

2011

## A Statistical Design for Testing Transgenerational Genomic Imprinting in Natural Human Populations

Yao Li  
*West Virginia University*

Yunqian Guo  
*Beijing Forestry University*

Jianxin Wang  
*Beijing Forestry University*

Wei Hou  
*University of Florida*

Myron N. CHang  
*University of Florida*

*See next page for additional authors*

Follow this and additional works at: [https://researchrepository.wvu.edu/faculty\\_publications](https://researchrepository.wvu.edu/faculty_publications)

---

### Digital Commons Citation

Li, Yao; Guo, Yunqian; Wang, Jianxin; Hou, Wei; CHang, Myron N.; Liao, Duanping; and Wu, Rongling, "A Statistical Design for Testing Transgenerational Genomic Imprinting in Natural Human Populations" (2011). *Faculty & Staff Scholarship*. 2736.

[https://researchrepository.wvu.edu/faculty\\_publications/2736](https://researchrepository.wvu.edu/faculty_publications/2736)

This Article is brought to you for free and open access by The Research Repository @ WVU. It has been accepted for inclusion in Faculty & Staff Scholarship by an authorized administrator of The Research Repository @ WVU. For more information, please contact [ian.harmon@mail.wvu.edu](mailto:ian.harmon@mail.wvu.edu).

---

**Authors**

Yao Li, Yunqian Guo, Jianxin Wang, Wei Hou, Myron N. CHang, Duanping Liao, and Rongling Wu

# A Statistical Design for Testing Transgenerational Genomic Imprinting in Natural Human Populations

Yao Li<sup>1,2,9</sup>, Yunqian Guo<sup>1,9</sup>, Jianxin Wang<sup>1</sup>, Wei Hou<sup>3</sup>, Myron N. Chang<sup>3</sup>, Duanping Liao<sup>4</sup>, Rongling Wu<sup>1,4\*</sup>

**1** Center for Computational Biology, Beijing Forestry University, Beijing, People's Republic of China, **2** Department of Statistics, West Virginia University, Morgantown, West Virginia, United States of America, **3** Department of Biostatistics, University of Florida, Gainesville, Florida, United States of America, **4** Department of Public Health Sciences, Penn State College of Medicine, Hershey, Pennsylvania, United States of America

## Abstract

Genomic imprinting is a phenomenon in which the same allele is expressed differently, depending on its parental origin. Such a phenomenon, also called the parent-of-origin effect, has been recognized to play a pivotal role in embryological development and pathogenesis in many species. Here we propose a statistical design for detecting imprinted loci that control quantitative traits based on a random set of three-generation families from a natural population in humans. This design provides a pathway for characterizing the effects of imprinted genes on a complex trait or disease at different generations and testing transgenerational changes of imprinted effects. The design is integrated with population and cytogenetic principles of gene segregation and transmission from a previous generation to next. The implementation of the EM algorithm within the design framework leads to the estimation of genetic parameters that define imprinted effects. A simulation study is used to investigate the statistical properties of the model and validate its utilization. This new design, coupled with increasingly used genome-wide association studies, should have an immediate implication for studying the genetic architecture of complex traits in humans.

**Citation:** Li Y, Guo Y, Wang J, Hou W, Chang MN, et al. (2011) A Statistical Design for Testing Transgenerational Genomic Imprinting in Natural Human Populations. PLoS ONE 6(2): e16858. doi:10.1371/journal.pone.0016858

**Editor:** Vladimir Brusic, Dana-Farber Cancer Institute, United States of America

**Received:** November 1, 2010; **Accepted:** December 31, 2010; **Published:** February 25, 2011

**Copyright:** © 2011 Li et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work is supported by the Changjiang Scholars Award, "One-thousand Person Plan" Award, National Natural Science Foundation of China (Grant No.30900854) and Beijing Forestry University Young Scientist Fund (Grant No. Blx2w8003). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: rwu@bjfu.edu.cn

<sup>9</sup> These authors contributed equally to this work.

## Introduction

Genomic imprinting arises from a gene when either the maternally or paternally derived copy of it is expressed while the other copy is silenced [1,2]. Caused by epigenetic modifications such as DNA methylation established during gametogenesis and maintained throughout somatic development in the offspring, genetic imprinting has been shown to play a pivotal role in regulating the formation, development, function, and evolution of complex traits and diseases [3,4,5,6,7,8,9,10]. While most studies of genetic imprinting focus on the epigenetic and molecular mechanisms of this phenomenon [7,11], the number and distribution of imprinted genes and their epistatic interactions for quantitative traits are poorly understood, limiting the scope of our inference about the effects of imprinting genes on the diversity of biological traits or processes. Several authors have started to use genome-wide association and linkage studies to identify the regions of the genome that contain imprinted sequence variants and further understand the epigenetic variation of complex traits [12,13,14,15].

In a series of recent studies, Cheverud, Wolf, and colleagues categorized genetic imprinting into different types based on the pattern of its expression, i.e., maternal expression, paternal expression, bipolar dominance, polar overdominance, and polar underdominance [14,15]. With a three-generation F<sub>2</sub> design, they identified these types of imprinted quantitative trait loci (iQTL)

affecting body weight and growth in mice, displaying much more complex and diverse effect patterns than previously assumed. A different design based on reciprocal backcrosses was proposed to test and estimate the distribution of iQTL responsible for physiological traits related to endosperm development in maize [16]. By modeling identical-by-descent relationships in multiple related families of canines, Liu et al. [13] derived a random effect model based on linkage analysis to genome-wide scan for the existence of iQTL that affect canine hip dysplasia. In a recent study, Wang et al. [9] used reciprocal F<sub>2</sub> designs to identify the additive and dominant effects of iQTLs and their interactions with imprinting effects for hyperoxic acute lung injury survival time in mice. These authors also explore the transgenerational inheritance of iQTLs.

While epigenetic marks resulting in genomic imprinting can be generally stable in an organism's lifetime, they may undergo reprogramming, i.e., a faithful clearing of the epigenetic state established in the previous generation, in the new generation during gametogenesis and early embryogenesis [17,18,19]. However, a growing body of evidence since the early 1980s indicates that genes may escape such reprogramming and, thus, inherit their imprinting effects into next generations [20,21,22,23,24,25]. Two fundamental questions will naturally arise from this discovery: how common are imprinted genes of this type and how strong is the evidence for their existence in humans and other organisms? If epigenetic changes through imprinted

genes can be inherited across generations, this would significantly alter the way we think about the inheritance of phenotype [26,27]. Such transgenerational epigenetic inheritance, i.e., modifications of the chromosomes that pass to the next generation through gametes, may be related with health and diseases with a mechanism for transmitting environmental exposure information that alters gene expression in the next generation(s) [28]. The identification of imprinted loci displaying transgenerational epigenetic inheritance will be greatly helpful for addressing the two questions mentioned above, in a quest to elucidate the detailed genetic architecture of complex traits and diseases.

The motivation of this study is to develop a novel strategy for identifying imprinted genes for a quantitative trait and understanding the transgenerational changes of their effects with a three-generation family design by sampling multiple unrelated nuclear families, each composed of the grandfather, grandmother, father, mother, and grandchildren, from a natural population. This transgenerational design contains information about how alleles at different loci co-transmit during meiosis from one generation to next and, thus, has been widely used for genetic linkage analysis [29,30]. By tracing the inheritance of alleles at a gene(s) from a paternal or maternal parent, this design allows the characterization of parent-of-origin of alleles and provides a powerful way to estimate genetic imprinting effects. Because only genotypes can be observed, we formulate a mixture model to specify allelic configurations in terms of parental origins of the alleles. The EM algorithm is implemented to estimate the effects of imprinted genes and their changes across generations. A testing procedure is proposed to study the pattern of transgenerational epigenetic inheritance. The statistical behavior of the model is examined through simulation studies.

## Results

Simulation studies were performed to examine the statistical behavior of the model. A three-generation design is simulated which include a certain number of first-generation families sorted into 9 mating types (as shown in Table S1) according to the genotype frequencies. Assume that the allele frequencies of a gene are 0.6 and 0.4 in a natural population at Hardy-Weinberg equilibrium. Our simulation will focus on the investigation of the impacts of different sampling strategies and heritabilities on parameter estimation and model power. For a given sample size, two sampling strategies are simulated, (1) a large family number and small family size, and (2) a small family number and large family size.

The first strategy samples 200 unrelated grandfathers and 200 unrelated grandmothers, who marry to form 200 the first-generation families. Each first-generation family is assumed to have one son who, as the father, form a second-generation family with the mother from the natural population. There is one child for each second-generation family. This allocation results in a total of 1000 subjects. All members in the design are typed for the gene, but only the fathers and offspring of the third generation are phenotyped for a normally distributed trait. The second strategy samples 50 unrelated grandfathers and 50 unrelated grandmothers. In each first-generation family, 3 sons are simulated, forming 150 second-generation families in which 4 children are assumed. This strategy also results in 1000 subjects.

Different genetic effects of the gene, additive, dominant, and imprinting, are simulated for the second- and third-generations using the designed shown in Tables S2 and S3. Two different heritability levels, 0.1 and 0.4, are simulated for each generation, from which variances are determined. Table 1 tabulates the estimates of population and quantitative genetic parameters from the three-generation design. As expected, allele frequency can be very well estimated. The model provides reasonable estimation accuracy and precision for all genetic parameters under different sampling strategies, even for a modest heritability level. Under both strategies, the model has great power (0.85 or higher) to test the significance of individual genetic effects, additive, dominant, and imprinting, expressed in different generations. The model is also powerful to detect differences of genetic effects between two consecutive generations. More interesting, the difference of imprinting effect between different generations, i.e., transgenerational inheritance of genetic imprinting, can be discerned with power 0.80 using our statistical design.

One major aim of this study is to estimate the change of genetic effects over generation. Although our model has great power to detect the transgenerational change of genetic effects, its false positive rates should also be assessed. We conducted an additional simulation study to address this issue by simulating a SNP that has the same genetic effects between the two generations. The model detects a small proportion of simulation replicates (<6%) which displays transgenerational differences in all types of genetic effects including additive, dominant, and imprinting. This suggests that the model has a small type I error rate for detecting the transgenerational difference of overall genetic effects. We particularly tested the type I error rate for the transgenerational difference of genetic imprinting, which is reasonably small (<8%).

The haplotype model is also examined through simulation studies. We simulated two SNPs with a recombination fraction of

**Table 1.** The maximum likelihood estimates (MLEs) of additive ( $a$ ), dominant ( $d$ ), and imprinting effects ( $i$ ) of a functional SNP on a complex trait in parental ( $F$ ) and offspring ( $O$ ) generations under two different strategies.

Genetic Parameter	True Value	Strategy 1		Strategy 2	
		$H^2 = 0.1$	$H^2 = 0.4$	$H^2 = 0.1$	$H^2 = 0.4$
$a_F$	1.0	1.0198(0.0236)	1.0028(0.0112)	1.0387(0.0333)	0.9975(0.0125)
$d_F$	0.6	0.5702(0.0340)	0.6184(0.0144)	0.6272(0.0392)	0.6046(0.0185)
$i_F$	0.6	0.6024(0.0306)	0.5991(0.0093)	0.5897(0.0328)	0.6037(0.012)
$a_O$	1.0	0.9779(0.0429)	0.9889(0.0167)	1.0136(0.0243)	1.0082(0.0108)
$d_O$	1.5	1.5218(0.0462)	1.4397(0.0242)	1.5465(0.0389)	1.4962(0.0138)
$i_O$	1.0	0.9853(0.0393)	0.9885(0.0138)	1.0272(0.0271)	1.0212(0.0119)

The estimates are the means of MLEs obtained from 200 simulation replicates, with standard errors given in parentheses.  
doi:10.1371/journal.pone.0016858.t001

$r=0.05$  that are segregating in a human population. Of the four haplotypes, one is assumed to function as a risk haplotype. The remaining is collectively called the non-risk haplotype. The genetic values of composite diploypes constituted by risk and non-risk haplotypes include the additive ( $a$ ), dominant ( $d$ ), and imprinting ( $i$ ) genetic effects. We assume that some of these effects are different, and the others are the same between the parental and offspring generations. Combinations of different heritabilities between the two generations are simulated.

Table 2 gives the results of simulation for different heritabilities and sample sizes (all subjects used). Overall, all parameters can be estimated reasonably well. As expected, the precision of parameter estimation increases with heritability and sample size. The additive genetic effects in both generations can well be estimated with a modest sample size (say 400) for a small heritability (0.1). More sample sizes (say 800) are needed to provide a good estimate for genetic imprinting effects for a small heritability. To well estimate dominant genetic effects, an even larger sample size (say 2000) is required for the same level of heritability.

## Discussion

The traditional view of quantitative trait expression analysis assumes that the maternally and paternally derived alleles of each gene are expressed simultaneously at a similar level. However, this view is violated by a growing body of evidence that alleles are expressed from only one of the two parental chromosomes [1,2]. This so-called genetic imprinting or parent-of-origin effect has been thought to play a pivotal role in regulating the phenotypic variation of a complex trait [3,4,6,8,9,12,13,14,15]. With the discovery of more imprinting genes involved in trait control through molecular and bioinformatics approaches, we will be in a position to elucidate the genetic architecture of quantitative variation for various organisms including humans.

Recent evidence shows that epigenetic inheritance in humans may experience a transgenerational change. This would represent a significant shift in our current understanding of inheritance and disease aetiology. Despite the development of new technologies that are reducing the time and cost of genotyping by several orders of magnitude [31,32], the understanding of the underlying

genetic events will be challenging. In this article, we present a computational model for identifying the genomic imprinting effect of genes on quantitative phenotypes and transgenerational change of genomic imprinting using a multigenerational sampling design for human families. The model formulates a general framework for testing the difference of genetic effects between different generations. By including multiple SNPs, the model was extended to estimate genomic imprinting and its transgenerational change expressed at the haplotype level. Although several models have been developed to estimate genomic imprinting for binary disease traits [33,34], our model is among the first for estimating genetic imprinting operational in regulating the variation of quantitative traits and is certainly the first of its kind that can discern the transgenerational change of genetic imprinting.

Although no real data were analyzed for the moment, this model presents a conceptual design by which new data can be collected according to the sampling strategy proposed and then analyzed by the computational algorithm derived. Based on computer simulation, the model should display convincing statistical properties in parameter estimation and test and can be applied to a practical data set. However, several issues need to be addressed when the model is attempted to solve broader genetic questions. First, the maternal effects that cause parent-of-origin effects of alleles may be confounded with imprinting effects [35], which should be separated by developing a proper design in order to better study the patterns of gene expression and evolutionary dynamics.

Second, this study assumes the unisex (sons) produced from the first-generation family. One can also assume daughters with no change of the model, allowing the test of genomic imprinting between mother and offspring. In fact, our model can involve both sexes so that in the second generation sex-specific genetic effects can be characterized. If the sexes in the third generation are considered, the model can be extended to study the transgenerational changes of gene-sex interactions. Third, it is possible that part of parental genotypes are missing in practice. To infer genomic imprinting using such data sets, a multi-hierarchical mixture model can be derived to estimate the missing parental genotypes based on observed offspring genotypes. Fourth,

**Table 2.** Simulation results for transgeneration imprinting effects comparisons.

$N$	$H^2$	First Generation Parameters				Second Generation Parameters			
		$r=0.05$	$a_F=1$	$d_F=0.5$	$i_F=0.4$	$a_O=1$	$d_O=1.5$	$i_O=0.6$	
400	0.1	0.1	0.0561(0.0049)	0.9936(0.0185)	0.5228(0.0263)	0.3824(0.0195)	1.0182(0.0266)	1.4761(0.0405)	0.6592(0.0269)
	0.1	0.4	0.0450(0.0049)	0.9975(0.0162)	0.4690(0.0279)	0.4437(0.0184)	0.9860(0.0106)	1.4936(0.0174)	0.5937(0.0115)
	0.4	0.1	0.0639(0.0055)	1.0011(0.0082)	0.5032(0.0107)	0.4013(0.0082)	1.0210(0.0277)	1.5275(0.0383)	0.5631(0.0285)
	0.4	0.4	0.0681(0.0054)	1.0137(0.0075)	0.5072(0.0115)	0.3916(0.0083)	0.9939(0.0108)	0.4920(0.0144)	0.5975(0.0110)
800	0.1	0.1	0.0486(0.0044)	1.0003(0.0134)	0.4610(0.0183)	0.4162(0.0126)	1.0079(0.0194)	1.5219(0.0284)	0.6322(0.0175)
	0.1	0.4	0.0461(0.0040)	0.9956(0.0135)	0.4933(0.0180)	0.3796(0.0114)	1.0049(0.0075)	1.4978(0.0112)	0.5874(0.0072)
	0.4	0.1	0.0516(0.0041)	1.0047(0.0046)	0.5074(0.0077)	0.3915(0.0055)	0.9916(0.0083)	1.5027(0.0091)	0.6002(0.0073)
	0.4	0.4	0.0567(0.0036)	1.0011(0.0056)	0.5023(0.0080)	0.3976(0.0061)	0.9773(0.0077)	1.5069(0.0104)	0.5926(0.0083)
2000	0.1	0.1	0.0516(0.0032)	1.0109(0.0079)	0.4951(0.0116)	0.4059(0.0089)	1.0053(0.0122)	1.5095(0.0150)	0.5878(0.0119)
	0.1	0.4	0.0536(0.0029)	1.0078(0.0094)	0.5283(0.0107)	0.4038(0.0099)	1.0017(0.0042)	1.5011(0.0064)	0.5912(0.0053)
	0.4	0.1	0.0488(0.0027)	0.9996(0.0034)	0.5076(0.0044)	0.4064(0.0036)	0.9830(0.0115)	1.4997(0.0152)	0.6001(0.0138)
	0.4	0.4	0.0545(0.0028)	1.0009(0.0033)	0.5043(0.0047)	0.3986(0.0033)	0.9993(0.0050)	1.5012(0.0064)	0.5996(0.0048)

The genetic design scenarios are chosen as the combination of different heritabilities and sample sizes. They are:  $H_1^2=0.1/0.4$ ,  $H_2^2=0.1/0.4$ ,  $n=400,800,2000$ .  
doi:10.1371/journal.pone.0016858.t002

although a basic premise of epigenetic processes was that, once established, these marks were maintained through rounds of mitotic cell division and stable for the life of the organism, several recent studies have shown that at some loci the epigenetic state can be altered by the environment [36]. The questions are how common are genes of this type and how strong is the evidence for their existence in humans? The development of our design and model will help to address these biological questions of fundamental importance in elucidating the genetic architecture of complex traits.

**Methods**

**Sampling Strategies**

Suppose there is a natural human population at Hardy–Weinberg equilibrium (HWE) from which a panel of three-generation families, each composed of the grandfather, grandmother, father, mother, and grandchildren, are sampled. Each member in a family is typed for single nucleotide polymorphisms (SNPs) from the human genome. Consider a quantitative trait affected by a SNP with two alleles *A* in a frequency of *p* and *a* in a frequency of *q*, leading to three genotypes *AA*, *Aa*, and *aa* with the frequencies of *p*<sup>2</sup>, *2pq*, and *q*<sup>2</sup>, respectively. In the grandparent generation, these three genotypes are mating randomly to produce nine cross types (Table S1). Given a cross type, the genotypes of sons or daughters can be inferred. Here we first assume one sex (say son) in the second generation, although both sexes can be considered. The sons from a family serve as the father to mate with the females as the mother derived from a natural population, with genotypes, *AA*, *Aa*, and *aa*, characterized by frequencies *p*<sup>2</sup>, *2pq*, and *q*<sup>2</sup>, respectively. Each of such second-generation families produces a certain number of grandchildren. The genotype frequencies in the third generation are derived according to Mendel’s first law.

According to this design, the grandfathers and grandmothers are founders whose parents are unknown. Alleles of sons from a first-generation family can be traced directly or indirectly, but the females used to generate the second-generation family are the founders with the unknown origin of alleles. For this reason, we will measure the phenotype for sons from the first-generation families and grandchildren from the second-generation families. This design will allow us to characterize imprinting effects of a gene in the second- and third-generations.

**Genetic Models**

There are three genotypes, *AA*, *Aa*, and *aa*, for a biallelic gene according to Mendelian segregation pattern. Considering the parent-of-origin of alleles, these genotypes are described by four configurations, *A|A* (coded as 2), *A|a* (coded as 1), *a|A* (coded as 1’), and *a|a* (coded as 0), where symbol | is used to separate the maternally- (left) and paternally-derived alleles (right). The genotypic values of the four configurations in two different generations are defined as follows:

Configuration	Paternal	Offspring
<i>A A</i>	$\mu_F^F = \mu_F + a_F$	$\mu_2^O = \mu_O + a_O$ ,
<i>A a</i>	$\mu_1^F = \mu_F + d_F + i_F$	$\mu_1^O = \mu_O + d_O + i_O$ ,
<i>a A</i>	$\mu_{1'}^F = \mu_F + d_F - i_F$	$\mu_{1'}^O = \mu_O + d_O - i_O$ ,
<i>a a</i>	$\mu_0^F = \mu_F - a_F$	$\mu_0^O = \mu_O - a_O$ ,

where  $\mu_F$  and  $\mu_O$  are the overall means of the paternal and offspring generations, *a<sub>F</sub>*, *d<sub>F</sub>*, and *i<sub>F</sub>* are the additive, dominant and imprinting genetic effects of the gene in the parental

generation, and *a<sub>O</sub>*, *d<sub>O</sub>*, and *i<sub>O</sub>* are the additive, dominant and imprinting genetic effects of the gene in the offspring generation.

The difference in the genetic architecture of a complex trait between two different generations is described as

$$\Delta_a = a_F - a_O, \tag{2}$$

$$\Delta_d = d_F - d_O, \tag{3}$$

$$\Delta_i = i_F - i_O. \tag{4}$$

By testing whether these differences are equal to zero jointly or individually, we can determine the transgenerational changes of the pattern of genetic control. If a significant imprinting effect is detected, we can test the type of genetic imprinting, i.e., parental or maternal dominance, by incorporating the imprinting models of Cheverud et al. [14].

**Estimation**

The grandfather and grandmother in the first generation from a natural population constitutes 3 × 3 = 9 mating types for three genotypes. For the *j*th first-generation mating type listed in Table S1 (*j* = 1, ..., 9), let *N<sub>j</sub>* denote the family number of this mating type. Each first-generation family may have one or multiple sons who serve the father of the second generation. Those families in the second generation with the father derived from the *j*th first-generation mating type and the mother of a particular genotype from the natural population are summed together, denoted by *N<sub>jl</sub><sup>M</sup>*, for mother genotype *l* (*l* = 2 for *AA*, 1 for *Aa*, and 0 for *aa*). Thus, we have a total of *N<sub>l</sub><sup>M</sup>* = ∑<sub>*j*=1</sub><sup>9</sup> *N<sub>jl</sub><sup>M</sup>* second-generation mothers who carry genotype *l*.

It is not difficult to derive the maximum likelihood estimate of allele frequency from the three-generation family design as

$$p = \frac{4N_1 + 3(N_2 + N_4) + 2(N_3 + N_5 + N_7) + (N_6 + N_8) + 2N_2^M + N_1^M}{\sum_{j=1}^9 4N_j + 2(N_2^M + N_1^M + N_0^M)}$$

$$q = \frac{4N_8 + 3(N_5 + N_7) + 2(N_2 + N_4 + N_6) + (N_1 + N_3) + 2N_0^M + N_1^M}{\sum_{j=1}^9 4N_j + 2(N_2^M + N_1^M + N_0^M)}$$

The male individuals from the first generation are typed for the marker, with four distinct configurations, *A|A* (2), *A|a* (1), *a|A* (1’), and *a|a* (0). Let *n<sub>jk</sub><sup>F</sup>* denote the cumulative number of male individuals (as the father for the second generation) bearing configuration *k* (*k* = 2, 1, 1’, 0) from *n<sub>j</sub>* first-generation families. In the third generation, only genotypes rather than configurations can be observed. We use *N<sub>jkls</sub><sup>O</sup>* to denote the number of children who carry genotype *s* (*s* = 2, 1, 0) from a second-generation family with father *k* (from the *j*th first-generation mating type) and mother *l* from a natural population. The phenotypic values measured are expressed as *y<sub>jki</sub><sup>F</sup>* (*i* = 1, ..., *n<sub>jk</sub><sup>F</sup>*) for the second-generation fathers and *y<sub>jkls</sub><sup>O</sup>* (*i* = 1, ..., *N<sub>jkls</sub><sup>O</sup>*) for the third-generation children. Both *y<sub>jki</sub><sup>F</sup>* and *y<sub>jkls</sub><sup>O</sup>* are assumed to follow a normal distribution with mean depending on genotypes and residual variances  $\sigma_F^2$  and  $\sigma_O^2$ , respectively.

Since offspring genotypes depend on parental genotypes, the log-likelihood of paternal and offspring parameters given marker (**M**) and phenotypic (*y*) data from the three generations is decomposed into two components, one related to the paternal

parameters and the second related to the offspring parameters given the paternal parameters, expressed as

$$L(\mathbf{\Omega}_F, \mathbf{\Omega}_O | y_F, y_O, \mathbf{M}_F^{(1)}, \mathbf{M}_M^{(1)}, \mathbf{M}_F^{(2)}, \mathbf{M}_M^{(2)}, \mathbf{M}_O^{(3)}) = L(\mathbf{\Omega}_F | y_F, \mathbf{M}_F^{(1)}, \mathbf{M}_M^{(1)}, \mathbf{M}_F^{(2)}) + L(\mathbf{\Omega}_O | y_O, \mathbf{M}_F^{(1)}, \mathbf{M}_M^{(1)}, \mathbf{M}_F^{(2)}, \mathbf{M}_O^{(3)}), \tag{5}$$

where  $\mathbf{\Omega}_F = (\mu_F, a_F, d_F, i_F, \sigma_F^2)$  are the paternal parameters and  $\mathbf{\Omega}_O = (\mu_O, a_O, d_O, i_O, \sigma_O^2)$  are the offspring parameters. Maximizing joint likelihood (5) is equivalent to maximizing its two likelihood components independently. The estimates of parameters  $\mathbf{\Omega}_F$  that maximize the first component can be obtained with the EM algorithm. In the E step, the posterior probability with which the double heterozygote father of the second generation from the 5th first-generation mating type in Table S1 has a particular configuration is calculated by

$$\Phi_{51i}^F = \frac{\frac{1}{2}f_1(y_{51i}^F)}{\frac{1}{2}f_1(y_{51i}^F) + \frac{1}{2}f_{1'}(y_{51'i}^F)} \text{ and } \Phi_{51'i}^F = \frac{\frac{1}{2}f_{1'}(y_{51'i}^F)}{\frac{1}{2}f_1(y_{51i}^F) + \frac{1}{2}f_{1'}(y_{51'i}^F)}. \tag{6}$$

In the M step, the genotypic values of configurations and variance are calculated by

$$\begin{aligned} \mu_2^F &= \frac{\sum_{i=1}^{N_{12}^F} y_{12i}^F + \sum_{i=1}^{N_{22}^F} y_{22i}^F + \sum_{i=1}^{N_{42}^F} y_{42i}^F + \sum_{i=1}^{N_{52}^F} y_{52i}^F}{N_{12}^F + N_{22}^F + N_{42}^F + N_{52}^F}, \\ \mu_1^F &= \frac{\sum_{i=1}^{N_{21}^F} y_{21i}^F + \sum_{i=1}^{N_{31}^F} y_{31i}^F + \sum_{i=1}^{N_{51}^F} \Phi_{51i}^F y_{51i}^F + \sum_{i=1}^{N_{61}^F} y_{61i}^F}{N_{21}^F + N_{31}^F + \sum_{i=1}^{N_{51}^F} \Phi_{51i}^F + N_{61}^F}, \\ \mu_{1'}^F &= \frac{\sum_{i=1}^{N_{41'}^F} y_{41'pi}^F + \sum_{i=1}^{N_{51'}^F} \Phi_{51'i}^F y_{51'i}^F + \sum_{i=1}^{N_{71'}^F} y_{71'i}^F + \sum_{i=1}^{N_{81'}^F} y_{81'i}^F}{N_{41'}^F + \sum_{i=1}^{N_{51'}^F} \Phi_{51'i}^F + N_{71'}^F + N_{81'}^F}, \\ \mu_0^F &= \frac{\sum_{i=1}^{N_{50}^F} y_{50i}^F + \sum_{i=1}^{N_{60}^F} y_{60i}^F + \sum_{i=1}^{N_{80}^F} y_{80i}^F + \sum_{i=1}^{N_{90}^F} y_{90i}^F}{N_{50}^F + N_{60}^F + N_{80}^F + N_{90}^F}, \\ \sigma_F^2 &= \frac{1}{\sum_{j=1}^9 N_j} \left\{ \sum_{i=1}^{N_{12}^F} (y_{12i}^F - \mu_2^F)^2 + \sum_{i=1}^{N_{22}^F} (y_{22i}^F - \mu_2^F)^2 + \sum_{i=1}^{N_{21}^F} (y_{21i}^F - \mu_1^F)^2 + \sum_{i=1}^{N_{31}^F} (y_{31i}^F - \mu_1^F)^2 + \sum_{i=1}^{N_{42}^F} (y_{42i}^F - \mu_2^F)^2 + \sum_{i=1}^{N_{41'}^F} [(y_{41'i}^F - \mu_1^F)^2 + \sum_{i=1}^{N_{52}^F} (y_{52i}^F - \mu_2^F)^2 + \sum_{i=1}^{N_{51}^F} (y_{51i}^F - \mu_1^F)^2 + \sum_{i=1}^{N_{51'}^F} (y_{51'i}^F - \mu_1^F)^2 + \sum_{i=1}^{N_{50}^F} (y_{50i}^F - \mu_0^F)^2 + \sum_{i=1}^{N_{61}^F} (y_{61i}^F - \mu_1^F)^2 + \sum_{i=1}^{N_{60}^F} (y_{60i}^F - \mu_0^F)^2 + \sum_{i=1}^{N_{71'}^F} (y_{71'i}^F - \mu_1^F)^2 + \sum_{i=1}^{N_{81'}^F} (y_{81'i}^F - \mu_1^F)^2 + \sum_{i=1}^{N_{80}^F} (y_{80i}^F - \mu_0^F)^2 + \sum_{i=1}^{N_{90}^F} (y_{90i}^F - \mu_0^F)^2 \right\}. \tag{7} \end{aligned}$$

The EM algorithm can also be implemented to estimate genetic parameters  $\mathbf{\Omega}_O$  in the third generation that maximize the second component in (5). In the E step, the posterior probability with which the double heterozygote offspring of the third generation derived from the combination of two double heterozygote parents in the second generation has a particular configuration is calculated by

$$\Phi_{jkli}^O = \frac{\frac{1}{4}f_1(y_{jkli}^O)}{\frac{1}{4}f_1(y_{jkli}^O) + \frac{1}{4}f_{1'}(y_{jkli}^O)} \text{ and } \Phi_{jk1'i}^O = \frac{\frac{1}{4}f_{1'}(y_{jkli}^O)}{\frac{1}{4}f_1(y_{jkli}^O) + \frac{1}{4}f_{1'}(y_{jkli}^O)}. \tag{8}$$

In the M step, the genotypic values of configurations and variance are calculated by

$$\begin{aligned} \mu_2^O &= \frac{\sum_{j=1}^9 \sum_{k=0}^2 \sum_{l=0}^2 \sum_{i=1}^{N_{jk12}^O} y_{jk12i}^O \zeta_{jk12i}}{\sum_{j=1}^9 \sum_{k=0}^2 \sum_{l=0}^2 N_{jk12}^O}, \quad k=0,1',1,2; l=0,1,2, \\ \mu_1^O &= \frac{\sum_{j=1}^9 \sum_{k=0}^2 \sum_{l=0}^2 \sum_{i=1}^{N_{jk11}^O} \Phi_{jk11i}^O y_{jk11i}^O \zeta_{jk11i}}{\sum_{j=1}^9 \sum_{k=0}^2 \sum_{l=0}^2 \sum_{i=1}^{N_{jk11}^O} \Phi_{jk11i}^O \zeta_{jk11i}}, \quad k=0,1',1,2; l=0,1,2, \\ \mu_{1'}^O &= \frac{\sum_{j=1}^9 \sum_{k=0}^2 \sum_{l=0}^2 \sum_{i=1}^{N_{jk11'}^O} \Phi_{jk11'i}^O y_{jk11'i}^O \zeta_{jk11'i}}{\sum_{j=1}^9 \sum_{k=0}^2 \sum_{l=0}^2 \sum_{i=1}^{N_{jk11'}^O} \Phi_{jk11'i}^O \zeta_{jk11'i}}, \quad k=0,1',1,2; l=0,1,2, \\ \mu_0^O &= \frac{\sum_{j=1}^9 \sum_{k=0}^2 \sum_{l=0}^2 \sum_{i=1}^{N_{jk10}^O} y_{jk10i}^O \zeta_{jk10i}}{\sum_{j=1}^9 \sum_{k=0}^2 \sum_{l=0}^2 N_{jk10}^O}, \quad k=0,1',1,2; l=0,1,2, \\ \sigma_O^2 &= \frac{1}{\sum_{j=1}^9 \sum_{k=0}^2 \sum_{l=0}^2 (N_{jk12}^O + N_{jk11}^O + N_{jk10}^O)} \times \sum_{j=1}^9 \sum_{k=0}^2 \sum_{l=0}^2 \left\{ \sum_{i=1}^{N_{jk12}^O} \zeta_{jk12i} (y_{jk12i}^O - \mu_2^O)^2 + \sum_{i=1}^{N_{jk11}^O} \zeta_{jk11i} [\Phi_{jk11i}^O (y_{jk11i}^O - \mu_1^O)^2 + \Phi_{jk11'i}^O (y_{jk11'i}^O - \mu_{1'}^O)^2 + \sum_{i=1}^{N_{jk10}^O} \zeta_{jk10i} (y_{jk10i}^O - \mu_0^O)^2] \right\}, \quad k=0,1',1,2; l=0,1,2, \end{aligned}$$

where  $\zeta_{jk12i}$ ,  $\zeta_{jk11i}$ , and  $\zeta_{jk10i}$  are the indicator variables that are defined as 1 if offspring  $i$  in the third generation from the combination of father  $k$  from the  $j$ th first-generation mating type and mother  $l$  from the natural population has genotype  $AA$ ,  $Aa$ , and  $aa$ , respectively, and 0 otherwise. The EM steps are iterated between equations (6) and (7) to obtain the MLEs of  $\mathbf{\Omega}_F$  and between equations (8) and (9) to obtain the MLEs of  $\mathbf{\Omega}_O$ .

### Hypothesis Tests

It is imperative to know whether there exists a significant association between a specific SNP and a complex trait and how a significant SNP triggers an additive, dominant, or imprinting effect on the trait. To test for the overall significant association of SNP genotype and trait phenotype, we generate the following hypotheses:

$$H_0 : a_F = d_F = i_F = a_O = d_O = i_O = 0$$

$H_1$  : At least one of these equalities above does not hold.

The log-likelihood ratio under the null and alternative hypotheses is calculated. Since the null hypothesis contains a nuisance parameter, allele frequency, this log-likelihood ratio test statistic may have an unclear distribution. For this reason, the critical threshold for claiming the existence of a significant SNP is determined from permutation tests [37]. If our interest is in testing whether there is an additive, dominant, or imprinting effect, the null hypothesis should be  $H_0 : a_F = a_O = 0$ ,  $H_0 : d_F = d_O = 0$ , and  $H_0 : i_F = i_O = 0$ , respectively. Because each of these null hypotheses is nested within its alternative, the log-likelihood ratio test statistic can be thought to asymptotically follow a  $\chi^2$ -distribution for a large sample size.

The transgenerational changes of different genetic effects can also be tested. The null hypotheses used to test whether the additive, dominant, and imprinting effects display significant changes from one generation to next are expressed as  $H_0 : \Delta_a = 0$ ,  $H_0 : \Delta_d = 0$ , and  $H_0 : \Delta_i = 0$ , respectively. These null hypotheses can be considered singly or jointly, in order to better study the transgenerational changes of the genetic architecture of a trait.

### Haplotyping Model

Recent molecular surveys suggest that the human genome contains many discrete haplotype blocks that are sites of closely located SNPs [38,39,40]. Each block may have a few common haplotypes which account for a large proportion of chromosomal variation. Between adjacent blocks are there large regions, called hotspots, in which recombination events occur with high frequencies. Several algorithms have been developed to identify a minimal subset of SNPs, i.e., tagging SNPs, that can characterize the most common haplotypes [41]. The number and type of tagging SNPs within each haplotype block can be determined prior to association studies. In this section, we will derive a model for detecting the association between haplotypes constructed by alleles at a set of SNPs and complex traits.

For the simplicity of our description, consider two SNPs **A** (with two alleles *A* and *a*) and **B** (with two alleles *B* and *b*). They form four haplotypes *AB*, *Ab*, *aB*, and *ab*, of which one that is distinct from the rest three is defined as a risk haplotype *W* and all the others are defined as a non-risk haplotype *w* [42]. Risk and non-risk haplotypes from the maternal and paternal parents generate four composite diplotypes, *W|W*, *W|w*, *w|W*, and *w|w*, whose genotypic values are described by the additive (*a*), dominant (*d*), and imprinting genetic effects (*i*). Cheng et al. [43] and Wang et al. [44] proposed a two- and three-SNP model for estimating and testing genetic imprinting effects in a natural population, respectively. Wu et al.'s procedure [45] allows the choice of an optimal number and combination of risk haplotypes within a multiallelic model framework. Here, we adopted Cheng et al.'s two-SNP model to estimate haplotype imprinting genetic effects and their transgenerational change.

In this example, four haplotypes *AB*, *Ab*, *aB*, and *ab* have frequencies denoted as  $p_{11}$ ,  $p_{10}$ ,  $p_{01}$ , and  $p_{00}$ , respectively. The two SNPs yield nine joint genotypes, *AABB* (coded as 1), *AABb* (coded as 2), ..., *aabb* (coded as 9), which are actually observed. Each subject must bear one of these genotypes, and the parents in each family will be one of  $9 \times 9 = 81$  possible genotype by genotype combinations. If each parent for a combination is homozygous for both SNPs, their offspring will have one genotype. As long as one parent is heterozygous for one SNP, the offspring will have two or more genotypes. However, only when both SNPs are heterozygous for at least one parent, the genotype frequencies of offspring will be determined by the recombination fraction between the markers (*r*). Tables S2 and S3 show the structure and frequencies of mother by father genotype combinations under random mating and their offspring genotype frequencies in the second and third generation, respectively. For a double heterozygote *AaBb*, its observed genotype may be derived from two possible diplotypes, *AB|ab* (with the relative proportion of  $\phi = \frac{p_{11}p_{00}}{p_{11}p_{00} + p_{10}p_{01}}$ ) or *Ab|aB* (with the relative proportion of  $1 - \phi = \frac{p_{10}p_{01}}{p_{11}p_{00} + p_{10}p_{01}}$ ). Each of these two diplotypes produce four haplotypes *AB*, *Ab*, *aB*, and *ab*, whose frequencies are expressed as

		Haplotype			
Diplotype	Proportion	<i>AB</i>	<i>Ab</i>	<i>aB</i>	<i>ab</i>
<i>AB ab</i>	$\phi$	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$
<i>Ab aB</i>	$1-\phi$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$

A similar likelihood (5) can be formulated for haplotype models. A complicated EM algorithm is derived to estimate haplotype frequencies using the parental information. Let  $N_{ij}$  denote the observation of mating type between genotype *i* for one parent and genotype *j* for the second parent. In the E step, calculate the proportion of a diplotype for a heterozygous genotype for a particular mating design by

$$\psi_1 = \frac{(1-\phi)r}{\omega_1}$$

$$\psi_2 = \frac{\phi r}{\omega_2}$$

$$\psi_3 = \frac{(1-\phi)r + \phi r}{\omega_1 + \omega_2}$$

$$\psi_4 = \frac{(1-\phi)^2 r^2 + \phi(1-\phi)r(1-r)}{\omega_1^2}$$

$$\psi_5 = \frac{\phi^2 r^2 + \phi(1-\phi)r(1-r)}{\omega_2^2}$$

$$\psi_6 = \frac{\phi(1-\phi)r^2 + [\phi^2 + (1-\phi)^2]r(1-r)}{2\psi_1\omega_2}$$

$$\psi_7 = \frac{\phi^2 r^2 + 2\phi(1-\phi)r(1-r) + (1-\phi)^2 r^2}{2(\omega_1^2 + \phi_2^2)},$$

where  $\omega_1 = (1-\phi)r + \phi(1-r)$ ,  $\omega_2 = \phi r + (1-\phi)(1-r)$ .

In the M step, estimate the haplotype frequencies and recombination fraction by

$$p_{11} = \{4N_{11} + 3(N_{12} + N_{21} + N_{14} + N_{41}) + 2(N_{22} + N_{44} + N_{13} + N_{31} + N_{16} + N_{61} + N_{17} + N_{71} + N_{18} + N_{81} + N_{19} + N_{91} + N_{24} + N_{42}) + N_{23} + N_{32} + N_{26} + N_{62} + N_{27} + N_{72} + N_{28} + N_{82} + N_{29} + N_{92} + N_{34} + N_{43} + N_{46} + N_{64} + N_{47} + N_{74} + N_{48} + N_{84} + N_{49} + N_{94} + \phi[3(N_{15} + N_{51}) + 2(N_{25} + N_{52} + N_{45} + N_{54}) + N_{35} + N_{53} + N_{56} + N_{65} + N_{57} + N_{75} + N_{58} + N_{85} + N_{59} + N_{95}] + (1-\phi)[2(N_{15} + N_{51}) + N_{25} + N_{52} + N_{45} + N_{54}] + 2\phi N_{55}\} / (4 \sum_{i=1, j=1}^{9,9} N_{ij}),$$

$$p_{10} = \{4N_{33} + 3(N_{23} + N_{32} + N_{36} + N_{63}) + 2(N_{22} + N_{66} + N_{13} + N_{31} + N_{26} + N_{62} + N_{34} + N_{43} + N_{37} + N_{73} + N_{38} + N_{83} + N_{39} + N_{93}) + N_{12} + N_{21} + N_{16} + N_{61} + N_{24} + N_{42} + N_{27} + N_{72} + N_{28} + N_{82} + N_{29} + N_{92} + N_{46} + N_{64} + N_{67} + N_{76} + N_{68} + N_{86} + N_{69} + N_{96} + \phi[2(N_{35} + N_{53}) + N_{25} + N_{52} + N_{56} + N_{65}] + (1-\phi)[3(N_{35} + N_{53}) + 2(N_{25} + N_{52} + N_{56} + N_{65}) + N_{15} + N_{51} + N_{45} + N_{54} + N_{57} + N_{75} + N_{58} + N_{85} + N_{59} + N_{95}] + 2(1-\phi)N_{55}\} / (4 \sum_{i=1, j=1}^{9,9} N_{ij}),$$

$$p_{01} = \{4N_{77} + 3(N_{47} + N_{74} + N_{78} + N_{87}) + 2(N_{44} + N_{88} + N_{17} + N_{71} + N_{27} + N_{72} + N_{37} + N_{73} + N_{67} + N_{76} + N_{48} + N_{84} + N_{79} + N_{97}) + N_{41} + N_{18} + N_{81} + N_{24} + N_{42} + N_{28} + N_{82} + N_{34} + N_{43} + N_{38} + N_{83} + N_{46} + N_{64} + N_{49} + N_{94} + N_{68} + N_{86} + N_{89} + N_{98} + \phi[2(N_{57} + N_{75}) + N_{45} + N_{54} + N_{58} + N_{85}] + (1-\phi)[3(N_{57} + N_{75}) + 2(N_{45} + N_{54} + N_{58} + N_{85}) + N_{15} + N_{51} + N_{25} + N_{52} + N_{35} + N_{53} + N_{56} + N_{65} + N_{59} + N_{95}] + 2(1-\phi)N_{55}\} / (4 \sum_{i=1, j=1}^{9,9} N_{ij}),$$

$$p_{00} = \{4N_{99} + 3(N_{69} + N_{96} + N_{89} + N_{98}) + 2(N_{66} + N_{88} + N_{19} + N_{91} + N_{29} + N_{92} + N_{39} + N_{93} + N_{49} + N_{94} + N_{68} + N_{86} + N_{79} + N_{97}) + N_{16} + N_{61} + N_{18} + N_{81} + N_{26} + N_{62} + N_{28} + N_{82} + N_{36} + N_{63} + N_{38} + N_{83} + N_{46} + N_{64} + N_{48} + N_{84} + N_{67} + N_{76} + N_{78} + N_{87} + \phi[3(N_{59} + N_{95}) + 2(N_{56} + N_{65} + N_{58} + N_{85}) + N_{15} + N_{51} + N_{25} + N_{52} + N_{35} + N_{53} + N_{45} + N_{54} + N_{57} + N_{75}] + (1-\phi)[2(N_{59} + N_{95}) + N_{56} + N_{65} + N_{58} + N_{85}] + 2\phi N_{55}\} / (4 \sum_{i=1, j=1}^{9,9} N_{ij}).$$

$$r = \{\psi_1(N_{51}^1 + N_{51}^5 + N_{52}^1 + N_{52}^6 + N_{53}^2 + N_{53}^6 + N_{54}^1 + N_{54}^8 + N_{56}^2 + N_{56}^9 + N_{57}^4 + N_{57}^8 + N_{58}^4 + N_{58}^9 + N_{59}^9 + N_{59}^9 + N_{15}^1 + N_{15}^5 + N_{25}^1 + N_{25}^6 + N_{35}^2 + N_{35}^6 + N_{45}^1 + N_{45}^8 + N_{65}^2 + N_{65}^9 + N_{75}^4 + N_{75}^8 + N_{85}^4 + N_{85}^9 + N_{95}^9 + N_{95}^9) + \psi_2(N_{51}^2 + N_{51}^4 + N_{52}^3 + N_{52}^4 + N_{53}^3 + N_{53}^5 + N_{54}^2 + N_{54}^7 + N_{56}^3 + N_{56}^8 + N_{57}^5 + N_{57}^7 + N_{58}^6 + N_{58}^7 + N_{59}^6 + N_{59}^8 + N_{15}^2 + N_{15}^4 + N_{25}^3 + N_{25}^4 + N_{35}^3 + N_{35}^5 + N_{45}^2 + N_{45}^7 + N_{65}^3 + N_{65}^8 + N_{75}^5 + N_{75}^7 + \psi_3(N_{52}^2 + N_{52}^5 + N_{54}^4 + N_{54}^5 + N_{56}^5 + N_{56}^6 + N_{58}^5 + N_{58}^8 + N_{25}^2 + N_{25}^5 + N_{45}^4 + N_{45}^5 + N_{85}^6 + N_{85}^7 + N_{95}^6 + N_{95}^8) + \psi_4(N_{55}^1 + N_{55}^9) + N_{65}^5 + \psi_5(N_{55}^3 + N_{55}^7) + \psi_6(N_{55}^2 + N_{55}^4 + N_{55}^4 + N_{55}^6 + N_{55}^8) + \psi_7 N_{55}^5\} / (\sum_{j=1}^9 N_{5j} + \sum_{i=1}^9 N_{i5} - N_{55})$$

In the M step, the equations for estimating additive, dominant, imprinting effects expressed in paternal and offspring generations are also derived. The E and M steps are iterated until the estimates converge to a stable value. These stable values are the maximum likelihood estimates (MLEs) of parameters. The estimated haplotype frequencies and recombination fraction are embedded into a mixture model for estimating genotypic values and variances for different generations.

### Supporting Information

**Table S1** A three-generation family design used to study transgenerational inheritance. (PDF)

**Table S2** A three-generation family design showing how to produce the second generation by mating different genotypes of grandfathers and grandmothers sampled from a natural population. (PDF)

**Table S3** A three-generation family design showing how to produce the second generation by mating different genotypes of grandfathers and grandmothers sampled from a natural population. (PDF)

## Author Contributions

Conceived and designed the experiments: RW. Performed the experiments: JW DL. Analyzed the data: YL YG WH JW MNC. Contributed reagents/materials/analysis tools: YL WH JW. Wrote the paper: RW YL.

## References

- Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2: 21–32.
- Wilkins JF, Haig D (2003) What good is genomic imprinting: The function of parent-specific gene expression. *Nat Rev Genet* 4: 359–368.
- Itier JM, Trempe G, Léonard JF, Multon MC, et al. (1998) Imprinted gene in postnatal growth role. *Nature* 393: 125–126.
- Li LL, Keverne EB, Aparicio SA, Ishino F, et al. (1999) Regulation of maternal behaviour and offspring growth by paternally expressed Peg3. *Science* 284: 330–333.
- Isles AR, Wilkinson LS (2000) Imprinted genes, cognition and behaviour. *Trend Cogn Sci* 4: 309–318.
- Constancia M, Kelsey G, Reik W (2004) Resourceful imprinting. *Nature* 432: 53–57.
- Wood AJ, Oakey RJ (2006) Genomic imprinting in mammals: Emerging themes and established theories. *PLoS Genet* 2(11): e147.
- Wilkinson LS, Davies W, Isles AR (2007) Genomic imprinting effects on brain development and function. *Nat Rev Neurosci* 4: 1–19.
- Wang CG, Wang Z, Luo JT, Li Q, et al. (2010) A model for transgenerational imprinting variation in complex traits. *PLoS ONE* 5(7): e11396.
- Frost JM, Moore GE (2010) The importance of imprinting in the human placenta. *PLoS Genet* 6(7): e1001015.
- Sha K (2008) A mechanistic view of genomic imprinting. *Ann Rev Genom Hum Genet* 9: 197–216.
- De Koning DJ, Rattnik AP, Harlizius B, Arendonk JAM, et al. (2000) Genome-wide scan for body composition in pigs reveals important role of imprinting. *Proc Natl Acad Sci U S A* 97: 7947–7950.
- Liu T, Todhunter RJ, Wu S, Hou W, et al. (2007) A random model for mapping imprinted quantitative trait loci in a structured pedigree: An implication for mapping canine hip dysplasia. *Genomics* 90: 276–284.
- Cheverud JM, Hager R, Roseman C, Fawcett G, et al. (2008) Genomic imprinting effects on adult body composition in mice. *Proc Natl Acad Sci U S A* 105: 4253–4258.
- Wolf JB, Cheverud JM, Roseman C, Hager R (2008) Genome-wide analysis reveals a complex pattern of genomic imprinting in mice. *PLoS Genet* 4: e1000091.
- Li YC, Coelho CM, Liu T, Wu S, et al. (2007) A statistical strategy to estimate maternal-zygotic interactions and parent-of-origin effects of QTLs for seed development. *PLoS ONE* 3: e3131.
- Morgan HD, Santos F, Green K, Dean W, et al. (2005) Epigenetic reprogramming in mammals. *Hum Mol Genet* 14: R47–R58.
- Sasaki H, Matsui Y (2008) Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* 9: 129–140.
- Tal O, Kisdí E, Jablonka E (2010) Epigenetic contribution to covariance between relatives. *Genetics* 184: 1037–1050.
- McGrath J, Solter D (1984) Inability of mouse blastomere nuclei transferred to enucleated zygotes to support development in vitro. *Science* 226: 1317–1319.
- Surani MA, Barton SC, Norris ML (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* 308: 548–550.
- Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 23: 314–318.
- Cropley JE, Suter CM, Beckman KB, Martin DI (2006) Germ-line epigenetic modification of the murine Avy allele by nutritional supplementation. *Proc Natl Acad Sci U S A* 103: 17308–17312.
- Skinner MK (2008) What is an epigenetic transgenerational phenotype? F3 or F2. *Reprod Toxic* 25: 2–6.
- Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL (2006) Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environ Health Perspect* 114: 567–572.
- Whitelaw NC, Whitelaw E (2008) Transgenerational epigenetic inheritance in health and disease. *Curr Opin Genet Dev* 18: 273–279.
- Youngson NA, Whitelaw E (2008) Transgenerational epigenetic effects. *Ann Rev Genom Hum Genet* 9: 233–257.
- Pembrey ME, Bygren LO, Kaati G, Edvinsson S, et al. (2006) Sex-specific, male-line transgenerational responses in humans. *Europ J Hum Genet* 14: 159–166.
- Wu RL, Ma CX, Casella G (2007) *Statistical Genetics of Quantitative Traits: Linkage, Map, and QTLs*. New York: Springer.
- Li Q, Wu RL (2009) A multilocus model for constructing a linkage disequilibrium map in human populations. *Stat Appl Genet Mol Biol* 8, Iss. 1, Article 18.
- Chan EY (2005) Advances in sequencing technology. *Mutant Res* 573: 13–40.
- Beckmann JS, Estivill X, Antonarakis SE (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* 8: 639–646.
- Weinberg CR, Wilcox AJ, Lie RT (1998) A log-linear approach to case-parent triad data: Assessing effects of disease genes that act directly or through maternal effects, and may be subject to parental imprinting. *Am J Hum Genet* 62: 969–978.
- Cordell HJ, Barratt BJ, Clayton DG (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.
- Hager R, Cheverud JM, Wolf JB (2008) Maternal effects as the cause of parent-of-origin dependent effects that mimic genomic imprinting. *Genetics* 178: 755–762.
- Jirtle RL, Skinner MK (2007) Environmental epigenomics and disease susceptibility. *Nat Rev Genet* 8: 253–262.
- Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963–971.
- Dawson E, Abecasis GR, Bumpstead S, Chen Y, et al. (2002) A first-generation linkage disequilibrium map of human chromosome. *Nature* 418: 544–548.
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, et al. (2002) The structure of haplotype blocks in the human genome. *Science* 296: 2225–2229.
- Patil N, Bero AJ, Hinds DA, Barrett WA, et al. (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 294: 1719–1723.
- Zhang K, Deng M, Chen T, Waterman MS, Sun F (2002) A dynamic programming algorithm for haplotype block partitioning. *Proc Natl Acad Sci U S A* 99: 7335–7339.
- Wu RL, Lin M (2008) *Statistical and Computational Pharmacogenomics*. London: Chapman & Hall/CRC.
- Cheng Y, Berg A, Wu S, Li Y, Wu RL (2009) Computing genetic imprinting expressed by haplotypes. *Method Mol Biol* 573: 189–212.
- Wang CG, Cheng Y, Liu T, Li Q, et al. (2008) A computational model for sex-specific genetic architecture of complex traits in humans. *Mol Pain* 4: 13.
- Wu S, Yang J, Wang CG, Wu RL (2007) A general quantitative genetic model for haplotyping a complex trait in humans. *Curr Genom* 8: 343–350.