists.

**Supplementary data**

Supplementary data are available at *Bioinformatics* online.

**Conflict of interest**

The authors of the manuscript declare no conflict of interest.

**Funding**

This study is partially funded by NIH R00HL122515 and DK135868.

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Description** | **Values** |
| Number of Relationships | Defined as the number of kinship relationships to consider in the analysis. | 100, 500, 1000, 5000, 10000, 50000, 100000 |
| Fuzziness | The degree of fuzziness allowed in the model. | 0, 1, 2, 5, 10 |
| Phenotypic Naïve | A Boolean parameter representing whether subjects are prioritized based on their phenotypic information. | FALSE, TRUE |
| Phenotype | Tested with three phenotype configurations (pheno1: binary; pheno2: categorical with 3 categories, pheno3: numerical). | pheno1, pheno2, pheno3 |

Table 1. Simulation parameters used to run KDPS.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Phenotype** | **Schizophrenia** | **Acute Myocardial Infarction** | **Multiple**  **Sclerosis** | **Alcohol Drinking (Never)** |
|  | UK Biobank ID | 130874 | 131298 | 131042 | 20117 |
| **Full Sample** | Sample Size | 502420 | 502420 | 502420 | 502420 |
| Case | 1356 | 22635 | 2500 | 22382 |
| Control | 501064 | 479785 | 499920 | 480038 |
| Prevalence | 2.70E-03 | 4.51E-02 | 4.98E-03 | 4.45E-02 |
| **Phenotypic Naïve Selection** | Sample Size | 427803 | 427797 | 427795 | 427793 |
| Case | 1184 | 19662 | 2175 | 19815 |
| Control | 426619 | 408135 | 425620 | 407978 |
| Prevalence | 2.77E-03 | 4.60E-02 | 5.08E-03 | 4.63E-02 |
| **Phenotypic Aware Selection** | Sample Size | 427817 | 427819 | 427817 | 427822 |
| Case | 1324 | 21836 | 2439 | 21547 |
| Control | 426493 | 405983 | 425378 | 406275 |
| Prevalence | 3.09E-03 | 5.10E-02 | 5.70E-03 | 5.04E-02 |
| % More Case Included | 11.8% | 11.1% | 12.1% | 8.7% |

Table 2. Testing results of removing related subjects in UK Biobank regarding four phenotypes of interest.

A diagram of a graph

Description automatically generated with medium confidence

Figure 1. Algorithm flowchart and performance benchmark of KDPS.

**References**

Abraham, Kuruvilla Joseph, and Diaz, Clara, “Identifying Large Sets of Unrelated Individuals and Unrelated Markers,” *Source Code for Biology and Medicine*, 9/1 (2014), 6

Akbari, Ali, Vitti, Joseph J., Iranmehr, Arya, Bakhtiari, Mehrdad, Sabeti, Pardis C., Mirarab, Siavash, et al., “Identifying the Favored Mutation in a Positive Selective Sweep,” *Nature Methods*, 15/4 (2018), 279–82

Ali-Khan, Sarah E., and Daar, Abdallah S., “Admixture Mapping: From Paradigms of Race and Ethnicity to Population History,” *The HUGO Journal*, 4/1–4 (2010), 23–34

Booker, Tom R., Jackson, Benjamin C., and Keightley, Peter D., “Detecting Positive Selection in the Genome,” *BMC Biology*, 15/1 (2017), 98

Bycroft, Clare, Freeman, Colin, Petkova, Desislava, Band, Gavin, Elliott, Lloyd T., Sharp, Kevin, et al., “The UK Biobank Resource with Deep Phenotyping and Genomic Data,” *Nature*, 562/7726 (2018), 203–9

Chang, Christopher C., Chow, Carson C., Tellier, Laurent Cam, Vattikuti, Shashaank, Purcell, Shaun M., and Lee, James J., “Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets,” *GigaScience*, 4 (2015), 7

Charlson, Fiona J., Ferrari, Alize J., Santomauro, Damian F., Diminic, Sandra, Stockings, Emily, Scott, James G., et al., “Global Epidemiology and Burden of Schizophrenia: Findings From the Global Burden of Disease Study 2016,” *Schizophrenia Bulletin*, 44/6 (2018), 1195–1203

Edwards, Todd L., and Li, Chun, “Optimized Selection of Unrelated Subjects for Whole-Genome Sequencing Studies of Rare High-Penetrance Alleles,” *Genetic Epidemiology*, 36/5 (2012), 472–79

Feng, Rui, Leckman, James F., and Zhang, Heping, “Linkage Analysis of Ordinal Traits for Pedigree Data,” *Proceedings of the National Academy of Sciences of the United States of America*, 101/48 (2004), 16739–44

Glazner, Chris, and Thompson, Elizabeth Alison, “Improving Pedigree-Based Linkage Analysis by Estimating Coancestry among Families,” *Statistical Applications in Genetics and Molecular Biology*, 11/2 (2012) <http://dx.doi.org/10.2202/1544-6115.1718>

Hruby, Adela, and Hu, Frank B., “The Epidemiology of Obesity: A Big Picture,” *PharmacoEconomics*, 33/7 (2015), 673–89

Hu, Hao, Roach, Jared C., Coon, Hilary, Guthery, Stephen L., Voelkerding, Karl V., Margraf, Rebecca L., et al., “A Unified Test of Linkage Analysis and Rare-Variant Association for Analysis of Pedigree Sequence Data,” *Nature Biotechnology*, 32/7 (2014), 663–69

de Jager, Deon, Swarts, Petrus, Harper, Cindy, and Bloomer, Paulette, “Friends and Family: A Software Program for Identification of Unrelated Individuals from Molecular Marker Data,” *Molecular Ecology Resources*, 17/6 (2017) <https://pubmed.ncbi.nlm.nih.gov/28503747/> [accessed 28 February 2024]

Manichaikul, Ani, Mychaleckyj, Josyf C., Rich, Stephen S., Daly, Kathy, Sale, Michèle, and Chen, Wei-Min, “Robust Relationship Inference in Genome-Wide Association Studies,” *Bioinformatics* , 26/22 (2010), 2867–73

Ong, Kanyin Liane, Stafford, Lauryn K., McLaughlin, Susan A., Boyko, Edward J., Vollset, Stein Emil, Smith, Amanda E., et al., “Global, Regional, and National Burden of Diabetes from 1990 to 2021, with Projections of Prevalence to 2050: A Systematic Analysis for the Global Burden of Disease Study 2021,” *The Lancet*, 402/10397 (2023), 203–34

Ottman, R., “Gene-Environment Interaction: Definitions and Study Designs,” *Preventive Medicine*, 25/6 (1996), 764–70

Rawla, Prashanth, Sunkara, Tagore, and Barsouk, Adam, “Epidemiology of Colorectal Cancer: Incidence, Mortality, Survival, and Risk Factors,” *Przeglad Gastroenterologiczny*, 14/2 (2019), 89–103

Spielman, R. S., McGinnis, R. E., and Ewens, W. J., “Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-Dependent Diabetes Mellitus (IDDM),” *The American Journal of Human Genetics*, 52/3 (1993), 506–16

Su, Shu-Yi, Kasberger, Jay, Baranzini, Sergio, Byerley, William, Liao, Wilson, Oksenberg, Jorge, et al., “Detection of Identity by Descent Using Next-Generation Whole Genome Sequencing Data,” *BMC Bioinformatics*, 13 (2012), 121

Talantseva, Oksana I., Romanova, Raisa S., Shurdova, Ekaterina M., Dolgorukova, Tatiana A., Sologub, Polina S., Titova, Olga S., et al., “The Global Prevalence of Autism Spectrum Disorder: A Three-Level Meta-Analysis,” *Frontiers in Psychiatry / Frontiers Research Foundation*, 14 (2023), 1071181

Taliun, Daniel, Harris, Daniel N., Kessler, Michael D., Carlson, Jedidiah, Szpiech, Zachary A., Torres, Raul, et al., “Sequencing of 53,831 Diverse Genomes from the NHLBI TOPMed Program,” *Nature*, 590/7845 (2021), 290–99

Tian, Jingru, Zhang, Dingyao, Yao, Xu, Huang, Yaqing, and Lu, Qianjin, “Global Epidemiology of Systemic Lupus Erythematosus: A Comprehensive Systematic Analysis and Modelling Study,” *Annals of the Rheumatic Diseases*, 82/3 (2023), 351–56

Tönnies, Thaddäus, Kahl, Sabine, and Kuss, Oliver, “Collider Bias in Observational Studies,” *Deutsches Arzteblatt International*, 119/7 (2022), 107–22

Uffelmann, Emil, Huang, Qin Qin, Munung, Nchangwi Syntia, de Vries, Jantina, Okada, Yukinori, Martin, Alicia R., et al., “Genome-Wide Association Studies,” *Nature Reviews Methods Primers*, 1/1 (2021), 1–21

Virolainen, Samuel J., VonHandorf, Andrew, Viel, Kenyatta C. M. F., Weirauch, Matthew T., and Kottyan, Leah C., “Gene-Environment Interactions and Their Impact on Human Health,” *Genes and Immunity*, 24/1 (2023), 1–11

Voight, Benjamin F., and Pritchard, Jonathan K., “Confounding from Cryptic Relatedness in Case-Control Association Studies,” *PLoS Genetics*, 1/3 (2005), e32

Winkler, Thomas W., Justice, Anne E., Cupples, L. Adrienne, Kronenberg, Florian, Kutalik, Zoltán, Heid, Iris M., et al., “Approaches to Detect Genetic Effects That Differ between Two Strata in Genome-Wide Meta-Analyses: Recommendations Based on a Systematic Evaluation,” *PloS One*, 12/7 (2017), e0181038

Yan, Ping, Qi, Feng, Bian, Lanzheng, Xu, Yajuan, Zhou, Jing, Hu, Jiajie, et al., “Comparison of Incidence and Outcomes of Neuroblastoma in Children, Adolescents, and Adults in the United States: A Surveillance, Epidemiology, and End Results (SEER) Program Population Study,” *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 26 (2020), e927218

Zeng, Z. B., “Precision Mapping of Quantitative Trait Loci,” *Genetics*, 136/4 (1994), 1457–68