UNIVERSITY OF COPENHAGEN

Analysis of Case-Parent Trios Using a Loglinear Model with Adjustment for Transmission Ratio Distortion

Huang, Lam Opal; Infante-RIvard, Claire; Labbe, Aurélie

Analysis of Case-Parent Trios Using a Loglinear Model with Adjustment for Transmission Ratio Distortion

Lam O. Huang[1]*, Claire Infante-Rivard[1] and Aurélie Labbe[1, 2, 3]

[1] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, QC, Canada, [2] Department of Psychiatry, McGill University, Montréal, QC, Canada, [3] Douglas Mental Health University Institute, Montréal, QC, Canada

Transmission of the two parental alleles to offspring deviating from the Mendelian ratio is termed Transmission Ratio Distortion (TRD), occurs throughout gametic and embryonic development. TRD has been well-studied in animals, but remains largely unknown in humans. The Transmission Disequilibrium Test (TDT) was first proposed to test for association and linkage in case-trios (affected offspring and parents); adjusting for TRD using control-trios was recommended. However, the TDT does not provide risk parameter estimates for different genetic models. A loglinear model was later proposed to provide child and maternal relative risk (RR) estimates of disease, assuming Mendelian transmission. Results from our simulation study showed that case-trios RR estimates using this model are biased in the presence of TRD; power and Type 1 error are compromised. We propose an extended loglinear model adjusting for TRD. Under this extended model, RR estimates, power and Type 1 error are correctly restored. We applied this model to an intrauterine growth restriction dataset, and showed consistent results with a previous approach that adjusted for TRD using control-trios. Our findings suggested the need to adjust for TRD in avoiding spurious results. Documenting TRD in the population is therefore essential for the correct interpretation of genetic association studies.

Keywords: Transmission Ratio Distortion, meiotic drive, family-based association analysis, log-linear model, case-parent triad, case-parent trios, intrauterine growth restriction, intrauterine growth retardation

INTRODUCTION

Transmission Ratio Distortion (TRD) occurs when the transmission of alleles from a heterozygous parent to the offspring statistically deviates from the Mendelian Law of Inheritance. TRD results from disruptive mechanisms occurring during gametic and embryonic development (Huang et al., 2013), including germline selection (Hastings, 1991), meiotic drive (Pardo-Manuel de Villena and Sapienza, 2001), gametic competition (Zöllner et al., 2004), embryo lethality (Zöllner et al., 2004), and imprint resetting error (Naumova et al., 2001; Yang et al., 2008). The presence of TRD leads to spurious conclusions in association studies.

A recent study uses a Bayesian framework to model TRD in boars and piglets and was shown to achieve appealing statistical performance (Casellas et al., 2014). In humans, individuals unselected for phenotype have been studied to detect TRD in the general population, such as in the Framingham Heart study (Paterson et al., 2009; Meyer et al., 2012), the Centre d'Etude du Polymorphisme Humain (Naumova et al., 2001; Yang et al., 2008), the HapMap project (The International HapMap Consortium, 2005), and the 1000 Genomes Project (Auton et al., 2005).

In family-based study design, Transmission Disequilibrium test (TDT; Spielman et al., 1993) is among the most well-known linkage disequilibrium tests. It is a McNemar test of transmitted vs. untransmitted alleles from parents to an affected child. It was originally developed to test both linkage and association at a marker locus by studying case-parent trios. The usage of TDT became wide-spread since its inception because of its simplicity and robustness to population stratification. There have been multiple extensions of TDT to address multi-allelic loci (Sham and Curtis, 1995; Wilson, 1997; Lazzeroni and Lange, 1998), multiple marker loci (Lazzeroni and Lange, 1998), quantitative traits (Allison, 1997; Rabinowitz, 1997; Xiong et al., 1998), nuclear family with multiple affected children (Martin et al., 1997) and unaffected siblings (Lazzeroni and Lange, 1998), pedigrees (Sham and Curtis, 1995), late-onset diseases (Spielman and Ewens, 1998), and imprinting effect (Hu et al., 2007).

In some studies, case and control populations were analyzed separately to detect a difference in transmission (Friedrichs et al., 2006; Shoubridge et al., 2012). To address the possible presence of TRD in the studied population, Spielman et al. (1993) analyzed both case/control-trios separately using the TDT. True association was then assessed using a Pearson's Chi-square test. Deng and Chen (2001) proposed a TDT statistic that is the sum of TDT statistics for case/control-trios for similar purpose. Previously, we also suggested a modified TDT statistics where the two diagonal counts in McNemar test are multiplied by t and $(1-t)$, respectively, where t is the transmission ratio of the minor allele in control-trios (Labbe et al., 2013).

Other statistical measures have also been proposed to study affected offspring, such as Binomial exact test (Dean et al., 2006; Yang et al., 2008), Pearson's Chi-square test (Imboden et al., 2006; Bettencourt et al., 2008), multipoint non-parametric linkage (NPL) test (Paterson and Petronis, 1999; Paterson et al., 2003), Mann-Whitney U-test (De Rango et al., 2007), and multivariate logistic model (Yang et al., 2008). These methods only give statistical significance of linkage and association, but do not estimate the disease relative risk (RR). Relative risk is considered as an important information because it measures the difference in risk between individuals of different genotypes.

The family-based association test (FBAT; Lazzeroni and Lange, 1998; Rabinowitz and Laird, 2000) and likelihood methods that use case-trios to construct conditional logistic (Cordell et al., 2004), unconditional logistic (Weinberg, 1999), and loglinear models (Weinberg et al., 1998; Sinsheimer et al., 2003; Gjessing and Lie, 2006; Kistner et al., 2006, 2009) have also been used in family-based studies. In particular, Weinberg et al. proposed a loglinear model to detect an association between a marker and disease (Weinberg et al., 1998). This model estimates a RR of disease for the offspring, assuming Mendelian transmission. Unlike the other tests and models, it has a probability component that can be easily extended to adjust for TRD. Our proposed method uses the transmission ratio of a minor allele in control-trios, obtained from an external dataset such as HapMap (The International HapMap Consortium, 2005), 1000 Genomes Project phase 3 data (Auton et al., 2005), and

family units in Framingham Heart Study (2008). These datasets are publically available and include healthy trios, which provide transmission ratio of alleles from parents to child, can be used to account for TRD through an offset in the model. There are others consortia with genome-wide data, but they are based mostly on unrelated individuals (Cavalli-Sforza, 2005; Prüfer et al., 2014), a few trios (Drmanac et al., 2010), large pedigrees (Drmanac et al., 2010; T2D-GENES Consortium TD-G, 2016) or diseased individuals (The Cancer Genome Atlas, 2016; T2D-GENES Consortium TD-G, 2016), which are neither adequate nor appropriate for our study on TRD.

This extended loglinear model was validated through extensive simulation studies and applied to an intrauterine growth restriction (IUGR) case-control study augmented with a case/control-trio study (Infante-Rivard et al., 2002; Infante-Rivard and Weinberg, 2005), investigating the role of thrombophilic genes in IUGR. The current literature in support of the association between thrombophilia and IUGR is inconsistent. We explored the possible role of TRD in these inconsistencies.

MATERIALS AND METHODS

We investigated the association between a bi-allelic codominant disease susceptibility locus (DSL) and a disease, of which individuals express distinct disease risk associated with each of the three possible genotypes at the DSL. We defined genotype by the number of copies of the minor allele.

Loglinear Model by Weinberg et al. (1998)

The loglinear model proposed by Weinberg et al. (1998) assumes Mendelian transmission and mating symmetry, but not Hardy-Weinberg Equilibrium (HWE). We considered the simpler form of this model with only child genotype parameters.

In this model, the response variable is the number of trios for the 15 mother-father-child (*MFC*) genotype categories (**Table 1**). These 15 categories can be subdivided into six parental mating types. Covariates entering the model include two indicator variables for child genotypes 1 and 2, and five for mating types. The model which includes an intercept and an offset, is described as:

$$log\{E[n_{MFC}|D]\} = \rho_6 + \sum_{j=1}^{5} \rho_j I_{[S=j]} + log(2) I_{[MFC=111]}$$
$$+ \beta_1 I_{[C=1]} + \beta_2 I_{[C=2]} \quad (1)$$

n_{MFC} is the number of trios with genotypes *MFC*, and *D* is the disease status of the child. The $\rho_j + \rho_6$ terms are the regression coefficients for the first five parental mating types; ρ_6 is the intercept for the 6th mating type $MF = 00$; β_1 and β_2 are the regression coefficients for child genotypes 1 and 2, where $\beta_1 = log(R_1)$ and $\beta_2 = log(R_2)$. R_1 and R_2 are the RR with respect to genotype 0. This model 1, operates under the assumption of Mendelian transmission [derived in Appendix Derivation of Model 1 (Without TRD Offset) and 2 (With TRD Offset) and Table 6 in Supplementary Materials].

TABLE 1 | Relative risk, stratum frequency, and probability of transmission (TRD or Mendelian) for Case-parent trios study design.

Stratum	*MFC* genotype	Stratum frequency under HWE	Probability of transmission (τ_{MFC}) under TRD	Probability of transmission (τ_{MFC}) under Mendelian	Relative risk
1	222	p^4	1	1	R_2
2	212	$2p^3(1-p)$	t	1/2	R_2
	211		$1-t$	1/2	R_1
	122		t	1/2	R_2
	121		$1-t$	1/2	R_1
3	201	$p^2(1-p)^2$	1	1	R_1
	021		1	1	R_1
4	112	$4p^2(1-p)^2$	t^2	1/4	R_2
	111		$2t(1-t)$	1/2	R_1
	110		$(1-t)^2$	1/4	1
5	101	$2p(1-p)^3$	t	1/2	R_1
	100		$1-t$	1/2	1
	011		t	1/2	R_1
	010		$1-t$	1/2	1
6	000	$(1-p)^4$	1	1	1

Loglinear Model with Adjustment for TRD

Without the assumption of Mendelian transmission, model 1 can be generalized into:

$$log\{E[n_{MFC}|D]\} = \xi_6 + \sum_{j=1}^{5} \xi_j I_{[S=j]} + log\,\tau_{MFC} + \beta_1 I_{[C=1]}$$
$$+ \beta_2 I_{[C=2]} \qquad (2)$$

where τ_{MFC} is the transmission offset $P[C|MF]$, $\xi_j + \xi_6$ terms ($j = 1-5$) are the regression coefficients for the first five mating types, and ξ_6 is the intercept corresponding to the 6th mating type. The coefficients β_1 and β_2 are as defined in model 1. This model 2 accounts for TRD [derived in Appendix and Table 6: Derivation of Model 1 (Without TRD Offset) and 2 (With TRD Offset) and Table 6 in Supplementary Materials].

The offset τ_{MFC} depends on the TRD ratio t, defined as the transmission probability of a minor allele from a heterozygous parent to the child. This leads to a different offset in each *MFC* genotype category. The parameter t can take on values different from 0.5, and $t = 0.5$ corresponds to Mendelian transmission, in which case models 1 and 2 are equivalent [see Appendix and Table 6: Derivation of Model 1 (Without TRD Offset) and 2 (With TRD Offset) and Table 6 in Supplementary Materials].

We fitted both loglinear models (1) and (2) to obtain estimates R_1 and R_2, and their corresponding Z-test p-values. To assess significance of the association between the disease and the DSL, a Likelihood Ratio Test (LRT) was used [see Appendix Non-Central Chi-Square Likelihood for Model 1 (Without TRD Offset) and Model 2 (With TRD Offset) for the distribution of the LRT under the null and alternative hypotheses].

Simulation Study

A simulation study was set up for different TRD scenarios, where RR parameters, p-values, LRT p-values, Type 1 error, and power were compared between the 2 models, and the true t was used in model 2. A sensitivity analysis was also carried out to test the impact on RR estimates and power when an incorrect t is used.

Simulation Setup

We considered a causal locus with no recombination. Disease prevalence is 0.1 for low penetrant common disease, and 0.01 for high penetrant rare disease. 100,000 trios were generated where 500 case-trios were sampled. Parental genotypes at the DSL were generated under HWE assuming a minor allele frequency (MAF) 0.1. The parameter t was specified between 0.1 and 0.9. Offspring were assigned to diseased or non-diseased phenotypes using risk associated with genotypes 0, 1, and 2, as f_0, f_1, and f_2, respectively. The simulation was repeated 100 times and averaged RR estimates, p-value of the averaged Z statistics for RR and p-value of the averaged LRT statistics are reported.

Measuring Impact of TRD on Association Statistics

We compared the RR, 95% CI, p-value and LRT p-value of both models under two scenarios: (1) a common disease associated of low penetrance at $f_0 = 0.1$, $f_1 = 0.11$, $f_2 = 0.15$, and (2) a rare disease of high penetrance at $f_0 = 0.1$, $f_1 = 0.5$, $f_2 = 0.5$. In scenario (2), a dominant model was assumed. To measure the inflation in RR and LRT p-values in model 1, we computed the log ratio of RR and LRT p-values in model 1 vs. 2. We also varied f_1 fixing $f_2 = 0.15$ to describe the corresponding inflation of LRT p-values. To assess the inflation of Type 1 error, we set the penetrance factors to $f_0 = f_1 = f_2 = 0.1$ assuming no association while varying t from 0.1 to 0.9, using sample sizes of 100, 300,

TABLE 2 | Relative risk with 95% CI, *P*-values, and likelihood ratio test *P*-values of models 1 (Unadjusted) and 2 (Adjusted) for a low penetrance common disease.

t	Model	R_1	95% CI	*P*-value	R_2	95%CI	*P*-value	LRT *P*-value
0.3	1	0.47	0.33, 0.65	6.00E-06	0.25	0.06, 1.08	0.07	2.85E-06
	2	1.09	0.78, 1.51	0.59	1.34	0.30, 5.84	0.51	0.28
0.5	1	1.10	0.81, 1.51	0.53	1.40	0.51, 3.89	0.43	0.26
	2	1.10	0.81, 1.51	0.53	1.40	0.51, 3.89	0.43	0.26
0.7	1	2.52	1.78, 3.57	2.00E-07	8.01	3.18, 20.2	8.27E-06	6.57E-10
	2	1.08	0.76, 1.53	0.7	1.47	0.58, 3.70	0.42	0.25

R_1, RR of cases carrying 1 copy of disease allele; R_2, RR of cases carrying two copies of disease allele; Simulated with t = 0.3, 0.5, and 0.7 and population parameters: p = 0.1, f_0 = 0.1, f_1 = 0.11, f_2 = 0.15.

and 500. Finally, we evaluated the power of both models to detect a true association signal in the presence of TRD, by setting $f_0 = 0.1, f_1 = 0.2, f_2 = 0.3$, varying t from 0.1 to 0.9 in the simulation. Critical value for declaring significance was $\alpha = 0.05$.

Sensitivity Analysis

The assumption in the simulation study was that true t is known. We examined the consequences of a misspecification of t on the RR estimates and the power, simulating three scenarios with true association signal, $f_0 = 0.1, f_1 = 0.2, f_2 = 0.3$, and true $t = 0.3, 0.5$, or 0.7. For each scenario, model 2 was fitted with the offset τ_{MFC} calculated using a selected t varying between 0.1 and 0.9. We then evaluated the log ratio of RR and power obtained from model 2 using selected t-values vs. true t that adjust for TRD.

Application of Models 1 and 2 to a Real Dataset

We applied our model to the IUGR study described previously (Sapru et al., 2009; Kvasnicka et al., 2012). Cases were below 10th percentile according to weight whereas controls were selected at the same hospital and measured at or above the 10th percentile. DNA was obtained from parents of both cases and controls. The investigation pertained to the role of thrombophilic genes in IUGR. We examined six thrombophilic genes: Coagulation Factor XIII, A1 polypeptide (*F13A1*), Plasminogen activator inhibitor type 1 (*PAI-1*), Methylenetetrahydrofolate reductase variant A1298C (*MTHFR A1298C*), Methylenetetrahydrofolate reductase variant C677T (*MTHFR C677T*), Coagulation Factor V (*F5*), and Coagulation Factor II (*F2*). We computed the MAF using all complete trios and t using control-trios. We compared our extended model 2 with another method proposed by Infante-Rivard and Weinberg (2005) to quantify the extent of TRD in the same IUGR population, specifically for *F5*. The difference between our model 2 and the model used in Infante-Rivard and Weinberg (2005) is that the former inserts t as an offset in the loglinear model fitted with case-trios only, while the latter uses both case- and control-triosadding an interaction term between child genotype and case status.

This study was carried out in accordance with the recommendations of Le Comité d'éthique de la recherche,

Centre Hospitalier Universitaire, Hôpital Sainte-Justine, Montréal, Québec, Canada. The protocol was approved by the same committee.

TABLE 3 | Relative risk with 95% CI, *P*-values, and likelihood ratio test *P*-values of models 1 (Unadjusted) and 2 (Adjusted) for a high penetrance rare disease.

t	Model	$R_{1/2}$	95%CI	*P*-value	LRT *P*-value
0.3	1	2.44	1.20, 4.94	0.014	0.025
	2	5.71	2.82, 11.57	1.29E-06	8.62E-07
0.5	1	5.58	2.55, 12.21	1.55E-05	6.55E-07
	2	5.58	2.55, 12.21	1.55E-05	6.55E-07
0.7	1	13.73	4.99, 37.79	1.57E-07	2.62E-13
	2	5.87	2.13, 16.16	0.000504	2.23E-05

$R_{1/2}$, RR of cases carrying one or two copies of disease allele; Simulated with t = 0.3, 0.5, and 0.7 and population parameters: p = 0.01, f_0 = 0.1, f_1 = 0.5, f_2 = 0.5; Data is fitted with a dominant genotype model.

RESULTS

Simulation Study

Inflation of RR Estimates and LRT *P*-values

When the transmission ratio was Mendelian, models 1 and 2 yielded the same RR and 95%CI (**Tables 2**, **3**). When testing $t = 0.3$ where the disease allele is under-transmitted, the RR for model 1 was attenuated excluding 1 in the 95% CI, whereas RR estimates, *p*-values and LRT *p*-values were restored in model 2. Similarly, for $t = 0.7$, the RR for model 1 were inflated and this inflation was removed under model 2. The RR inflation ratio changes exponentially with respect to t, implying that even small deviation from $t = 0.5$ can lead to a substantial inflation (**Figure 1A**). The slope of RR ratio for R_2 was double that of R_1, showing that TRD affected R_2 more severely than R_1. In **Figure 1B**, when TRD is not adjusted for, the significance of the LRT *p*-values was inflated when t deviates from 0.5.
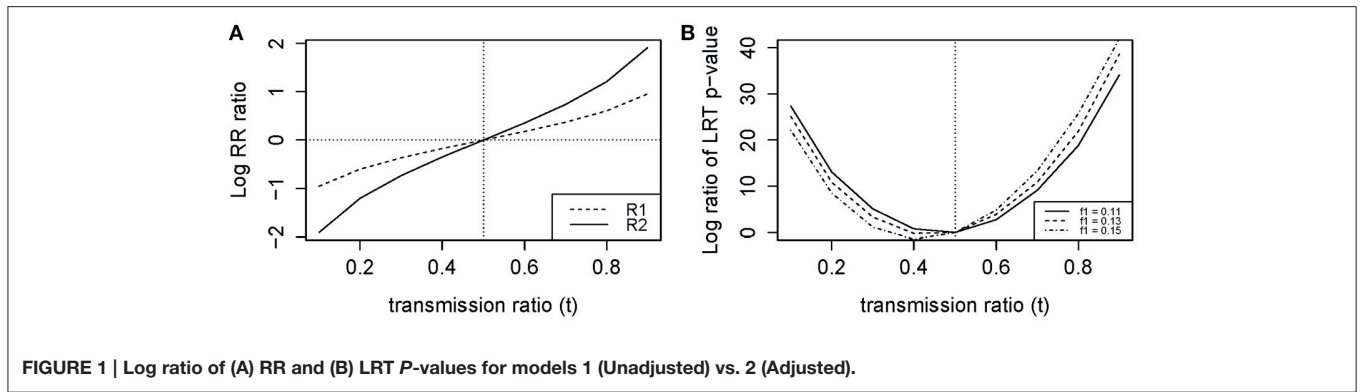
FIGURE 1 | Log ratio of (A) RR and (B) LRT *P*-values for models 1 (Unadjusted) vs. 2 (Adjusted).
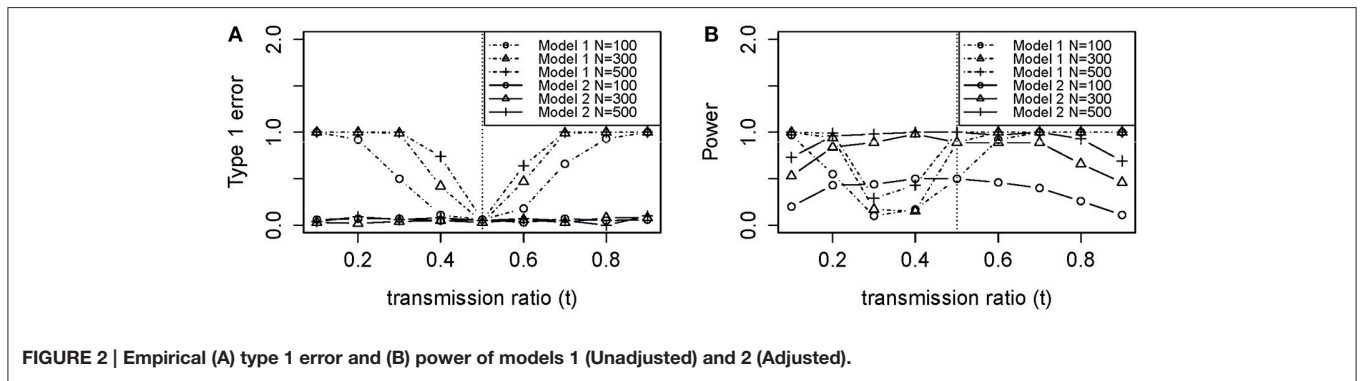

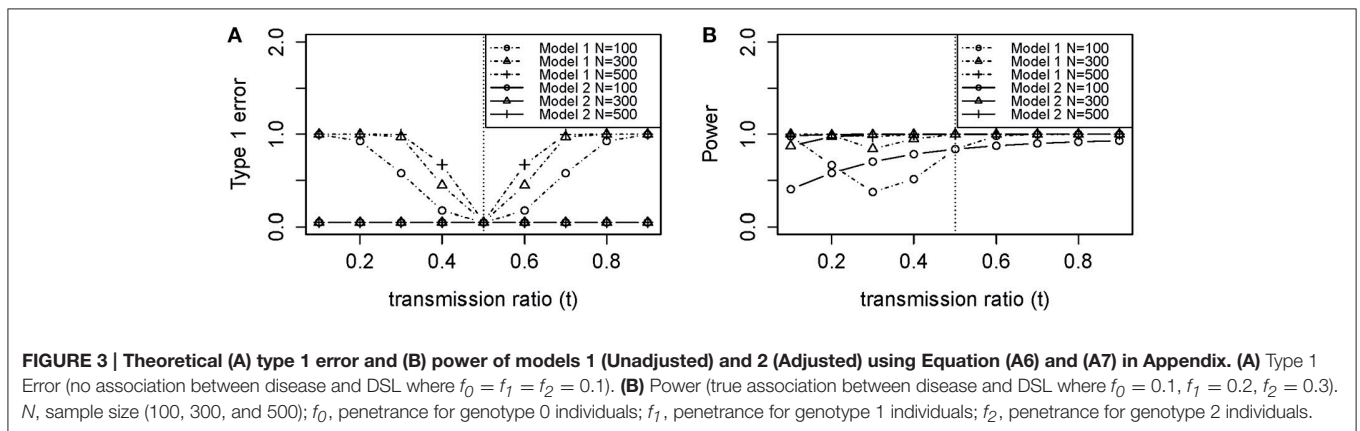FIGURE 2 | Empirical (A) type 1 error and (B) power of models 1 (Unadjusted) and 2 (Adjusted).

FIGURE 3 | Theoretical (A) type 1 error and (B) power of models 1 (Unadjusted) and 2 (Adjusted) using Equation (A6) and (A7) in Appendix. (A) Type 1 Error (no association between disease and DSL where $f_0 = f_1 = f_2 = 0.1$). (B) Power (true association between disease and DSL where $f_0 = 0.1$, $f_1 = 0.2$, $f_2 = 0.3$). N, sample size (100, 300, and 500); f_0, penetrance for genotype 0 individuals; f_1, penetrance for genotype 1 individuals; f_2, penetrance for genotype 2 individuals.

Inflation of Type 1 Error

Figure 2A shows the empirical Type 1 Error we observed by fitting the loglinear model which is similar to our theoretical results in **Figure 3A**. Type 1 Error of the TRD-adjusted model 2 remained the same across all *t*-values, and were exactly the same for all sample sizes. Type 1 Error for model 2 does not depend on sample size or *t*, meaning that this model is robust to the effect of TRD when the null hypothesis is true. In **Figure 2A**, Type 1 Error for the unadjusted model 1 increased as *t* deviated from 0.5 which led to a false inflation of the association signals.

Power Loss

Power for sample size $n = 100$ was poor in **Figure 2B**, with or without TRD. We also noticed that model 2 gave relatively stable power in the range of *t*, while model 1 power suffered from the effect of TRD. However, when *t* was lower than 0.2 or >0.5, model 1 power was greater than that of model 2. This is because a strong TRD actually inflates the power of detecting an association signal in either direction. Power for model 2 decreased slightly when *t* > 0.7, which suggested that the TRD offset overcompensates the inflation in power. However, a TRD ratio as large as 0.9 is rare, but even when $t = 0.8$, the power was still

maintained around 0.8 for sample sizes of 300 and 500. Therefore, the power for model 2 was still adequate for a t between 0.2 and 0.8. Relatively consistent results were obtained between theoretical power (**Figure 3B**) and empirical power (**Figure 2B**).

Sensitivity Analysis: Inflation in RR Estimates

We observed that using an under-estimated t-value in model 2 led to inflation, while an over-estimated t led to attenuation for R_1 (**Figure 4**). We also noted that the inflation and attenuation of the log RR ratio was linear, which means exponential in arithmetic scale. When the difference between the true and selected t was ± 0.1, the inflation ratio lied between $10^{0.25} = 1.78$ and $10^{-0.25} = 0.56$ for R_1. When the difference was greater than ± 0.1, the inflation ratio became more pronounced. The slope of the log RR ratio curve for R_2 was twice (not shown) that of R_1 in **Figure 4**. Therefore, the inflation or attenuation in R_2 was more severe than in R_1. Results from our model 2 were highly sensitive to an incorrect input of t-value.

Sensitivity Analysis: Attenuation and Inflation in Power

In **Figures 5A,B**, for $t = 0.3$ and 0.5, the power to detect true association was completely restored when the selected t was equal to the true t. However, setting the selected and true at $t = 0.7$ (**Figure 5C**), the power for detecting true association was not completely restored, consistent with what we observed previously in power analysis. There was a decrease in power when true signal is partially canceled by the selected t. We see that power was also highly sensitive to incorrect t.

Application to a Case-Control, Case-, and Control-Parent Trio Study of IUGR

The MAF calculated from all complete trios in our sample was 23.8% for *F13A1*, 46.4% for *PAI-1*, 27.1% for *MTHFR A1298C*, 28.9% for *MTHFR C677T*, 2.92% for *F5*, and 1.68% for *F2* (**Tables 3, 4**). Except for *MTHFR A1298C*, all MAF were close to the expected range from the literature (Kawamura et al., 1989; Ulvik et al., 1998; Ariens et al., 2002; Sapru et al., 2009; Alfirevic et al., 2010; Kvasnicka et al., 2012). Discrepancies were likely due
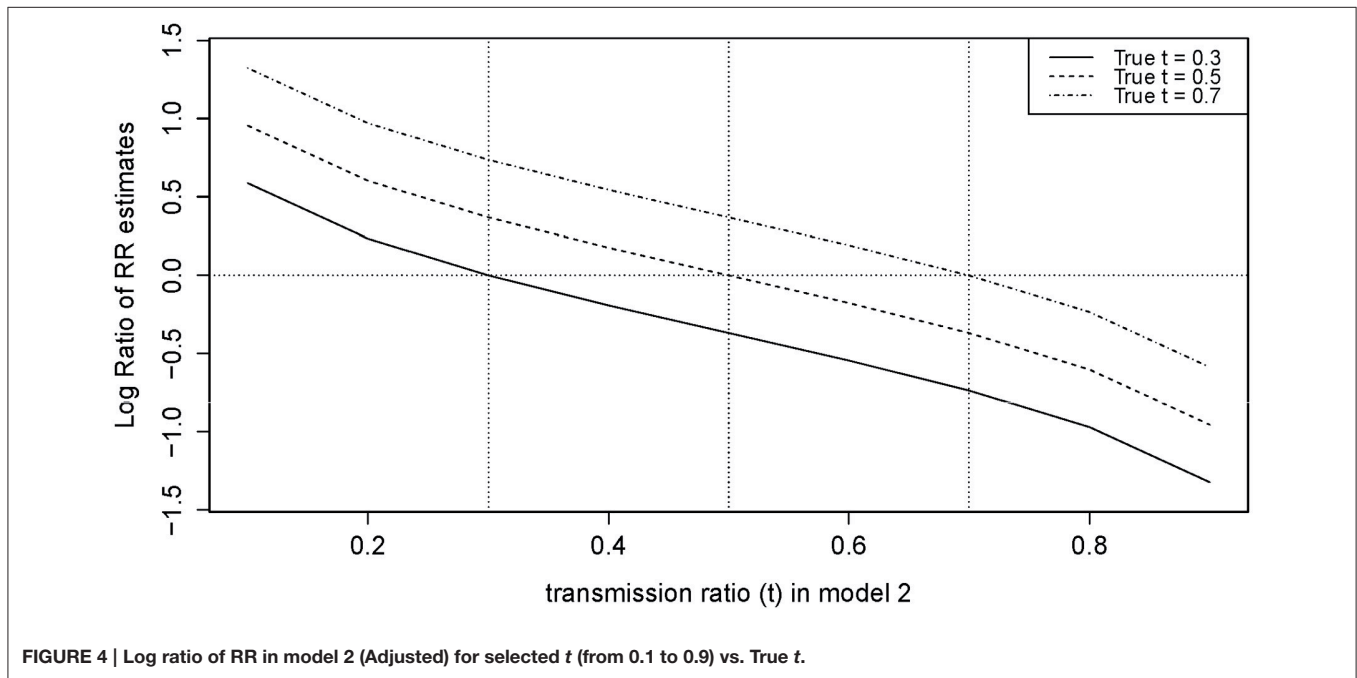

FIGURE 4 | Log ratio of RR in model 2 (Adjusted) for selected t (from 0.1 to 0.9) vs. True t.
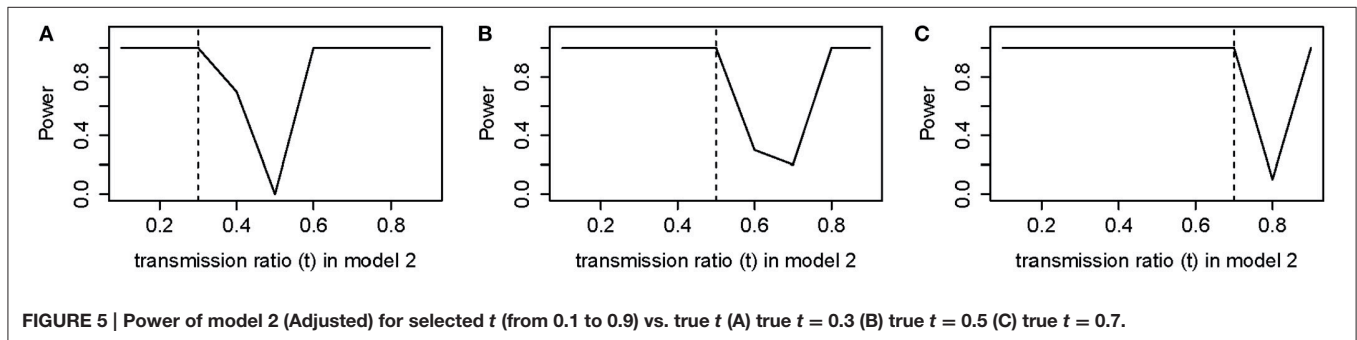

FIGURE 5 | Power of model 2 (Adjusted) for selected t (from 0.1 to 0.9) vs. true t (A) true $t = 0.3$ (B) true $t = 0.5$ (C) true $t = 0.7$.

TABLE 4 | Relative risk with 95% CI, *P*-values, and LRT *P*-values of models 1 (Unadjusted) and 2 (Adjusted) for 4 thrombopilic genes (*F13A1, PAI-1, MTHFR A1298C,* and *MTHFR C677T*), With MAF and transmission ratio (*t*), on an intrauterine growth restriction dataset collected from a Canadian hospital between 1998 and 2000.

Gene	Model	MAF	*t*	R_1	95%CI	R_1 *P*-value	R_2	95%CI	R_2 *P*-value	LRT *P*-value
F13A1	1	0.24	0.54	0.97	0.66, 1.43	0.89	1.41	0.68, 2.94	0.354	0.57
	2			0.82	0.56, 1.21	0.32	1.01	0.48, 2.1	0.98	0.55
PAI-1	1	0.46	0.49	0.80	0.49, 1.30	0.37	0.97	0.52, 1.82	0.93	0.53
	2			0.83	0.51, 1.35	0.46	1.06	0.57, 1.98	0.86	0.53
MTHFR A1298C	1	0.27	0.45	0.84	0.60, 1.19	0.34	0.78	0.40, 1.52	0.46	0.58
	2			1.04	0.74, 1.47	0.82	1.18	0.60, 2.31	0.63	0.89
MTHFR C677T	1	0.29	0.50	0.95	0.67, 1.35	0.8	0.75	0.39, 1.43	0.38	0.67
	2			0.94	0.67, 1.34	0.75	0.73	0.38, 1.40	0.34	0.65

R_1, RR of cases carrying one copy of disease allele; R_2, RR of cases carrying two copies of disease allele.

Application to 6 IUGR Genes

We see in **Table 4** that *F13A1, PAI-1,* and *MTHFR C677T* all had transmission ratios around 0.5. *MTHFR A1298C* had slightly lower transmission of the disease allele with $t = 0.45$. However, *F5* and *F2* had transmission deviate significantly from the Mendelian ratio with $t = 0.36$ and 0.11 (**Table 5**). RR from the loglinear model showed noassociation for *F13A1, PAI-1, MTHFR A1298C,* and *MTHFR C677T* variants (**Table 4**), similar to previous reports (Infante-Rivard et al., 2002, 2005). Due to the small number of genotype 2 cases for *F5* and *F2*, these two genes were analyzed under a dominant model. We see that for *F5*, conclusion on RR, *p*-values and LRT *p*-values are reversed from model 1 to model 2, suggesting a deleterious effect of the minor allele. For *F2*, we observed the opposite trend. The change in risk after adjustment for TRD was coherent with the expected effects from these variants given that they are known to affect placental circulation and thus potentially fetal growth.

Comparison with TRD Analysis in Infante-Rivard and Weinberg (2005) on FV Gene

Infante-Rivard and Weinberg (2005) found in their study that both *F5* and *F2* exhibited evidence of TRD, as well as *MTHFR A1298C* but to a lesser extent, which is consistent with our estimation from control-trios (**Tables 4**, **5**). The authors used six more strata from control-trios together with an interaction term between child genotype and case status. A gene-dosage model ($R_2 = R_1^2$) was used implicitly to adjust for TRD; the RR for cases was estimated to be 3.59. We fitted model 2 using a gene-dosage model, and obtained a RR estimate of 2.88 with 95% CI: 1.31, 6.35. This result is in the range of the estimate from Infante-Rivard and Weinberg (2005). The number of trios included in these two analyses was different as Infante-Rivard and Weinberg (2005) used the LEM software with built-in EM algorithm for missing data whereas we only used complete trios. This shows that results from our extended loglinear model 2, which adjusts

to the fact that the samples were genetically heterogeneous with ~25% being black.

TABLE 5 | Relative risk With 95% CI, *P*-values, LRT *P*-values of models 1 (Unadjusted) and 2 (Adjusted) for 2 thrombopilic genes (*F5* and *F2*), with MAF, transmission ratio (*t*) and Number of Genotype 2 Cases (G2), on an intrauterine growth restriction dataset collected from a Canadian Hospital Between 1998 and 2000.

Gene	Model	MAF	*t*	G2	$R_{1/2}$	95%CI	*P*-value	LRT *P*-value
F5	1	0.03	0.36	2	1.29	0.57, 2.93	0.54	0.53
	2				2.35	1.04, 5.33	0.04	0.039
F2	1	0.02	0.11	0	0.31	0.11, 0.85	0.023	0.014
	2				2.5	0.91, 6.82	0.074	0.1

$R_{1/2}$, RR of cases carrying one or two copies of disease allele; Data is fitted with a dominant genotype model.

for TRD were comparable to those from the augmented model proposed in Infante-Rivard and Weinberg (2005).

The method proposed by Infante-Rivard and Weinberg (2005) requires fitting the loglinear model with actual control-trios, which is not required in our method where the transmission ratio of the minor allele is obtained through publicly available datasets. Therefore, less recruitment effort is needed leading to lower study cost. This difference is more significant for genome-wide studies where large samples are required.

Both models can include the same covariates. However, since control-trios are directly fitted in the model proposed by Infante-Rivard and Weinberg (2005), each covariate included in the model will lead to 2°of freedom loss because an interaction between case status (0, control; 1, case) and the covariate itself also has to be added. This leads to a faster decline in degrees of freedom than our method. The difference will further be magnified when other more complicated covariates, such as the mother-fetal interaction effect, are included in the model. Each of the four mother-fetal interaction covariates requires an additional interaction term with the case status.

The loglinear model proposed by Infante-Rivard and Weinberg (2005) allows missing data while our method requires

complete trios only. The former has the advantage of using trios with missing parental genotypes, and hence does not need to discard trios with incomplete information. Currently, there is no immediate plan to augment our R-package for missing data, but it is possible in the future to address this issue using EM algorithm and include it as an option in our R-package. The loglinear model with control-trios has the advantage of adjusting for TRD without knowing the extent of distortion, and hence, remains a gold standard when the transmission ratio of the minor allele is not available.

DISCUSSION

Studies using animal models can potentially provide new insights in handling the phenomenon of TRD. TRD is much less studied in humans. In most genetic association studies in the current literature TRD remains largely unaccounted for. We previously reviewed a number of human studies on TRD (Naumova et al., 2001; Pardo-Manuel de Villena and Sapienza, 2001; Zöllner et al., 2004; Hanchard et al., 2005; The International HapMap Consortium, 2005; Paterson et al., 2009) and discussed the various methods and study designs in detecting TRD (Huang et al., 2013).

Here, we extend a model used for family-based association studies, accounting for TRD. Our simulation study showed that when TRD is unaccounted for as in model 1, the RR is inflated or attenuated exponentially. Power and Type 1 error also suffered greatly. Using a real dataset where the *F5* gene was studied as a determinant of IUGR, we validated our model in comparison with an approach using control trios (Infante-Rivard and Weinberg, 2005). However, we noted that the accuracy of our results depended on the correct TRD offset used in model 2. If we conduct a study with less well-known DSL and diseases, it is unlikely that we will have information on the TRD factor. Nevertheless, by leveraging on studies such as the HapMap project (The International HapMap Consortium, 2005), the 1000 Genomes Project (Auton et al., 2005), or the Framingham Heart Study (Framingham Heart Study, 2008), it may be possible to obtain such information.

The LEM software developed by van Den Oord and Vermunt (2000) that was used by Infante-Rivard and Weinberg (2005) to fit a loglinear model that takes into account of missing data. We compared RR estimates obtained from LEM and our models in the absence of TRD, and they were similar in values. HAPLIN, a software developed by Gjessing and Lie also studies case-parent-trios, which estimates the effect of multi-allelic markers or haplotype for single- and double-dose maternal and fetal haplotype (Gjessing and Lie, 2006). There are other software developed for studying case-parent trios such as TRANSMIT (Clayton and Jones, 1999),

which can handle multi-locus haplotypes and missing parental information, and GASSOC (Schaid, 1996), which accommodate multi-allelic markers. These software do not readily have a component to adjust for TRD. However, we implemented the model 2 with TRD offset in an R package (named TRD) available on the Comprehensive R Archive Network (CRAN).

Currently, there is no comprehensive knowledge on TRD in the human genome. As TRD can inflate or attenuate an association signal, with large sets of SNPs being tested, results can be severely biased leading to spurious conclusions. Since TRD over generations leads to reduced mutational diversity in the genome, many of these TRD loci contain rare variants which are currently intensively researched. When transmission counts are small, even a slight distortion could lead to major impact on the outcome of the studies. Given what we observed in our simulation study, sequencing a control population to identify and quantify the extent of TRD in the human genome would seem necessary. Incorporating this information in the analysis of genetic association studies provides more accurate and valid estimates. Therefore, we suggest that knowledge of TRD in genomic databases is essential to determine the relevance of genes in various diseases.

AUTHOR CONTRIBUTIONS

The research question for this manuscript was conceived by LH. AL reviewed and approved the conceived research question. CI acquired and provided the data used in this manuscript. LH developed, implemented and applied the method for simulation studies and real data analysis, and wrote the R software package "TRD." AL contributed to a revision of the statistical model. LH drafted the manuscript. AL and CI reviewed it critically for important intellectual content. LH, AL, and CI all approved the final version to be published. LH, AL, and CI all agreed to be accountable for all aspects of work in ensuring the questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2016.00155

REFERENCES

Alfirevic, Z., Simundic, A.-M., Nikolac, N., Sobocan, N., Alfirevic, I., Stefanovic, M., et al. (2010). Frequency of factor II G20210A, factor V Leiden, MTHFR C677T and PAI-1 5G/4G polymorphism

in patients with venous thromboembolism: croatian case control study. *Biochem. Med.* 20, 229–235. doi: 10.11613/BM.2010.028

Allison, D. B. (1997). Transmission-disequilibrium tests for quantitative traits. *Am. J. Hum. Genet.* 60, 676–690.

Ariëns, R. A., Lai, T. S., Weisel, J. W., Greenberg, C. S., and Grant, P. J. (2002). Role of factor XIII in fibrin clot formation and effects of genetic polymorphisms. *Blood* 100, 743–754. doi: 10.1182/blood.V100.3.743

Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., et al. (2005). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393

Bettencourt, C., Fialho, R. N., Santos, C., Montiel, R., Bruges-Armas, J., Maciel, P., et al. (2008). Segregation distortion of wild-type alleles at the Machado-Joseph disease locus: a study in normal families from the Azores islands (Portugal). *J. Hum. Genet.* 53, 333–339. doi: 10.1007/s10038-008-0261-7

Casellas, J., Manunza, A., Mercader, A., Quintanilla, R., and Amills, M. (2014). A flexible bayesian model for testing for transmission ratio distortion. *Genetics* 198, 1357–1367. doi: 10.1534/genetics.114.169607

Cavalli-Sforza, L. L. (2005). The human genome diversity project: past, present and future. *Nat. Rev. Genet.* 6, 333–340. doi: 10.1038/nrg1579

Clayton, D., and Jones, H. (1999). Transmission/disequilibrium tests for extended marker haplotypes. *Am. J. Hum. Genet.* 65, 1161–1169. doi: 10.1086/302566

Cordell, H. J., Barratt, B. J., and 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. *Genet. Epidemiol.* 26, 167–185. doi: 10.1002/gepi.10307

Dean, N. L., Loredo-Osti, J. C., Fujiwara, T. M., Morgan, K., Tan, S. L., Naumova, A. K., et al. (2006). Transmission ratio distortion in the myotonic dystrophy locus in human preimplantation embryos. *Eur. J. Hum. Genet.* 14, 299–306. doi: 10.1038/sj.ejhg.5201559

Deng, H. W., and Chen, W. M. (2001). The power of the transmission disequilibrium test (TDT) with both case-parent and control-parent trios. *Genet. Res.* 78, 289–302. doi: 10.1017/S001667230100533X

De Rango, F., Dato, S., Bellizzi, D., Rose, G., Marzi, E., Cavallone, L., et al. (2007). A novel sampling design to explore gene-longevity associations: the ECHA study. *Eur. J. Hum. Genet.* 16, 236–242. doi: 10.1038/sj.ejhg.5201950

Drmanac, R., Sparks, A. B., Callow, M. J., Halpern, A. L., Burns, N. L., Kermani, B. G., et al. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 327, 78–81. doi: 10.1126/science.1181498

Framingham Heart Study (2008). *Data Repository: dbGaP* Available online at: http://www.ncbi.nlm.nih.gov/bioproject/76025

Friedrichs, F., Brescianini, S., Annese, V., Latiano, A., Berger, K., Kugathasan, S., et al. (2006). Evidence of transmission ratio distortion of DLG5 R30Q variant in general and implication of an association with Crohn disease in men. *Hum. Genet.* 119, 305–311. doi: 10.1007/s00439-006-0133-1

Gjessing, H. K., and Lie, R. T. (2006). Case-parent triads: estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. *Ann. Hum. Genet.* 70(Pt 3), 382–396. doi: 10.1111/j.1529-8817.2005.00218.x

Hanchard, N., Rockett, K., Udalova, I., Wilson, J., Keating, B., Koch, O., et al. (2005). An investigation of transmission ratio distortion in the central region of the human MHC. *Genes Immun.* 7, 51–58. doi: 10.1038/sj.gene.6364277

Hastings, I. M. (1991). Germline selection: population genetic aspects of the sexual/asexual life cycle. *Genetics* 129, 1167–1176.

Hu, Y. Q., Zhou, J. Y., and Fung, W. K. (2007). An extension of the transmission disequilibrium test incorporating imprinting. *Genetics* 175, 1489–1504. doi: 10.1534/genetics.106.058461

Huang, L. O., Labbe, A., and Infante-Rivard, C. (2013). Transmission ratio distortion: review of concept and implications for genetic association studies. *Hum. Genet.* 132, 245–263. doi: 10.1007/s00439-012-1257-0

Imboden, M., Swan, H., Denjoy, I., Van Langen, I. M., Latinen-Forsblom, P. J., Napolitano, C., et al. (2006). Female predominance and transmission distortion in the long-QT syndrome. *N. Engl. J. Med.* 355, 2744–2751. doi: 10.1056/NEJMoa042786

Infante-Rivard, C., Rivard, G. E., Guiguet, M., and Gauthier, R. (2005). Thrombophilic polymorphisms and intrauterine growth restriction. *Epidemiology* 16, 281–287. doi: 10.1097/01.ede.0000158199.64871.b9

Infante-Rivard, C., Rivard, G. E., Yotov, W. V., Génin, E., Guiguet, M., Weinberg, C., et al. (2002). Absence of association of thrombophilia polymorphisms with intrauterine growth restriction. *N. Engl. J. Med.* 347, 19–25. doi: 10.1056/NEJM200207043470105

Infante-Rivard, C., and Weinberg, C. R. (2005). Parent-of-origin transmission of thrombophilic alleles to intrauterine growth-restricted newborns and

transmission-ratio distortion in unaffected newborns. *Am. J. Epidemiol.* 162, 891–897. doi: 10.1093/aje/kwi293

Kawamura, Y., Endo, K., Koizumi, M., Watanabe, Y., Saga, T., Konishi, J., et al. (1989). Gadolinium-phthalein complexone as a contrast agent for hepatobiliary MR imaging. *J. Comput. Assist. Tomogr.* 13, 67–70. doi: 10.1097/00004728-198901000-00014

Kistner, E. O., Infante-Rivard, C., and Weinberg, C. R. (2006). A method for using incomplete triads to test maternally mediated genetic effects and parent-of-origin effects in relation to a quantitative trait. *Am. J. Epidemiol.* 163, 255–261. doi: 10.1093/aje/kwj030

Kistner, E. O., Shi, M., and Weinberg, C. R. (2009). Using cases and parents to study multiplicative gene-by-environment interaction. *Am. J. Epidemiol.* 170, 393–400. doi: 10.1093/aje/kwp118

Kvasnicka, J., Hájková, J., Bobciková, P., Kvasnicka, T., Duskova, D., Poletinová, S., et al. (2012). [Prevalence of thrombophilic mutations of FV Leiden, prothrombin G20210A and PAl-1 4G/5G and their combinations in a group of 1450 healthy middle-aged individuals in the Prague and Central Bohemian regions (results of FRET real-time PCR assay)]. *Cas. Lek. Cesk.* 151, 76–82.

Labbe, A., Huang, L., and Infante-Rivard, C. (2013). "Transmission ratio distortion: a neglected phenomenon with many consequences in genetic analysis and population genetics," in *Epigenetics and Complex Traits*, eds A. K. Naumova and C. M. T. Greenwood (New York, NY; Heidelberg; Dordrecht; London: Springer), 265–285.

Lazzeroni, L. C., and Lange, K. (1998). A conditional inference framework for extending the transmission/disequilibrium test. *Hum. Hered.* 48, 67–81. doi: 10.1159/000022784

Martin, E. R., Kaplan, N. L., and Weir, B. S. (1997). Tests for linkage and association in nuclear families. *Am. J. Hum. Genet.* 61, 439–448. doi: 10.1086/514860

Meyer, W. K., Arbeithuber, B., Ober, C., Ebner, T., Tiemann-Boege, I., Hudson, R. R., et al. (2012). Evaluating the evidence for transmission distortion in human pedigrees. *Genetics* 191, 215–232. doi: 10.1534/genetics.112.139576

Naumova, A. K., Greenwood, C. M., and Morgan, K. (2001). Imprinting and deviation from Mendelian transmission ratios. *Genome* 44, 311–320. doi: 10.1139/g01-013

Pardo-Manuel de Villena, F., and Sapienza, C. (2001). Nonrandom segregation during meiosis: the unfairness of females. *Mamm. Genome* 12, 331–339. doi: 10.1007/s003350040003

Paterson, A. D., Petronis, A. (1999). Transmission ratio distortion in females on chromosome 10p11 p15. *Am. J. Med. Genet.* 88, 657–661. doi: 10.1002/(SICI)1096-8628(19991215)88:6<657::AID-AJMG15>3.0.CO;2-#

Paterson, A. D., Waggott, D., Schillert, A., Infante-Rivard, C., Bull, S. B., Yoo, Y. J., et al. (2009). Transmission-ratio distortion in the Framingham Heart Study. *BMC Proc.* 3(Suppl. 7):S51. doi: 10.1186/1753-6561-3-s7-s51

Paterson, A. D., Sun, L., and Liu, X. Q. (2003). Transmission ratio distortion in families from the Framingham Heart Study. *BMC Genet.* 4(Suppl. 1):S48. doi: 10.1186/1471-2156-4-S1-S48

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. 505, 43–49. doi: 10.1038/nature12886

Rabinowitz, D. (1997). A transmission disequilibrium test for quantitative trait loci. *Hum. Hered.* 47, 342–350. doi: 10.1159/000154433

Rabinowitz, D., and Laird, N. (2000). A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum. Hered.* 50, 211–223. doi: 10.1159/000022918

Sapru, A., Hansen, H., Ajayi, T., Brown, R., Garcia, O., Zhuo, H., et al. (2009). 4G/5G polymorphism of plasminogen activator inhibitor-1 gene is associated with mortality in intensive care unit patients with severe pneumonia. *Anesthesiology* 110, 1086–1091. doi: 10.1097/ALN.0b013e3181a1081d

Schaid, D. J. (1996). General score tests for associations of genetic markers with disease using cases and their parents. *Genet. Epidemiol.* 13, 423–449. doi: 10.1002/(SICI)1098-2272(1996)13:5<423::AID-GEPI1>3.0.CO;2-3

Sham, P. C., and Curtis, D. (1995). An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann. Hum. Genet.* 59(Pt 3), 323–336. doi: 10.1111/j.1469-1809.1995.tb00751.x

Shoubridge, C., Gardner, A., Schwartz, C. E., Hackett, A., Field, M., and Gecz, J. (2012). Is there a Mendelian transmission ratio distortion of the

c.429_452dup(24bp) polyalanine tract ARX mutation? *Eur. J. Hum. Genet.* 20, 1311–1314. doi: 10.1038/ejhg.2012.61

Sinsheimer, J. S., Palmer, C. G., and Woodward, J. A. (2003). Detecting genotype combinations that increase risk for disease: maternal-fetal genotype incompatibility test. *Genet. Epidemiol.* 24, 1–13. doi: 10.1002/gepi.10211

Spielman, R. S., and Ewens, W. J. (1998). A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am. J. Hum. Genet.* 62, 450–458. doi: 10.1086/301714

Spielman, R. S., McGinnis, R. E., and Ewens, W. J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* 52, 506.

T2D-GENES Consortium TD-G (2016). *Type 2 Diabetes Genetic Exploration by Next-Generation Sequencing in Multi-Ethnic Samples (T2D-GENES) Consortium.*

The Cancer Genome Atlas (2016). *Data Repository: TCGA Data Portal.* Available online at: https://tcga-data.nci.nih.gov/docs/publications/tcga/

The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature* 437, 1299–1320. doi: 10.1038/nature04226

Ulvik, A., Ren, J., Refsum, H., and Ueland, P. M. (1998). Simultaneous determination of methylenetetrahydrofolate reductase C677T and factor V G1691A genotypes by mutagenically separated PCR and multiple-injection capillary electrophoresis. *Clin. Chem.* 44, 264–269.

van Den Oord, E. J., and 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. *Am. J. Hum. Genet.* 66, 335–338. doi: 10.1086/302708

Weinberg, C. R. (1999). Methods for Detection of Parent-of-Origin Effects in Genetic Studies of Case-Parents Triads. *Am. J. Hum. Genet.* 65, 229–235. doi: 10.1086/302466

Weinberg, C. R., Wilcox, A. J., and 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. *Am. J. Hum. Genet.* 62, 969–978. doi: 10.1086/301802

Wilson, S. R. (1997). On extending the transmission/disequilibrium test (TDT). *Ann. Hum. Genet.* 61(Pt 2), 151–161. doi: 10.1017/S0003480097006040

Xiong, M. M., Krushkal, J., and Boerwinkle, E. (1998). TDT statistics for mapping quantitative trait loci. *Ann. Hum. Genet.* 62(Pt 5), 431–452. doi: 10.1046/j.1469-1809.1998.6250431.x

Yang, L., Andrade, M. F., Labialle, S., Moussette, S., Geneau, G., Sinnett, D., et al. (2008). Parental effect of DNA (Cytosine-5) methyltransferase 1 on grandparental-origin-dependent transmission ratio distortion in mouse crosses and human families. *Genetics* 178, 35–45. doi: 10.1534/genetics.107.081562

Zöllner, S., Wen, X., Hanchard, N. A., Herbert, M. A., Ober, C., and Pritchard, J. K. (2004). Evidence for extensive transmission distortion in the human genome. *Am. J. Hum. Genet.* 74, 62–72. doi: 10.1086/381131