Genetic Epidemiology

OFFICIAL JOURNAL
INTERNATIONAL GENETIC
EPIDEMIOLOGY SOCIETY
www.geneticepi.org

WILEY

Benchmarking statistical methods for analyzing parent–child dyads in genetic association studies

Debashree Ray[1,2] | Candelaria Vergara[1] | Margaret A. Taub[2] |
Genevieve Wojcik[1] | Christine Ladd-Acosta[1] | Terri H. Beaty[1] | Priya Duggal[1]

[1]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

[2]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

Correspondence
Debashree Ray, Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA.
Email: dray@jhu.edu

Abstract

Genetic association studies of child health outcomes often employ family-based study designs. One of the most popular family-based designs is the case–parent trio design that considers the smallest possible nuclear family consisting of two parents and their affected child. This trio design is particularly advantageous for studying relatively rare disorders because it is less prone to type 1 error inflation due to population stratification compared to population-based study designs (e.g., case–control studies). However, obtaining genetic data from both parents is difficult, from a practical perspective, and many large studies predominantly measure genetic variants in mother–child dyads. While some statistical methods for analyzing parent-child dyad data (most commonly involving mother–child pairs) exist, it is not clear if they provide the same advantage as trio methods in protecting against population stratification, or if a specific dyad design (e.g., case–mother dyads vs. case–mother/control–mother dyads) is more advantageous. In this article, we review existing statistical methods for analyzing genome-wide marker data on dyads and perform extensive simulation experiments to benchmark their type I errors and statistical power under different scenarios. We extend our evaluation to existing methods for analyzing a combination of case–parent trios and dyads together. We apply these methods on genotyped and imputed data from multiethnic mother–child pairs only, case–parent trios only or combinations of both dyads and trios from the Gene, Environment Association Studies consortium (GENEVA), where each family was ascertained through a child affected by nonsyndromic cleft lip with or without cleft palate. Results from the GENEVA study corroborate the findings from our simulation experiments. Finally, we provide recommendations for using statistical genetic association methods for dyads.

KEYWORDS
dyads, family-based GWAS, hybrid design, log-linear models, mother–child pairs, parent–offspring design, transmission disequilibrium, trios

---

Terri H. Beaty and Priya Duggal contributed equally to this study.

---

1 | INTRODUCTION

Studies of genetic associations of most human traits and diseases focus on population-based designs (e.g., case–control studies, cohort studies, or data from biobanks) especially for complex and heterogeneous disorders, where both genetic and environmental risk factors are likely involved. For rare disorders, this often requires amassing cases sampled from multiple populations, which creates the possibility of type I error inflation due to confounding (as termed by epidemiologists) or population stratification (as referred to by geneticists). Family-based designs, on the other hand, play an important role in the investigation of genetic underpinnings of low frequency or rare disorders (e.g., birth defects) (Benyamin et al., 2009). The case–parent trio design is one of the most popular family-based designs which consists of an affected child (i.e., the proband) and both parents. Statistical methods focused on transmission of variants within families, such as in case–parent trio design, protect against population stratification (Schwender et al., 2012). However, genetic information on both parents is often not available because it is more difficult to recruit biological fathers, which often leads to an abundance of information on mother–child dyads (Shi et al., 2008). As more multiethnic studies become possible, the value of family-based methods that are less prone to population stratification are key for studies of childhood diseases. It is essential to define the best suited analytical approach for each alternative family-based scenario, especially in the presence of a predominance of dyads in available samples.

Although statistical methodologies for population-based genome-wide association studies (GWAS) continue to evolve, less ongoing attention has been devoted to methods for family-based studies. While there exist methods for analyzing dyads, it is not clear if one method is consistently more advantageous over another for a given design, and if they have the same robustness against population stratification as full case–parent trios. Furthermore, methods for analyzing dyads were mostly developed 20 or more years ago when type I error performance and statistical power were evaluated at nominal significance levels instead of the more stringent genome-wide levels used now. Recently, Hecker et al. (2019) compared type I error and power of methods for general pedigree designs under different founder genotype distribution schemes but only considered the case–parent trio design among nuclear family designs with no hybrid designs. Gjerdevik et al. (2020) compared the relative efficiency of different hybrid designs with the case–control and the case–parent trio designs using direct and indirect effects under log-linear models implemented in the HAPLIN program. However, Gjerdevik et al. (2020) focused on log-linear models

only, power/sample size comparisons without any type I error calibration or compute time comparison, and were restricted to a homogenous racial/ethnic group.

In this paper, we benchmark multiple open-source popular statistical methods across both dyad and trio designs on multiethnic samples in terms of their compute times, type I error control, and statistical power for identifying common variant associations at stringent significance levels. We focus on identifying the direct effects of common variants in offspring on their phenotype and not on the indirect effects of parental genotypes. Using large-scale simulations, we compare and evaluate methods for three nuclear family designs (case–parent trios, case–mother dyads, and a combination of case–parent trios and case–mother dyads), plus two related hybrid designs (case–parent/control–parent trios, case–mother/control–mother dyads) under different parameter settings. Then, we apply these methods to genotyped and imputed data from multiethnic mother–child pairs only or trios only or combinations of both dyads and trios from the Gene, Environment Association Studies consortium (GENEVA).

2 | MATERIAL AND METHODS

2.1 | Model and notation

In the following, we consider a GWAS of individuals collected under different nuclear-family study designs, particularly case–parent trios and case–parent dyads (typically mother–child pairs). All individuals are genotyped/imputed or sequenced genome-wide at p genetic variants (where p is of the order of millions), and their disease status of interest is measured. For simplicity, we only consider data on biallelic single nucleotide polymorphisms (SNPs). We are interested in testing for the association between an SNP and the disease status. In the following sections, we provide a brief overview of existing methods for analyzing different nuclear family and related hybrid designs.

2.2 | Methods for case-parent trios

Consider a study of n case–parent trios, where n affected (case) offspring are ascertained from a population, and the affected offspring along with their parents are genotyped/imputed or sequenced. Whether information on parents' disease status is needed depends on the statistical method used. This is the simplest family-based design and existing approaches for analyzing such data include variations of the transmission disequilibrium test and mating-type-stratified regression approaches.

2.2.1 | Transmission disequilibrium test (TDT)

The TDT was originally proposed to assess if one allele at an SNP is transmitted from heterozygous parents to their affected offspring more often than that expected under strict Mendelian transmission (Spielman et al., 1993). This would indicate that the SNP being tested (or marker locus) is both linked and associated with a causal SNP (or disease-susceptibility locus [DSL]) for the disease. Essentially, the TDT is a nonparametric test that does not require assumptions about the disease model or disease distribution in the population (Laird & Lange, 2006). Note, although the TDT does not require prior specification of a disease model, it implicitly assumes multiplicative effects of alleles (Fallin et al., 2002). It does not use parental phenotype information even when available. It is also referred to as the allelic TDT, and it boils down to McNemar's test for a 2×2 contingency table that cannot accommodate covariates or provide estimates of relative risks (RRs). For a large enough sample size, the TDT statistic has a χ^2 distribution with 1 degree of freedom (df) under the null hypothesis of no association *or* no linkage between the marker SNP and an unobserved DSL.

Although the TDT was originally proposed to test for linkage in the presence of association, it is now typically used as a test for association (Laird & Lange, 2006) where the goal is to detect a disease-associated SNP by considering alleles at several markers in a region sequentially rather than a single marker SNP in linkage disequilibrium (LD) with an unobserved DSL underlying the disease phenotype (Fallin et al., 2002). Presence of linkage but no association between the marker SNP and a DSL results in no association between the disease and the marker SNP (Laird & Lange, 2006). The presence of both linkage and association between the marker SNP and a DSL indicates the presence of an association between the disease and the marker SNP being tested. Henceforth, in general, any reference to the null hypothesis will refer to no association between the disease and the marker SNP. In particular, for the TDT-like methods, this null hypothesis will mean no association between the marker SNP and a DSL in the presence of linkage between them.

2.2.2 | Genotypic TDT (gTDT)

The gTDT compares the case's transmitted genotype at the SNP to the set of all possible genotypes the case could have inherited from the parental genotypes (Schaid, 1999; Self et al., 1991). Unlike the TDT, the gTDT considers individuals as units of analysis; can accommodate

family-level covariates; assumes some prespecified genetic inheritance model; and yields estimated RRs with their standard errors (Fallin et al., 2002). The gTDT uses a conditional logistic regression model and involves a numeric-likelihood maximization that can be computationally burdensome at a genome-wide level. However, recently derived closed-form parameter estimates enable rapid genome-wide application of the gTDT when testing additive, dominant or recessive effects using Wald's test (Schwender et al., 2012). The gTDT statistic has an asymptotic 1-df χ^2 distribution under the null hypothesis. Like TDT, the gTDT too does not use parental phenotype information.

2.2.3 | Generalized disequilibrium test (GDT)

This is a generalization of TDT-like family-based association method proposed to take advantage of larger pedigree information (Chen et al., 2009). It assesses the genotypic differences between all discordant relative pairs; can model covariates and uses a score test, where the score is obtained from the quasi-likelihood function for a conditional logistic regression model. Unlike the TDT or the g TDT, the GDT uses parental phenotype information. Under the null hypothesis, the GDT statistic has an asymptotic $N(0,1)$ distribution that does not depend on the inheritance model for the DSL. For case–parent trios, a special case of the GDT called the GDT-PO is used where the score is a weighted sum of genotypic differences between phenotypically discordant parent–child pairs.

2.2.4 | Mating-type-stratified conditional likelihood approach

Similar to the genotype RR modeling of Schaid and Sommer (1993) and the retrospective likelihood-based approach of UNPHASED (Dudbridge, 2008), Fan et al. (2013) derived the conditional-likelihood of parental mating type and offspring genotype at an SNP given the affected status of offspring under a specific inheritance model assuming Hardy–Weinberg equilibrium (HWE) and random mating in the parental generation. This approach does not use parental phenotype information, can handle missing parental data without imputation, and can be applied to data on trios, dyads, and monads (henceforth, it is referred to as the 'TDM'). The likelihood function is modeled using the two unknown genotypic relative risk (GRR) parameters and the minor allele frequency (MAF) at the marker SNP, obtained

using the Newton–Raphson method. Additive, dominant, recessive or multiplicative effects may be assumed, and the resultant likelihood ratio test (LRT) statistic for the TDM has an approximate 1-df χ^2 distribution under the null hypothesis. One need not assume any inheritance model (i.e., assume an unrestricted model) and the resultant TDM statistic has an approximate 2-df χ^2 distribution under the null hypothesis.

2.2.5 | Log-linear modeling approach

The log-linear approach generalizes the TDT to include orthogonal tests of offspring versus maternal genetic factors, and can accommodate different risk conferred by a single copy versus two copies of a risk allele (Weinberg et al., 1998). This approach does not use parental phenotype information, provides RR estimates for offspring genotype and is generalizable to a wide range of causal scenarios. It lists all possible trio genotypes stratified by parental mating type and applies a Poisson regression to the expected counts of the different trio genotypes conditional on the affected status of the offspring. Inferences about the association are carried out using asymptotically χ^2-distributed LRT statistic. The χ^2df under the null is dictated by the number of GRR parameters, which in turn depends on the inheritance model and the causal scenario assumed. van den Oord and Vermunt (2000) described how this log-linear approach can be implemented in LEM, a general computer program for analyzing categorical data. Later, a dedicated genetics software package, HAPLIN, was developed to implement such log-linear models not only for biallelic variants but also for multiallelic variants and other generalizations (Gjerdevik et al., 2019, Gjessing & Lie, 2006). Using LEM or HAPLIN, one can test for offspring effects only using a 2-df test, maternal effects only in a separate 2-df test or both offspring and maternal effects in a 4-df test. P values from the 2-df test of offspring effects are directly comparable to p values from other TDT-like methods. While these 2-df and 4-df tests assume no particular inheritance model, one can also implement a 1-df test in HAPLIN assuming multiplicative effects of allele like the TDT.

2.3 | Methods for case–mother dyads

In any study of nuclear families, genetic measurements are frequently missing for one parent. More often than not, fathers are missing: they can be harder to recruit, and paternity is inherently harder to be confident of than maternity (Shi et al., 2008). Consider a study of n case–mother dyads, where n affected (case) offspring are sampled from the population, and the affected offspring and their mothers are genotyped/imputed or sequenced. One might come up with a straightforward approach of applying the TDT (or any method for case–parent trio data) on such family pairs for whom the genotype of the father at the SNP of interest can be unambiguously inferred. However, this process of selectively including only unambiguous dyads and discarding ambiguous ones can lead to invalid inference due to biases that depend heavily on allele frequencies (Curtis & Sham, 1995).

2.3.1 | TDT-like approaches

The first appropriate methodological development for the analysis of nuclear family data with missing genetic information on one parent was the 1-TDT (Sun et al., 1999). It examines the difference between the GRRs of heterozygotes versus homozygotes (two possible choices) by using all heterozygous parent–homozygous offspring and homozygous parent–heterozygous offspring pairs (Sun et al., 1998, 1999). Under the null hypothesis, these two GRRs are expected to be the same and the 1-TDT test statistic has an approximate $N(0,1)$ distribution. If the total number of afore-mentioned parent–child pairs is not large, an exact p value can also be calculated under the binomial distribution. Like the TDT, the 1-TDT does not use parental phenotype information. The GDT-PO can also be directly applied to data with one missing parent. Unlike the 1-TDT, GDT-PO examines all heterozygous parent–homozygous offspring and homozygous parent–heterozygous offspring among available parents that are unaffected (recall, GDT only uses phenotypically discordant pairs). For a data set where all offspring are affected and all their parents are unaffected, 1-TDT and GDT-PO become identical. The GDT-PO statistic has an asymptotic $N(0,1)$ distribution under the null hypothesis.

2.3.2 | Mating-type-stratified likelihood approaches

The TDM is another approach that may be applied to data on parent–child dyads alone and gives a 1-df or a 2-df χ^2 test statistic under the null depending on whether a specific inheritance model is assumed or not. The log-linear approach, too, is flexible enough to handle missing genetic data on parents via the Expectation–Maximization (EM) algorithm (Weinberg, 1999), and can be implemented using programs such as LEM or HAPLIN.

2.4 | Methods for case–mother dyads and case–parent trios combined

When conducting family studies, it is not always possible to collect only families of one structure. In practice, we may not have only case–parent trios or only case–mother dyads but combinations of both. Suppose we have n_1 complete case–parent trios and n_2 incomplete trios where, without loss of generality, the fathers are missing. Note, it does not matter here if the fathers or the mothers are missing since we are only assessing the direct effects of inherited genotypes of offspring on their disease status.

2.4.1 | TDT-like approaches

Instead of ignoring one set of families depending on whether the sample size n_1 is larger than n_2 or not, Sun et al. (1999) proposed applying the TDT on all complete trios and the 1-TDT on all dyads and then combining the two statistics. The resultant combined statistic, denoted TDT_{com}, has an asymptotic $N(0,1)$ distribution. However, there is currently no software for TDT_{com}. One can instead implement the 1-TDT or the GDT-PO on all parent–child pairs without discarding any families.

2.4.2 | Mating-type-stratified likelihood approaches

As described before, the TDM and the log-linear models can also be used in this scenario.

2.5 | Methods for case–parent/control–parent trios

Genetic association studies, whether family-based (e.g., case–parent trio design) or population-based (e.g., case–control design), have their own strengths and limitations; see Weinberg and Umbach (2005) for a comprehensive summary of their advantages and disadvantages. A hybrid design bringing the strengths of case–parent trio and case–control designs into a single analytic framework is the case–parent/control–parent trio design. Consider a study of n trios, where n_1 affected (case) and $n_2 = n - n_1$ unrelated unaffected (control) offspring are sampled from the population, and all sampled offspring and their parents are genotyped/imputed or sequenced. Although control–parent trios are generally easier to recruit and, along with case–parent trios, can help guard against spurious signals due to segregation distortion, they are typically either not recruited or discarded from analysis even when available because the TDT or the gTDT is only applicable to trios with affected offspring (Deng & Chen, 2001).

2.5.1 | TDT-like approaches

A straightforward approach is to apply the TDT on case–parent trios alone (refer to this as TDT_D) and on control–parent trios separately (refer to this as TDT_C, which, in contrast to TDT_D, can be viewed as a test of transmission of the "nonrisk" allele at the DSL rather than the "risk" allele), and then combine these two independent tests into a new test TDT_{D+C} (Deng & Chen, 2001). This TDT_{D+C} statistic has an asymptotic 2-df χ^2 distribution under the null hypothesis. Deng and Chen (2001) additionally proposed TDT_{DC}, a contingency table association test of allele transmissions (from heterozygous parents) with disease status in unrelated offspring. This TDT_{DC} statistic has an approximate 1-df χ^2 distribution under the null hypothesis and does not require equal numbers of case–parent and control–parent trios. Neither of these two TDT-like tests uses parental phenotype information.

2.5.2 | Mating-type-stratified likelihood approaches

A log-linear model can be used to combine the family-based case–parents-trio component and the population-based parent–parent component, thus not requiring genetic data on the control offspring (Weinberg & Umbach, 2005). The requirement of only the parental genotypes and the case genotypes provides a distinct advantage of log-linear models over TDT-like approaches for such hybrid designs. It assumes the disease is rare for the offspring of each parental genotype combination, mating symmetry, and Mendelian proportions in the population. It neither assumes HWE nor random mating. In the presence of population structure, one can generalize this log-linear model to include a disease-status by total-number-of-parental–alleles interaction term or five additional disease-status by mating-type interaction terms (note, six distinct unordered parental mating types are possible here). However, "[a] direct consequence of preferring the enlarged model is that the control portion of the data will not contribute to inference related to the risk parameters, and the population-based component, in effect, becomes statistically irrelevant" (Weinberg & Umbach, 2005).

2.6 | Methods for case–mother/control–mother dyads

Consider a study of n dyads, where n_1 affected (case) offspring are ascertained, and $n_2 = n - n_1$ unaffected (control) offspring are sampled. Genotype data are available on offspring and their mothers. There are fewer methods for this hybrid dyad design. To our knowledge, no TDT-like method has been proposed for this design but the log-linear approach using LEM or HAPLIN is applicable (Shi et al., 2008).

2.7 | Simulation experiments

We first simulated $n = 1000$ trios using the LE program (Chen & Deng, 2001, Chen et al., 2009) as follows, and for the relevant scenarios detailed below, we removed fathers from trios to obtain mother–child pairs (dyads). The LE program is a general program for simulating pedigrees with only 1 DSL (causal SNP) at a time and requires the following parameters as input: disease prevalence, disease allele frequency, genotypic penetrances, number of families, and structure of the pedigree. For simulating phenotypes, we set the parental phenotypes as 0 (control) always, and offspring as 1 (case) or 0 depending on whether we simulated a case– or a control–family. For genotypes, we simulated multiple independent replicates of only one biallelic causal SNP to ensure the independence of SNPs needed to estimate type I error rate and statistical power. The main results are described for a fixed MAF at this causal SNP to ensure a fair comparison of type I error and power across methods keeping all other parameters fixed (additional results for varying MAFs are included in the Supporting Information). We simulated only offspring GRR effects and assume different GRR values to generate both null and nonnull SNPs. For our type I error and power analyses, we simulated 1 million null SNPs (GRR = 1) and 10,000 nonnull SNPs (GRR=2), respectively. In other words, we generated either 1 million or 10,000 independent and identically distributed replicates of 1,000 families for a single SNP. For nonnull SNPs, we assume GRR = 2 under four different inheritance models (additive, multiplicative, dominant, and recessive). Note, for GRR = 1, the inheritance model does not influence the values of genetic penetrance, and hence data generated under different inheritance models are all identical. Our choices of other parameters, such as disease prevalence, MAF and subgroup-specific sample sizes, are described later for specific scenarios.

We evaluated two classes of methods: TDT-like methods (TDT, gTDT, 1-TDT, GDT-PO, TDT_{DC}, TDT_{D+C}) and the mating-type-stratified likelihood approaches (HAPLIN). While the TDM and the log-linear modeling using LEM are also candidates for mating-type-stratified likelihood approaches, we excluded both. In many analyses, TDM faced issues with matrix inversion preventing successful model fit, leading to missing results for many SNPs. We also faced several roadblocks in implementing the current LEM executable genome-wide on a Unix cluster. Note, not all methods in each class are applicable for a given study design. We simulated two scenarios involving samples from either one or two homogenous ancestral populations mimicking situations without or with population stratification. In all our analyses, we specified an additive model for implementing gTDT, regardless of the genetic inheritance model we used to generate data. Implementation of TDT, 1-TDT, GDT-PO, TDT_{DC}, and TDT_{D+C} do not require prespecification of a genetic model since they implicitly assume multiplicative effects of alleles. For HAPLIN, the use of 2-df and 4-df tests implies no specific inheritance model is assumed. While we benchmarked type I error of methods for all simulation settings, only statistical power for the homogenous group was used for benchmarking since not all methods could maintain type I error in the presence of population stratification. We used QQ plots as well as type I error estimates with 95% confidence interval (CI) to evaluate type I error control at stringent levels, and compared power estimates calculated at the conventional genome-wide level (5×10^{-8}). The type I error and the power estimates at a fixed significance level (α) are calculated as the proportion of SNPs with p values $< \alpha$ from the null and the nonnull data, respectively. An approximate asymptotic 95% CI for such an estimate ($\hat{\alpha}$) is calculated as $[\hat{\alpha} - 1.96\sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{N}}, \hat{\alpha} + 1.96\sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{N}}]$, where N is the total number of independent SNPs from which $\hat{\alpha}$ is estimated.

2.7.1 | Scenario 1: One homogenous genetic ancestry

All the parents were simulated from a single homogenous ancestral population with disease prevalence of 30%. While family-based designs are best suited for disorders with rare prevalence, we simulated data for a common disease prevalence to ease simulation time and computational resources needed for sampling millions of case families across several combinations of parameter values. A fixed MAF of 10% is assumed for the causal

SNP. Under this scenario, both type I error and power are compared for all methods.

Scenario 1A: Case–mother dyads. The fathers from 1,000 case–parent trios are removed to obtain 1,000 case–mother dyads. Note, although we removed the fathers for this parent–child design, the mothers could have been removed instead and the inference on direct effects of offspring genotype from these methods would remain unchanged.

Scenario 1B: Case–mother dyads and case–parent trios combined. We removed fathers from the first 750 case–parent trios to obtain 750 case–mother dyads (75% of the data set), leaving the remaining 250 case–parent trios as is (25% of the data set).

Scenario 1C: Case–mother/control–mother dyads. We generated 500 case–parent trios and 500 control–parent trios independently. Removing the father from each trio resulted in 500 case–mother dyads and 500 control–mother dyads, which were analyzed together. For power calculations, besides this 50:50 case–control ratio among offspring, we also explored 70:30 and 30:70 ratios of case–control families.

2.7.2 | Scenario 2: Two distinct genetic ancestry groups

This scenario considers the existence of population substructure between families in the sample. We simulated the parents of 500 families from one homogenous ancestral population with a disease prevalence of 30%, and the parents of the other 500 families from a separate ancestral population with a lower disease prevalence of 15%. The causal SNP was simulated to have an MAF of 10% and 3%, respectively in the two populations. We analyzed a pooled sample for all methods. Additionally, for the mating-type-stratified likelihood approach, we applied HAPLIN to each ancestry group separately and then meta-analyzed using Fisher's p value combination method (Fisher, 1925, Ray et al., 2016). Under this scenario, we compared type I error rates only.

Scenario 2A: Case–mother dyads. Fathers from all 1,000 case–parent trios (500 from each ancestral population) were removed to obtain 1,000 case–mother dyads.

Scenario 2B: Case–mother dyads and case–parent trios combined. From each ancestral group, we removed fathers of the first 75% of the case–parent trios to obtain 750 case–mother dyads in total and combined this with the remaining 250 case–parent trios. This resulted in 375 case–parent trios and 125 case–mother dyads from each ancestral population.

Scenario 2C: Case–mother/control–mother dyads. We generated 500 case–parent trios and 500 control–parent trios independently. Among case–parent trios, 250 were simulated for each ancestral population. Similarly for the control–parent trios. Removing the father from each trio gave us 500 case–mother and 500 control–mother dyads in a genetically heterogeneous sample.

2.8 | Application to GENEVA data on orofacial clefts

In GENEVA, case–parent trios were ascertained through cases with an isolated, nonsyndromic orofacial cleft (i.e., cleft lip; cleft palate; or cleft lip with palate). They were largely recruited through surgical treatment centers by multiple investigators from Europe (Norway and Denmark), the United States (Iowa, Maryland, Pennsylvania, and Utah) and Asia (People's Republic of China, Taiwan, South Korea, Singapore, and the Philippines) over several years (Beaty et al., 2010). Type of cleft, sex, race, family history, and common environmental risk factors were collected through direct maternal interview. Genotyping on the Illumina Human610 Quadv1_B array with 589,945 SNPs was performed at the Center for Inherited Disease Research (https://cidr.jhmi.edu/). As part of two recent publications (Ray et al., 2021, Zhang et al., 2021), trio-aware phasing and reimputation using the 1000 Genomes Phase 3 release 5 reference panel were performed. Among GENEVA participants who were reimputed and used in these two articles, we restricted our analysis to the participants ascertained through nonsyndromic CL/P. We used "hard" genotype calls: if the calls had uncertainty >0.1 (i.e., genotype-likelihoods <0.9), they were treated as missing; the rest were regarded as observed genotype calls. All imputed SNPs were filtered to exclude any with $R^2 < 0.3$. All variants are on the forward strand.

We analyzed genotyped/imputed SNPs only and employed the following quality control measures using PLINK 1.9 (Chang et al., 2015): all SNPs with MAF < 5% and any showing deviation from HWE at $p < 10^{-6}$ among parents were excluded; all genotyped SNPs with missingness >5% and Mendelian error rate >5% were also removed. Additionally, all trios with per-trio Mendelian error rate >5% were dropped. Our final GENEVA analytical data set contained 5,204,784 autosomal SNPs, including both observed and imputed SNPs having MAF > 5% among parents, for 1,487 multiethnic complete case–parent trios. Of these 1,487 complete trios, 891 trios were of Asian ancestry (including Malays from Singapore) and 575 were of European ancestry. The remaining 22 trios were from other racial/ethnic groups. Among 2,974 parents, 560 had missing phenotype information. There were 534 female and 953 male CL/P probands in total.

We analyzed three separate designs: case–parent trios only (considered here as the gold standard); case–mother dyads only; and trios and dyads combined. For the trios only data set, we analyzed all 1,487 case–parent trios (a multiethnic sample). For the dyads only data set, we removed the father from each trio and analyzed the resulting 1,487 case–mother pairs. For the combined data set, we removed the father from 75% of all trios within each racial/ethnic group (consistent with the simulations for scenario 2B detailed above), yielding 1,116 multiethnic case–mother pairs and 371 multiethnic case–parent trios. We compared findings from the dyads only and the combined designs against the trios only design. Note, there are no control families in GENEVA, and hence no case–parent/control–parent or case–mother/control–mother design could be considered. We analyzed the complete GENEVA data set using TDT-like methods. We applied HAPLIN to the Asian and the European groups separately and then meta-analyzed results from the 2-df test for offspring genotype effects using Fisher's p value combination. We excluded the 4-df combined test of offspring or maternal effects in this comparison since we cannot rule out the influence of maternal genes on risk of clefts in offspring (Jugessur et al., 2010, Shi et al., 2012).

For each analysis, we defined independent loci by clumping all the genome-wide significant SNPs ($p < 5 \times 10^{-8}$) in a ± 500 kb span and with LD $r^2 > 0.2$ into a single genetic locus. We used the SNP2GENE function of FUMA (v1.3.6b; Watanabe et al., 2017) for clumping and mapping each locus to the gene nearest to the lead SNP. The index SNP for each locus was the most significant SNP. Since we performed multiethnic analysis, we separately used 1000G Phase 3 EUR and EAS as reference populations for LD calculation. For a given analysis, we found independent hits to be the same regardless of the ancestry of the reference group chosen for LD calculation. We defined the bounds of a locus as the minimum of lower bounds and the maximum of upper bounds across both ancestry groups. All genomic coordinates are given in NCBI Build 37/UCSC hg19.

3 | RESULTS

3.1 | Simulation experiments: Type I error

3.1.1 | Scenario 1A: One homogenous ancestry group, case–mother dyads

The QQ plots of all methods, except 1-TDT and GDT-PO, reside within the 95% CI for the expected distribution of p values, indicating their correct type I error control (Figure 1). The 1-TDT and GDT-PO performed well at nominal error levels ($\lambda_{1-\text{TDT}} = 1$ and $\lambda_{\text{GDT-PO}} = 1$) but showed some inflation at stringent error levels as indicated by the data points outside the 95% CI. The corresponding gold standards for all methods (i.e., methods applied to case–parent trios) maintained correct type I error, although the TDT-like methods were slightly conservative at stringent levels. The type I error estimates and their 95% CIs provide a granular view of type I error control of each method at specific significance levels (Table S1). For instance, the TDT-like methods were slightly conservative at all levels for case–parent trios while they were less conservative at nominal levels and more inflated at stringent levels for case–mother dyads. On the other hand, HAPLIN showed similar type I error control at all levels for case–mother dyads and case–parent trios. As expected, when all offspring are affected and all parents are unaffected, the 1-TDT and the GDT-PO results are identical for case–mother dyads. While these findings were based on SNPs with a fixed MAF, we also simulated data on null SNPs with varying MAFs and found qualitatively similar results (Table S2).

3.1.2 | Scenario 1B: One homogenous ancestry group, case–mother dyads and case–parent trios combined

All methods performed similar to the corresponding gold standard methods (Figure 1). Specifically, the TDT-like methods were slightly inflated at nominal levels and slightly conservative at more stringent levels (Table S1). HAPLIN showed better type I error control at more stringent levels compared to nominal levels. These results were robust to varying MAFs of the null SNPs (Table S2).

3.1.3 | Scenario 1C: One homogenous ancestry group, case-mother/control-mother dyads

The QQ plots of HAPLIN, the only applicable method in this scenario, showed well-controlled type I error (Figure 2). For the corresponding case-parent/control-parent trio data, both HAPLIN and the TDT-like methods maintained correct type I error. Type I error estimates and corresponding 95% CIs, however, indicate the TDT-like methods were slightly conservative at all levels while HAPLIN was slightly inflated at nominal levels (Table S3). Results were robust to varying SNP allele frequencies (Table S4).
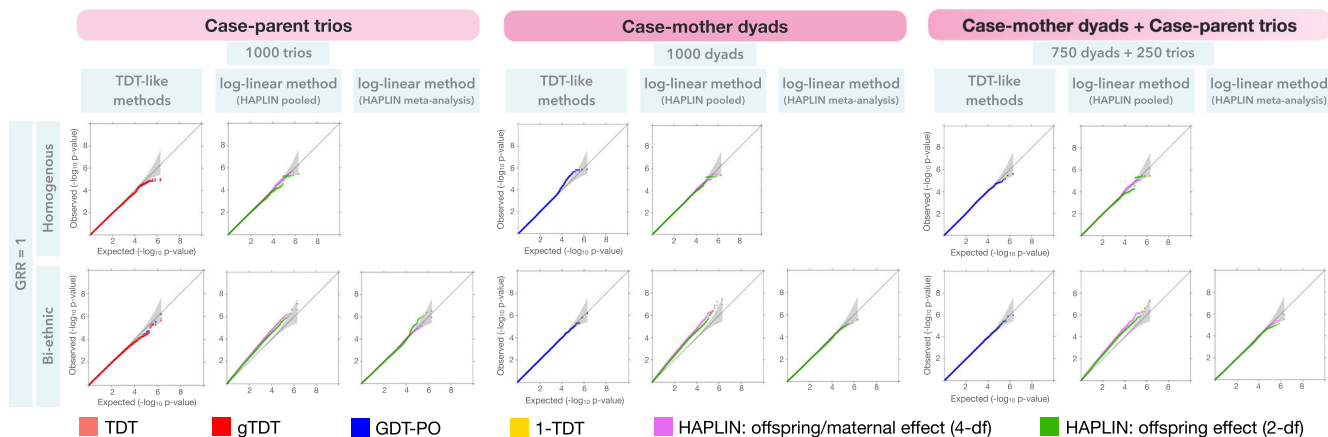
FIGURE 1 Type I error performance of the different combinations of methods and nuclear-family designs at stringent significance levels. Results are based on simulated data on 1,000 families from either one or two homogenous racial/ethnic groups with 1 million null SNPs. For one homogenous sample, a common disease prevalence of 30% and MAF 10% was simulated. For biethnic data, the second homogenous group had a disease prevalence of 15% and MAF 3%. All offspring were affected, and all parents were unaffected. Observed ($-\log_{10} p$ values) are plotted on the y-axis and Expected($-\log_{10} p$ values) on the x-axis of these QQ plots. The gray shaded region in each QQ plot represents a conservative 95% confidence interval for the expected distribution of p values. GDT, generalized disequilibrium test; gTDT, genotypic TDT; MAF, minor allele frequency; SNP, single nucleotide polymorphism; TDT, transmission disequilibrium test

3.1.4 | Scenario 2A: Two distinct ancestry groups, case–mother dyads

All TDT-like methods maintained appropriate type I error as indicated by their QQ plots lying within the 95% CI for the expected distribution of p values (Figure 1 and Table S1). HAPLIN showed inflated type I error even at nominal significance levels ($\lambda_{\text{HAPLIN-2df}} = 1.34$ and $\lambda_{\text{HAPLIN-4df}} = 1.60$). Inflation in HAPLIN was exacerbated in the presence of varying SNP allele frequencies (Table S2). These observations were reflected for the gold standards as well. When HAPLIN was applied to each ancestry group separately and then meta-analyzed, type I error was well-controlled at nominal ($\lambda_{\text{HAPLIN-2df}} = 1.02$ and $\lambda_{\text{HAPLIN-4df}} = 1.00$) as well as stringent levels (Figure 1). However, for the corresponding gold standard, the QQ plots indicated type I error was well-controlled for the 4-df test but was inflated at stringent levels for the 2-df test.

3.1.5 | Scenario 2B: Two distinct ancestry groups, case–mother dyads and case–parent trios combined

As before, all TDT-like methods exhibited well-controlled type I error rate while HAPLIN showed significant type I error inflation (Figure 1 and Table S1) particularly for varying SNP allele frequencies (Table S2). Meta-analysis of HAPLIN applied to ancestry-stratified data gave well-controlled type I error for both tests

(Figure 1). We also simulated a skewed distribution of dyads and trios within each ancestral group; results indicated robustness of TDT-like methods to population substructure while HAPLIN showed even greater inflation if data are not stratified and meta-analyzed (Figure S1). It is worth noting that with real data sets one may not have clearly distinct groups to stratify and then meta-analyze.

3.1.6 | Scenario 2C: Two distinct ancestry groups, case–mother/control–mother dyads

HAPLIN is the only method that could be evaluated here, and it showed considerable type I error inflation when applied to the pooled sample (Figure 2 and Table S3), particularly for varying SNP allele frequencies (Table S4). Meta-analysis of HAPLIN applied to ancestry-stratified data gave well-controlled type I error; no different from the corresponding gold standard (Figure 2). On the other hand, the TDT-like methods applicable only for the case–parent/control–parent design were slightly conservative. Note, we also simulated a skewed distribution of case and control offspring within each ancestral group, with all 500 case families coming from one ancestry group and all 500 control families from the other. Our results indicated robustness of TDT-like methods to extremely skewed distribution of case- and control-families across ancestral groups while HAPLIN showed extreme type I error inflation (Figure S2).
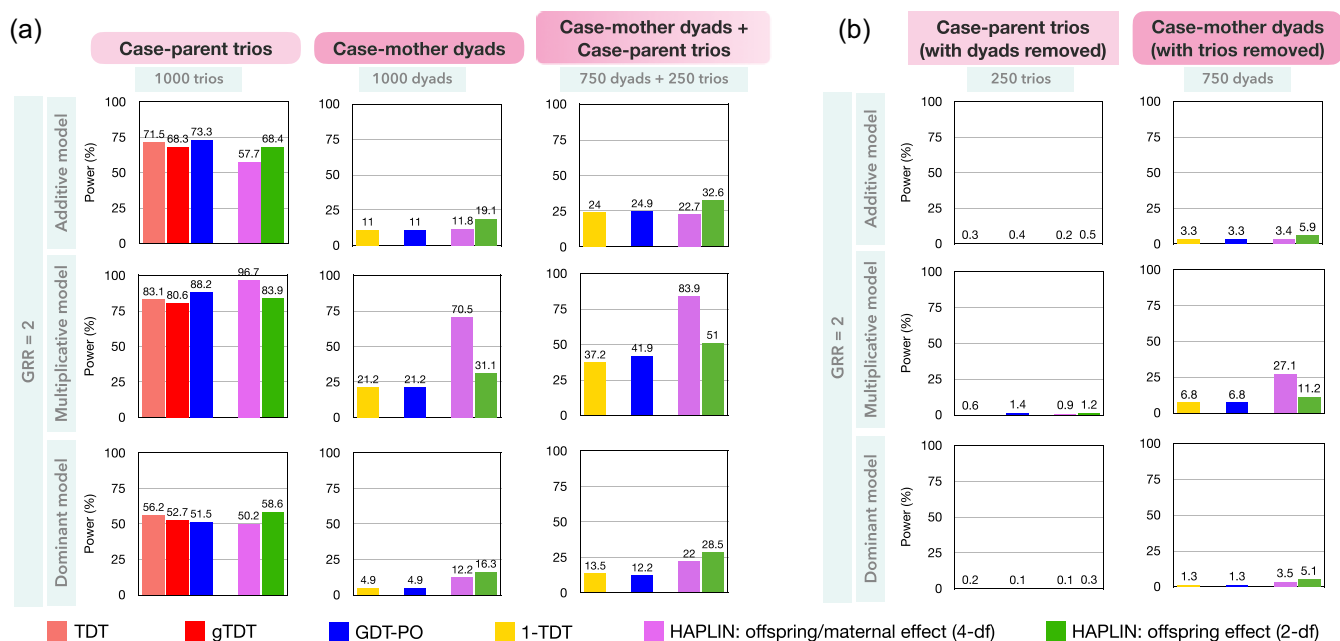
FIGURE 2 Type I error performance of the different combinations of methods and hybrid-family designs at stringent significance levels. Results are based on simulated data on 1,000 families from either one or two homogenous racial/ethnic groups with 1 million null SNPs. For one homogenous group, a common disease prevalence of 30% and MAF 10% was simulated. For biethnic data, the second homogenous group had a disease prevalence of 15% and MAF 3%. Case-to-control ratio among offspring was 50:50, and all parents were unaffected. Observed ($-\log_{10} p$ values) are plotted on the y-axis and Expected ($-\log_{10} p$ values) on the x-axis of these QQ plots. The gray shaded region in each QQ plot represents a conservative 95% confidence interval for the expected distribution of p values. GDT, generalized disequilibrium test; gTDT, genotypic TDT; MAF, minor allele frequency; SNP, single nucleotide polymorphism; TDT, transmission disequilibrium test

3.2 | Simulation experiments: Power

3.2.1 | Scenario 1A: One homogenous ancestry group, case–mother dyads

All the TDT-like methods had the similar statistical power to detect phenotype–genotype association (Figure 3a). HAPLIN was more powerful and, depending on the inheritance model, the 4-df test of offspring/maternal effects showed improved power over the 2-df test of offspring effects alone. The relative power of these methods was qualitatively similar across different MAFs and underlying inheritance model (Figure S3). For case–parent trios, the power estimates of these methods were not qualitatively very different at a given MAF and inheritance model; however, the relative behavior of these power curves showed slight variation with the inheritance model. All methods had reduced power for case–mother dyads compared to those for case–parent trios, as expected.

3.2.2 | Scenario 1B: One homogenous ancestry group, case–mother dyads, and case–parent trios combined

For a fixed number of families, all methods had improved power over-analyzing case–mother dyads alone (Figure 3a and S3), again as expected. Further, discarding any type of family from a mixed family design like this resulted in substantial loss of statistical power (Figure 3b).

3.2.3 | Scenario 1C: One homogenous ancestry group, case–mother/control–mother dyads

HAPLIN, the only applicable method, was nearly as powerful in the dyad design as in the full trio design (Figure 4). As the proportion of cases among offspring decreases, HAPLIN's power also decreases. The 4-df test

FIGURE 3 Statistical power for the different combinations of methods and nuclear-family designs at genome-wide significance level (5×10^{-8}). Results are based on simulated data on 1,000 families from one homogenous racial/ethnic group with 10,000 nonnull SNPs at MAF 10% at the casual SNP, and a common disease prevalence of 30%. Results for data simulated using the recessive inheritance model are not shown due to nearly zero power of these methods at the chosen parameter values. All offspring were affected, and all parents were unaffected. (a) Comparison of designs with the same number of families of different compositions. (b) Comparison of the combined analysis of 750 case–mother dyads and 250 case–parent trios against the scenarios when either all dyads or all trios are removed from analysis. GDT, generalized disequilibrium test; gTDT, genotypic TDT; MAF, minor allele frequency; SNP, single nucleotide polymorphism; TDT, transmission disequilibrium test

for offspring/maternal effect was unexpectedly more powerful than the 2-df test of offspring genotypic effect for a multiplicative inheritance model despite the larger df. Additionally, HAPLIN unexpectedly achieved better power for case–mother/control–mother dyads than case–parent/control–parent trios when the case–control ratio among offspring is low. These are, however, artifacts arising from this simulation experiment not satisfying the rare disease assumption (Figure S4). For the corresponding gold-standard design (i.e., case–parent/control–parent design), there are competing methods although they provided less power than HAPLIN. These findings were qualitatively similar across a range of MAFs (Figure S5). Between the two TDT-like methods, TDT_{D+C} was uniformly more powerful than TDT_{DC} across different MAFs regardless of the inheritance model and the case–control distribution among offspring, despite the higher df of TDT_{D+C}. In fact, the power difference between TDT_{D+C} and TDT_{DC} increased substantially with decreasing prevalence (Figure S4) and increasing MAF (Figure S5) across different inheritance models when the number of case families is at least as large as the number of control families. This relative behavior of TDT_{D+C} and TDT_{DC} power curves

contradicts what Deng and Chen (2001) found likely because their conclusion was based on a high disease prevalence, equal numbers of case- and control-families only, dominant effects only, an MAF of 50% (which, from an evolutionary genetics perspective, rarely happens for a disease allele), and at a nominal significance level of 5%.

3.3 | Application to GENEVA data on orofacial clefts

3.3.1 | Case–parent trios

To establish a gold standard, we first analyzed the complete case–parent trio data (pooled sample) using several TDT-like methods (TDT, gTDT, and GDT-PO). As expected, the TDT and the gTDT gave nearly identical results with both replicating known signals for CL/P (Beaty et al., 2016, Dixon et al., 2011, Ray et al., 2021) at the conventional genome-wide threshold ($p < 5 \times 10^{-8}$): 1p22.1 (*ABCA4/ARHGAP29*), 1q32.2 (*IRF6*), 8q24 (gene desert), 17p13.1 (*NTN1*), 18q12.1 (*TTR*, not a known cleft-associated region and could be spurious), and 20q12 (*MAFB*) (Figure 5). Further, 3p11.1 (*EPHA3*), 8q21.3
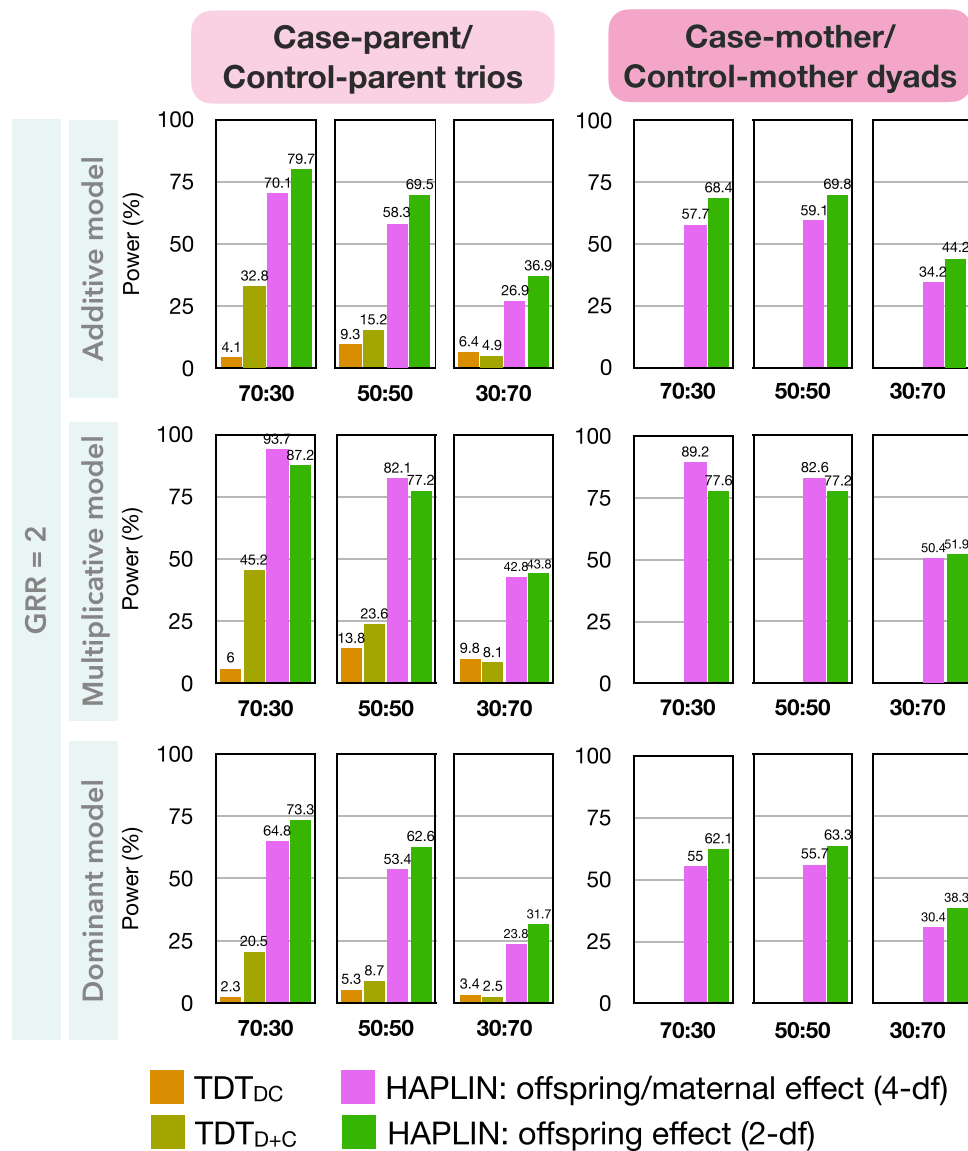
FIGURE 4 Statistical power for different combinations of methods and hybrid-family designs at genome-wide significance level (5×10^{-8}). Results are based on simulated data on 1,000 families from one homogenous racial/ethnic group with 10,000 nonnull SNPs at MAF 10% at the causal SNP, and a common disease prevalence of 30%. Results for data simulated using the recessive inheritance model are not shown due to nearly zero power of these methods at the chosen parameter values. Case-to-control ratio among offspring is either 70:30, 50:50 or 30:70. All parents are unaffected. GDT, generalized disequilibrium test; gTDT, genotypic TDT; MAF, minor allele frequency; SNP, single nucleotide polymorphism; TDT, transmission disequilibrium test

(*DCAF4L2*), and 10q25.3 (*SHTN1*) yielded suggestive significance ($p < 10^{-6}$). The GDT-PO signals were somewhat attenuated compared to those from TDT/gTDT, and it failed to replicate the signals at/near genes *TTR*, *DCAF4L2*, and *SHTN1*. This may be due to the reduced sample sizes when only phenotypically discordant parent–child pairs contributed to the test statistic since some parents had more subtle "microforms" or missing phenotype information.

We additionally analyzed the complete trios using HAPLIN. Unlike TDT-like methods, HAPLIN was not immune to type I error inflation due to population stratification (Figure S6). So, we analyzed Asian and European groups separately using HAPLIN and then meta-analyzed the results from the 2-df test for offspring effects (Figure 5). HAPLIN detected a new signal at 6p22.1 (*TRIM26*, $p = 2.1 \times 10^{-8}$), which has no known relevance to CL/P and could be spurious. It failed to
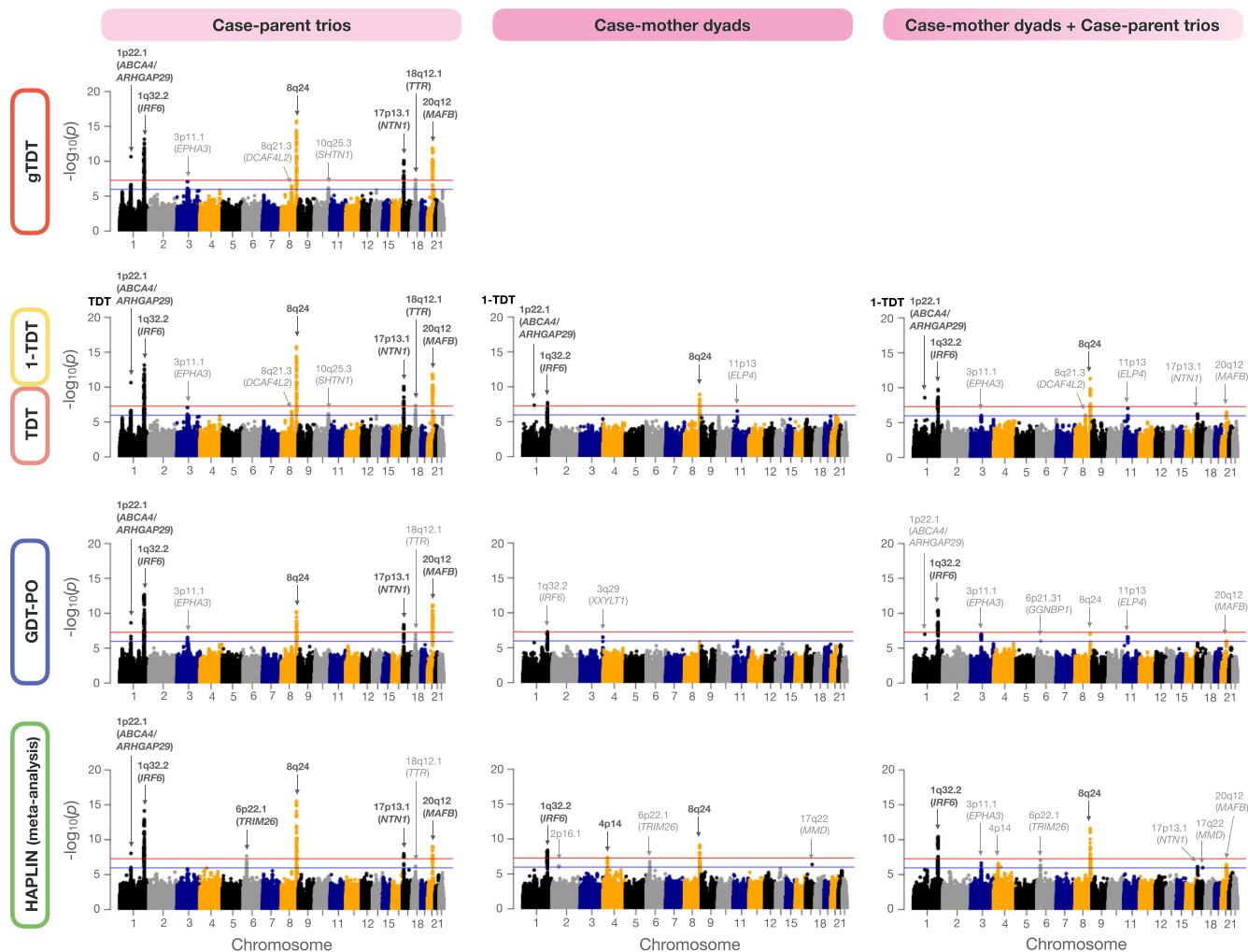
FIGURE 5 Manhattan plots for the different combinations of methods and nuclear family designs from the multiethnic GENEVA study on CL/P. The gTDT and the TDT are applicable to case–parent trio design ($n = 1487$) only. The 1-TDT (a generalization of TDT) is applicable to both case–mother dyad design ($n = 1487$) and the combined case–mother dyad case–parent trio design ($n_1 = 371$ trios, $n_2 = 1116$ dyads). The GDT-PO and HAPLIN methods are applicable to all three designs. Here, HAPLIN (2-df test of offspring genotypic effect) was applied on each racial/ethnic group separately and then meta-analyzed. The red and blue horizontal lines in each plot correspond to the genome-wide (5×10^{-8}) and a suggestive (10^{-6}) significance levels, respectively. The genome-wide significant loci for each method-design pair are annotated in dark gray and the suggestively significant loci in light gray. The gene names provided are labels for the genetic loci based on nearest gene mapping approach and do not necessarily represent causal genes. GDT, generalized disequilibrium test; gTDT, genotypic TDT; TDT, transmission disequilibrium test

detect the *EPHA3* signal, which could be due to reduced sample size for the meta-analysis compared to the pooled analysis.

3.3.2 | Case–mother dyads

The 1-TDT and the GDT-PO showed considerably reduced power to detect genetic associations compared to the case–parent trio analysis, as expected (Figure 5). At genome-wide significance, 1-TDT identified only the *ABCA4/ARHGAP29*, *IRF6* and 8q24 signals compared to

TDT/gTDT on complete trios. GDT-PO failed to identify any genome-wide significant signals. This lack of power for GDT-PO compared to 1-TDT is presumably in part due to smaller sample sizes and missing phenotype information in some mothers. HAPLIN, when meta-analyzed over Asian and European groups, detected only the *IRF6* and 8q24 signals. It also detected an intergenic region at 4p14 ($p = 4.2 \times 10^{-8}$), which may be spurious as it was not detected by the HAPLIN meta-analysis of complete trios and has not been previously reported in the cleft literature. It is possible our HAPLIN meta-analysis results were still inflated due to population
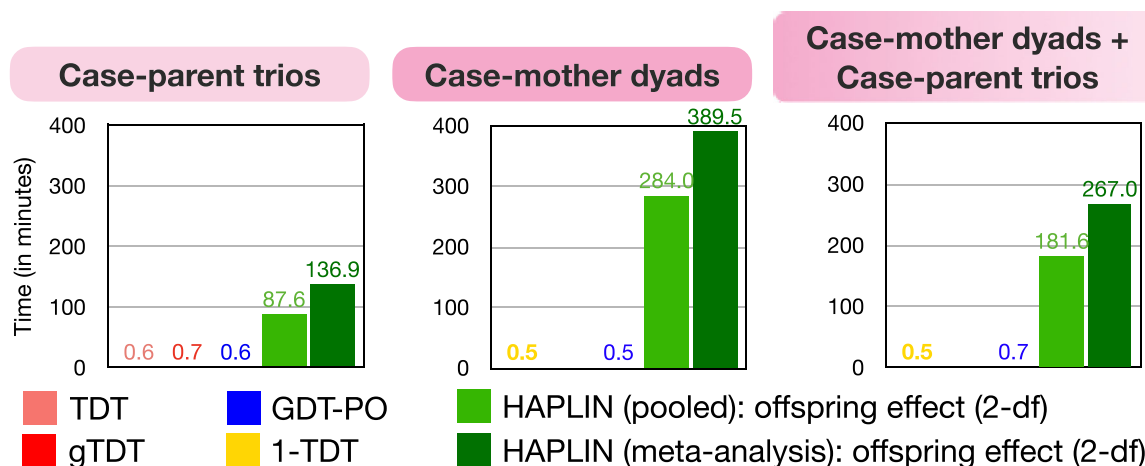
FIGURE 6 Compute times for the different combinations of methods and nuclear family designs from the multiethnic GENEVA study on CL/P. Results are based on a subset of the genetic data: 8,015 genotyped/imputed SNPs in the region chr8:128344410-132105518 that includes the known cleft locus 8q24. GDT, generalized disequilibrium test; gTDT, genotypic TDT; SNP, single nucleotide polymorphism; TDT, transmission disequilibrium test

stratification because each racial/ethnic group was not completely homogenous (participants were sampled from various countries). At the suggestive significance level, a few additional regions were detected, some of which may be spurious since they were not found in the corresponding analysis of the complete trios: 11p13 by 1-TDT; *IRF6* and 3q29 by GDT-PO; 2p16.1, 6p22.1 and 17q22 by HAPLIN.

3.3.3 | Case–mother dyads and case–parent trios combined

The 1-TDT and the GDT-PO showed improved power over-analyzing the same number of families consisting of case–mother dyads alone (Figure 5). At the genome-wide significance level, 1-TDT identified only the *ABCA4/ARHGAP29*, *IRF6* and 8q24 signals while GDT-PO identified only the *IRF6* signal when compared to TDT/gTDT signals for complete trios. HAPLIN, when meta-analyzed over Asian and European groups, detected only the *IRF6* and 8q24 signals. Interestingly, HAPLIN did not detect the possibly spurious signal at 4p14 at the genome-wide threshold that it identified from case–mother dyads alone. At the suggestive significance level, 1-TDT detected the *EPHA3*, *DCAF4L2*, 11p13 (possibly spurious), *NTN1* and *MAFB* signals; the GDT-PO detected the *ABCA4/ARHGAP29*, *EPHA3*, 6p21.31 (possibly spurious), 8q24, 11p13 (possibly spurious) and *MAFB* signals; and finally, HAPLIN detected the *EPHA3*, 4p14 (possibly spurious), 6p22.1 (possibly spurious), *NTN1*, 17q22 (possibly spurious) and the *MAFB* signals.

3.4 | Comparison of compute times

We use genetic data on 8,015 SNPs in the region 8q24:128344410-132105518, which includes the known gene desert region strongly associated with risk to CL/P. Figure 6 shows the compute times on an Oracle Virtual Machine 6.1.24 (64 bit) with Intel® Xeon® CPU E3-1270 v6 @3.80 GHz processor and 8 GB of RAM. Across all three designs, the TDT-like methods had comparable compute times of <1 minute, which included loading and necessary formatting of data. HAPLIN, when applied on the pooled sample, took at least 90 minutes to run for the case–parent trio design and took 2-3 fold more time for study designs with at most one missing parent. For multiethnic samples, application of HAPLIN on ancestry-stratified data and subsequent meta-analysis using Fisher's method required nearly 1.5-fold increased time compared to HAPLIN applied on the pooled sample. For genome-wide scans, one can reduce HAPLIN's compute time by using its built-in parallel implementation if multiple cores are available.

4 | DISCUSSION

In this article, we reviewed and benchmarked existing open-source statistical methods for the genome-wide analysis of parent–child dyads against those available for case–parent trios. We considered case–mother dyads alone, a combination of case–mother dyads and case–parent trios, and combinations of case–mother/control–mother dyads. We compared these study designs against

their corresponding gold standards: either case–parent trios alone or case–parent/control–parent trios. We used extensive simulation experiments, and array-based genetic data on trios ascertained through a child affected by CL/P from the GENEVA study with parental phenotypes when available. Although we used the methods for parent–child dyads exclusively on mother–child dyads, one could use them on father–child dyads, and theoretically on some combination of father–child and mother–child dyads assuming only direct effects of offspring genotypes are of interest and there is no mating asymmetry in the population.

This study was partly motivated by the Environmental influences on Child Health Outcomes (ECHO) study, a cohort collaboration that seeks to identify environmental and genetic exposures relevant to child development and disease. Genetic array data will be paired with a catalog of early childhood or maternal outcomes on more than 30,000 individuals in the United States comprising mother–child–father trios and mother–child pairs across these multiethnic ECHO cohorts. Given the expected genetic diversity of these cohorts, it is imperative to evaluate methods that can be applied individually to dyads or trios (or some combination) while taking maximal advantage of the family-based structure and information whenever possible. Similarly, it is essential to define the analytical approach best suited for each family-based design, especially since dyads are easier to recruit. Recommendations from this study will not only help identify the best strategies to use in diverse studies like the ECHO study, but also provide guidelines for other dyad design-based studies.

4.1 | Recommendations

For case–parent trios, the TDT-like approaches (e.g., TDT, gTDT, and GDT-PO) inherently circumvent the confounding due to population stratification by distinguishing between transmitted and nontransmitted parental alleles either in a contingency table framework or in a conditional regression framework. The non-transmitted parental alleles in a case–parent trio design—also referred to as the case–parental control design (Sun et al., 1998)—serve as matched genetic controls even when random mating and HWE assumptions are not met (Weinberg et al., 1998). Consequently, generalizations of these methods to accommodate a missing parent in the case–mother dyad design (e.g., 1-TDT, GDT-PO) or to accommodate control–parent families (e.g., TDT_{D+C}, TDT_{DC}) also protect against confounding due to population stratification.

For multiancestry data consisting of either case–mother dyads alone or a combination of case–mother dyads and case–parent trios, both 1-TDT and GDT-PO are useful. We recommend GDT-PO since it can model covariates in a regression framework. If there are many affected parents or parents with missing phenotypes, we recommend 1-TDT since, in this scenario, GDT-PO considers only phenotypically discordant relative pairs and hence less powerful than 1-TDT. If the data set consists of one homogeneous genetic ancestry group, we recommend a log-linear approach (e.g., HAPLIN) as it is often more powerful than 1-TDT and GDT-PO. In particular, the 2-df test for offspring genotype effects alone under a log-linear model using LRT tends to be more powerful than the 1-df TDT under a dominant or a recessive genetic model since LRT uses information about the joint transmission from pairs of parents, rather than accounting for the parental transmissions of individual alleles (Weinberg, 1999; Weinberg et al., 1998). If the data set is multiancestry but consists of identifiable genetic ancestry groups, we recommend using a log-linear approach on each homogeneous subgroup and then meta-analyzing the results. However, caution should be exercised in interpreting findings from these log-linear approaches on multiethnic data since it is often impossible to ensure each racial/ethnic group in any real data set is truly homogenous. It is important to highlight that we did not include covariates to adjust for heterogeneity within each group or in the analysis of pooled sample.

For multiancestry data consisting of case–parent/control–parent trios, either TDT_{DC} or TDT_{D+C} may be used when there are many more control–families than case–families and the disease prevalence is not rare. Otherwise, we recommend using TDT_{D+C} since it showed improved power over TDT_{DC} in our experiments regardless of disease prevalence, disease allele frequency and inheritance model. Both these methods control type I error even when case– and control–parent trios come from different ancestral populations. Unfortunately, we do not have any recommendation if the data consist of multiancestry case–mother/control–mother dyads. A meta-analysis of results from a log-linear approach like HAPLIN can be used if the case–mother/control–mother dyads (or the case–parent/control–parent trios) come from one or more identifiable homogenous genetic ancestry groups. In this case, HAPLIN is usually considerably more powerful than TDT-like approaches; however, few populations are truly homogeneous and for rare diseases, it is often necessary to draw case-families from multiple racial/ethnic groups. We provide a summary of our recommendations in Table 1.

TABLE 1 Recommended methods for testing direct effects of offspring genotypes on risk of a disease in different nuclear-family related study designs.

Type of nuclear family/ hybrid design	Applicable method (s)	Recommendation	
		Type of analysis	Method
Case–parent trios	TDT, gTDT,1-TDT, GDT-PO, HAPLIN	1 genetic ancestry group	HAPLIN
		Multiple genetic ancestry groups	gTDT
Case–mother dyads	1-TDT, GDT-PO, HAPLIN	1 genetic ancestry group	HAPLIN
		Multiple genetic ancestry groups	GDT-PO, if there are covariates; 1-TDT, if there are many parents affected or with missing disease status
Case–mother dyads + case–parent trios	1-TDT, GDT-PO, HAPLIN	1 genetic ancestry group	HAPLIN
		Multiple genetic ancestry groups	GDT-PO, if there are covariates; 1-TDT, if there are many parents affected or with missing disease status
Case–parent/ control–parent trios	TDT_{DC}, TDT_{D+C}, HAPLIN	1 genetic ancestry group	HAPLIN
		Multiple genetic ancestry groups	TDT_{D+C}
Case–mother/ control–mother dyads	HAPLIN	1 genetic ancestry group	HAPLIN
		Multiple genetic ancestry groups	No recommendation

Abbreviations: GDT, generalized disequilibrium test; gTDT, genotypic TDT; TDT, transmission disequilibrium test.

4.2 | Study limitations

We only considered methods from an extensive literature search with open-source implementation and an available manual. Other general classes of methods that could be used in one or more of our designs are family-based association test (FBAT), linear mixed models (LMM), and models based on generalized estimating equations (GEE). The FBAT approach generalizes the TDT non-parametrically to incorporate non-inbred pedigrees and can additionally handle missing parents (more generally missing founders), complex phenotypes, multiallelic markers, and arbitrary genetic models (Hecker et al., 2019, Laird & Lange, 2006). It is exactly the TDT test when considering biallelic markers under an additive model from case–parent trios, and in the case of a missing parent 1-TDT is the non-parametric generalization of TDT. We did not use FBAT in our analyses since we did not consider any of the other generalizations implemented in FBAT. Variance component models, and more generally LMM, incorporate phenotypic information of parents unlike most TDT-like methods, and incorporate familial relationships through a covariance matrix reflecting kinship between pairs of individuals (Chen & Abecasis, 2007; Eu-Ahsunthornwattana et al., 2014). Chen et al. (2011) was one of the first to recommend the use of GEE over variance component strategy in family-based studies of dichotomous phenotypes. To our knowledge, both the LMM and the GEE approaches are more commonly used in GWAS of secondary phenotypes from extended pedigree data (Ngwa et al., 2022, Suktitipat et al., 2012). The focus of this study is on a dichotomous trait—the case–control status used to ascertain probands—and common markers. We have not considered methods for secondary phenotypes, particularly quantitative traits (Ewens et al., 2008, Laird & Lange, 2008), or methods tailored to rare variants, which are increasingly becoming available via high-throughput whole exome or whole genome sequencing techniques (Hecker et al., 2020).

We did not assess any indirect effects (e.g., maternal effect, parent-of-origin effect, imprinting) or any interaction effect (e.g., maternal–fetal genotype interactions, gene–environment interaction, epistasis) (Ainsworth et al., 2011). In this context, EMIM, a tool to estimate the afore-mentioned indirect and interaction effects using multinomial modeling of case–parent trios, case–mother dyads, and/or monads (Howey & Cordell, 2012) could be considered. We used the TDT-like methods exclusively as association tests (Laird & Lange, 2006), and did not explore their type I error rates separately for the other possible null hypothesis scenarios under the original "no linkage or no association" composite null hypothesis (Hecker et al., 2019, Laird & Lange, 2008). We also did not consider monads (Fan et al., 2013) or any other pedigree structure (Chen

et al., 2009, Hecker et al., 2019). We explored three different mother–child pair designs and their corresponding trio designs; other hybrid designs are certainly possible (Gjerdevik et al., 2020, Vermeulen et al., 2009) but are beyond the scope of this paper. We considered biallelic SNPs only (not considering multiallelic or haplotype effects; Cordell et al., 2004, Gjessing & Lie, 2006). Our simulation framework is simple and do not reflect the usual complex genetic architecture underlying many disorders; however, we used a similar framework and parameter choices as used by many others (Chen et al., 2009, Deng & Chen, 2001, Hecker et al., 2019). Our simulations did not involve any confounders since most TDT-like methods cannot accommodate covariate effects. We focused on either one or two homogenous genetic ancestry groups to consider effects of population substructure but did not consider the full range of admixture.

Nonetheless, it is important to bear in mind that we have undertaken the first attempt at benchmarking these popular methods at more stringent levels across both dyad and trio designs, across a multitude of modeling approaches and under different data types and structure. Here we provide some practical guidelines for an appropriate selection of methods to use in different potential scenarios present in consortium studies such as ECHO, or any other nuclear family-based study designs.

WEB RESOURCES

gTDT: https://www.bioconductor.org/packages/release/bioc/html/trio.html

TDT, 1-TDT, GDT-PO: https://www.chen.kingrelatedness.com/software/GDT/index.shtml

TDT$_{D+C}$, TDT$_{DC}$: https://github.com/RayDebashree/TDT-like-tests

HAPLIN: https://cran.r-project.org/web/packages/Haplin/index.html

ORCID

Debashree Ray http://orcid.org/0000-0002-0979-2935

REFERENCES

Ainsworth, H. F., Unwin, J., Jamison, D. L., & Cordell, H. J. (2011). Investigation of maternal effects, maternal-fetal interactions and parent-of-origin effects (imprinting), using mothers and their offspring. *Genetic Epidemiology*, *35*, 19–45.

Beaty, T. H., Marazita, M. L., & Leslie, E. J. (2016). Genetic factors influencing risk to orofacial clefts: Today's challenges and tomorrow's opportunities. *F1000Research*, *5*, 2800.

Beaty, T. H., Murray, J. C., Marazita, M. L., Munger, R. G., Ruczinski, I., Hetmanski, J. B., Liang, K. Y., Wu, T., Murray, T., Fallin, M. D., Redett, R. A., Raymond, G., Schwender, H., Jin, S. C., Cooper, M. E., Dunnwald, M., Mansilla, M. A., Leslie, E., Bullard, S., ... Scott, A. F. (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nature Genetics*, *42*, 525–529.

Benyamin, B., Visscher, P. M., & McRae, A. F. (2009). Family-based genome-wide association studies. *Pharmacogenomics*, *10*, 181–190.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(7), 8. eCollection 2015.

Chen, M. H., Liu, X., Wei, F., Larson, M. G., Fox, C. S., Vasan, R. S., & Yang, Q. (2011). A comparison of strategies for analyzing dichotomous outcomes in genome-wide association studies with general pedigrees. *Genetic Epidemiology*, *35*, 650–657.

Chen, W. M., & Abecasis, G. R. (2007). Family-based association tests for genomewide association scans. *American Journal of Human Genetics*, *81*, 913–926.

Chen, W. M., & Deng, H. W. (2001). A general and accurate approach for computing the statistical power of the transmission disequilibrium test for complex disease genes. *Genetic Epidemiology*, *21*, 53–67.

Chen, W. M., Manichaikul, A., & Rich, S. S. (2009). A generalized family-based association test for dichotomous traits. *American Journal of Human Genetics*, *85*, 364–376.

Cordell, H. J., Barratt, B. J., & Clayton, D. G. (2004). Case/pseudocontrol analysis in genetic association studies: A unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. *Genetic Epidemiology*, *26*, 167–185.

Curtis, D., & Sham, P. C. (1995). A note on the application of the transmission disequilibrium test when a parent is missing. *American Journal of Human Genetics*, *56*, 811–812.

Deng, H. W., & Chen, W. M. (2001). The power of the transmission disequilibrium test (TDT) with both case-parent and control-parent trios. *Genetical Research*, 78, 289–302.

Dixon, M. J., Marazita, M. L., Beaty, T. H., & Murray, J. C. (2011). Cleft lip and palate: Understanding genetic and environmental influences. *Nature Reviews Genetics*, 12, 167–178.

Dudbridge, F. (2008). Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Human Heredity*, 66, 87–98.

Eu-Ahsunthornwattana, J., Miller, E. N., Fakiola, M., Wellcome Trust Case Control, Consortium, Jeronimo, S. M., Blackwell, J. M., & Cordell, H. J. (2014). Comparison of methods to account for relatedness in genome-wide association studies with family-based data. *PLoS Genetics*, 10, e1004445.

Ewens, W. J., Li, M., & Spielman, R. S. (2008). A review of family-based tests for linkage disequilibrium between a quantitative trait and a genetic marker. *PLoS Genetics*, 4, e1000180.

Fallin, D., Beaty, T., Liang, K. Y., & Chen, W. (2002). Power comparisons for genotypic vs. allelic TDT methods with >2 alleles. *Genetic Epidemiology*, 23, 458–4.

Fan, R., Lee, A., Lu, Z., Liu, A., Troendle, J. F., & Mills, J. L. (2013). Association analysis of complex diseases using triads, parent-child dyads and singleton monads. *BMC Genetics*, 14, 78.

Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver and Boyd.

Gjerdevik, M., Gjessing, H. K., Romanowska, J., Haaland, Ø. A., Jugessur, A., Czajkowski, N. O., & Lie, R. T. (2020). Design efficiency in genetic association studies. *Statistics in Medicine*, 39, 1292–1310.

Gjerdevik, M., Jugessur, A., Haaland, Ø. A., Romanowska, J., Lie, R. T., Cordell, H. J., & Gjessing, H. K. (2019). Haplin power analysis: A software module for power and sample size calculations in genetic association analyses of family triads and unrelated controls. *BMC Bioinformatics*, 20, 165–3.

Gjessing, H. K., & Lie, R. T. (2006). Case-parent triads: Estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. *Annals of Human Genetics*, 70, 382–396.

Hecker, J., Laird, N., & Lange, C. (2019). A comparison of popular TDT-generalizations for family-based association analysis. *Genetic Epidemiology*, 43, 300–317.

Hecker, J., Townes, F. W., Kachroo, P., Laurie, C., Lasky-Su, J., Ziniti, J., Cho, M. H., Weiss, S. T., Laird, N. M., & Lange, C. (2020). A unifying framework for rare variant association testing in family-based designs, including higher criticism approaches, SKATs, and burden tests. *Bioinformatics*, 36, 5432–5438.

Howey, R., & Cordell, H. J. (2012). PREMIM and EMIM: Tools for estimation of maternal, imprinting and interaction effects using multinomial modelling. *BMC Bioinformatics*, 13, 149.

Jugessur, A., Shi, M., Gjessing, H. K., Lie, R. T., Wilcox, A. J., Weinberg, C. R., Christensen, K., Boyles, A. L., Daack-Hirsch, S., Nguyen, T. T., Christiansen, L., Lidral, A. C., & Murray, J. C. (2010). Maternal genes and facial clefts in offspring: A comprehensive search for genetic associations in two population-based cleft studies from Scandinavia. *PLoS One*, 5, e11493.

Laird, N. M., & Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. *Nature Reviews Genetics*, 7, 385–394.

Laird, N. M., & Lange, C. (2008). Family-based methods for linkage and association analysis. *Advances in Genetics*, 60, 219–252.

Ngwa, J. S., Yanek, L. R., Kammers, K., et al. (2022). Secondary analyses for genome-wide association studies using expression quantitative trait loci. *Genetic Epidemiology*. https://doi.org/10.1002/gepi.22448

van den Oord, E. J., & Vermunt, J. K. (2000). Testing for linkage disequilibrium, maternal effects, and imprinting with (In)complete case-parent triads, by use of the computer program LEM. *American Journal of Human Genetics*, 66, 335–338.

Ray, D., Pankow, J. S., & Basu, S. (2016). USAT: A unified score-based association test for multiple phenotype-genotype analysis. *Genetic Epidemiology*, 40, 20–34.

Ray, D., Venkataraghavan, S., Zhang, W., Leslie, E. J., Hetmanski, J. B., Weinberg, S. M., Murray, J. C., Marazita, M. L., Ruczinski, I., Taub, M. A., & Beaty, T. H. (2021). Pleiotropy method identifies genetic overlap between orofacial clefts at multiple loci from GWAS of multi-ethnic trios. *PLoS Genetics*, 17, e1009584.

Schaid, D. J. (1999). Likelihoods and TDT for the case-parents design. *Genetic Epidemiology*, 16, 250–260.

Schaid, D. J., & Sommer, S. S. (1993). Genotype relative risks: Methods for design and analysis of candidate-gene association studies. *American Journal of Human Genetics*, 53, 1114–1126.

Schwender, H., Taub, M. A., Beaty, T. H., Marazita, M. L., & Ruczinski, I. (2012). Rapid testing of SNPs and gene-environment interactions in case-parent trio data based on exact analytic parameter estimation. *Biometrics*, 68, 766–773.

Self, S. G., Longton, G., Kopecky, K. J., & Liang, K. Y. (1991). On estimating HLA/disease association with application to a study of aplastic anemia. *Biometrics*, 47, 53–61.

Shi, M., Murray, J. C., Marazita, M. L., Munger, R. G., Ruczinski, I., Hetmanski, J. B., Wu, T., Murray, T., Redett, R. J., Wilcox, A. J., Lie, R. T., Jabs, E. W., Wu-Chou, Y. H., Chen, P. K., Wang, H., Ye, X., Yeow, V., Chong, S. S., Shi, B., … Beaty, T. H. (2012). Genome wide study of maternal and parent-of-origin effects on the etiology of orofacial clefts. *American journal of medical genetics. Part A*, 158A, 784–794.

Shi, M., Umbach, D. M., Vermeulen, S. H., & Weinberg, C. R. (2008). Making the most of case-mother/control-mother studies. *American Journal of Epidemiology*, 168, 541–547.

Spielman, R. S., McGinnis, R. E., & Ewens, W. J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics*, 52, 506–516.

Suktitipat, B., Mathias, R. A., Vaidya, D., Yanek, L. R., Young, J. H., Becker, L. C., Becker, D. M., Wilson, A. F., & Fallin, M. D. (2012). The robustness of generalized estimating equations for association tests in extended family data. *Human Heredity*, 74, 17–26.

Sun, F., Flanders, W. D., Yang, Q., & Khoury, M. J. (1998). A new method for estimating the risk ratio in studies using case-parental control design. *American Journal of Epidemiology*, 148, 902–909.

Sun, F., Flanders, W. D., Yang, Q., & Khoury, M. J. (1999). Transmission disequilibrium test (TDT) when only one parent

is available: the 1-TDT. *American Journal of Epidemiology*, *150*, 97–104.

Vermeulen, S. H., Shi, M., Weinberg, C. R., & Umbach, D. M. (2009). A hybrid design: Case-parent triads supplemented by control-mother dyads. *Genetic Epidemiology*, *33*, 136–144.

Watanabe, K., Taskesen, E., vanBochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*, 1826–5.

Weinberg, C. R. (1999). Allowing for missing parents in genetic studies of case-parent triads. *American Journal of Human Genetics*, *64*, 1186–1193.

Weinberg, C. R., & Umbach, D. M. (2005). A hybrid design for studying genetic influences on risk of diseases with onset early in life. *American Journal of Human Genetics*, *77*, 627–636.

Weinberg, C. R., Wilcox, A. J., & Lie, R. T. (1998). A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *American Journal of Human Genetics*, *62*, 969–978.

Zhang, W., Venkataraghavan, S., Hetmanski, J. B., Leslie, E. J., Marazita, M. L., Feingold, E., Weinberg, S. M., Ruczinski, I.,

Taub, M. A., Scott, A. F., Ray, D., & Beaty, T. H. (2021). Detecting gene-environment interaction for maternal exposures using case-parent trios ascertained through a case with non-syndromic orofacial cleft. *Frontiers in Cell and Developmental Biology*, *9*, 621018.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.