# Inference of maternal allele inheritance via hierarchical Bayesian model in noninvasive prenatal diagnosis

Chunxuan Shao[1,2]

[1]Division of Theoretical Systems Biology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany, [2]Bioquant Center, University of Heidelberg, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany

## Abstract

Noninvasive prenatal diagnosis (NIPD) poses a promising solution for detecting genetic alterations in fetus genome. However, inference the maternal allele inheritance of monogenic autosomal recessive disease is still challenging. We perform the null hypothesis testing of haplotype frequency using a hierarchical Bayesian model to deduce the allele inheritance. The Bayesian approach, which does not depend on the fetus DNA proportion in maternal plasma, provides accurate estimations on real and simulated datasets, moreover, it is most robust than current methods in analyzing noisy or even erroneous datasets.

## Introduction

The discovery of free fetal DNA in maternal plasma starts a new era of non-invasive prenatal diagnosis (NIPD).[1] The placenta origin cell-free DNA could make up to 19% of total cell-free DNA in plasma, providing necessary materials for enormous applications.[2] For example, sex determination by RT-PCR could achieve near 100% sensitivity and specificity.[3,4] More recently, with the advent of next generation sequencing, NIPD of aneuploidy and subchromosomal abnormalities are feasible or even recommended for high risk patients in the clinical practices. [5–7]

Based on the causative mutations identified in monogenic autosomal recessive diseases, NIPD is of great value in clinical diagnoses and treatments by determining the parental alleles inheritance in fetus genome.[10]

However, it is still challenging to infer the inheritance of heterogeneous maternal alleles in recessive disease. Relied on the imbalance of maternal haplotypes in maternal plasma DNA mixed with fetus DNA, relative haplotype dosage analysis (RHDO)[8–10], haplotype counting analysis[11] and hidden Markov model (HMM)[12] have been employed in the maternal alleles inference. However, the proportion of fetus DNA of total cell-free DNA in mother plasma, which is a key parameter in above methods, might not be accurately estimated.

In the framework of RHDO, the sequential probability ratio test (SPRT)[8] is employed to perform the hypothesis testing that the haplotype frequency of heterozygous SNPs is not significantly different from 0.5, and the over-/under-representation of heterozygous SNPs suggest which allele is inherited in fetus genome. The data and procedure used by SPRT reminds us the classical coin bias problem in Bayesian applications, where several coins minted in a factory are tossed and the head/tail data are used to deduce the mint bias of the factory. A Head biased factory tends to produce in more heads in the tossing procedure. Analogous to coin bias problem, we could consider SNPs as coins, the haplotype of heterozygous maternal SNPs as the mint, the haplotype frequency as the mint bias. The null hypothesis is that the haplotype frequency is equal to 0.5, indicating there is no fetus DNA in maternal plasma. We then deduce the maternal allele inheritance via the null hypothesis testing procedure based on the relative location between the region of practical equivalence (ROPE) and high density interval (HDI). We showed that the Bayesian approach provides informative results and has excellent performance on real and simulated data. More importantly, this approach does not depend on the fetus DNA fraction in plasma, thus it is more robust to noisy or even erroneous datasets. This Bayesian approach is implemented in an R package (maIHB, https://github.com/ccshao/maIHB) to facilitate its usage.

## Methods

### Relative haplotype dosage analysis

Relative haplotype dosage analysis (RHDO), which makes use of next generation sequencing data of heterozygous maternal SNPs and

homozygous paternal SNPs flanking interested alleles, aims to identify the subtle haplotype imbalance caused by the mixture of fetus DNA and thereafter deduces maternal alleles inheritance in fetus genome[9]. Briefly, type α or type β SNPs are selected from maternal haplotypes (HapI and HapII) based on their identity to corresponding paternal alleles, where maternal SNPs in HapI that are same to the paternal SNPs are identified as type α SNPs, and maternal SNPs in HapII that are same to the paternal SNPs are identified as type β SNPs. Then the sequential probability ratio test (SPRT),[8,9] which cumulatively tests the haplotype imbalance against predefined threshold, is used to determine if the proportion of either haplotype is significantly different from 0.5. This method has been extended to the scenarios where both maternal and paternal SNPs are heterozygous[9].

**Hierarchical Bayesian model**

Thinking of a classical application of Bayesian analysis: several coins, which are independently minted in the same factory, are tossed a certain number of times to calculate the bias, i.e., the probability that coins comes up heads. The bias values are likely to be different for each coin but will be close to the mint bias. The hierarchical Bayesian model is showed to be a mature and powerful tool in estimating the distribution of mint bias from coins tossing data via specifying the dependence of coin bias on mint bias with hyperparameters[13]. It is interesting to consider the RHDO analysis in the simple coin flipping setting, where the plasma DNA is regard as the “mint”, type α or type β SNPs are “coins”, read counts of SNPs on HapI and HapII are numbers of heads and tails of tossing respectively, and the proportion of HapI or HapII could be viewed as the “mint bias”. If there is no fetus DNA in maternal plasma, the estimated distributions of HapI and HapII will be centralized on 0.5, and shift to higher value or lower value if fetus DNA are mixed. We refer the proportion of HapI as HapBias in the following analysis hereafter.

The structure of hierarchical Bayesian model is summarized in Figure 1 in a bottom-up way. For each SNP, the read counts assigned to SNPs could be viewed as numbers of heads or tails generated in a series of Bernoulli process with a SNP specific bias $\theta_s$ respectively, which is commonly modeled by a beta distribution with mode $\omega$ and concentration $\kappa$, i.e., beta($\omega$($\kappa$-2)+1,

(1-$\omega$)($\kappa$-2)+1). The individual value of $\theta_s$ will be near to $\omega$, and $\kappa$ controls how close $\theta_s$ is to $\omega$. Thus $\omega$ represents the HapBias, and $\kappa$ expresses the certainty on the dependence between $\theta_s$ and $\omega$. The interested hyperparameter $\omega$ is further modeled by a beta distribution, and $\kappa$-*2* is usually modeled by a gamma distribution. We set the prior distribution of $\omega$ to be beta(2, 2) to reflect the fact that the HapBias is close to 0.5 and let the data play a major role in determining the posterior distribution. The Markov chain Monte Carlo simulation is used to approximate the posterior distribution of $\theta_s$, $\kappa$ and $\omega$.

Two outputs are possible for type α SNPs based analysis: HapBias is very close to 0.5 or HapBias is much larger than 0.5, and the latter scenario indicates that HapI is overrepresented and hence inherited by the fetus. Similarly, two outputs are possible for type β based analysis: HapBias is very close to 0.5 or HapBias is much smaller than 0.5, which indicates HapI is underrepresented and HapII is inherited by the fetus. The significance of over/under-represention of haplotype frequency is determined by the null hypothesis testing based on the relative location between high density interval (HDI) of the distribution of HapBias $\omega$ and region of practical equivalence (ROPE), which described in the next section.

**Null hypothesis testing in Bayesian analysis**

In the context of the allele inheritance deducing, the null condition is that the HapBias $\omega$ is equal to 0.5. In Bayesian analysis, a common approach to perform the null hypothesis testing is to examine the relative location between highest density interval of posterior distribution (HDI) and region of practical equivalence (ROPE).[13,14] Parameter values inside HDI have higher probability density comparing with values outside this region, and sum to a defined probability (95% in our analysis).[13] ROPE, or alternatively named indifference zone, is popularized by Freedman et al. as the counterpart to the single null value used in frequentist field.[15] ROPE means the parameter values within this range are recognized to be practically equivalent for the null value. We denote ($\Delta_L$, $\Delta_H$) as the limits of HDI and ($\delta_L$, $\delta_H$) as the limits of ROPE. Six types of relation location between HDI and ROPE exist (Fig. 2).[16] We reject the null hypothesis that the HapBias is equal to 0.5 if there is no intersection between the HDI and the ROPE, and accept the null hypothesis if the ROPE

completely contains the HDI. We make no decision for the other three types of relations and need to collect more data.

**Limits of ROPE of the null hypothesis**

The limits of ROPE vary according to practical purposes and overall considerations of experiments. In principle, wider ROPEs lead to the acceptance of null value, and vice versa for narrower ROPEs. We calculated the ROPE limits from the perspective of next-generation sequencing error. Assuming the error rate of sequencing platform is *e*, the total read counts for a single SNP on both HapI and HapII is *C*. In the null condition, the lower and upper limits of read counts of a SNP on HapI is *C/2 - C*e/2* and *C/2 + C*e/2*, respectively. Therefore, independent of the sequencing depth, the proportion of reads on HapI of a single SNP lies in the range [1/2 – e/2, 1/2 + e/2]. As the sample principle applies to all SNPs, the range is considered to be true for the haplotype as well. Thus the range [1/2 – e/2, 1/2 + e/2] is used as the ROPE limits of HapBias of the null hypothesis.

**Simulation**

We simulated the read counts data of haplotypes via the following steps: 1) we focused on the type α SNPs in the scenarios that the HapI was overrepresented; 2) the proportion of fetus DNA was set to be *p*; 3) the sequencing depth is set to be *C* for all SNPs, here we set *C* to be 200 4) we used 500 SNPs and 1000 SNPs in the simulations; 5) the read counts for SNPs on HapI and HapII are (1+p)*C/2 and (1-p)*C/2, respectively; 6) we sampled data from the pool of HapI and HapII with replacement to reflect the stochastic noise in the real data. We simulated another dataset with *p* was set to be 0 to reflect that there is no fetus DNA in the maternal plasma.

## Results

**Application of the hierarchical Bayesian model on read data sets**

We evaluated the hierarchical Bayesian model on noninvasive prenatal data sets to access its accuracy. The data are from two β-thalassemia patient families where SNPs flanking interested genes are sequenced to generate read counts on HapI and HapII.[9] To obtain the posterior distribution of HapBias, we performed 100,000 MCMC simulations with four chains (i.e., four individual simulation with difference initiation value). The ROPE of (0.49, 0.51)

(the rationale will be discussed in a later section) and the 95% HDI of HapBias posterior distribution were used to perform null hypothesis testings. In the first family the fetus inherited the HapII, and there are five type α and 46 type β SNPs available. The posterior distribution of HapBias using type β SNP demonstrated a 95% HDI region with limits of (0.435, 0.478), lied completely downstream of the ROPE, indicating HapI is underrepresented and HapII inherited in fetus, consistent with the publication results (Fig. 3A). In fact, the simulation results suggested that there is no HapBias value locating in the ROPE. On the other hand, the posterior distribution of HapBias based on type α SNPs showed that the ROPE completely fall within the HDI region, thus no conclusion could be drawn (Supplementary Fig. 1).

As Majority of analyzed SNPs in both parents in the second family are heterozygous, additional tagging SNPs were used to identify 24 type α SNPs.[9] The hierarchical Bayesian model revealed that the ROPE lied completely upstream of 95% HDI, indicating the HapI was inherited in the fetus, consistent with known results (Fig. 3B). Taken together, the hierarchical Bayesian model correctly deduced the maternal allele inheritance and represented the results in an intuitive way.

**Examination of Markov chain Monte Carlo simulation**

It is important to check the quality of MCMC simulations used in the Bayesian model. We employed various visualization methods to examine two critical characters, convergence and accuracy, in above examples. The trace plot, which shows the parameter in simulated chain steps, is useful to provide an overview of the convergence.[17,18] If MCMC simulations converged for the posterior distribution, the parameter value estimated from different chains should largely overlap. The existing of orphan chains are alarms of lack of convergence. The overlapping of trace plots suggested the MCMC simulations converged in both family one data and family two data (Fig. 4A, Supplementary Fig. 2A). Meanwhile, we summarized the parameter value sampled in individual chains as density plot. The overlapping of density plot suggests convergence of simulation (Fig. 4B, Supplementary Fig. 2B). In addition, we could see there are slightly difference of 95% HDI among chains. As the limits of HDI will converge to the same value in the condition of infinite sample sizes, the difference of HDI provides hints on how the finite samples

sizes influence posterior estimation. In other words, the largely overlapping of 95% HDI in our results suggests that the sample sizes are sufficient. The shrink factor, which reflects the ratio between between-chain variance and within-chain variance during simulation, is a commonly used numerical measurement for convergence. The shrink factor is equal to one if all chains converged, and larger than one if orphan chains exist.[19] Practically, the MCMC simulations with the shrink factor greater than 1.1 indicate lack of convergence.[13] The mean and 97.5% quantile value of shrink factors gets close to 1 very quickly after the initial burn-in period in our data (Fig. 4C, Supplementary Fig. 2C), indicating the convergence is archived.

Another goal is to generate enough simulation steps to accurately represent the parameter distribution. One problem in the simulation is clumpy chains, in which successive steps are partially redundant in exploring the parameter space, and do not provide independent estimation. The autocorrelation plot, calculated as the average correlation between data and k steps ahead data, provides hits on level of clumpiness. Based on the value of autocorrelation, we could further calculate the effective sample size (ESS), which summarize how many completely non-autocorrelated steps in our simulation.[20] Heuristically, an ESS of 10,000 is recommended for many MCMC simulations.[13] In our results, we could see autocorrelation value gets close to 0 very quickly for k great than 6, while 37881 ESS were achieved (Fig. 4D, ESS is 34887 in Supplementary Fig. 2D). Together with the overlapping of HDI of individual chains in Figure 4B, we could see that the simulated posterior distributions have sufficient accuracy for tested data sets.

**Calculation the limits of the region of practical equivalence**

We discuss the rationality of determining ROPE in this section. ROPE, which is alternatively named “indifference zone” or “range of equivalence”, means the parameter values within this range are considered to be practically equivalent for the null value[15]. Small difference between null value and estimated value are most likely caused by sampling variation or data error, rather than true difference. In our work, the null hypothesis is that HapI and HapII are equally presented and HapBias is 0.5. For the data generated by next generation sequencing methods, we assumed that the ROPE limits are mainly influenced by sequencing errors,[21] while higher sequencing error rates

lead to wider ROPE. Currently the error rate of next-generation sequencing platform is around $10^{-2}$,[21,22] and an error rate of 0.303% has been reported in one NIPD study.[8] We applied a ROPE of (0.49, 0.51) in the tested data sets, reflecting an error rate of 2% (See Methods). Interestingly, even with this rather conservative ROPE limits, we still successfully inferred the allele inheritance in the real data sets, suggesting the sensitivity of the hierarchical Bayesian model.

**Simulation**

To further explore the performance of the hierarchical Bayesian model, we generated the sequencing data *in silico*. Briefly, we sampled the SNPs counts for HapI and HapII according to the proportion of fetus DNA from a pool of reads, with 500 and 1000 SNPs used, respectively. In the condition of 3% fetus DNA, the Bayesian approach correctly detected the maternal allele inherited in all samples using 1000 SNPs, while the detection rate reduced to 76% using 500 SNPs, suggesting the benefits of larger data (Supplementary Table 1). SPRT could correctly deduce the maternal allele inheritance in 80% of simulated samples with as low as 0.8% fetus DNA using 1000 SNPs, showing a higher sensitivity (Supplementary Table 2). Taken together, both Bayesian approach and SPRT have great sensitivity in deducing the true maternal allele inheritance.

We next asked how robust these methods are in the existence of measurement errors. The fetus DNA proportion is a key parameter in SPRT procedure as it affects the low and up-boundary of the test statistic under null hypothesis. However, the fetus DNA proportion might not be accurately estimated. We counts data of 500 SNPs in the conditions that the true fetus DNA proportion was set to be 0% or 10%, and the measured fetus DNA proportion varied from 1% to 20%. In the scenarios that the true fetus DNA proportion is 0%, SPRT provided the wrong inference in all tested data sets as it relied on the fetus DNA proportion for the significance test (Figure 5A). In the condition that the true fetus DNA proportion is 10% but the estimated value is 20%, SPRT inferred 61% of test samples inherited HapII or undetermined inheritance (Supplementary Table 3). On the contrast, the Bayesian approach does not depend on the fetus DNA proportion. When the Bayesian approach applied to the simulated data with 0% true fetus DNA

proportion, the HDI of HapBias posterior distribution fell completely within the ROPE, indicating that the HapBias was equal to 0.5 and there is no fetus DNA in the sequenced data (Fig. 5B).

**Discussion**

Although several methods are available to analyze the cell-free DNA sequencing data, inference the inheritance of heterogeneous maternal alleles in fetus genome from NIPD sequencing data is a challenging task.[23] Haplotypes based approaches which uses the imbalance representation of SNPs introduced by fetus DNA provide convincing results[9,12,24]. However, the performance of these methods relies on the value of fetus DNA proportion, which might be not feasible to accurately estimate. We introduced a hierarchical Bayesian model under the framework of RHDO that is independent of fetus DNA proportion to inference the inherited maternal alleles in fetus genome. To test the null hypothesis that the haplotype frequency is equal to 0.5, we established the posterior distribution of Hapbias (proportion of HapI) based on high throughput sequencing data of individual SNPs, and compared the HDI of HapBias distribution with predefined ROPE to determine which allele is inherited in fetus genome. This method correctly deduced maternal allele inheritances on experimental data, and the results are summarized as intuitive graphs, together with various diagnoses plots.

We evaluated the performance of the Bayesian approach and another widely used approach, SPRT, on simulated data. Both methods could successfully identify the inherited alleles in all samples containing as low as 3% fetus DNA, and SPRT performs well even in samples with 0.8% fetus DNA. Moreover, we found that the Bayesian approach is more robust in conditions where the fetus DNA proportion is not correctly measured. In the extreme case where there is actually no fetus DNA in the plasma but incorrectly recorded, SPRT could provide misleading test results. In contrast, the Bayesian model showed that the HDI of HapBias posterior distribution falls completely within ROPE, corresponding to the hypothesis that there is no fetus DNA in the plasma. Taken together, the hierarchical Bayesian model is a robust tool in deducing the allele inheritance in NIPD.

## Acknowledgement

We thank Dr. Y M Dennis Lo and Dr. JIANG Peiyong for kindly providing the SRPT codes.

## Conflict of interest

The authors declare that they have no conflict of interest.

## REFERENCES

1. Lo YD, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, et al. Presence of fetal DNA in maternal plasma and serum. The Lancet. 1997;350(9076):485–487.

2. Lun FMF, Chiu RWK, Allen Chan KC, Yeung Leung T, Kin Lau T, Dennis Lo YM. Microfluidics Digital PCR Reveals a Higher than Expected Fraction of Fetal DNA in Maternal Plasma. Clin Chem. 2008 Aug 14;54(10):1664–72.

3. Galbiati S, Smid M, Gambini D, Ferrari A, Restagno G, Viora E, et al. Fetal DNA detection in maternal plasma throughout gestation. Hum Genet. 2005 Jul;117(2–3):243–8.

4. Devaney SA, Palomaki GE, Scott JA, Bianchi DW. Noninvasive Fetal Sex Determination Using Cell-Free Fetal DNA: A Systematic Review and Meta-analysis. JAMA [Internet]. 2011 Aug 10 [cited 2016 Apr 18];306(6). Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2011.1114

5. Zhao C, Tynan J, Ehrich M, Hannum G, McCullough R, Saldivar J-S, et al. Detection of fetal subchromosomal abnormalities by sequencing circulating cell-free DNA from maternal plasma. Clin Chem. 2015 Apr;61(4):608–616.

6. Yin A, Peng C, Zhao X, Caughey BA, Yang J, Liu J, et al. Noninvasive detection of fetal subchromosomal abnormalities by semiconductor sequencing of maternal plasma DNA. PNAS. 2015 Nov;112(47):14670–14675.

7. Ferres MA, Hui L, Bianchi DW. Antenatal Noninvasive DNA Testing: Clinical Experience and Impact. Amer J Perinatol. 2014 Jun 24;31(7):577–82.

8. Lo YMD, Chan KCA, Sun H, Chen EZ, Jiang P, Lun FMF, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med. 2010 Dec;2(61):61ra91.

9. Lam KWG, Jiang P, Liao GJW, Chan KCA, Leung TY, Chiu RWK, et al. Noninvasive Prenatal Diagnosis of Monogenic Diseases by Targeted Massively Parallel Sequencing of Maternal Plasma: Application to β-Thalassemia. Clin Chem. 2012 Sep;58(10):1467–1475.

10. New MI, Tong YK, Yuen T, Jiang P, Pina C, Chan KCA, et al. Noninvasive Prenatal Diagnosis of Congenital Adrenal Hyperplasia Using Cell-Free Fetal DNA in Maternal Plasma. J Clin Endocrinol Metab. 2014 Jun;99(6):E1022–E1030.

11. Fan HC, Gu W, Wang J, Blumenfeld YJ, El-Sayed YY, Quake SR. Non-invasive prenatal measurement of the fetal genome. Nature. 2012 Jul 4;487(7407):320–4.

12. Kitzman JO, Snyder MW, Ventura M, Lewis AP, Qiu R, Simmons LE, et al. Noninvasive whole-genome sequencing of a human fetus. Sci Transl Med. 2012 Jun;4(137):137ra76.

13. Kruschke JK. Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan. Edition 2. Boston: Academic Press; 2015. 759 p.

14. Hobbs BP, Carlin BP. Practical Bayesian Design and Analysis for Drug and Device Clinical Trials. J Biopharm Stat. 2007 Dec 10;18(1):54–80.

15. Freedman LS, Spiegelhalter DJ. The Assessment of the Subjective Opinion and its Use in Relation to Stopping Rules for Clinical Trials. The Statistician. 1983 Mar;32(1/2):153.

16. Berry SM. Bayesian adaptive methods for clinical trials [Internet]. Boca Raton: CRC Press; 2011 [cited 2016 Apr 19]. Available from: http://www.crcnetbase.com/doi/book/10.1201/EBK1439825488

17. Schafer JL. Analysis of incomplete multivariate data. 1. ed., 1. CRC Press reprint. Boca Raton: Chapman & Hall/CRC; 2000. 430 p. (Monographs on statistics and applied probability).

18. Enders CK. Applied missing data analysis. New York: Guilford Press; 2010. 377 p. (Methodology in the social sciences).

19. Brooks SP, Gelman A. General Methods for Monitoring Convergence of Iterative Simulations. J Comput Graph Stat. 1998 Dec;7(4):434–55.

20. Kass RE, Carlin BP, Gelman A, Neal RM. Markov Chain Monte Carlo in Practice: A Roundtable Discussion. Am Stat. 1998 May;52(2):93–100.

21. Fox EJ, Reid-Bayliss KS, Emond MJ, Loeb LA. Accuracy of next generation sequencing platforms. Gener Seq Appl [Internet]. 2014 [cited 2016 Apr 20];1. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4331009/

22. Victoria X, Blades N, Ding J, Sultana R. Estimation of Sequencing Error Rates in Short Reads. 2012 [cited 2016 Apr 20]; Available from: https://pdfs.semanticscholar.org/4e2e/093491c5a03f70cd00fcd2f4d3657e54f1b8.pdf

23. Chan LL, Jiang P. Bioinformatics analysis of circulating cell-free DNA sequencing data. Clin Biochem. 2015 Oct;48(15):962–75.

24. Lun FM, Tsui NB, Chan KA, Leung TY, Lau TK, Charoenkwan P, et al. Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. Proc Natl Acad Sci. 2008;105(50):19920–19925.

## Figures Legends

**Figure 1.** A scheme describes the hierarchical Bayesian model used to infer maternal alleles inheritance. The read counts for HapI and HapII are assumed to be generated in a Bernoulli process with a SNP specific bias $\theta_s$, which depends on a beta distribution with hyperparameter $\omega$ and $\kappa$. The parameter $\omega$, which is referred as HapBias as well, has a prior beta distribution, whereas $\kappa$ has prior gamma distribution. Names of probability distribution are highlighted with red.

**Figure 2.** The Null Hypothesis testing based on ROPE and HDI. ($\Delta_L$, $\Delta_H$) and ($\delta_L$, $\delta_H$) are the limits of HDI and ROPE of 0.5, respectively. We reject the null hypothesis if there are no intersection between HDI and ROPE (highlighted in red). If ROPE contains HDI completely, we accept hypothesis that HapBias is equal to 0.5 (highlighted in orange). We make no decision for three other conditions (highlighted in blue).

**Figure 3.** Posterior distribution of HapBias by the hierarchical Bayesian approach. (A) 95%HDI locates completely downstream of ROPE region for type β SNP in family one; 0% of posterior parameter value locates within the ROPE. (B) 95%HDI locates completely upstream of ROPE region for type α SNP in family two, 1% of posterior parameter value locates within the ROPE.

**Figure 4.** Diagnosis plots of HapBias distribution for type β SNP in family one. (A) The trace plot shows the sampled HapBias value of four independent chains in MCMC simulations. (B) HapBias density distributions of four chains. The limits of 95% HDI are labeled. (C) The trend of shrink factor along increased iteration in simulations, median and 95% quantile values are showed. (D) Autocorrelation calculated with various lags for four chains. Total ESS is showed in the plot. Lag, the number of iterations between the chain and the superimposed copy.

**Figure 5.** Compare the performance of SPRT and Bayesian model on erroneous data. (A) The SPRT classification incorrectly shows that HapII is inherited in fetus even there is no fetus DNA in sample. (B) Hierarchical Bayesian model correctly shows that the true HapBias is equal to 0.5, indicating there is no fetus DNA in samples

**Figure 1.**

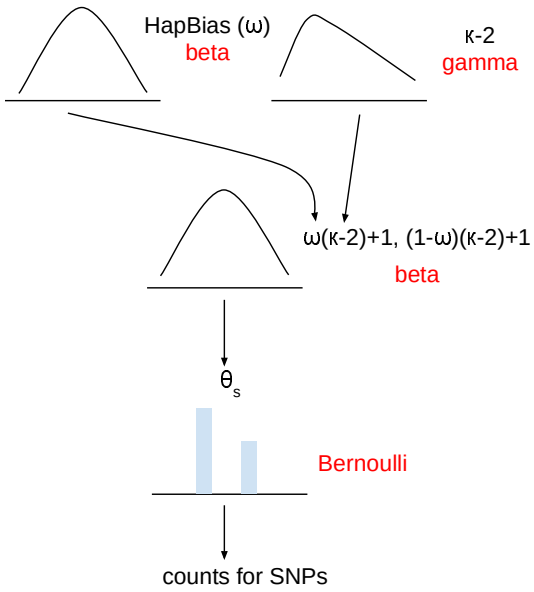

**Figure 2**

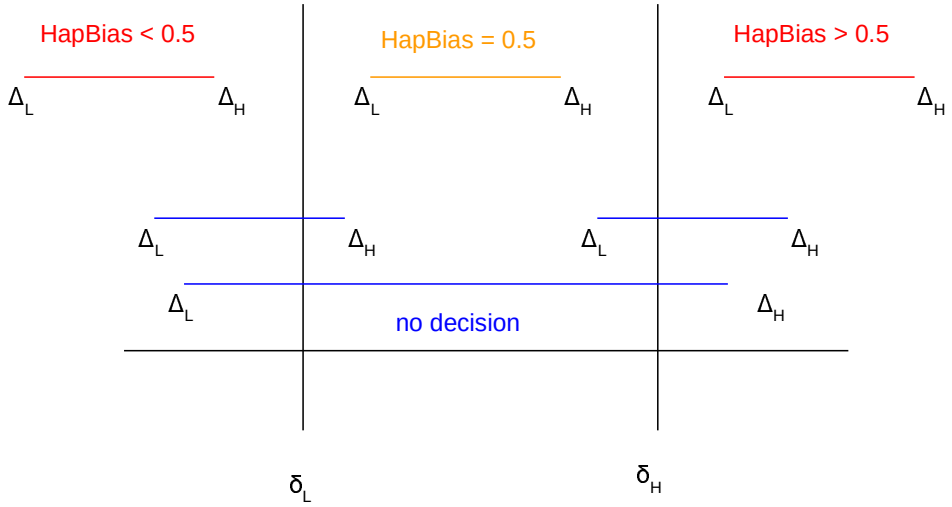

**Figure 3**

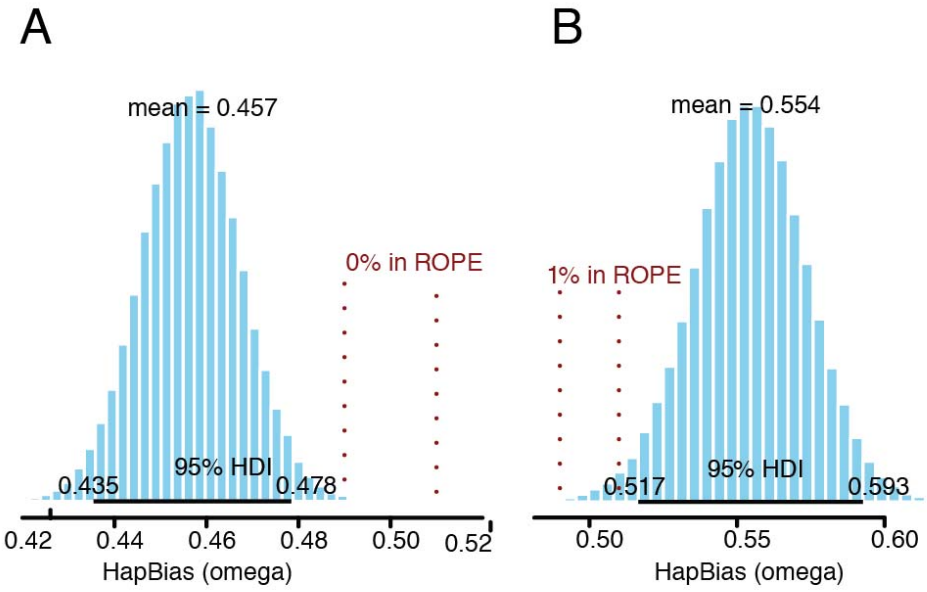

**Figure 4**

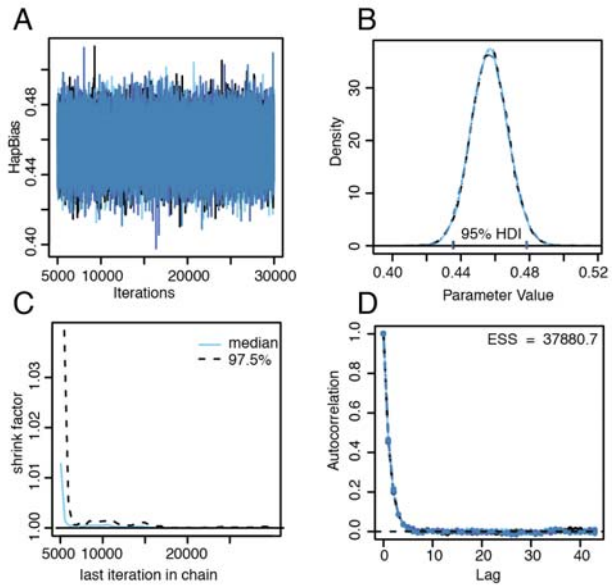

**Figure 5**