A comprehensive predictive method for low fetal fraction in noninvasive prenatal screening

Liang Hu

Longgang Maternity&Child Healthcare Hospital of Shenzhen

Yuanyuan Pei

Longgang Maternity&Child Healthcare Hospital of Shenzhen

Xiaojin Luo

Longgang Maternity&Child Healthcare Hospital of Shenzhen

Lijuan Wen

Longgang Maternity&Child Healthcare Hospital of Shenzhen

Hui Xiao

Shenzhen longgang People's Hospital

Jinxing Liu

Longgang Maternity&Child Healthcare Hospital of Shenzhen

Dongxue Li

Longgang Maternity&Child Healthcare Hospital of Shenzhen

Liping Wu

Longgang Maternity&Child Healthcare Hospital of Shenzhen

Jian Ran

Longgang Maternity&Child Healthcare Hospital of Shenzhen

Gaochi Li

Longgang Maternity&Child Healthcare Hospital of Shenzhen

Xiaohang Chen

Longgang Maternity&Child Healthcare Hospital of Shenzhen

Li Du

Guangdong Maternal and Child Hospital

Xiulan Zhou

Longgang Maternity&Child Healthcare Hospital of Shenzhen

Fengxiang Wei (✉ haowei727499@163.com)

Longgang Maternity&Child Healthcare Hospital of Shenzhen <https://orcid.org/0000-0001-7559-2504>

Methodology

Keywords: NIPS, cell free DNA, fetal fraction, prediction model, random forests

Posted Date: April 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-22603/v1>

Abstract

Background

The study was a retrospective cohort analysis based on the results of noninvasive prenatal screening (NIPS), complete blood count, thyroxine test and Down's syndrome screening in first or second trimester from 14043 pregnant women. Random forests algorithm was applied to predict the low fetal fraction of cell free DNA (with FF lower than 10th percentile) through individual and laboratory information. Performance of the model was evaluated and compared to prediction using maternal weight. To investigate factors associated with lower FF in the NIPS and to develop a new predictive method for low FF before NIPS.

Results

Of 14043 cases, maternal weight, RBC, HGB and free T3 were significantly negative correlated with FF while gestation age, free T4, PAPP-A, AFP, uE3 and β -hCG were significantly positive correlated with FF. Compared to prediction using maternal weight as isolated parameter, the model has a higher area under curve (AUC) of Receiver Operating Characteristic (ROC) and overall accuracy.

Conclusions

The comprehensive predictive method based on combined multiple factors was more effective than single-factor model in low FF status prediction. This method can provide more information for clinical choice and pre-test quality control of NIPS.

Background

Noninvasive prenatal screening (NIPS) is a method of fetal chromosomal aneuploidy screening mainly based on next generation sequencing (NGS). It has been validated in multiple clinical cohorts that NIPS is highly sensitive and specific for patients at increased risk of T13, T18 and T21 aneuploidies [1–3]. It also has potential application value in the prenatal screening of copy number variables (CNV) and single nucleotide variables (SNV) by target sequencing and extending the sequencing depth [4, 5]. Recently, NIPS is a widespread first-tier prenatal screening method used for high risk pregnancies of aneuploidy [6].

Cell free DNA in maternal plasma is the sequencing target of NIPS and derived from both mother and fetus. While z scores are used for evaluating risk of aneuploidy of chromosome 13, 18 and 21, fetal fraction (FF) of cell free DNA is also considered critical to the results [7, 8]. FF is based on the sequencing results of NIPS and is calculated using chrY concentration or algorithms of generalized linear regression such as seqFF [9]. It is confirmed that FF lower than 3.5%–4% is a key factor may cause no-call reports or false positive and negative results [10, 11]. The American College of Medical Genetics and Genomics (ACMG) supported that a clearly visible fetal fraction should be included on NIPS reports [11]. It is significant to determine whether FF is too low to ensure the quality of NIPS results.

However, it is a challenge to predict the FF before NIPS and provide the predictive FF as a reference information for avoiding no-call or unreliable reports. Several studies suggested that individual profiles such as maternal

weight and gestation age were factors that influence FF[12, 13]. Significant obesity and gestation age below 9–12 weeks are linked to low FF and high probability of unreliable results, consequently not recommended for NIPS. FF is also speculative to be associated with placental development and volume which partially relates to chromosomal aneuploidy and Down's serum screening in first and second trimester[14]. Nevertheless, because of the insufficient knowledge about mechanism of generation and degradation of cell free DNA, the relationships between FF and these variables are not significant enough to predict whether FF is too low accurately by an isolated factor[15].

The aim of this study was to investigate the factors associated with lower fetal fraction and to better predict the low FF status. This study was based on the data collection which contains NIPS information and the results of complete blood count, thyroxine test and Down's serum screening in first and second trimester in our center. Random forest algorithm, an ensemble machine learning algorithm commonly used in various prediction scenarios and appropriate for data with high dimension and collinearity, was applied for comprehensive analysis of clinical and individual profiles[16]. The performance of prediction was evaluated and compared with single-factor predictive models using maternal weight to provide a more effective reference for the choice of NIPS clinical application.

Results

Profiles of study population

In this dataset, 10th percentile of FF was 5.1%. The characteristics and profiles of 14043 NIPS participants with normal ($FF > 5.1\%$) or low ($FF \leq 5.1\%$) fetal fraction status were described in Table.1. Compared to normal FF status group (NFS), the gestation age and maternal weight at first and second Down's syndrome screening is higher in low FF status group (LFS), while no significant difference was observed in gestation age, gravidity and parity between the two groups. Positive rates of T21 and Sex Chromosomal Aneuploidies (SCA) were 0.71% and 1.28% in LFS group respectively. By comparison, positive rates of T21 (0.14%) and SCA (0.38%) were significantly lower in NFS group. It should be noted that NIPS results were gathered posteriorly to NIPS sampling, so that this information was excluded in following analysis. To demonstrate the data profile of extreme low ff, specimens with low FF induced no-call ($FF < 3.5\%$) reports were shown in Table 1.

Distribution of each lab measurement and clinical information in this dataset was shown in Table.2. In this dataset, 11076(78.87%) specimens contained NA values which were enriched in RBC (64.32%) and HGB (64.32%), while 2967(21.13%) specimens were completely observed. All distributions of each variant were not Gaussian distribution ($p < 0.05$).

Associations Between Laboratory Measurements And Fetal Fraction

To reveal the relationships between laboratory measurements and fetal fraction, Pearson correlation coefficient were calculated for fetal fraction values. In addition, AUC of ROC and adjusted odds ratios (OR) were obtained for low or normal fetal fraction status (Table.2). The difference of each measurement in low and normal fetal fraction status group was showed in Fig. 1A. There was no significant linear association between all the serum markers and fetal fraction ($r < 0.2$). however, some variables were predictive and suggestive for fetal fraction status. In low

fetal fraction status group, RBC, HGB and free T3 were significantly higher while free T4, PAPP-A, AFP, uE3 and β -hCG in first and second trimester were significantly lower. AUCs and adjusted ORs showed that free T3, RBC and HGB were relative risk factors of low FF status and PAPP-A, AFP, uE3 and β -hCG in first and second trimester were protective factors (Fig. 2A). Remarkably, higher Maternal weight in first and second trimester was significantly associated with low FF status. In contrast, these significant associations between TSH and maternal age with FF status was not observed. Gestation age is positive related to FF weakly ($r = 0.2184$), but not obviously predictive for FF status (AUC = 0.5204) in our dataset.

NA Value Imputation

The k-nearest neighbors (kNN) algorithm was applied in NA value imputation to make the dataset complete. Median of each variant before and after Not Available (NA) value imputation were shown in Table 2. For each variable, Mann-Whitney U test showed no significant difference was observed before and after imputation ($p > 0.05$). Density plots showed that except RBC and HGB, the distribution density curve after imputation was nearly close to the distribution density curve before imputation, indicating there was no significant impact on data distribution characteristics induced by NA value imputation (Fig. 1B). The distribution transformation in RBC and HGB might be caused by too high proportion of NA values.

Performance Evaluation Of Predictive Model

According to the correlation analysis hereinbefore, TSH and Maternal age were excluded for model training. Gestation age, maternal weight in first and second trimester, RBC, HGB, free T3, free T4, PAPP-A, AFP, uE3 and β -hCG in first and second trimester were input into predictive model as dependent variables. After 10 iterations of model training and validation, an average of ROC and PR curve for predicted FF status was showed in Fig. 2B. Compared to maternal weight, the AUC of random forest model (0.7022) was significantly higher ($p < 0.001$), which indicated that the model was more effective than maternal weight in first trimester (0.6595) and second trimester (0.6508) in prediction of low FF status. PR curve also showed that random forest model had better performance than maternal weight. In the predicted low FF status group, FF was significantly lower than in predicted normal FF status group (Fig. 2C). The overall accuracy of the predictive model is 0.8532, higher than the overall accuracy of the prediction through maternal weight in first (0.8083) and second trimester (0.8076).

Discussion

The amount of fetal DNA in maternal plasma sample is affected by multiple factors which make challenges in prediction of low fetal DNA fraction before NIPS. In this study, to predict the existence of low FF status, we collected 14 laboratory test results and individual information, and fitted a regression model based on random forests algorithm. In our cohort, we confirmed that the model was more effective in predicting low FF status than using isolated maternal weight and was robust against incomplete observations.

Fetal fraction is a varied and complex biological indicator which influenced by individual difference and laboratory factors. It has been reported that overweight pregnant women are at higher risk of test failure due to low fetal fraction. This might be due to a dilutional effect and higher release of maternal cfDNA from adipose cells into the systemic circulation[17]. Gestation age is another relevant factor of fetal fraction. Wang E. et.al suggested that

cfDNA rises by almost 1% per week after 20 weeks of gestation, but only 0.1% per week between 10 and 21 week of gestation[13]. It has been suggested that low placental volume could also lead to a low fetal fraction. G. Ashoor et al reported that fetal fraction increases with maternal serum level of free β -hCG and PAPP-A which reflected placental volume and development[18, 19]. Besides, fetal numbers, fetal aneuploidies and maternal smoking are reported to be associated with lower fetal fraction[20, 21]. It has also been found that the maternal exercise leads to a decrease in the fetal fraction, as the level of the maternal cfDNA increases directly after physical activity[22]. It should be noted that fetal fraction could also be affected by sample transport or laboratory work flow[23]. In our cohort, gestation age, maternal weight in first and second trimester, serum markers of Down's syndrome screen such as PAPP-A, AFP, uE3 and β -hCG were all weakly correlated with FF, which was confirmed by previous studies. In addition, this was a novel report of the negative correlation between RBC, HGB, free T3 and free T4 with fetal fraction.

Although it is difficult to reliably predict the value of fetal fraction without quantitative experiment. Instead of quantitatively predicting fetal fraction, we suggested that qualitative prediction of low FF status binarized from fetal fraction is practicable. Recently, isolated information with a constant cut-off value such as maternal weight and gestation age is used for evaluation of low FF status and protection from unreliable reports. In this study, we confirmed that prediction based on combined clinical and laboratory indicators using machine learning algorithm is more effective. For the positive prediction, sequencing shorter cfDNA fragments or cfDNA enrichment can be used to improve the FF of NIPS[24]. Furthermore, there was a large number of missing values in the raw clinical data set in which directly filtering of missing values will result in insufficient specimens for subsequent analysis. Incomplete data is common in clinic and cannot be avoided. Therefore, we adopted missing value imputation to reduce the impact of missing values and increase the robustness of the model.

There are some limitations in our study. First, it was a single-center study and should be validated in multi-center cohorts with larger sample size because of diversities in sample size or laboratory methodology. Second, although the performance of predictive model was obviously promoted compared to maternal weights-based prediction, there were still capacity of improvements on performance to make it more clinical applicable. We have a hypothesis that besides the information used in this study, there are unknown or unadopted factors and variables correlated with FF. For example, the information of vanished twin which had impact on fetal fraction was unavailable and excluded in the model. So, a larger sample size is needed for better fitting the model and more clinical or lab test information should be collected and added into the comprehensive predictive model.

Conclusions

In conclusion, in this study we reported a new method based on comprehensive clinical and laboratory information for predicting the relatively low fetal fraction status in NIPS. It was conformed that the prediction model was more effective than prediction using maternal weight as independent variable. This study was an application of machine learning in prenatal screen and would provide more reference information for clinical choices of NIPS.

Methods

Population selection

This study was a retrospective analysis focused on a total of 14043 pregnant women who received prenatal care in Shenzhen Longgang District Women and Children Healthcare Hospital from November 2017 to April 2019. Inclusion criteria consisted of: 1. pregnant women underwent NIPS between 12 weeks and 24 weeks of gestation 2. underwent first trimester serum screening or second trimester serum screening as well as at least one of the following clinical laboratory tests: complete blood count (CBC) test and thyroxin test. 3. singleton pregnancy. Each participant in our study was written informed and consented that laboratory and clinic information would be used in anonymous and non-profit scientific studies. The study was approved by the ethics committee of Shenzhen Longgang District Women and Children Healthcare Hospital before implementation.

NIPS And Fetal Fraction Calculation

For each participant, 5 ml maternal whole blood sample was collected using EDTA-K2 tubes (BD, UK). Cell free maternal plasma was separated and purified by centrifugation at 1600 g (10 min, 4 °C) for whole blood and at 13000 g (10 min, 4 °C) for plasma sequentially. DNA extraction and library construction were performed for each sample following the instruction of NIFTY chromosomal abnormality test kit (BGI, Wuhan, China)[10]. Library qualification was performed using Qubit 3.0 (Thermo, USA). 48 libraries were pooled into one mixed library which was single-end sequenced on BGI SEQ500 sequencing platform (BGI, Wuhan, China) with 35 read length and 9.7M average reads number of each sample.

After reads alignment to reference genome hg19 (bwa-0.7.11)[25], PCR duplication removing (SAMtools-1.9)[26], reads counting for 30 kb bins (bedtools-2.18.0)[27] and GC correction based on LOWESS regression, fetal fraction was calculated based on the method of chromosome Y(chrY) and SeqFF[28, 29]. Firstly, FF of all male fetal samples were obtained by chrY method. Then SeqFF model including an elastic regression and weighted rank selection criterion (WRSC) was fitted using male samples. Finally, FF of both male and female samples were calculated using the SeqFF model as the final outcome for following prediction. Standalone chrY fraction reference and SeqFF model was established in our lab to fit the BGI sequencing platform. In this study, the relative low FF status was defined as the FF was lower than 10th percentile of FF in all samples. While relative low FF status was a logical variable, the existence of the status was defined as 1 and non-existence as 0.

Association Between FF And Clinical Variables

To construct a machine learning model predicting the existence of low fetal fraction, we collected clinical and individual information which listed in Table.1 for each participant. NA value was accepted to enhance the robustness of the model. Maternal weight at first trimester or second trimester was included separately instead of BMI because height of stature was not accessible in this dataset.

For each variable, Pearson correlation coefficient with FF and AUC of ROC curve for predicting low FF status was calculated to explore the impact to FF. To investigate the risk factors of low FF status, adjusted OR and 95% confidence intervals (CIs) for each variable were calculated and adjusted by maternal age and gestation age. In addition, all specimens were divided into higher group and lower group of each variables, Mann-Whitney U was performed to assess the significance of FF in the two groups.

Random Forest Model And K-fold Cross Validation

Data pretreatment was performed before fitting prediction model. Firstly, clinical laboratory test results and information were normalized using z-score method after replacing outliers with NA values. Secondly, NA values were imputed using kNN method. Finally, the dataset was split randomly and equally into 10 folds. For each iteration, 7 folds were selected as training set which contains 70% specimens and the rest of 3 folds as test set which contains 30% specimens[30].

We established a supervised regression model using R package randomForest 4.6–14 which implements random forests algorithm introduced by Breiman et.al[31]. All clinical test results and information of training set were input as independent variables while relative low FF status as the response variable. Then, the fitted model was validated by test set to predict relative low FF status of which output ranged in [0,1]. Samples with the predicted value lower than 10th percentile were marked as predictive relative low FF status.

Overall accuracy, ROC curves and precision recall (PR) curves were selected as the performance evaluation metrics. To compare this comprehensive predictive method with single-variable based predictive method, we also defined another prediction method based on maternal weight in first and second trimester: samples with maternal weight higher than 90th percentile were marked as predictive relative low FF status. We compared the effectiveness of the predictive model in classifying low and normal FF status against maternal weight in first and second trimester.

Statistical analysis

Statistical analysis was performed using R-3.4.4 (<https://www.r-project.org/>). NA value imputation was performed using R package of DMwR (0.4.1) [32]. Kolmogorov-Smirnov test was applied for determining whether continuous variables could be fitted with a Gaussian distribution. Student's t test and Mann-Whitney U test were performed to compare differences of Gaussian and non-Gaussian distributed continuous variables, respectively. Chi square test and fisher exact test were used for categorical variables. Delong test was used for comparing the performance of two ROC curves. A probability value (p-value) of < 0.05 was considered statistically significant.

Abbreviations

NIPS: noninvasive prenatal screening; FF: fetal fraction of cell free DNA; AUC: area under curve; ROC: Receiver Operating Characteristic; NGS: next generation sequencing; CNV: copy number variables; SNV: single nucleotide variables;ACMG: The American College of Medical Genetics and Genomics; NFS: normal FF status; LFS: low FF status; SCA: Sex Chromosomal Aneuploidies; OR: odds ratios; kNN: k-nearest neighbors; NA: Not Available; CBC: complete blood count; WRSC: weighted rank selection criterion; CIs: confidence intervals; PR: precision recall.

Declarations

Ethics approval and consent to participate

The study was approved by the ethics committee of Shenzhen Longgang District Women and Children Healthcare Hospital before implementation.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The presented work was supported by grants of Shenzhen Science and Technology Innovation Commission (Grants numbers: JCYJ20160427114320284 and JCYJ20180305125647151) and grants of Shenzhen Longgang Science and Technology Innovation Commission (Grants numbers: LGKCZSYS2018000010). **Authors' contributions**

LH wrote the manuscript, developed the prediction methodology. JL, GL, LD, DL and LW collected the clinical information. JR, XL and YP performed data cleaning. XC, LW, HX and YP developed the fetal fraction estimation. XZ and FW oversaw the project.

Acknowledgements

The authors would like to thank for Song Xu and Zhuyan Huang from Shenzhen Longgang District Women and Children Healthcare Hospital for data collection; Hsi-Yuan Huang from Chinese University of Hong Kong, Shenzhen for statistical guidance; and all women who consented to participate in this study.

References

1. Sago H, Sekizawa A, Japan N. Nationwide demonstration project of next-generation sequencing of cell-free DNA in maternal plasma in Japan: 1-year experience. *Prenat Diagn*. 2015;35(4):331–6.
2. Noh JJ, Ryu HM, Oh SY, Choi SJ, Roh CR, Kim JH. A two-year experience of non-invasive prenatal testing (NIPT) at an urban tertiary medical center in South Korea. *Taiwan J Obstet Gynecol*. 2019;58(4):545–51.
3. Cheung SW, Patel A, Leung TY. Accurate description of DNA-based noninvasive prenatal screening. *N Engl J Med*. 2015;372(17):1675–7.
4. Liang D, Cram DS, Tan H, Linpeng S, Liu Y, Sun H, et al. Clinical utility of noninvasive prenatal screening for expanded chromosome disease syndromes. *Genet Med*. 2019;21(9):1998–2006.
5. Lv W, Li Z, Wei X, Zhu H, Teng Y, Zhou M, et al. Noninvasive fetal genotyping in pregnancies at risk for PKU using a comprehensive quantitative cSMART assay for PAH gene mutations: a clinical feasibility study. *BJOG*. 2019;126(12):1466–74.
6. Palomaki GE, Kloza EM, O'Brien BM, Eklund EE, Lambert-Messerlian GM. The clinical utility of DNA-based screening for fetal aneuploidy by primary obstetrical care providers in the general pregnancy population. *Genet Med*. 2017;19(7):778–86.
7. Committee Opinion No. 640: Cell-Free DNA Screening For Fetal Aneuploidy. *Obstet Gynecol*. 2015;126(3):e31–7.

8. Balslev-Harder M, Richter SR, Kjaergaard S, Johansen P. Correlation between Z score, fetal fraction, and sequencing reads in non-invasive prenatal testing. *Prenat Diagn.* 2017;37(9):943–5.
9. Hestand MS, Bessem M, van Rijn P, de Menezes RX, Sie D, Bakker I, et al. Fetal fraction evaluation in non-invasive prenatal screening (NIPS). *Eur J Hum Genet.* 2019;27(2):198–202.
10. Jiang F, Ren J, Chen F, Zhou Y, Xie J, Dan S, et al. Noninvasive Fetal Trisomy (NIFTY) test: an advanced noninvasive prenatal diagnosis methodology for fetal autosomal and sex chromosomal aneuploidies. *BMC Med Genomics.* 2012;5:57.
11. Gregg AR, Skotko BG, Benkendorf JL, Monaghan KG, Bajaj K, Best RG, et al. Noninvasive prenatal screening for fetal aneuploidy, 2016 update: a position statement of the American College of Medical Genetics and Genomics. *Genet Med.* 2016;18(10):1056–65.
12. Kinnings SL, Geis JA, Almasri E, Wang H, Guan X, McCullough RM, et al. Factors affecting levels of circulating cell-free fetal DNA in maternal plasma and their implications for noninvasive prenatal testing. *Prenat Diagn.* 2015;35(8):816–22.
13. Wang E, Batey A, Struble C, Musci T, Song K, Oliphant A. Gestational age and maternal weight effects on fetal cell-free DNA in maternal plasma. *Prenat Diagn.* 2013;33(7):662–6.
14. Wataganara T, Metzenbauer M, Peter I, Johnson KL, Bianchi DW. Placental volume, as measured by 3-dimensional sonography and levels of maternal plasma cell-free fetal DNA. *Am J Obstet Gynecol.* 2005;193(2):496–500.
15. Tamminga S, van Maarle M, Henneman L, Oudejans CB, Cornel MC, Sistermans EA. Maternal Plasma DNA and RNA Sequencing for Prenatal Testing. *Adv Clin Chem.* 2016;74:63–102.
16. Liaw A. MW. Classification and Regression by randomForest. *R News.* 2002;2(3):18–22.
17. Haghiac M, Vora NL, Basu S, Johnson KL, Presley L, Bianchi DW, et al. Increased death of adipose cells, a path to release cell-free DNA into systemic circulation of obese women. *Obesity (Silver Spring).* 2012;20(11):2213–9.
18. Ashoor G, Syngelaki A, Poon LC, Rezende JC, Nicolaides KH. Fetal fraction in maternal plasma cell-free DNA at 11–13 weeks' gestation: relation to maternal and fetal characteristics. *Ultrasound Obstet Gynecol.* 2013;41(1):26–32.
19. Scott FP, Menezes M, Palma-Dias R, Nisbet D, Schluter P, da Silva Costa F, et al. Factors affecting cell-free DNA fetal fraction and the consequences for test accuracy. *J Matern Fetal Neonatal Med.* 2018;31(14):1865–72.
20. Suzumori N, Ebara T, Yamada T, Samura O, Yotsumoto J, Nishiyama M, et al. Fetal cell-free DNA fraction in maternal plasma is affected by fetal trisomy. *J Hum Genet.* 2016;61(7):647–52.
21. Rava RP, Srinivasan A, Sehnert AJ, Bianchi DW. Circulating fetal cell-free DNA fractions differ in autosomal aneuploidies and monosomy X. *Clin Chem.* 2014;60(1):243–50.
22. Schlutter JM, Hatt L, Bach C, Kirkegaard I, Kolvraa S, Uldbjerg N. The cell-free fetal DNA fraction in maternal blood decreases after physical activity. *Prenat Diagn.* 2014;34(4):341–4.
23. Qiao L, Mao J, Liu M, Liu Y, Song X, Tang H, et al. Experimental factors are associated with fetal fraction in size selection noninvasive prenatal testing. *Am J Transl Res.* 2019;11(10):6370–81.
24. Qiao L, Yu B, Liang Y, Zhang C, Wu X, Xue Y, et al. Sequencing shorter cfDNA fragments improves the fetal DNA fraction in noninvasive prenatal testing. *Am J Obstet Gynecol.* 2019;221(4):345. e1- e11.

25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
27. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
28. Kim SK, Hannum G, Geis J, Tynan J, Hogg G, Zhao C, et al. Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts. *Prenat Diagn*. 2015;35(8):810–5.
29. Hudecova I, Sahota D, Heung MM, Jin Y, Lee WS, Leung TY, et al. Maternal plasma fetal DNA fractions in pregnancies with low and high risks for fetal chromosomal aneuploidies. *PLoS One*. 2014;9(2):e88484.
30. Jung Y, Hu J. A K-fold Averaging Cross-validation Procedure. *J Nonparametr Stat*. 2015;27(2):167–79.
31. Breiman L. Random Forests. *Mach Learn*. 2001;45(1):5–32.
32. Torgo L. Data Mining with R, learning with case studies. Chapman and Hall/CRC. 2010.

Tables

Table.1 Characteristic of the study population

*P < 0.05, compared to normal FF status group. SCA: Sex Chromosomal Aneuploidies.

Table.2 NA imputation and association with FF and clinical variables.

Characteristic	Overall	Normal FF status	Low FF status	Low FF induced no-call
Observations [N (%)]	14043	13340	703	68
Maternal age at delivery(years)	29(27-32)	29(27-32)	30(27-33) *	29(27-34)
>=35 [N (%)]	1975(14.06)	1856(13.91)	119(16.93) *	16(23.53)
<35 [N (%)]	12068(85.94)	11484(86.09)	584(83.07)	52(76.47)
Gestation age at lab test(days)				
1st Down's syndrome screening	88(86-91)	89(86-91)	88(87-90)	89(86-91)
2nd Down's syndrome screening	118(115-122)	118(115-122)	118(115-122)	119(114.5-121.5)
NIPT	115(98-123)	115(98-123)	115(101-120)	98(114-120)
Thyroxine test	115(75-124)	115(75-125)	114(75-121)	115(74-137)
Complete blood count	113(93-121)	113(93-121)	112(99-118)	96(112-118)
Maternal weight at Down's syndrome screening (Kg)				
1st trimester	53(48-58.5)	52.9(48-58)	58(52-64.5)	57.5(52.4-64.9)
2nd trimester	54(49.4-59.8)	54(49.1-59.2)	59(53-65.6)	57.9(52.7-66)
Gravidity [N (%)]				
<3	8206(58.43)	7836(58.74)	423(60.17)	39(57.35)
>=3	5837(41.57)	5504(41.26)	280(39.83)	29(42.65)
Parity [N (%)]				
0	6517(46.41)	5982(44.84)	342(48.65)	31(45.58)
1	6845(48.74)	6675(50.04)	325(46.23)	26(38.24)
>1	681(4.85)	683(5.12)	36(5.12)	11(16.18)
NIPT result [N (%)]				
Trisomy 21	24(0.17)	19(0.14)	5(0.71) *	
Trisomy 18	9(0.06)	7(0.05)	2(0.28)	
Trisomy 13	2(0.014)	2(0.015)	0(0.00)	
SCA	60(0.43)	51(0.38)	9(1.28) *	
Insertions and deletions	138(0.98)	129(0.97)	9(1.28)	

es	NA numbers (%)	Before NA imputation	After NA imputation	r	AUC	OR
ne ng						
G)	2662(18.96)	57.12(36.13- 90.88)	57.65(38.64- 88.90)	0.1372	0.5837(0.5676- 0.5998)	0.7485(0.693- 0.8063)
2-)	2662(18.96)	3834.83(2472.23- 5681.68)	4036.46(2738.46- 5583.26)	0.1259	0.6032(0.5877- 0.6186)	0.6848(0.6345- 0.7375)
ne ng						
G)	3309(23.56)	15.70(10.07- 25.52)	16.00(10.82- 25.99)	0.1234	0.5903(0.5746- 0.6061)	0.7321(0.6664- 0.8002)
)	3309(23.56)	37.29(29.51- 47.02)	37.99(31.10- 45.73)	0.1403	0.6052(0.5897- 0.6208)	0.7136(0.6549- 0.7756)
.)	3626(25.82)	5.22(4.23-6.49)	5.27(4.50-6.21)	0.0678	0.5544(0.5386- 0.5702)	0.8501(0.7902- 0.9128)
n						
)	1787(12.73)	1.27(0.83-1.81)	1.31(0.89-1.76)	-0.0259	0.5067(0.4909- 0.5225)	1.0091(0.9493- 1.0673)
Γ3)	1803(12.84)	4.29(3.96-4.64)	4.33(4.01-4.62)	-0.1101	0.5923(0.5768- 0.6078)	1.2076(1.146- 1.277)
Γ4)	1805(12.85)	11.58(10.73- 12.51)	11.51(10.79- 12.38)	0.0155	0.5338(0.5178- 0.5498)	0.9106(0.8514- 0.972)
te						
	9032(64.32)	3.79(3.55-4.07)	3.79(3.67-3.94)	-0.1023	0.5991(0.5833- 0.6148)	1.3433(1.2338- 1.4613)
	9032(64.32)	114(108-121)	113.52(110.03- 117.00)	-0.1092	0.5911(0.5752- 0.6071)	1.3612(1.2379- 1.4982)
al	0(0.00)	29(27-32)	29(27-32)	-0.0561	0.4822(0.4662- 0.4981)	1.0342(0.9777- 1.0939)
on	0(0.00)	115(98-123)	115(98-123)	0.2184	0.5204(0.5056- 0.5353)	0.8682(0.8154- 0.923)
al at						

ne
ng

er	2662(18.96)	53.0(48.0-58.5)	53.0(48.5-58.3)	-0.2451	0.6629(0.6475-0.6783)	1.7177(1.6241-1.8169)
er	3309(23.56)	54.0(49.4-59.8)	54.0(49.7-59.6)	-0.2444	0.6676(0.6523-0.6829)	1.7565(1.6569-1.8626)

*P < 0.05, compared to normal FF status group. TSH: thyroid stimulating hormone; T3: triiodothyronine; T4: thyroxine; β-hCG: β-human chorionic gonadotropin; PAPP-A: pregnancy-associated plasma protein-A; AFP: alpha-fetoprotein; uE3: unconjugated estriol; RBC: red blood cell; HGB: hemoglobin.

Figures

Figure 1

Characteristic of clinical variables and NA imputation. (A): FF in low and normal groups of observed clinical variables. (B): Distribution before and after NA imputation of each variable.

Figure 2

Association between FF with clinical variables and performance evaluation of prediction model. (A) OR and P value of each clinical variable for low FF status. (B): Fetal fraction in predicted normal FF status group and in predicted low FF status group. (C): ROC and PR curve of comprehensive predictive model and prediction using maternal weight.