Human Genetics (2020) 139:1107–1117
https://doi.org/10.1007/s00439-020-02162-2

ORIGINAL INVESTIGATION

Joint modeling of eQTLs and parent-of-origin effects using an orthogonal framework with RNA-seq data

Shirong Deng[1] · James Hardin[2] · Christopher I. Amos[3] · Feifei Xiao[2]

Received: 8 October 2019 / Accepted: 1 April 2020 / Published online: 8 April 2020

Abstract

Extensive studies have been conducted on the analysis of genome function, especially on the expression quantitative trait loci (eQTL). These studies offered promising results for characterization of the functional sequencing variation and understanding of the basic processes of gene regulation. Parent of origin effect (POE) is an important epigenetic phenomenon describing that the expression of certain genes depends on their allelic parent-of-origin and it is known to play important roles in human complex diseases. However, traditional eQTL mapping approaches do not allow for the detection of imprinting, or they focus on modeling the additive genetic effect thereby ignoring the estimation of the dominance genetic effect. In this study, we proposed a statistical framework to test the additive and dominance genetic effects of the candidate eQTLs along with detection of the POE with a functional model and an orthogonal model for RNA-seq data. We demonstrated the desirable power and preserved Type I errors of the methods in most scenarios, especially the orthogonal model with un-biased estimation of the genetic effects and over-dispersion of the RNA-seq data. The application to a HapMap project trio dataset validated existing imprinting genes and discovered two novel imprinting genes with potential dominance genetic effect and *RB1* and *IGF1R* genes. This study provides new insights into the next generation statistical modeling of eQTL mapping for better understanding of the genetic architecture underlying the mechanisms of gene expression regulation.

Introduction

With the completion the 1000 Genomes Project (Genomes Project et al. 2015), an unprecedented wealth of knowledge has been accumulated for understanding the variations at the human DNA level. However, little of this DNA-level knowledge has been translated into understanding the mechanisms of human diseases. Gene expression quantitative trait locus (eQTL) mapping is one of the most promising approaches to fill this gap, which aims to explore the genetic basis of gene expression (Cookson et al. 2009). Among the eQTL techniques, *cis*-eQTL mapping is the most commonly used technique to map local eQTLs on the same chromosome of the gene. To date, many statistical methods for eQTL mapping have been developed, however, the modeling of imprinting is typically ignored in these methods.

Imprinting is a type of parent-of-origin effect (POE) that the expression of certain genes depends on their allelic parent-of-origin. As such, the same alleles transmitted from the mother have different expression levels on transcripts compared with those transmitted from the father. Consequently, the influences on the phenotype between the two types of heterozygotes are different, as so-called parent-of-origin effect (POE). There are at least 80 imprinted genes discovered in humans, many of which are involved in embryonic and placental growth and development (Perry et al. 2014). Studies have suggested that POE is an important contributor to phenotypic variation in human complex diseases and may explain some of the "hidden" heritability. An earlier study showed that for type II diabetes, a variant of SNP rs2334499 in chromosome region 11p15 was protective when maternally transmitted, whereas it conferred risk when paternally transmitted (Kong et al. 2009). Important roles of POEs

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00439-020-02162-2) contains supplementary material, which is available to authorized users.

✉ Feifei Xiao
xiaof@mailbox.sc.edu

[1] School of Mathematics and Statistics, Wuhan University, Wuhan, China

[2] Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Room 449, 915 Greene St., Columbia, SC 29201, USA

[3] Department of Quantitative Sciences, Baylor College of Medicine, Houston, TX 77030, USA

are also implicated in type I diabetes, breast cancer and other carcinomas (Kong et al. 2009; Wallace et al. 2010). In the past few years, there were few approaches that modeled POEs while searching for eQTLs with RNA-seq data. The only report was from a study recently conducted by Zhabotynsky et al. which proposed to jointly model genetic effect and POE focusing on modeling the allele specific expression (ASE) (Zhabotynsky et al. 2019).

In recent years, the inclusion of dominance in animal genomic models has been proposed by several researchers (Duenk et al. 2017; Ertl et al. 2014; Su et al. 2012; Xiang et al. 2018). From the theory of quantitative genetics, statistical additive genetic effects are obtained from average allele substitution effects, whereas dominance genetic effects reflect the deviation of the genotypic values of the heterozygotes and the expected midpoint of the two homozygotes. In quantitative genetics, the partition of the variance in statistical components is due to additivity. Dominance does not reflect the biological effect of the genes, but it is most useful for prediction, selection, and evolution (Huang and Mackay 2016).

Multi-collinearity, however, is an important issue arising from modeling multiple genetic effects. To achieve straightforward model selection and variance component analysis, uncorrelated estimation of the additive and dominance effects is necessary. To achieve this goal, in our study, we developed an orthogonal model to jointly evaluate the effect from both additive and dominance genetic effects along with the detection of POE in eQTL mapping for RNA sequencing read count data. To evaluate gene expression levels, RNA sequencing (RNA-seq) technology has recently become a widely used high-throughput technology to assess the gene expression abundance, especially in discovery of novel eQTLs (Ellis et al. 2013).

Genetic imprinting affects complex diseases through regulating the gene expression and can reveal an important component of heritable variation that remains "hidden" in traditional complex trait studies. In this study, we hypothesized that POEs contribute to regulating gene expression along with the main allelic effect (i.e., additive and dominance effects) from the gene. Accordingly, we developed a statistical framework to test the main allelic effects of the candidate eQTLs along with the detection of POE with a natural model and an orthogonal model. Intensive simulations were conducted to evaluate the methods. We also applied the methods to an existing HapMap project trio dataset to validate the reported imprinting genes and identify novel *cis*-eQTLs for these genes.

Methods

The stat-POE and func-POE methods

Our methods were developed from a basic model of eQTL mapping of a single gene with RNA-seq data that are read counts. Therefore, we consider a single gene and study the association of its expression with the jth candidate eQTL. Let y_i be the total read counts mapped to this gene in the ith sample, where $i = 1, \ldots, n$ and n is the sample size. We model y_i using the negative binomial(NB) distribution as they are sparse count data. The NB distribution allows over-dispersion (the variance exceeds the mean) estimation. Let $f_{\mathrm{NB}}(y_i;\mu_i,\phi)$ be the probability mass function for a NB distribution with mean μ_i and dispersion parameter ϕ:

$$f_{\mathrm{NB}}(y_i;\mu_i,\phi) = \frac{\Gamma(y_i+1/\phi)}{y_i!\Gamma(1/\phi)}\left(\frac{1}{1+\phi\mu_i}\right)^{1/\phi} \times \left(\frac{\phi\mu_i}{1+\phi\mu_i}\right)^{y_i}, y_i = 0,1,2,\ldots \quad (1)$$

where $\Gamma(\cdot)$ is the gamma function. It's easy to find that the variance $\mathrm{Var}(y_i) = \mu_i + \phi\mu_i^2$, in which $\phi\mu_i^2$ is the over-dispersion part. As the over-dispersion parameter ϕ converges to 0, $f_{NB}(y_i;\mu_i,\phi)$ converges to $f_p(y_i;\mu_i) = \mu_i^{y_i}e^{-\mu_i}/y_i!$, which is the probability mass function for Poisson distribution with mean parameter μ_i. Let $\mathbf{X}_i$ be a set of p covariates and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ be the regression coefficients, and $\boldsymbol{\beta}_G = (R, a, d, l)'$ be the genetic effects from genotypes (G) of the eQTL on Y, where R is the baseline, a, d and l are the additive, dominance and imprinting effects from G, respectively. The covariate effect of $G = G_i$ and covariates $\mathbf{X} = \mathbf{x}_i$ on the gene expression, can be formulated through the following log-linear regression model

$$\log(\mu_i) = \boldsymbol{\beta}'\mathbf{x}_i + \omega(G_i, \boldsymbol{\beta}_G), \quad (2)$$

where $\omega(G_i, \boldsymbol{\beta}_G)$ is the function reflecting the genetic effects.

For a bi-allelic locus, let the major and minor alleles of the jth candidate eQTL as A_1 and A_2, respectively. The genotype G takes four possible values $\vec{A}_1A_1$,$\vec{A}_1A_2$, $\vec{A}_2A_1$ and $\vec{A}_2A_2$, the first allele of which with arrow denotes the paternal allele and the second allele denotes the one originated from the maternal side. We use p_{11}, p_{12}, p_{21} and p_{22} to denote genotype frequencies in the population, and use M to denote the number of variant allele A_2, which takes values of 0, 1, 1 and 2 for the four genotypes separately. $\overline{M} = 1 + p_{22} - p_{11}$ and $V = (p_{11} + p_{22}) - (p_{11} - p_{22})^2$ are the mean and variance of M.

For estimation of the genetic effects, there are different methods we can epress the genetic effect function $\omega(G_i, \boldsymbol{\beta}_G)$. Early in 2013, we proposed a unified orthogonal framework to model genetic variants displaying imprinting effects for quantitative traits (Xiao et al. 2013). We proposed two related methods for identifying genetic variants influences on quantitative traits with different characteristics, the statistical and functional POE methods. The statistical POE method in Xiao et al. (2013) was claimed to be partially orthogonal and allows for imprinting effect detection while

maintaining sufficient power for main allelic effects (i.e., the additive and dominance effects) in certain conditions. Motivated by that study, we here develop $\omega(G_i, \boldsymbol{\beta}_G)$ in a population-referenced formulation with an orthogonal model, termed Stat-POE model, generated from reparameterization procedure:

$$\omega(G_i, \boldsymbol{\beta}_G) = \begin{cases} R - \overline{M}a - 2p_{22}(p_{12}+p_{21})d/V, & \text{if } G_i = \vec{A}_1A_1 \\ R + \left(1-\overline{M}\right)a + 4p_{11}p_{22}d/V - 2p_{21}l/(p_{12}+p_{21}), & \text{if } G_i = \vec{A}_1A_2 \\ R + \left(1-\overline{M}\right)a + 4p_{11}p_{22}d/V + 2p_{12}l/(p_{12}+p_{21}), & \text{if } G_i = \vec{A}_2A_1 \\ R + \left(2-\overline{M}\right)a - 2p_{11}(p_{12}+p_{21})d/V, & \text{if } G_i = \vec{A}_2A_2 \end{cases} \tag{3}$$

With the orthogonality property, this model allows for uncorrelated estimation of the genetic effects including the additive, dominance and imprinting effects. Such orthogonal model also enables straightforward model comparison with nested genetic models (Alvarez-Castro and Carlborg 2007). Note that we continued to use the same terminology of the Stat-POE model as what we used in Xiao et al., 2013, although the Stat-POE model in this study as shown in Eq. (3) is a newly proposed model that is fully orthogonal.

For a functional model without the orthogonalization property but with a POE component, the genetic effect function $\omega(G_i, \boldsymbol{\beta}_G)$ can be expressed as

$$\omega(G_i, \boldsymbol{\beta}_G) = \begin{cases} R & \text{if } G_i = \vec{A}_1A_1 \\ R + a + d - l, & \text{if } G_i = \vec{A}_1A_2 \\ R + a + d + l, & \text{if } G_i = \vec{A}_2A_1 \\ R + 2a, & \text{if } G_i = \vec{A}_2A_2 \end{cases} \tag{4}$$

from which we obtain

$$\begin{aligned} a &= \tfrac{1}{2}\log\left(\mu_{i\vec{A}_2A_2}/\mu_{i\vec{A}_1A_1}\right), \\ d &= \tfrac{1}{2}\log\left(\frac{\mu_{i\vec{A}_1A_2}+\mu_{i\vec{A}_2A_1}}{\mu_{i\vec{A}_2A_2}+\mu_{i\vec{A}_1A_1}}\right), \\ l &= \tfrac{1}{2}\log\left(\mu_{i\vec{A}_2A_1}/\mu_{i\vec{A}_1A_2}\right), \end{aligned} \tag{5}$$

where $\mu_{iA_1A_1}, \mu_{iA_1A_2}, \mu_{iA_2A_1}$ and $\mu_{iA_2A_2}$ are the underlying means of the read counts for subjects with the four genotypes, respectively. The additive effect a measures the average fold change of gene expression between the two homozygotes; the dominance effect d measures the deviation of the heterozygotes from its additive expectation; and the imprinting effect l reflects the different effect from the two types of heterozygotes. Following the notations in Alvarez-Castro and Carlborg (2007), the model in Eq. (5) is defined as a functional POE (Func-POE) model, or a natural model since it uses natural effects of allele substitutions as parameters, mainly focusing on the biological properties (Alvarez-Castro and Carlborg 2007).

The general orthogonal and functional models have presented different properties for various application scopes in detecting epistasis, gene environment interactions and parent-of-origin effects in quantitative traits and qulitative traits (Ma et al. 2012; Xiao et al. 2013, 2014).

Parameter estimation and hypothesis testing

To estimate the genetic effects and POE, we can write the likelihood based on the data $(y_i, X_i, G_i)\,(i = 1, 2, ...N)$ as

$$\begin{aligned} &L(\boldsymbol{\beta}, \boldsymbol{\beta}_G, \phi; \{y_i, \mathbf{x}_i\}_{i=1}^N) \\ &= \prod_{i=1}^{N} f_{NB}(y_i; \mu_i(\boldsymbol{\beta}, \boldsymbol{\beta}_G), \phi)^{I_{NB}(y_1,\ldots,y_N)} \\ &\quad f_P(y_i; \mu_i(\boldsymbol{\beta}, \boldsymbol{\beta}_G))^{1-I_{NB}(y_1,\ldots,y_N)}, \end{aligned} \tag{6}$$

where $I_{NB}(y_1, \ldots, y_N)$ is an indicator function which is equal to 1 if a negative binomial distribution is used and 0 if a Poisson distribution is used.

With Eq. 2.4, the maximum likelihood estimator (MLE) of the model parameters $(\boldsymbol{\beta}_\phi, \boldsymbol{\beta}_G)$ with $\boldsymbol{\beta}_\phi \triangleq (\boldsymbol{\beta}', \phi)'$ can be estimated by the following iterative procedure.

1. Initialization: we first fit a null model using Poisson regression using the covariate $\mathbf{X}_i$, and estimate $\boldsymbol{\beta}$, using a Newton–Raphson optimization algorithm based on formulas given in Appendix A.1. Subsequently, a score test is conducted for the over-dispersion parameter ϕ where the hypothesis testing procedure is illustrated in Appendix A.2. If the p value of the score test is smaller than a cutoff value, e.g., $\alpha = 0.05$, we estimate a negative binomial regression model for which the regression parameters are denoted $\boldsymbol{\beta}_\phi$. The details of the iterative formulas for estinating $\boldsymbol{\beta}$ and ϕ are given in (A.10) and (A.11) in Appendix A.3, which are based on the iteratively re-weighted least squares method (Green 1984) and the Newton–Raphson iterative method, respectively.
2. Iteration: (a) given $\boldsymbol{\beta}$ or $\boldsymbol{\beta}_\phi$, we estimate $\boldsymbol{\beta}_G$ by the Newton–Raphson method illustrated in Appendix B; (b). Given $\boldsymbol{\beta}_G$, we estimate $\boldsymbol{\beta}$ by a Poisson regression with offsets $\omega(G_i, \boldsymbol{\beta}_G)$, or estimate $\boldsymbol{\beta}_\phi$ by a negative binomial regression with offsets $\omega(G_i, \boldsymbol{\beta}_G)$. The estimation for $\boldsymbol{\beta}$ under the Poisson regresion is the same as that in the initialization step with the first and second derivatives

given in Appendix B.6 for the Stat-POE model and Appendix B.8 for the Func-POE model, respectively. Under the negative binomial regression, the estimation method for $\boldsymbol{\beta}_{\phi}$ described in the the initialization step is also used here with the detailed forlumas given in Appendix B.5 for the Stat-POE model and in Appendix B.7 for the Func-POE model, respectively.

3. Termination: until iterate steps (1) and (2) estimate of all the parameters converge.

To assess whether each covariate in the model is significant on the read counts of the gene expression, statistical hypothesis testing will be performed. We constructed three testing methods including the likelihood ratio test (LRT), score test and Wald test as follows. For example, the of additive effect was tested using the hypotheses

$$H_0 : a = 0 \text{ vs } H_1 : a \neq 0. \tag{7}$$

Denote $\boldsymbol{\theta} = (\boldsymbol{\beta}_{\phi}^{'}, \boldsymbol{\beta}_{G}^{'})'$ for the Negative Binomial (NB) regression, or $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\beta}_{G}^{'})'$ for the Poisson regression, the unrestricted MLE and restricted MLE under (Appendix D.1) obtained by the algorithm given in the above section are denoted as $\widehat{\boldsymbol{\theta}}$ and $\hat{\tilde{\theta}}$, respectively. Without loss of generality, we put the parameter a in the first position of $\boldsymbol{\theta}$ and denote the other parameters as ξ, i.e. $\boldsymbol{\theta} = (a, \xi')'$. Then the score function for $\boldsymbol{\theta}$ is $U(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial l(\boldsymbol{\theta})}{\partial a} \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \xi} \end{bmatrix}$, and the expected fisher information matrix is $I(\boldsymbol{\theta}) = -\mathrm{E}\begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial a^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial a \partial \xi'} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \xi \partial a} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial a^2} \end{bmatrix} \triangleq \begin{bmatrix} I_{aa}(\boldsymbol{\theta}) & I_{a\xi}(\boldsymbol{\theta}) \\ I_{\xi a}(\boldsymbol{\theta}) & I_{\xi\xi}(\boldsymbol{\theta}) \end{bmatrix}$, where $l(\boldsymbol{\theta})$ is the log-likelihood function given as in Appendix B.1 for NB regression and the Stat-POE model, in Appendix B.2 for Poisson regression and statistical model, in Appendix B.3 for NB regression and the Func-POE model, and in Appendix B.4 for Poisson regression and the Func-POE model. The formulas of $U(\boldsymbol{\theta})$ and $I(\boldsymbol{\theta})$ are given in Appendix C. The LRT statistic is

$$T_L = 2[l(\widehat{\boldsymbol{\theta}}) - l()]. \tag{8}$$

According to the theory from Rao (2005), in our statistical setting, the score test statistic is defined as

$$T_S = \left(\frac{\partial l(\boldsymbol{\theta})}{\partial a}\right)^2 |_{\hat{\theta}} J_{aa}(\boldsymbol{\theta})|_{\hat{\theta}}. \tag{9}$$

where $J_{\mathrm{aa}}(\boldsymbol{\theta}) = (I_{aa}(\boldsymbol{\theta}) - I_{a\xi}(\boldsymbol{\theta}) I_{\xi\xi}^{-1}(\boldsymbol{\theta}) I_{\xi a}(\boldsymbol{\theta}))^{-1}$.

Moreover, the Wald test statistic is defined by:

$$T_W = \frac{\widehat{a}^2}{J_{aa}(\boldsymbol{\theta})|_{\widehat{\theta}}}. \tag{10}$$

Under H_0, the statistics T_L, T_S, and T_W all converge to χ^2_{1-} distributions. For a given significance level α, we reject H_0 when the observed value of the statistics are greater than $\chi^2_{1,1-\alpha}$. The process of the hypothesis testing for the other parameters can be implemented in a similar manner.

Simulations

To evaluate the performance of the proposed statistical methods in eQTL mapping with RNA-seq data, we carried out extensive simulation studies in realistic settings. First, we compared the statistical power of the Stat-POE and Func-POE methods in detecting the main allelic effects (i.e., the additive and dominance effects) and POE. We simulated y_i, the total number of read counts of a gene in the ith sample as being generated from a negative binomial distribution with $\mu_{iG} = \exp\left(0.1x_i + \omega\left(G_i, \beta_G\right)\right)$. The over-dispersion parameter = 0.2 and the covariate $\mathbf{X}$ was a continuous variable $\mathbf{X} \sim \mathrm{N}(0,1)$. To evaluate the performance of the methods in estimating both genetic effects and over-dispersion, we generated data with different sample sizes $N = 50$, 100, 200 and 500, respectively. Hardy–Weinberg Equilibrium (HWE) proportion was used so that the genotype frequencies in the samples were set at $[p_{11}, p_{12}, p_{21}, p_{22}] = [0.36, 0.24, 0.24, 0.16]$. In addition to the main scenarios of HWE proportions, non-HWE genotype proportions were also simulated that the proportions of two heterozygotes were different, $[p_{12}, p_{21}] = [0.20, 0.28]$ or $[p_{12}, p_{21}] = [0.28, 0.20]$. The over-dispersion parameter was set at empirical values that was 0.2 or 0.5. The additive effect α and dominance effect δ were both fixed at values of log (1.2) where the values of 1.2 reflected the fold change of the logarithm mean shift of the genotypic values, referring to Eq. (5). The POE parameter ι was set at log (1.1) or log (1.2), respectively.

Each simulation was replicated 500 times to evaluate the performance of the Stat-POE and Func-POE methods. Relative bias and mean square of errors (MSE) were calculated for each parameter in the different scenarios to evaluate the estimation accuracy. The estimation relative bias was defined as the difference between estimated value and the true parameter value and then divided by the true parameter value. We also used simulated data to quantify the statistical power and Type I error rates of the methods. To illustrate the performance of the proposed methods in detecting genetic effect terms and POE, the statistical power was calculated using a range of different critical values. Type I error was calculated under the null model where there was no genetic effect or POE for the three testing methods, the LRT, Wald and score tests.

Application to a HapMap RNA-seq dataset

Datasets

We used an RNA-seq dataset from 30 HapMap Caucasian samples obtained from the NCBI Bioproject (PRJNA385599). The samples were collected from lymphoblastoid cell lines from 15 males and 15 females. For most of these samples, the RNA reads were 150 bp paired-end reads, with an additional run with 75 bp paired-end reads. The median of the total number of reads for these 30 samples was approximately 20 million. All of these reads were mapped to hg38 human reference genome using Tophat2 (Zhabotynsky et al. 2019).

Since all of these samples were from children of family trios, the parents of these children were also part of the samples included in the 1000 Genomes Project (2012). For these 30 trios, the HapMap project genotyped about 3.9 million SNPs. Genotyping data of the 30 trios were used to obtain the phased genotype of the children. The phasing and imputation of these 30 trios were conducted by Zhabotynsky et al.'s study, from where the phased and imputed genotypes in our study were directly obtained (Zhabotynsky et al. 2019). Briefly, SHAPEIT2 (Delaneau et al. 2014) was used for phasing and IMPUTE2 (Howie et al. 2012) was used for imputation against the 1000 Genome reference panel containing 2504 individuals and ~ 82 million SNPs. Based on the phased and imputed SNPs, we had 6,211,048 imputed SNPs of high confidence in total, the ones with at least one heterozygote in the sample which were all informative.

Identification of imprinted genes and genes with dominance effect

We selected 22 known imprinted genes based on the list reported by a recent publication (Jadhav et al. 2019). These genes were selected because they had abundant expression in the 30 samples. The genes and related information are listed in Supplementary Table 1. For each potential imprinting gene, all SNPs in the gene coding region were defined as candidate *cis*-eQTLs. For each SNP and gene expression pair, the Stat-POE method was applied to detect candidate *cis*-eQTLs with additive, dominance and POE effects. Four covariates were adjusted in the model including the total read counts per individual and the first three principal components computed from the matrix of normalized expression to remove the effect from potential confounders. In all of the above hypothesis testing, the Benjamini–Hochberg (BH) method was used for multiple comparisons to adjust the p-values obtained from the LRTs (Benjamini and Hochberg 1995). We tested the POE of the previously reported imprinted genes in Supplementary Table 1 to evaluate the performance of our methods. For novel discovery of genetic effects of these potential imprinting genes, we tested the additive and dominance effects simultaneously.

Results

Simulations

The statistical power of the Stat-POE and Func-POE methods was investigated for POE at two scenarios with different levels of POE: (a) a small POE with $\iota = \log(1.1)$ and (b) a moderate POE with $\iota = \log(1.2)$. The results are shown in Figs. 1, 2 for these two scenarios, respectively. In both scenarios, the additive and dominance effects were fixed at log (1.2). In the simulations, to demonstrate the desirable performance of the Stat-POE method when the effect size was relatively too small to detect, we evaluated the methods at a fixed and small fold change in both overall allelic effects (i.e., additive and dominance effects). Consequently, even at a sample size of 50 with moderate over-dispersion ($\phi = 0.2$), the Stat-POE method presented around 70.8% power to detect genetic effect at 1.2-fold change in additive effect, corresponding to an effect size of $\log(1.2) = 0.18$ (Fig. 1). To detect POE at the fold change of 1.2, the Stat-POE and Func-POE methods both reached a statistical power of 83% with a reasonable sample size of 100 (Fig. 2). Even with a very small effect size from POE at a fold change of 1.1, corresponding to an effect size of $\log(1.1) = 0.10$, the methods yielded 61% power when the sample size was 200, and 91% when the sample size was 500 (Fig. 1). As expected, the Stat-POE method yielded the same power in detecting POE but a more desirable power in detecting main genetic effects compared to the Func-POE model (Figs. 1, 2). We also simulated the proportion of non-HWE that genotype frequencies of the two heterozygotes are unequal. The Stat-POE model always outperformed the Func-POE model in detecting additive effects, though it is not always the case for detection of the dominance effect (Supplementary Figs. 1, 2). In conclusion, the Stat-POE method outperformed the Func-POE method in most simulation scenarios and these two methods all achieved sufficient power for detection of POEs with a practical sample size for family data ($N = 100$).

With the simulated data, we also evaluated the estimation bias for all the parameters $(\beta, \beta_G,)$ estimated from the Stat-POE model. Table 1 shows that the estimation of all genetic effects achieved higher accuracy when sample size increased. Interestingly, the estimation of the covariate coefficient β and over-dispersion parameter ϕ was not notably affected by the sample sizes. Also, the estimation of genetic effects was not obviously affected by the value of the over-dispersion parameter. These results revealed the accurate

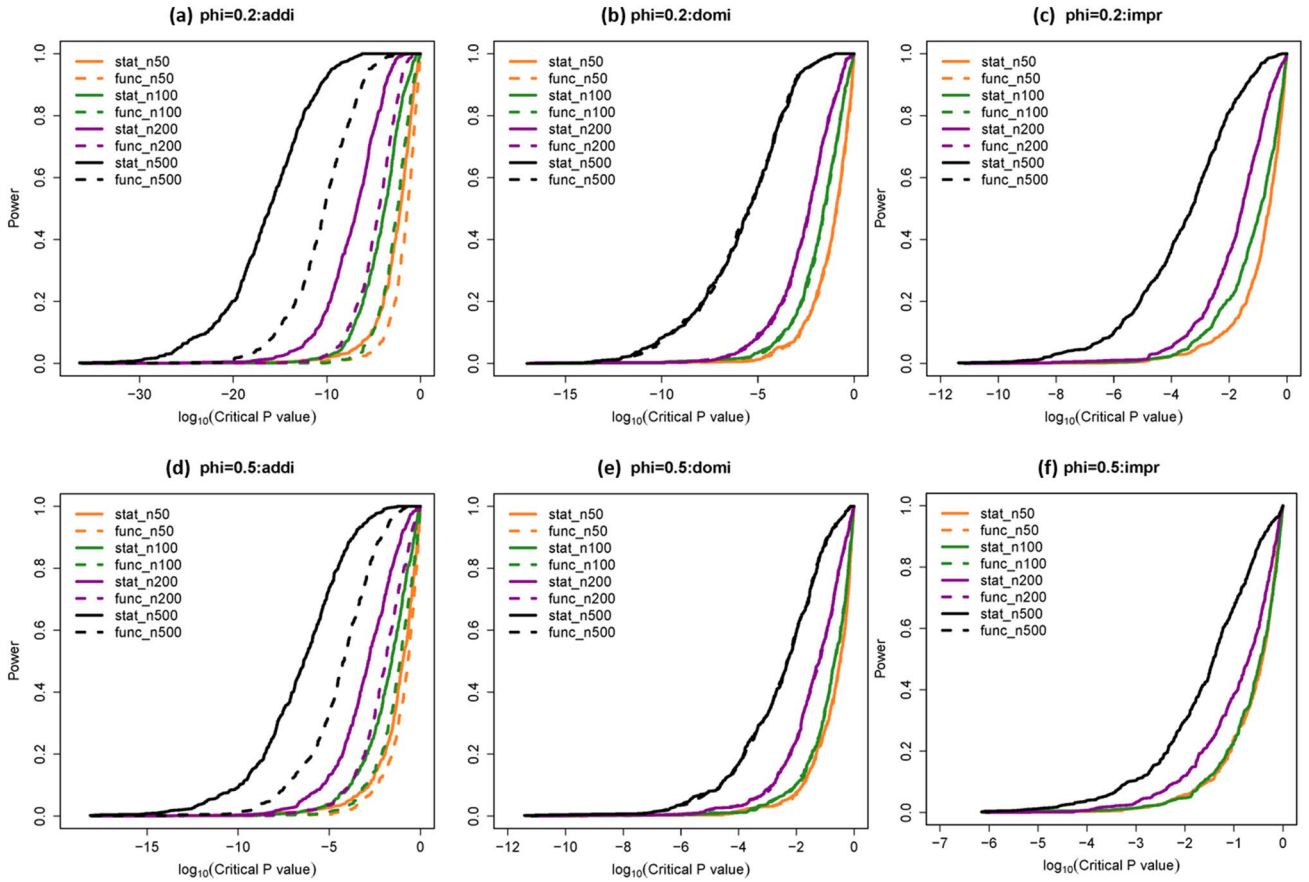

Fig. 1 Statistical power to detect additive, dominance and POE effect when **a**–**c** overdispersion $\varphi = 0.2$ and **d**–**e** $\varphi = 0.5$ for various samples sizes, using Stat-POE model (stat) or Func-POE model (func).The covariate coefficient $\beta = 0.1$, the sample size (n) was set at 50, 100, 200 and 500, respectively. *Addi* additive effect, *domi* dominant effect, *impr* imprinting effect, *stat* statistical model, *func* functional model. The additive effect $\alpha = \log(1.2)$, dominant effect $\delta = \log(1.2)$, imprinting effect $\iota = \log(1.1)$. The genotype frequencies in the samples were set at $[p_{11}, p_{12}, p_{21}, p_{22}] = [0.36, 0.24, 0.24, 0.16]$. Score test results are shown

and robust estimation of the covariates and over-dispersion parameters determined using the Stat-POE model. Moreover, large sample sizes and small over-dispersion ensured better overall performance of the proposed methods.

We observed the global trend of the type I error approaching the nominal level for all tests of both Stat-POE and Func-POE methods when sample size increases. The overall type I error rate of the LRT was closer to the nominal level than were the rates for the score and Wald tests (Table 2). Although there were slightly inflated false positives in detecting genetic effect and POEs when sample sizes were small, the type I error rates were close to nominal levels for relatively large sample sizes. Also, the score test achieved approximately equivalent performance with the LR tests given large sample sizes (for example, when $\phi = 0.5$, $N = 500$). Notably, the type I errors for detecting the genetic effects was comparable between the Stat-POE and Func-POE models in most scenarios.

Real data application to HapMap parent–child trio data

Using 30 children of the family trios from the HapMap project, we applied the proposed Stat-POE methods to estimate the additive and imprinting effects for 22 genes with previous evidence of imprinting. These selected genes were identified as imprinted genes using 296 phased trios from the 1000 Genomes Project and the Genome of the Netherlands participants (Jadhav et al. 2019).

With the proposed Stat-POE method, we identified 33 significant *cis*-eQTLs (with adjusted p values in additive

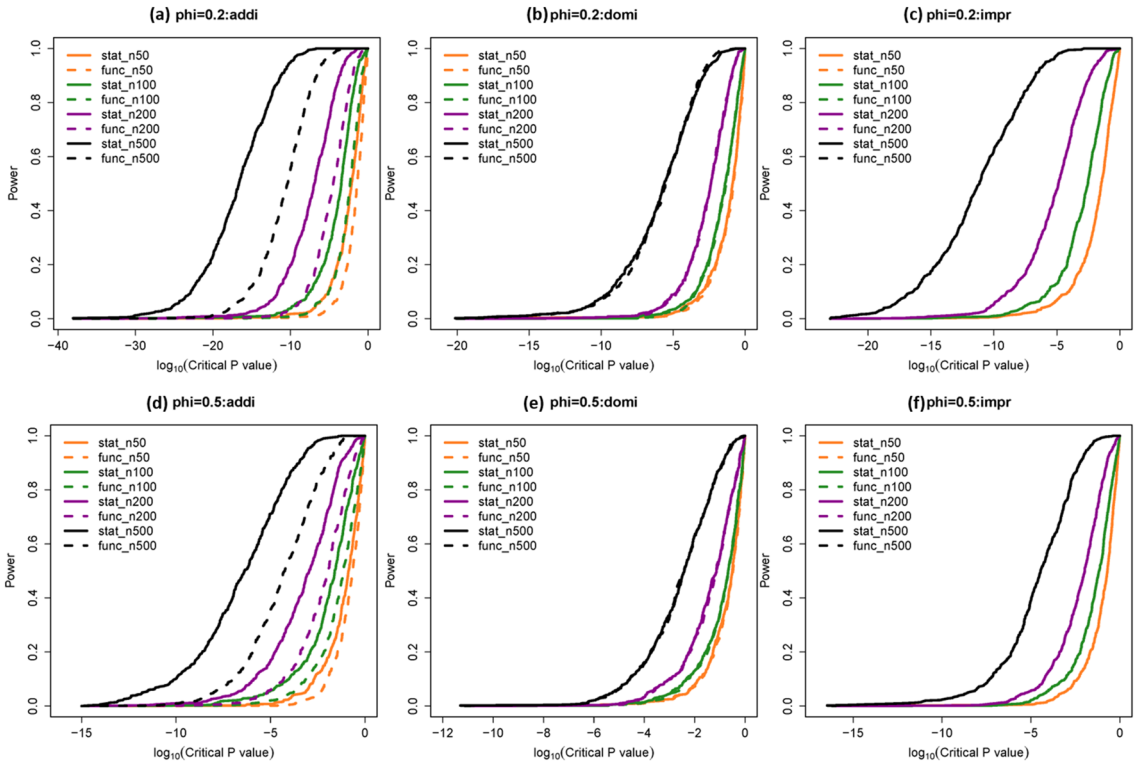

Fig. 2 Statistical power to detect additive, dominance and POE effect when **a**–**c** overdispersion $\varphi = 0.2$ and **d**–**e** $\varphi = 0.5$ for various samples sizes, using Stat-POE model (stat) or Func-POE model (func).The covariate coefficient $\beta = 0.1$, the sample size (n) was set at 50, 100, 200 and 500, respectively. *Addi* additive effect, *domi* dominant effect, *impr* imprinting effect, *stat* statistical model, *func* functional model. The additive effect $\alpha = \log(1.2)$, dominant effect $\delta = \log(1.2)$, imprinting effect $\iota = \log(1.2)$. The genotype frequencies in the samples were set at $[p_{11}, p_{12}, p_{21}, p_{22}] = [0.36, 0.24, 0.24, 0.16]$. Score test results were shown

Table 1 Simulation results with different sample sizes

Over-dispersion	True POE effect size	$N = 200$						$N = 500$					
		Bias			MSE			Bias			MSE		
		α	δ	ι	α	δ	ι	α	δ	ι	α	δ	ι
0.2	log(1.1)	− 0.98	1.96	− 0.47	0.21	0.44	0.20	1.15	− 0.83	2.35	0.09	0.18	0.07
0.2	log(1.2)	− 1.80	1.67	− 1.12	0.21	0.39	0.22	− 0.18	− 0.03	− 0.41	0.09	0.18	0.08
0.5	log(1.1)	− 3.48	− 1.57	− 6.65	0.49	1.08	0.58	− 2.47	0.46	− 1.84	0.20	0.40	0.19
0.5	log(1.2)	− 0.71	2.13	0.66	0.52	1.04	0.52	− 0.34	0.85	0.07	0.22	0.42	0.17

The estimation relative bias was defined as the difference between estimated value and true parameter value divided by the true parameter value. Relative bias and mean square of errors (MSE) have been multiplied by 100. for model parameters including additive effect (α), imprinting effect (ι), covariate (β), and dispersion parameter (φ)

effects < 0.05) for seven genes (*LPAR6*, *RB1*, *PXDC1*, *IGF1R*, *AC069277.2*, *IGF2BP3* and *SNRPN*) (Supplementary Table 2). Among them, most candidate *cis*-eQTLs presented maternal expression pattern in regulating the gene expression. In addition, we identified six genes with significant imprinting effects, which were *LPAR6*, *PER3*, *RB1*, *PXDC, IGF1R* and *IGF2BP3* with adjusted p values < 0.05 (Table 3). Among the significantly imprinting genes, the gene expression of *LPAR6* and *IGF1R* had significant regulation from the candidate *cis*-eQTLs rs11633209, rs728075 and rs7329291 in additive effect (adjusted p values $= 1.96 \times 10^{-66}$, 3.57×10^{-64} and 3.57×10^{-64}). Interestingly, we also discovered two novel genes that presented significant dominance effect in gene expression (Table 4), including *RB1* from multiple candidate *cis*-eQTLs (adjusted p value $= 3.02 \times 10^{-80}$) and *IGF1R* from SNP rs4965238

Table 2 Type I error of the methods in detecting genetic effects

Model		LRT			Wald test			Score test		
		α	δ	ι	α	δ	ι	α	δ	ι
$\phi = 0.2$										
$N = 50$	Stat	0.062	0.053	0.060	0.085	0.073	0.079	0.079	0.071	0.080
	Func	0.067	0.057	0.060	0.090	0.073	0.079	0.086	0.071	0.080
$N = 100$	Stat	0.047	0.044	0.053	0.056	0.063	0.068	0.058	0.061	0.069
	Func	0.051	0.042	0.053	0.064	0.064	0.068	0.059	0.061	0.069
$N = 200$	Stat	0.069	0.061	0.051	0.082	0.068	0.058	0.079	0.066	0.059
	Func	0.068	0.057	0.051	0.085	0.066	0.058	0.081	0.064	0.059
$N = 500$	Stat	0.046	0.047	0.050	0.058	0.059	0.068	0.056	0.060	0.069
	Func	0.051	0.048	0.050	0.065	0.060	0.068	0.062	0.060	0.069
$\phi = 0.5$										
$N = 50$	Stat	0.059	0.065	0.064	0.075	0.083	0.079	0.072	0.079	0.070
	Func	0.056	0.065	0.064	0.069	0.085	0.079	0.071	0.080	0.070
$N = 100$	Stat	0.055	0.049	0.056	0.061	0.063	0.063	0.061	0.059	0.065
	Func	0.052	0.048	0.056	0.060	0.059	0.063	0.062	0.053	0.065
$N = 200$	Stat	0.054	0.049	0.053	0.068	0.056	0.058	0.064	0.053	0.058
	Func	0.065	0.049	0.053	0.070	0.056	0.058	0.070	0.056	0.058
$N = 500$	Stat	0.058	0.056	0.045	0.062	0.062	0.052	0.062	0.062	0.051
	Func	0.057	0.058	0.045	0.064	0.062	0.052	0.063	0.061	0.051

Each simulation was replicated 1000 times

Stat Stat-POE model, *Func* Func-POE model, α additive effect, δ dominant effect, ι imprinting effect, ϕ dispersion parameter, *LRT* likelihood ratio test

Table 3 A list of genes with potential imprinting effects (p values < 0.05)

ID	Gene	Chr	SNP	Position	Allele	ATL	p value.a	p value.d	p value.l	Expression
ENSG00000049246	*PER3*	Chr1	rs67537516	7864693	G	A	0.96	1	0	Maternal
ENSG00000139687	*RB1*	Chr13	rs61949059	48923973	T	G	0.18	1	6.26E−82	Maternal
ENSG00000139687	*RB1*	Chr13	rs150098624	49001371	T	C	0.9	1	3.06E−80	Paternal
ENSG00000139687	*RB1*	Chr13	rs4258502	48989564	G	A	0.9	1	3.06E−80	Paternal
ENSG00000139687	*RB1*	Chr13	rs9568143	48996494	T	A	0.9	1	3.06E−80	Paternal
ENSG00000140443	*IGF1R*	Chr15	rs4965238	99456357	T	C	0.78	1.37E−67	1.12E−71	Paternal
ENSG00000140443	*IGF1R*	Chr15	rs11633209	99221205	A	T	1.96E−66	1	5.01E−71	Maternal
ENSG00000168994	*PXDC1*	chr6	rs113644426	3738592	C	T	0.98	1	4.14E−69	Paternal
ENSG00000139679	*LPAR6*	Chr13	rs4942796	49011882	T	C	0	1	5.81E−64	Maternal
ENSG00000139679	*LPAR6*	Chr13	rs728075	48999087	G	A	3.57E−64	1	5.81E−64	Maternal
ENSG00000139679	*LPAR6*	Chr13	rs7329291	48999148	G	A	3.57E−64	1	5.81E−64	Maternal
ENSG00000139679	*LPAR6*	Chr13	rs4942795	49003191	A	G	0	1	7.86E−64	Maternal
ENSG00000139679	*LPAR6*	Chr13	rs7998359	48983977	G	A	0	1	7.86E−64	Maternal
ENSG00000140443	*IGF1R*	Chr15	rs62023801	99295429	T	C	0.99	1	0.001	Maternal
ENSG00000140443	*IGF1R*	Chr15	rs6598236	99299001	G	A	1	1	0.001	Maternal
ENSG00000136231	*IGF2BP3*	Chr7	rs56259943	23400797	C	T	0.15	1	0.0137	Maternal
ENSG00000136231	*IGF2BP3*	Chr7	rs73073066	23447351	C	T	0.22	1	0.0207	Maternal

ATL alternative allele, *a* additive effect; d: dominance effect; l: imprinting effect; expression: imprinting status. The p values were adjusted by the BH multiple comparison method

(adjusted p value $= 1.37 \times 10^{-67}$). Among the identified genes presenting dominance effect in eQTL mapping, the *RB1* gene located on chromosome 13 was a tumor suppressor gene, the mutation inactivation of which has been found to be the cause of human cancer (Chinnam and Goodrich 2011). It was also found to be an imprinting gene earlier in

Table 4 A list of genes with potential dominance effects (p values < 0.05)

ID	Gene	Chr	SNP	Position	Allele	ATL	p value.a	p value.d	p value.l	Expression
ENSG00000139687	*RB1*	Chr13	rs9596041	48898297	C	T	0.9	0	1	
ENSG00000139687	*RB1*	Chr13	rs2122276	48893306	T	C	0.9	0	1	–
ENSG00000139687	*RB1*	Chr13	rs1529370	48903892	G	T	0.9	0	1	–
ENSG00000139687	*RB1*	Chr13	rs1427167	48885561	A	G	0.9	0	1	–
ENSG00000139687	*RB1*	Chr13	rs1449574	48887861	A	T	0.9	0	1	–
ENSG00000139687	*RB1*	Chr13	rs7490216	48897733	A	T	0.9	0	1	–
ENSG00000139687	*RB1*	Chr13	rs4141954	48895766	A	G	0.9	0	1	–
ENSG00000139687	*RB1*	Chr13	rs7985609	48945285	T	G	0	0	1	–
ENSG00000139687	*RB1*	Chr13	rs9568118	48887412	T	C	0.99	3.02E−80	1	–
ENSG00000139687	*RB1*	Chr13	rs1449576	48887803	C	T	0.99	3.02E−80	1	–
ENSG00000139687	*RB1*	Chr13	rs1449577	48887205	G	T	0.99	3.02E−80	1	–
ENSG00000139687	*RB1*	Chr13	rs6561483	48885946	T	C	0.99	3.02E−80	1	–
ENSG00000139687	*RB1*	Chr13	rs9596038	48889438	T	C	0.99	3.02E−80	1	–
ENSG00000139687	*RB1*	Chr13	rs6561482	48884687	A	G	0.99	3.02E−80	1	–
ENSG00000139687	*RB1*	Chr13	rs1834735	48885404	C	T	0.99	3.02E−80	1	–
ENSG00000139687	*RB1*	Chr13	rs9596040	48889980	T	C	0.99	3.02E−80	1	–
ENSG00000140443	IGF1R	Chr15	rs4965238	99456357	T	C	0.78	1.37E−67	1.12E−71	Paternal

The p values were adjusted by BH multiple comparison

ATL alternative allele, *a*: additive effect, *d* dominance effect, *l* imprinting effect; expression: imprinting status

2009 (Kanber et al. 2009). Interesting, *IGF1R* was the only gene that presented both dominance effect and imprinting effect from the candidate *cis*-eQTL rs4965238 (Table 4).

In conclusion, our real data application validated several existing imprinting genes. Additionally, we mapped candidate eQTLs for these imprinted genes using our proposed methods. More interestingly, we discovered that a few genes presented significant dominance effect which might be involved in tumorigenesis.

Discussion

This article stands on recent advances in genetic modeling for carrying out new methodological developments to the aid of the analysis of eQTL mapping with genetic imprinting detection. We developed two statistical methods. The Stat-POE model provides a solution that allows for additive-by-dominance genetic effects for *cis*-eQTL mapping with RNA-seq data. The Func-POE is an alternative method which focuses on biological interpretations. We demonstrated the desirable power and preserved Type I error of the methods in most scenarios with un-biased estimation of the genetic effects and over-dispersion of the RNA-seq data. The application to the HapMap project validated previously reported imprinting genes and discovered significant *cis*-eQTLs for these imprinted genes. More interestingly, we identified two novel imprinting genes with significant dominance effect.

In the parameter estimation and hypothesis testing, we implemented the Stat-POE and Func-POE with three different tests, including the LRT, score and Wald tests. Among these three tests, the score and Wald tests are known to have poorer performance and less reliable results with small sample sizes by comparing to LRT (Table 2). From theory and our simulations (results not shown), the score test usually outperformed the other two in statistical power. To achieve a well-balanced Type I error and statistical power in detecting these genetic effects, we will suggest users to use the LRT when sample sizes are relatively small and score test otherwise.

We developed two imprinting effect models with RNA-seq data, including the Stat-POE model and Func-POE model both of which are appropriate for estimation of the genetic effects. The commonly used functional approach (i.e., Func-POE) is based on the observed genotype instead of the population frequencies therefore the results from which are easier to interpret. The disadvantage is that the functional model generates non-orthogonal estimates of regression coefficients when dominance components are included in the model. In contrast, parameters from application of the Stat-POE model describe the variance components rather than allele substitution effects, so may be seen to be having a less clear interpretation but it renders more straightforward model selection. Indeed, these two models can be transformed to each other in the estimates of the parameters, but the test statistics varied in formulas. (Xiao

et al. 2013 Text S5). As a result, the orthogonal model is presenting better power than the functional model especially in detecting additive effect. To be noted, since the parameters are different biological properties and interpretations in the two models, the comparison of these two models should be understood in terms of the comparison of testing a genetic effect and/or the imprinting effects when the effect is existing, instead of the concrete values of the parameters. For example the orthogonal model presented increased power than the functional model when additive effect exists (Xiao et al. 2014).

Several alternative imprinting effect testing methods have been described in previous literature (Álvarez-Castro 2014; Palowitch et al. 2018; Wolf and Cheverud 2009; Xiao et al. 2013). The Xiao et al. (2013) was indeed the initial attempt of our reseach team to implement imprinting effect detection with a one-locus orthogonal model which provided both statistical (i.e., population-referenced) and functional (which are not population-referenced) formulations of the genetic effects. This imprinting model was shown to be orthogonal in certain conditions. Then Álvarez-Castro (2014) provided a formula similar to the model in our method (Eq. 3) for imprinting detection which was claimed to be fully orthogonal. However, their model in formula (10) was not completely accurate (might be due to typo or some other reasons). The developments by Wolf and Cheverud (2009) proposed a two-locus model that included epistatic interactions involving imprinting effects. They also provided a model (Wolf and Cheverud 2009, Appendix 2) with an explicit imprinting parameter that is orthogonal under the Hardy–Weinberg proportions. Nevertheless, none of the above methods provides explicit expressions for performing variance decompositions or addressing the hypothesis testing problems. In a more recent article of Palowitch et al. (2018), eQTL analyses were performed using a non-linear regression model for log-transformed expression, termed ACME (Additive Contributions, Multiplicative Error), assuming additive allelic effects on the original expression scale. Their count-based modeling approach through some transformations of expression is different from the traditional Poisson or NB generalized linear models for the count-based RNA-seq data. Also, it lacks the ability to test the imprinting effect and overdispersion of the RNA-seq data hence it cannot be fairly compared to our methods. It is worthwhile noticing that none of the above methods have been implemented the coding to be used with RNA-seq data for eQTL mapping.

This is the first time the performance of the natural and orthogonal (NOIA) models have been evaluated in RNA-seq data analyses. The NOIA method was proposed by Alvarez-Castro et al. in 2007 (Alvarez-Castro and Carlborg 2007) which was composed by a one-locus functional model and statistical/orthogonal model, with which we have extensively implemented to the estimation of statistical epistasis, gene-environmental interactions and imprinting effect in genotype–phenotype mapping for quantitative traits and qualitative traits (Ma et al. 2012; Xiao et al. 2013, 2014). It has been shown that the one-locus statistical model was orthogonal independent of whether HWE was satisfied or not for quantitative trait analysis (Alvarez-Castro and Carlborg 2007). The conclusion can be straightforwardly extended to the NB or Poisson regressions in our study after implementation of the imprinting effect detection. Through implementing the developed methods in RNA-seq data, we provide new insights into eQTL mapping with powerful accurate estimation of genetic effects, covariates and over-dispersion parameters, especially that the proposed Stat-POE model allows uncorrelated estimation of the genetic effects.

We also investigated the parameter estimation and hypothesis testing of the two models with NB regression and Poisson regression assumption of the read counts for different application scope (results not shown). NB regression is suggested when the over-dispersion is relatively large and Poisson regression is suggested to be used for small over-dispersion. For observed RNA-seq data with excessive zeros, for example, when estimated $\mu_{\vec{A_1A_2}}$ or $\mu_{\vec{A_2A_1}}$ in Eq. (5) equals to zero, we suggest adding one to the read counts to satisfy the hypothesis of data distribution. Fitting a zero-inflated Poisson regression is another promising direction to address such problems, which warrants a future research goal.

Still, this study has several limitations. First, family data such as trios are needed to obtain the genotypes of heterozygotes in the offspring. Imputation-based approaches might be useful for haplotype-based inference of the phase of the heterozygotes, such as BEAGLE (Browning and Browning 2009). Second, borrowing information from the whole samples will allow for more accurate modeling of the RNA-seq data. A third important direction would be incorporating the testing of allele specific gene expression (ASE) as conducted in Zhabotynsky's work so that we can extend our work to model both ASE and POE for candidate *cis*-eQTLs (Zhabotynsky et al. 2019). The decreased cost of RNA-seq technology and future studies in methodology are warranted to achieve a more powerful estimation of decomposed variance from different genetic components.

Acknowledgements The preparation of this manuscript was supported by the internal funding from the University of South Carolina for Dr. Feifei Xiao. We are extremely grateful for the generous support of Dr. Wei Sun and Dr. Zhabotynsky, who kindly provided the phased and imputed genotypes and pre-processed RNA-seq data in the application study. We also acknowledge Dr. Guoshuai Cai for his valuable inputs in the RNA-seq data statistical modeling and analysis.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Álvarez-Castro JM (2014) Dissecting genetic effects with imprinting. Front Ecol Evol. https://doi.org/10.3389/fevo.2014.00051

Alvarez-Castro JM, Carlborg O (2007) A unified model for functional and statistical epistasis and its application in quantitative trait Loci analysis. Genetics 176:1151–1167. https://doi.org/10.1534/genetics.106.067348

Benjamini Y, Hochberg Y (1995) controlling the false discovery rate—a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

Browning BL, Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 84:210–223. https://doi.org/10.1016/j.ajhg.2009.01.005

Chinnam M, Goodrich DW (2011) Rb1, development, and cancer. Curr Top Dev Biol 94:129–169. https://doi.org/10.1016/B978-0-12-380916-2.00005-X

Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. Nat Rev Genet 10:184–194. https://doi.org/10.1038/nrg2537

Delaneau O, Marchini J, Genomes Project C, Genomes Project C (2014) Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat Commun 5:3934. https://doi.org/10.1038/ncomms4934

Duenk P, Calus MPL, Wientjes YCJ, Bijma P (2017) Benefits of dominance over additive models for the estimation of average effects in the presence of dominance. G3 (Bethesda) 7:3405–3414. https://doi.org/10.1534/g3.117.300113

Ellis SE, Gupta S, Ashar FN, Bader JS, West AB, Arking DE (2013) RNA-Seq optimization with eQTL gold standards. BMC Genom 14:892. https://doi.org/10.1186/1471-2164-14-892

Ertl J, Legarra A, Vitezica ZG, Varona L, Edel C, Emmerling R, Gotz KU (2014) Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. Genet Sel Evol 46:40. https://doi.org/10.1186/1297-9686-46-40

Genomes Project C et al (2012) An integrated map of genetic variation from 1092 human genomes. Nature 491:56–65. https://doi.org/10.1038/nature11632

Genomes Project C et al (2015) A global reference for human genetic variation. Nature 526:68–74. https://doi.org/10.1038/nature15393

Green PJ (1984) Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. J R Stat Soc Ser B (Methodol) 46:149–192

Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 44:955. https://doi.org/10.1038/ng.2354

Huang W, Mackay TF (2016) The genetic architecture of quantitative traits cannot be inferred from variance component analysis. PLoS Genet 12:e1006421. https://doi.org/10.1371/journal.pgen.1006421

Jadhav B et al (2019) RNA-Seq in 296 phased trios provides a high-resolution map of genomic imprinting. BMC Biol 17:50. https://doi.org/10.1186/s12915-019-0674-0

Kanber D et al (2009) The Human Retinoblastoma Gene Is Imprinted. Plos Genet 5:e1000790. https://doi.org/10.1371/journal.pgen.1000790

Kong A et al (2009) Parental origin of sequence variants associated with complex diseases. Nature 462:868–U859. https://doi.org/10.1038/nature08625

Ma J et al (2012) Natural and orthogonal interaction framework for modeling gene-environment interactions with application to lung cancer. Hum Hered 73:185–194. https://doi.org/10.1159/000339906

Palowitch J, Shabalin A, Zhou YH, Nobel AB, Wright FA (2018) Estimation of cis-eQTL effect sizes using a log of linear model. Biometrics 74:616–625. https://doi.org/10.1111/biom.12810

Perry JRB et al (2014) Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. Nature 514:92. https://doi.org/10.1038/nature13545

Rao CR (2005) Score test: historical review and recent developments. In: Balakrishnan N, Nagaraja HN, Kannan N (eds) Advances in ranking and selection, multiple comparisons, and reliability. Statistics for Industry and Technology, Birkhäuser, Boston, pp 3–20. https://doi.org/10.1007/0-8176-4422-9_1

Su G, Christensen OF, Ostersen T, Henryon M, Lund MS (2012) Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PLoS ONE 7:e45293. https://doi.org/10.1371/journal.pone.0045293

Wallace C, Smyth DJ, Maisuria-Armer M, Walker NM, Todd JA, Clayton DG (2010) The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. Nat Genet 42:68–U85. https://doi.org/10.1038/ng.493

Wolf JB, Cheverud JM (2009) A framework for detecting and characterizing genetic background-dependent imprinting effects. Mamm Genome 20:681–698. https://doi.org/10.1007/s00335-009-9209-2

Xiao F, Ma J, Amos CI (2013) A unified framework integrating parent-of-origin effects for association study. PLoS ONE 8:e72208. https://doi.org/10.1371/journal.pone.0072208

Xiao F, Ma J, Cai G, Fang S, Lee JE, Wei Q, Amos CI (2014) Natural and orthogonal model for estimating gene-gene interactions applied to cutaneous melanoma. Hum Genet 133:559–574. https://doi.org/10.1007/s00439-013-1392-2

Xiang T, Christensen OF, Vitezica ZG, Legarra A (2018) Genomic model with correlation between additive and dominance effects. Genetics 209:711–723. https://doi.org/10.1534/genetics.118.301015

Zhabotynsky V, Inoue K, Magnuson T, Mauro Calabrese J, Sun W (2019) A statistical method for joint estimation of *cis*-eQTLs and parent-of-origin effects under family trio design. Biometrics. https://doi.org/10.1111/biom.13026