Title: Uncertainty of fetal fraction determination in Non-Invasive Prenatal Screening by highly polymorphic SNPs

Authors: Marian Grendár, Dušan Loderer, Zuzana Laučeková, Iveta Švecová, Michaela Hrtánková, Andrea Hornáková, Bálint Nagy, Pavol Žúbor, Zora Lasabová, Ján Danko

PII: S0168-1656(19)30143-9
DOI: https://doi.org/10.1016/j.jbiotec.2019.04.020
Reference: BIOTEC 8414

To appear in: *Journal of Biotechnology*

Received date: 29 January 2019
Revised date: 23 April 2019
Accepted date: 24 April 2019

Please cite this article as: Grendár M, Loderer D, Laučeková Z, Švecová I, Hrtánková M, Hornáková A, Nagy B, Žúbor P, Lasabová Z, Danko J, Uncertainty of fetal fraction determination in Non-Invasive Prenatal Screening by highly polymorphic SNPs, *Journal of Biotechnology* (2019), https://doi.org/10.1016/j.jbiotec.2019.04.020

Uncertainty of fetal fraction determination in Non-Invasive Prenatal Screening by highly polymorphic SNPs

Marian Grendár[1+*], Dušan Loderer[5+], Zuzana Laučeková[2], Iveta Švecová[2], Michaela Hrtánková[2], Andrea Hornáková[3], Bálint Nagy[4], Pavol Žúbor[2], Zora Lasabová[5], Ján Danko[2]

[1]Bioinformatics Unit, Biomedical Center Martin, Jessenius Faculty of Medicine, Comenius University in Bratislava, Malá Hora 4C, 036 01 Martin, Slovakia

[2]Department of Obstetrics and Gynecology, Jessenius Faculty of Medicine, Comenius University in Bratislava, Kollárova 2, 036 59 Martin, Slovakia

[3]Division of Molecular Medicine, Biomedical Center Martin, Jessenius Faculty of Medicine, Comenius University in Bratislava, Malá Hora 4C, 036 01 Martin, Slovakia

[4]Department of Human Genetics, Faculty of Medicine, University of Debrecen, Nagyerdei krt. 98, H-4032 Debrecen, Hungary

[5]Division of Oncology, Biomedical Center Martin, Jessenius Faculty of Medicine, Comenius University in Bratislava, Malá Hora 4C, 036 01 Martin, Slovakia

[+]These authors contributed equally.

[*]Corresponding author.

E-mail address: marian.grendar@jfmed.uniba.sk

Highlights

- Fetal fraction is a quality control parameter in NIPS
- There are several methods for determining fetal fraction, including snpFF
- snpFF is based on targeted sequencing of highly polymorphic SNPs
- Individual SNP-level variability of fetal fraction is rather high
- Uncertainty of snpFF increases with fetal fraction
- snpFF cannot reliably measure fetal fraction below 2.5%

Abstract

Fetal fraction and the chromosome representation are the two key quantities used in Non-Invasive Prenatal Screening (NIPS) to determine the aneuploidy status of a fetus. Several methods for fetal fraction determination have been proposed in the literature, including a class of the methods, denoted snpFF, based on high-coverage targeted sequencing of highly polymorphic Single Nucleotide Polymorphisms (SNPs). The variant of snpFF, investigated here, has similar properties as the other variants of snpFF. We point out that the variability of the individual informative SNPs-based estimates of fetal fraction increases with the increase of fetal fraction. At 4% fetal fraction the Inter-Quartile Range (IQR) of the individual estimates of fetal fraction is around 3% and it increases to 6% at 15% fetal fraction. snpFF cannot detect fetal fraction below 2.5% because the number of informative SNPs becomes too small, even zero.

Abbreviations: NIPS, Non-Invasive Prenatal Screening; cfDNA, cell free DNA; WGS, Whole Genome Sequencing; SNP, Single Nucleotide Polymorphism; BAF, B Allele Frequency; QC, Quality Control; IQR, Inter-Quartile Range

Keywords Fetal fraction, Single Nucleotide Polymorphism, NIPS

1. Introduction

The discovery of cell-free fetal DNA (cfDNA) in the plasma of pregnant women (Lo et al., 1997) has allowed the development of a wide range of Non-Invasive Prenatal Screening (NIPS) applications. Either the shallow-depth Whole Genome Sequencing (WGS)-based read-counting approach to NIPS or the targeted sequencing-based Single Nucleotide Polymorphism (SNP) approach to NIPS are used to detect aneuploidies, chromosomal microdeletions, microduplications and monogenic diseases.

In NIPS, fetal fraction and the chromosome representation are the two basic characteristics of a sample. They are used to determine the aneuploidy status of a fetus. Moreover, fetal fraction is used for the Quality Control (QC). If fetal fraction of a sample is below the Limit of Detection (LoD) of NIPS the sample is reported as 'no call' and recommended for a redraw. The LoD of NIPS can be determined experimentally (Fiorentino et al., 2016), by the coverage method (Fan and Quake, 2010) or by the *in silico* method LoD^q (Grendár et al., 2017a). The LoD^q method exploits the variability of the chromosome representation in euploid samples for determining the LoD. It may lead the LoD lower or higher than the customary 4% LoD, implied by the coverage method. Moreover, it leads different LoD for different chromosomes. The authors (Grendár et al., 2017a) also suggest utilizing both fetal fraction and the chromosome representation in the QC of NIPS, and specify the set of pairs of fetal fraction and the chromosome representation for which the aneuploidy status of a fetus cannot be determined.

There are several methods for fetal fraction determination available in the literature. They include the chromosome Y-based method ffY (Fan et al., 2008), the seqFF method (Kim et al., 2015) and the seqFFY method (Grendár et al., 2017a). These methods utilize the data from the

WGS, which are used also to obtain the chromosome representation of a sample. Unlike the ffY, seqFF, seqFFY methods, which are based on WGS, the class of methods that utilizes a set of highly polymorphic Single Nucleotide Polymorphisms (SNPs) - referred to as snpFF, hereafter - requires high-coverage targeted sequencing. Thus, snpFF incurs extra costs to NIPS. Several variants of snpFF were considered in the literature (Chu et al., 2010a; Jiang et al., 2012; Kang et al., 2016; Song et al., 2016; Jiang et al., 2016; Larson et al., 2018). They differ, among other things, in complexity of the method for determining the set of informative SNPs.

The present study aims to explore and quantify the uncertainty of a variant of snpFF, compare it with the seqFFY method, and discuss its advantages and disadvantages. We point out that the variability of the individual informative SNP-based values of fetal fraction, as quantified by the Inter-Quartile Range (IQR) of those values, is rather large and it increases with fetal fraction. SnpFF underestimates fetal fraction at the low end of the values (below approximately 7%), relative to seqFFY. A potentially useful property of snpFF is its inability to detect fetal fraction below approximately 2.5%, which is caused by the small (even zero) number of informative SNPs. Unlike the other methods for fetal fraction determination, snpFF thus may suggest absence of fetal cfDNA (Takoudes and Hamar, 2015).

2. Material and methods

2.1 Patient cohort

Twenty-four women were recruited from the Department of Obstetrics and Gynecology, Jessenius Faculty of Medicine, Comenius University, Slovakia. Blood samples (10 mL) were collected in BD Vacutainer® tubes (Becton Dickinson, USA) from pregnant women with male fetuses (n=10), female fetuses (n=10) and non-pregnant women (n=4). Pregnant women were at different gestation ages (7-19 weeks) and BMI (15-38), to cover wide range values of fetal fraction. After the venipuncture, all pregnant participant underwent invasive diagnostic procedures (chorionic villus sampling or amniocentesis). Informed consent was obtained from all individual participants.

2.2 Sample processing, DNA extraction and quantification

Plasma were extracted from 10 ml whole blood samples using a two-step centrifugation process (1520 g/10min/4°C and 12052 g/10min/4°C) to remove residual cells. All processed samples were stored at −80 °C until the DNA extraction. DNA was isolated from 2 mL of plasma using QIAamp DSP Virus Kit (Qiagen, Germany). The resulting cfDNA was eluted in 80 µL of elution solution. The extracted cfDNA was quantified on a Qubit Fluorimeter 3.0 (ThermoFisher Scientific) using the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific).

2.3 DNA size selection

DNA samples were size-selected using two-step process with AMPure XP beads (Beckman Coulter, US). To each sample of cfDNA (80 µL) was added 60 µL beads (ratio 1:0.75), mixed and incubated at room temperature for 5 minutes. After incubation were sample placed on a magnetic stand until the solution was clear (~7 minutes). The supernatant (140 µL) was

transferred to the new tube (to remove large cfDNA fragments). Then was added 259 µL beads (ratio 1:1.85), mixed and incubated for 5 minutes at room temperature. After incubation tubes were placed on magnetic stand and supernatant was removed and discard (to remove small cfDNA fragments). Beads were washed two times by freshly prepared 80% ethanol. After air-drying procedure to beads was added 65 µL of resuspension buffer (obtained cfDNA fragments ranging from 100 to 160bp).

2.4 Targeted sequencing and WGS

Obtained DNA fragments were used for two different fetal quantification assays. Total volume was divided to two part: 60 µL for WGS and 5 µL for high-depth targeted sequencing. Thus, the same fragments of cfDNA were used for preparing libraries for both the WGS and targeted sequencing. This way, the bias due to different library preparation was minimized.

2.4.1 WGS

The Illumina TruSeq® Nano DNA LT Sample Preparation Kit (Illumina, San Diego, CA) was used for library preparation. We followed manufacturer´s protocol from the step 2: Adenylate 3´ Ends. Final DNA libraries were sequenced on the Illumina HiSeq 1500 in Rapid Run Mode. The reads were aligned to human genome hg19 by bowtie2 aligner in BioConductor (Gentleman, 2004) using the library Rsamtools (Morgan, 2011). Duplicate and poorly aligned reads were filtered out using Rsamtools function ScanBamParam with the default setting resulting in a loss of ~17% of ~7 million reads. Reads were normalized by all autosomes and binned into 50 kbp bins. Fetal fraction was determined by the ffY method (Fan et al., 2008), the seqFF method (Kim et al., 2015) and the seqFFY method (Grendár et al., 2017a).

2.4.2 Custom panel - Targeted NGS

The design of the full-custom AmpliSeq™ Custom DNA Panel for Illumina® was performed by the Designstudio (https://designstudio.illumina.com, Illumina, San Diego, CA). The 67 SNPs from (Kim et al., 2015) were used. Library preparation was performed according to the manufacturer´s protocol. The amplicon size ranged from 125 to 140 bp with an average size of 135bp. The high-depth sequencing was performed on the Illumina MiSeq instrument. Sequencing reads with length 2 x 151bp were mapped to the human reference genome hg19 using BWA MEM aligner (Li, 2013). Sam files were converted into bam files by BioConductor (Gentleman, 2004) library Rsamtools (Morgan, 2011) and nucleotide frequencies were obtained by the GenomicAlignments library (Lawrence et al., 2013). All the other computing was done using R (R-Development-Core-Team, 2016). SNPs with coverage smaller than 500 were filtered out.

3. Theory

Fetal fraction can be estimated from SNPs at genomic positions that are homozygous in mother and heterozygous in fetus (Chu et al., 2010a). Denoting the maternal allele in uppercase and letting B be the minor allele, the seven joint maternal-fetal genotypes are: AAaa, AAab, ABaa,

ABab, ABbb, BBab, BBbb. If the respected genotype relative frequencies are n_{AAaa}, n_{AAab}, n_{ABaa}, n_{ABab}, n_{ABbb}, n_{BBab}, and n_{BBbb}, then fetal fraction can be estimated as $2n_{AAab}$ or as $2(0.5 - n_{ABaa})$, or as $2(n_{ABab} - 0.5)$ or as $2(1 - n_{BBab})$. Sequencing permits to determine the allelic read counts n_α, n_β, where α stands for A or a, β stands for B or b. The allelic frequencies are $f_\alpha = n_\alpha/n$, $f_\beta = n_\beta/n$, where n is the sequencing depth. The latter frequency is the minor allele frequency, denoted BAF. Instead of working with f_β it is possible to use $f_\beta' = \min(n_\alpha, n_\beta)/n$ (Jiang et al., 2012). Not every SNP is informative in a particular sample. It is not trivial to tell apart the informative and non-informative SNPs (Sparks et al., 2012). The main reason for this difficulty lays in the existence of the allelic bias of the sequencing technology (Chu et al., 2010b). There are several methods for identifying the informative SNPs. A probabilistic model of the allelic bias was proposed by (Jiang et al., 2012). The model involves a mixture of distributions. The model was recently extended by (Larson et al., 2018). The authors also proposed a more elaborated method of estimating the parameters of the model. Both models can be seen as a probability density-based clustering of SNPs. Kim et al. applied the K-means algorithm for the clustering (Kim et al., 2015). Clustering assigns each SNP into one of the seven joint genotypes, or the four genotypes, if the frequencies are folded around 0.5.

Another possible method of identifying the informative SNPs, suggested by (Sparks et al., 2012) rests on bounds on f_β'. Kang et al. delineate a region in the 2D space of minor and major allele frequencies to select the set of informative SNPs (Kang et al., 2016). Regardless of the method used for identifying informative SNPs, the relative frequencies of the informative SNPs are used to obtain a set of the values of fetal fraction. In some of the variants of snpFF the estimate of fetal fraction is obtained as the median (the mean, the weighted average) of the individual informative SNPs-based values of fetal fraction, provided the set is sufficiently large. The resulting method of fetal fraction determination will be referred to as snpFF.

Among the other methods of fetal fraction determination the most reliable (Peng and Jiang, 2017) appears to be the ffY method (Fan et al., 2008; Hudecova et al., 2014) which is based on chromosome Y. The method is applicable for the male carrying pregnancies, only. The seqFF method (Kim et al., 2015) can determine fetal fraction regardless of the fetal gender, however, the method underestimates fetal fraction at the lower and upper ends of the range, relative to ffY. The seqFFY method (Grendar et al., 2017a) borrows strength from both ffY and seqFF, and corrects the bias of seqFF. For male carrying pregnancies seqFFY is just the ffY method, as it is more reliable than seqFF. For female carrying pregnancies ffY cannot be applied, but ffY can be calibrated against seqFF, and the calibration curve can be used to predict the ffY value for the seqFF value of a female carrying pregnancy.

4. Results

The values of BAF f_β for a six samples are depicted on Fig. 1.

Besides the probability density mixture-based clustering (Benaglia et al., 2009) of f_β' for the 67 SNPs in a sample, we have applied also the K-means and K-medoids clustering algorithms (Maechler et al., 2018) for identifying the informative SNPs. However, we have achieved the best results by the bounds method of (Sparks et al., 2012). The number of informative SNPs identified by the method depends on the actual values of the bounds. Sparks et al. suggest using 1% and 20% bounds (Sparks et al., 2012). A too low value of the lower bound captures some of the non-informative SNPs, and leads too small estimates of fetal fraction. A too high value of the lower bound may lead a sparse set of informative SNPs and hence unreliable estimates. It is necessary to find a balance; in our data the best value of the lower bound was 1.25%. For the [1.25, 20]% bounds, the number of informative SNPs in each sample is depicted on Fig. 2. If the set of informative SNPs contained less than 5 SNPs, the snpFF estimate was not provided. There are two samples (20, 24) with no informative SNPs and three samples (3, 19, 23) with less than 5 informative SNPs. Among them, four samples (19, 20, 23, 24) are from non-pregnant females. The remaining sample for which snpFF fails to provide estimate of fetal fraction (i.e., sample 3) has the seqFFY estimate 1.08%. Note, that majority of samples have more than 10 informative SNPs.

Recall that the snpFF estimate of fetal fraction is based on the values of f_β' (or f_β) for the informative SNPs. The snpFF estimate is the median (or the mean) of the individual informative values of fetal fraction, denoted snpFF_i. The variability of snpFF_i is visualized by a boxplot on Fig. 3 for the 19 samples where snpFF can provide the estimate. The variability is rather large.

The variability is particularly important for the values of fetal fraction close to the Limit of Detection (LoD). The LoD of NIPS is typically around 4%. In our data set, there are two samples with fetal fraction in [4.5, 5.5]%, with the values of fetal fraction 4.9% and 4.7%. The first sample has 10 informative SNPs, the second one has 22 informative SNPs. The range of the values of ff implied by individual informative SNPs is [2.6, 14.6]%, [2.6, 39.0]% for the first, second sample; respectively. The lower and upper quartiles are [4.0, 7.4]%, [3.2, 6.8]%; respectively.

Crossplot of IQR of snpFF_i vs. snpFF, see Fig. 4, indicates that the IQR increases with the increase of fetal fraction. The association between IQR of snpFF_i and snpFF was quantified by the robust linear regression model (Rousseeuw et al., 2015) which had both the intercept (0.017) and slope (0.271) statistically significant (p-value 0.047, 0.0005; respectively. T-test of significance of predictor).

4.1. Comparison of snpFF and seqFFY methods

As it was already mentioned, the same cfDNA fragments were used to prepare library for WGS and targeted sequencing. This way, the bias that can be introduced due to differences in library preparation was minimized. The WGS was used to determine the ffY and seqFF estimates of fetal fraction, which in turn, were used to obtain the seqFFY estimates. These were compared with snpFF estimates in the 19 samples where the snpFF method provides estimates. As Fig. 5 demonstrates, snpFF overestimates fetal fraction for fetal fraction below 7% and overestimates it above 7%, relative to seqFFY.

5. Discussion

The snpFF class of methods for fetal fraction determination is based on the high-coverage targeted sequencing of highly polymorphic SNPs. The main bottleneck in snpFF is the determination of the set of informative SNPs. Several methods for determining the set of informative SNPs were proposed, leading different variants of the snpFF method. The set may be determined by the hierarchical bayesian density-based clustering, by the K-means or K-medoids clustering, by the region in the 2D space of minor and major allele frequencies, or by the bounds method, among other methods.

The bounds instance of snpFF, investigated here, estimates fetal fraction by the median of the individual informative SNPs-based values of fetal fraction. The individual values, $snpFF_i$, are obtained from the relative frequencies of the informative SNPs.

The variability of the individual informative $snpFF_i$ values is rather substantive. Already the first variant of snpFF considered by (Chu et al., 2010a) led the individual informative $snpFF_i$ values ranging from around 5% to almost 30%, when the true fetal fraction was known to be 10%. (Larson et al., 2018) report the values of BAF for 57 SNPs in 120 non-pregnant samples, where the BAF is expected to be 50%. The median values of BAF for the SNPs go from around 25% to around 75%; see Fig. 1 in (Larson et al., 2018).

We have quantified the uncertainty of snpFF by the IQR of the individual informative SNPs. The IQR increases from some 3% at 4% fetal fraction to 6% at 15% fetal fraction. Thus, the uncertainty is non-negligible.

The snpFF method was not included into the recent comparative study (van Beek et al., 2017) of several fetal fraction determination methods. However, the papers on various variants of snpFF usually include a comparison of the studied variant of snpFF with some other method of fetal fraction determination. Usually, the chromosome Y-based ffY method is used in this role; see (Larson et al., 2018; Song et al., 2016). Kim et al. compared the seqFF method with their variant of snpFF (Kim et al., 2015). seqFF overestimates fetal fraction relative to snpFF over the entire range of values; see Fig. 2C in (Kim et al., 2015). And snpFF underestimates fetal fraction relative to ffY; see Fig. 2B in (Kim et al., 2015). Song et al. (2016) have found that their variant of snpFF underestimates fetal fraction relative to ffY along the entire range of values. The underestimation of snpFF relative to ffY was also reported by (Larson et al., 2018). For their variant of snpFF, the underestimation occurs for values of ffY below 10%. This is in a good

agreement with our results where snpFF underestimates fetal fraction for seqFFY below 7%. Recall, that for the male carrying pregnancies the seqFFY is just the ffY method.

The snpFF cannot determine fetal fraction below approximately 2.5%, as there tend to be too small number of the informative SNPs to provide a reliable estimate. This can be seen as an advantage of the snpFF method, since the other methods for fetal fraction determination usually lead a positive (or even negative!) values of fetal fraction for non-pregnant samples.

6. Conclusions

The variant of snpFF method of fetal fraction determination investigated in the present communication has similar properties as the other variants of snpFF. In particular, it underestimates fetal fraction below ~7%, relative to the chromosome Y-based method ffY, as well as relative to the seqFFY method. Variability of the individual informative components $snpFF_i$ increases with the increase of fetal fraction. The IQR at 4% fetal fraction is around 3% and it increases to ~6% at 15% fetal fraction. snpFF cannot detect fetal fraction below 2.5%, because the number of informative SNPs becomes too small, even zero.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Conflicts of interest

None to declare.

Acknowledgments

This work was supported by the project 'Biomedical Center Martin' [ITMS 26220220187] co-financed from EU sources and by CEPV II. MG and ZL also acknowledge support from APVV-16-0066 grant (Slovak Research and Development Agency). ZL acknowledges support from VEGA 1/0380/18 grant.

References

Benaglia, T., Hunter, D.R., Young, D.S., 2009. mixtools : An R Package for Analyzing Finite. J Stat Soft 32, 1–29. https://doi.org/10.18637/jss.v032.i06

Chu, T., Bunce, K., Hogge, W.A., Peters, D.G., 2010a. A novel approach toward the challenge of accurately quantifying fetal DNA in maternal plasma. Prenat. Diagn. https://doi.org/10.1002/pd.2656

Chu, T., Bunce, K., Hogge, W.A., Peters, D.G., 2010b. Statistical considerations for digital approaches to non-invasive fetal genotyping. Bioinformatics 26, 2863–2866. https://doi.org/10.1093/bioinformatics/btq544

Dennis Lo, Y.M., Corbetta, N., Chamberlain, P.F., Rai, V., Sargent, I.L., Redman, C.W.G., Wainscoat, J.S., 1997. Presence of fetal DNA in maternal plasma and serum. Lancet 350, 485–487. https://doi.org/10.1016/S0140-6736(97)02174-0

Fan, H.C., Blumenfeld, Y.J., Chitkara, U., Hudgins, L., Quake, S.R., 2008. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc. Natl. Acad. Sci. 105, 16266–16271. https://doi.org/10.1073/pnas.0808319105

Fan, H.C., Quake, S.R., 2010. Sensitivity of noninvasive prenatal detection of fetal aneuploidy from maternal plasma using shotgun sequencing is limited only by counting statistics. PLoS One 5. https://doi.org/10.1371/journal.pone.0010439

Fiorentino, F., Bono, S., Pizzuti, F., Mariano, M., Polverari, A., Duca, S., Sessa, M., Baldi, M., Diano, L., Spinella, F., 2016. The importance of determining the limit of detection of non-invasive prenatal testing methods. Prenat. Diagn. 36, 304–311.

Gentleman, R., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80. https://doi.org/10.1186/gb-2004-5-10-r80

Grendár, M., Loderer, D., Lasabová, Z., Danko, J., 2017a. Method for determining the qualification of a sample for fetal aneuploidy determination in Non-Invasive Prenatal Testing. EP17203198.

Grendár, M., Loderer, D., Lasabová, Z., Danko, J., 2017b. A comment on “Comparing methods for fetal fraction determination and quality control of NIPT samples.” Prenat. Diagn. 37. https://doi.org/10.1002/pd.5151

Hudecova, I., Sahota, D., Heung, M.M.S., Jin, Y., Lee, W.S., Leung, T.Y., Lo, Y.M.D., Chiu, R.W.K., 2014. Maternal plasma fetal DNA fractions in pregnancies with low and high risks for fetal chromosomal aneuploidies. PLoS One 9. https://doi.org/10.1371/journal.pone.0088484

Jiang, P., Chan, K.C.A., Liao, G.J.W., Zheng, Y.W.L., Leung, T.Y., Chiu, R.W.K., Lo, Y.M.D., Sun, H., 2012. FetalQuant: Deducing fractional fetal DNA concentration from massively parallel sequencing of DNA in maternal plasma. Bioinformatics 28, 2883–2890. https://doi.org/10.1093/bioinformatics/bts549

Kang, X., Xia, J., Wang, Y., Xu, H., Jiang, H., Xie, W., Chen, F., Zeng, P., Li, X., Xie, Y., Liu, H., Huang, G., Chen, D., Liu, P., Jiang, H., Zhang, X., 2016. An advanced model to precisely estimate the cell-free fetal DNA concentration in maternal plasma. PLoS One 11. https://doi.org/10.1371/journal.pone.0161928

Kim, S.K., Hannum, G., Geis, J., Tynan, J., Hogg, G., Zhao, C., Jensen, T.J., Mazloom, A.R., Oeth, P., Ehrich, M., van den Boom, D., Deciu, C., 2015. Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts. Prenat. Diagn. 35, 810–815.

https://doi.org/10.1002/pd.4615

Larson, N.B., Wang, C., Na, J., Rowsey, R.A., Highsmith, W.E., Hoppman, N.L., Kocher, J.-P., Klee, E.W., 2018. Improving Single-Nucleotide Polymorphism-Based Fetal Fraction Estimation of Maternal Plasma Circulating Cell-Free DNA Using Bayesian Hierarchical Models. J. Comput. Biol. cmb.2018.0056. https://doi.org/10.1089/cmb.2018.0056

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., Carey, V.J., 2013. Software for Computing and Annotating Genomic Ranges. PLoS Comput. Biol. 9. https://doi.org/10.1371/journal.pcbi.1003118

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv Prepr. arXiv1303.3997.

Maechler M; Rousseeuw P; Struyf A; Hubert M; Hornik K, 2018. Cluster Analysis Basics and Extensions. R package version 2.0.7-1. CRAN. https://doi.org/10.1007/s11224-015-0664-y

Morgan, M., 2011. An introduction to Rsamtools. Dispon{\'\i}vel http//prs. ism. ac. jp/bioc/2.6/bioc/html/rsamtools. html. Acessado em outubro.

Peng, X.L., Jiang, P., 2017. Bioinformatics approaches for fetal DNA fraction estimation in noninvasive prenatal testing. Int. J. Mol. Sci. https://doi.org/10.3390/ijms18020453

R-Development-Core-Team, 2016. R Software 3.3.0. R A Lang. Environ. Stat. Comput.

Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-barrera, M., Verbeke, T., Koller, M., Maechler, M.M., 2015. robustbase: Basic Robust Statistics. R Packag.

Song, Y., Zhou, X., Huang, S., Li, X., Qi, Q., Jiang, Y., Liu, Y., Ma, C., Li, Z., Xu, M., Cram, D.S., Liu, J., 2016. Quantitation of fetal DNA fraction in maternal plasma using circulating single molecule amplification and re-sequencing technology (cSMART). Clin. Chim. Acta 456, 151–156. https://doi.org/10.1016/j.cca.2016.03.005

Sparks, A.B., Struble, C.A., Wang, E.T., Song, K., Oliphant, A., 2012. Noninvasive prenatal detection and selective analysis of cell-free DNA obtained from maternal blood: Evaluation for trisomy 21 and trisomy 18. Am. J. Obstet. Gynecol. 206. https://doi.org/10.1016/j.ajog.2012.01.030

Takoudes, T., Hamar, B., 2015. Performance of non-invasive prenatal testing when fetal cell-free DNA is absent. Ultrasound Obstet. Gynecol. https://doi.org/10.1002/uog.14715

van Beek, D.M., Straver, R., Weiss, M.M., Boon, E.M.J., Huijsdens-van Amsterdam, K., Oudejans, C.B.M., Reinders, M.J.T., Sistermans, E.A., 2017. Comparing methods for fetal fraction determination and quality control of NIPT samples. Prenat. Diagn. 37, 769–773. https://doi.org/10.1002/pd.5079

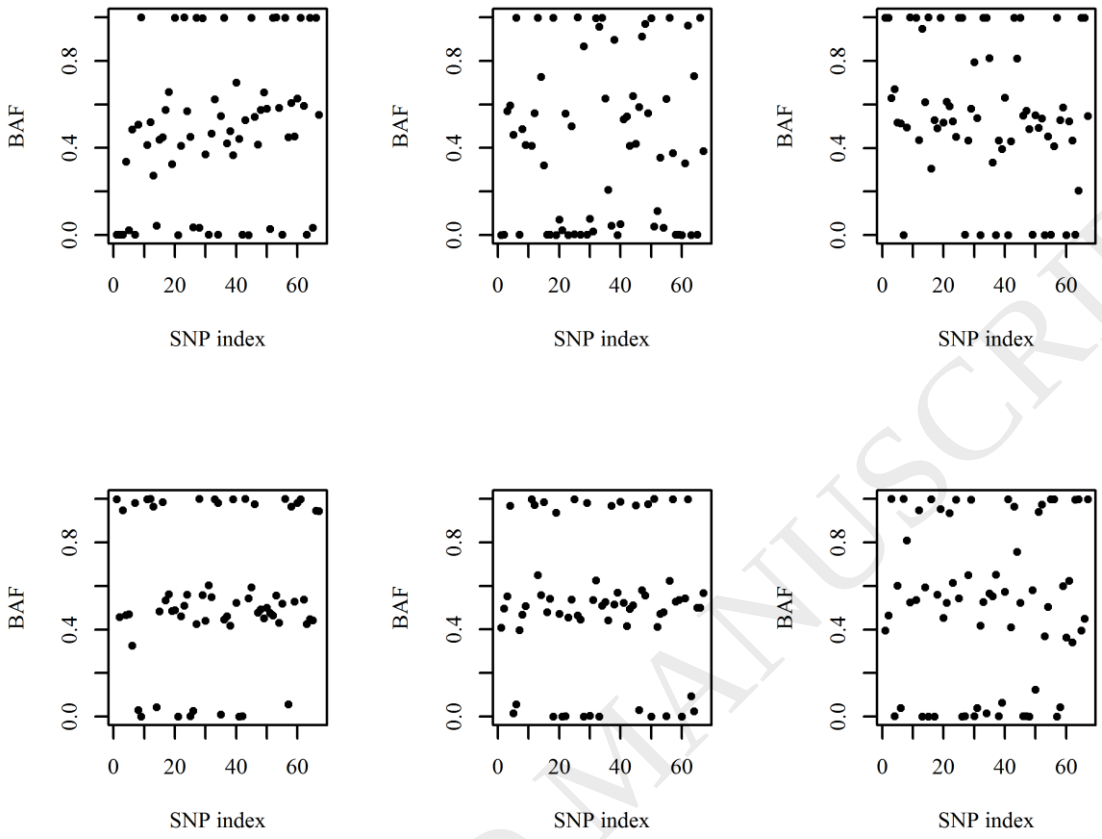

Fig 1. The BAF f_β for 67 SNPs in a subset of 6 samples.

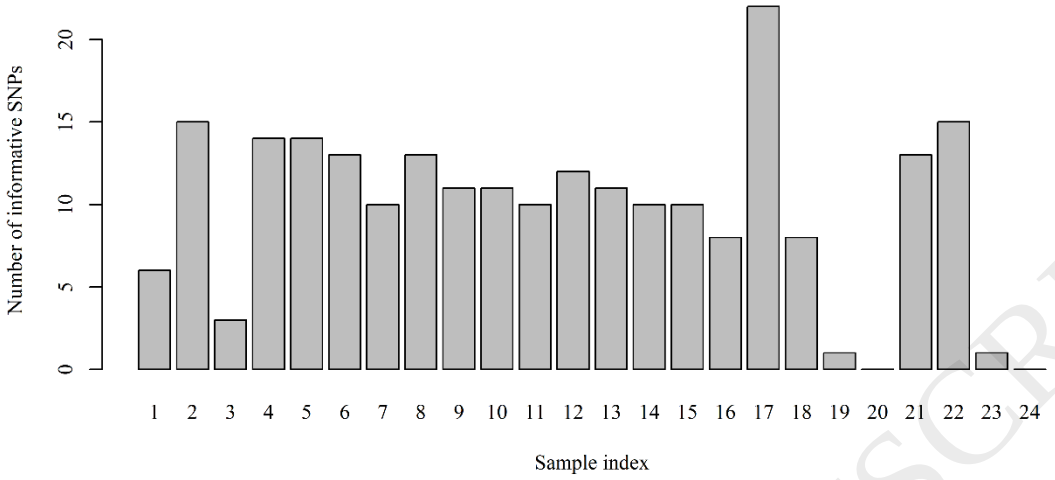

Fig 2. The number of informative SNPs in the samples.

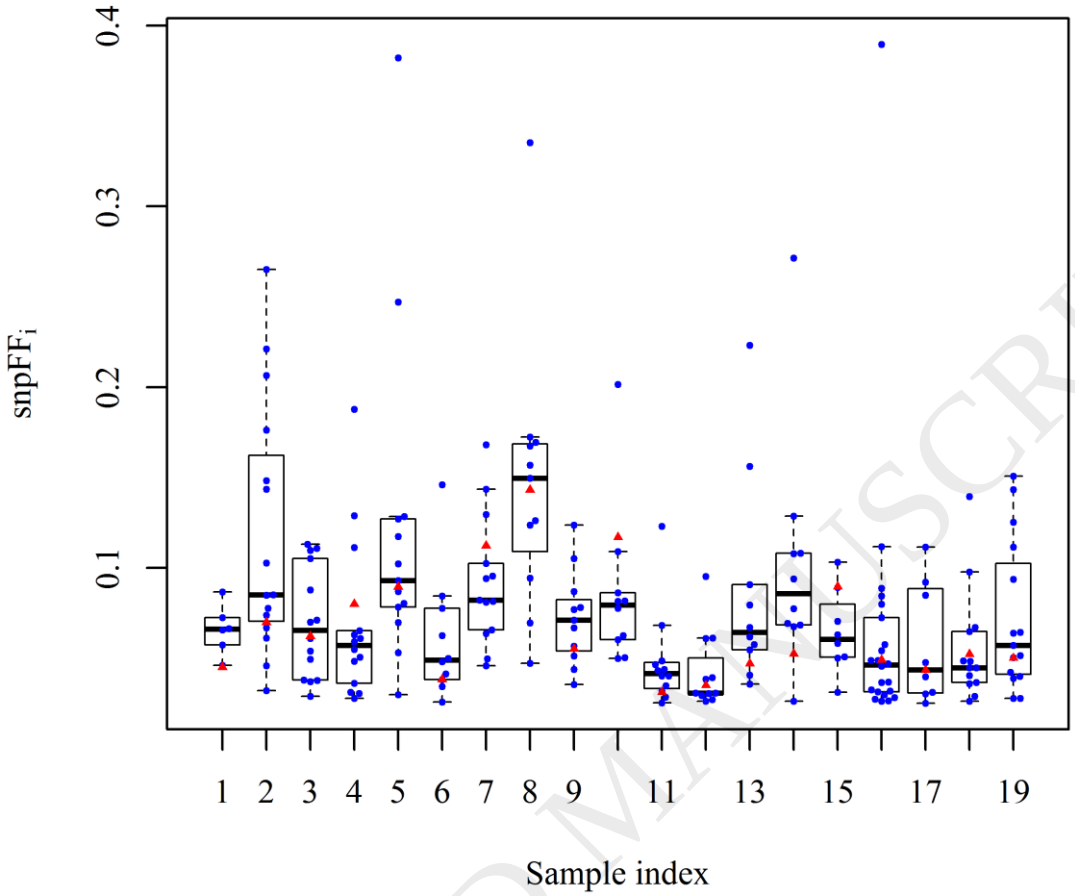

Fig 3. The values of snpFF$_i$ for the informative SNPs, overlaid by boxplot to visualize the median of snpFF$_i$ which is the snpFF estimate. The red triangle is the value of seqFFY estimate of fetal fraction.

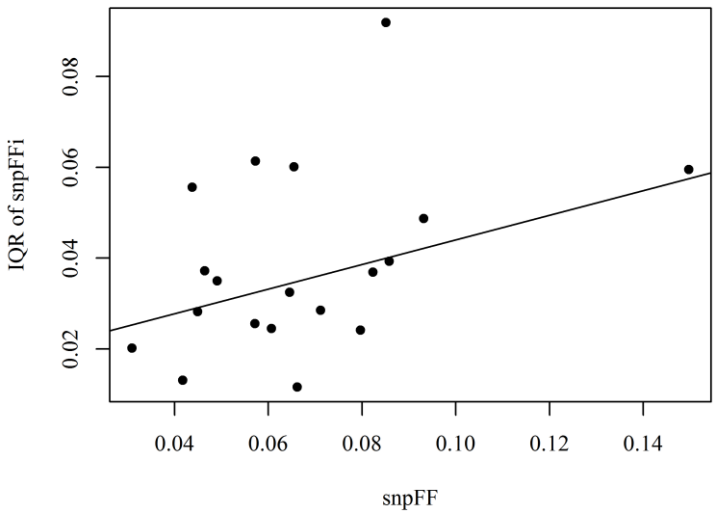

Fig 4. Crossplot of IQR of $snpFF_i$ vs. snpFF, for the 19 samples. The line was fitted by a robust linear regression method. IQR, Inter-Quartile Range.

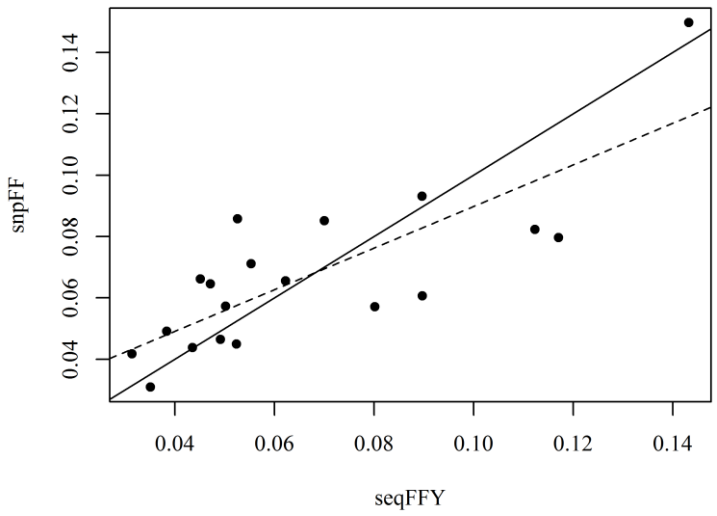

Fig 5. Crossplot of the snpFF estimates vs. the seqFFY estimates, for the 19 samples where snpFF provides estimates. To facilitate the comparison of the methods, the 45° line (solid line) and the OLS fit (dashed line) are added to the plot; see (Grendár et al., 2017b). The OLS fit has the intercept 0.022 and the slope 0.678 statistically significant (p-value 0.028, < 0.0001; respectively. T-test of significance of predictor) and the Adjusted R^2 = 0.611. OLS, Ordinary Least Squares.