Problem Solving Protocol

NiPTUNE: an automated pipeline for noninvasive prenatal testing in an accurate, integrative and flexible framework

Véronique Duboc†, David Pratella†, Marco Milanesio, John Boudjarane, Stéphane Descombes, Véronique Paquis-Flucklinger‡ and Silvia Bottini‡

Corresponding author. Silvia Bottini, Université Côte d’Azur, Center of Modeling, Simulation and Interactions, Hopital l’Archet 3, 151 route de Saint Antoine de Ginestiere, 06200, Nice, France. Tel.: +33(6)30566999; E-mail: silvia.bottini@univ-cotedazur.fr
†These authors contributed equally to this work.
‡These authors contributed as co-last authors.

Abstract

Noninvasive prenatal testing (NIPT) consists of determining fetal aneuploidies by quantifying copy number alteration from the sequencing of cell-free DNA (cfDNA) from maternal blood. Due to the presence of cfDNA of fetal origin in maternal blood, in silico approaches have been developed to accurately predict fetal aneuploidies. Although NIPT is becoming a new standard in prenatal screening of chromosomal abnormalities, there are no integrated pipelines available to allow rapid, accurate and standardized data analysis in any clinical setting. Several tools have been developed, however often optimized only for research purposes or requiring enormous amount of retrospective data, making hard their implementation in a clinical context. Furthermore, no guidelines have been provided on how to accomplish each step of the data analysis to achieve reliable results. Finally, there is no integrated pipeline to perform all steps of NIPT analysis. To address these needs, we tested several tools for performing NIPT data analysis. We provide extensive benchmark of tools performances but also guidelines for running them. We selected the best performing tools that we benchmarked and gathered them in a computational pipeline. NiPTUNE is an open source python package that includes methods for fetal fraction estimation, a novel method for accurate gender prediction, a principal component analysis based strategy for quality control and fetal aneuploidies prediction. NiPTUNE is constituted by seven modules allowing the user to run the entire pipeline or each module independently. Using two cohorts composed by 1439 samples with 31 confirmed aneuploidies, we demonstrated that NiPTUNE is a valuable resource for NIPT analysis.

Key words: prenatal testing; fetal aneuploidies prediction; fetal fraction; bioinformatic pipeline; benchmark
Introduction

The development of noninvasive prenatal testing (NIPT) revolutionized the landscape of prenatal diagnosis by offering to pregnant women a safe alternative to test for fetal aneuploidies such as the trisomy 21 (Down syndrome; for review see Tamminga et al. [1]). This test is based on the discovery of circulating cell-free DNA fragments (cfDNA) in maternal blood. Maternal plasma contains both maternal and fetal cfDNAs that are analyzed by low coverage next generation sequencing (NGS; [2–4]). The fetal fraction (ff) is the percentage of cfDNA in maternal plasma of fetoplacental origin (ff= fetal cfDNA/fetal cfDNA + maternal cfDNA). It increases during pregnancy with an average of 10–15% between 10 and 20 gestational weeks as assessed by direct measurement of fetal cfDNA in maternal blood [5, 6]. The ff depends on different maternal parameters including body mass index (BMI) and diseases [5, 7, 8] as well as on fetoplacental parameters (for review see Hui et al. [9]).

Reliability of NIPT is high but it depends on a sufficiently high ff. It has been proposed that the ff should be at least 4% to validate NIPT results [6]. Several bioinformatics tools have been developed for ff estimation with very different performances [10–12]. Early approaches to calculate the ff are based on the count of reads assigned to the Y chromosome [13], e.g. the DEFRAG algorithm [10]. Despite being the most accurate, unfortunately these methods are informative for male fetuses only. Two other approaches, independent of the sex of the fetus, have been developed: Seqff [14] and Sanefalcon [15]. The first one is based on the assumption that fetal and maternal fragments are not evenly distributed on the genome. The authors used a large cohort to pretrain a model able to detect these small differences in fragments length within a predefined bin resolution and to estimate the ff. The second approach is founded on the hypothesis that differential nucleosome packaging exists between fetal and maternal DNA, resulting in shorter fetal fragments. Thus the authors proposed to estimate the ff by exploiting the spatial distribution of reads on the estimated nucleosome profiles.

Several methods have also been proposed to predict fetal aneuploidies. Usually, they use a step of genome binning, a normalization technique and a Z-score calculation to assess the variation of the normalized fragment count in each bin. The comparison is performed with euploid reference samples [cn.MOPS [16] and CNVkit [17]], between sample variations [Wisecondor [18], WisecondorX [19]], or it is reference free (FREEC [20], QDNaSeq [21] and BiC-seq2 [22]). Although these tools enable the NIPT analysis, they are difficult to implement in clinical context because they require hard data processing and extensive bioinformatics competences. Furthermore, given the different technologies and algorithms implemented, results may vary quite dramatically and a quantification of this variation is still missing. As a result, there is currently no consensus between clinical laboratories regarding the methods to apply both for ff estimation and chromosomal aberration detection.

Here we report NiPTUNE, an accurate open source python package to perform NIPT in an accUrate, iNtegrative and flexiblE framework including several methods for fetal fraction estimation, a novel method for accurate gender prediction, a principal component analysis (PCA) based strategy for quality control and chromosomal aneuploidies prediction. NiPTUNE allows for the first time to perform all steps of NIPT analysis in one unique pipeline including fetal fraction estimation and aneuploidy detection. NiPTUNE is an open source python package, composed by seven modules that can be run independently or altogether for whole NIPT analysis, making this tool very flexible and easy to use. To test NiPTUNE performances we used two cohorts composed by 1439 samples with 31 confirmed aneuploidies demonstrating a sensitivity and specificity of 100%.

Materials and methods

Patient cohort

NIPT was performed on 377 samples from pregnant women at Nice university hospital from January 2017 to September 2018 (cohort 1) and 1062 samples at Marseille university hospital from January 2017 to December 2018 (cohort 2) after informed consent. The sequencing data resulting from these series were retrospectively used to validate our bioinformatics suite with consents for research (INDS—MR3310281119; R04-018 Nice and PADS20-53 Marseille). We included blood samples from two nonpregnant women as negative control. For further cohort description see Supplementary Table 1. Protocols of DNA extraction, library preparation and sequencing were identical for both cohorts.

DNA isolation

Maternal samples were collected in blood collection tubes from Streck (cfDNA BCT) or Roche Diagnostics (Cell-Free DNA Collection Tube) and centrifuged at 1600 g for 10 min to separate the plasma from the blood cells. Plasma was subsequently centrifuged at 16 000 g for 10 min. The supernatant was transferred to a new microcentrifuge tube and stored at –80°C until further processing. cfDNA was extracted from 4 ml of plasma using the QIAamp® Circulating Nucleic Acid kit (Qiagen®, Hilden, Germany) according to the manufacturer’s protocol. The DNA was eluted into a final volume of 35 μl of AVE buffer, and concentration was measured using Qubit dsDNA High Sensitivity Kit (Thermo Fisher Scientific) prior to storage –20°C.

Library preparation and sequencing

Shallow whole-genome sequencing of cfDNA was performed using either a Proton or an S5XL sequencer (Thermo Fisher Scientific®, Waltham, MA, USA), starting from 15 ng input of cfDNA. For library building, cfDNA samples were processed either manually or in semiautomated procedure with the Ion Plus fragment library kit and Ion Plus Core Library Module for AB Library Builder TM System respectively (Life Technologies-Thermo Fisher Scientific®, Waltham, MA, USA) using an optimized procedure [23]. Library concentrations were measured using Ion Library TagMan™ Quantitation Kit (Thermo Fisher Scientific). Equimolar concentrations (15pM) of each library were then automatically prepared and loaded on the chip (IonPI-TM Chip Kit V3 or Ion 540 Chip Kit) using the Ion Chef (Thermo Fisher Scientific®, Waltham, MA, USA). Preprocessing quality control, trimming and mapping to GRCh37 were performed using the Ion Torrent Suite *. Pipeline in use at the time in the diagnostic laboratory. Mapping to GRCh38 would now be recommended.

Statistical analysis

All statistics were performed using the software R. Violin plots were done using the library ggplot2 from R. Correlations were calculated using corr function from R.

Normalization and quality control

Aligned sequences were corrected for GC-content using the script gcc.py from Wisecondor. Each sample was divided in 1 Mb
supplementary content from the original manuscript:

**Fetal gender prediction**

Four methods were tested and results were compared with genders after birth available for 93% of cohort 1.

We defined a novel method, MagicY, that first quantify the proportion of reads of 7 Y-specific regions with respect to the number of reads on the Y chromosome:

\[
\text{MagicY} = \frac{\sum \# \text{reads specific region Chr Y}}{\# \text{reads on ChrY}}
\]

These 7 Y-specific regions are defined in [Supplementary Table 2](#) and were described by Skaletsky et al. [24]. Then, a gaussian mixture model with two gaussians is used to fit the distribution of the ratio of counts and to identify the threshold to separate male and female populations ([Supplementary Figure 1A](#)). We implemented the same gaussian mixture model using the function GaussianMixture from Python package sklearn (version 0.23.1) to predict gender on ff calculated by Bayindir [25] ([Supplementary Figure 1B and C](#)). WisecondorX quantifies the proportion of reads on the Y chromosome with respect to the total number of reads per sample. To identify the threshold to discriminate fetal gender, we projected MagicY threshold on y fraction counts ([Supplementary Figure 1D](#)). Finally, Defrag uses a KNN binary classification on a training set to classify samples as males or females [10].

**Tools for ff estimation**

Performances of four ff estimation tools were assessed on our two cohorts: Defrag.a and Defrag.b [10], Seqff [14] and Sanefalcon [15]. Tools characteristics are described in Table 1. Default parameters were used for Defrag.a, Defrag.b and Seqff. The original code of Sanefalcon was completely revised in order to improve its computational performances and to automatize the several steps needed to perform the calculations.

**iSanefalcon**

Here, we explain the major changes we have made to upgrade the original code in an optimized version. First, we migrated to Python3.6, removing the inner dependencies between Bash scripts and python scripts to make a more stand-alone application, updating the dependencies with the newer versions of supporting software. The stand-alone application can be executed on any platform, and it exploits the core Python libraries for concurrency and parallelization: we heavy parallelized all the steps, from the read-start positions extraction to the nucleosome profile computation. We introduced a dedicated class for managing all the file system operations: in this way there is no manual intervention to be carried out (as in the original version), but everything is managed at runtime. Finally, we reduced at the bare minimum the manual configuration steps that need to be addressed. This way, we made the application more robust and less error-prone. A simple configuration file is provided to set up all the locations and the most important parameters for the application. The novel workflow of Sanefalcon is further explained in [Supplementary Figure 2](#) and the code of iSanefalcon is freely available at: [https://github.com/UCA-MSI/iSanefalcon](https://github.com/UCA-MSI/iSanefalcon).

To demonstrate that iSanefalcon obtains the same results as the original implementation, we calculated ff for both cohorts with both versions, and correlation plots are showed in [Supplementary Figure 3](#). Around 300 samples for each cohort were used as training set, the rest as test, having cared to keep together samples from the same run. To provide ff for each sample, the procedure was applied five times for cohort 1 and two times for cohort 2. Good correlation values are reported for both confirming that the two versions of Sanefalcon do not show significant difference on ff calculation. However different trends are observed for cohort 1 only. To inspect this result, we used a mixed effect model with the equation:

\[
y = ax + b + r
\]

combining the linear model expected with a random effect r, that cause the offsets on Y axis for each group. Both the likelihood and Akaike information criterion (AIC) of linear and mixed model confirmed that the mixed model better suits the modeling for cohort 1 (linear model, Log-likelihood: \(-537.31\) AIC: \(819.906\) and mixed model, Log-likelihood: \(735.863\) AIC: \(-1463.727\)). The mixed effect is relevant only for cohort 1 ruling out that it is due to the new implementation of Sanefalcon. PCA analysis on both cohorts suggests the mixed effect for cohort 1 does not originate from the samples ([Figure 1](#)). Sanefalcon strategy is the only analysis done at read start position resolution, whereas the others are the bin level. This suggests that the read start position resolution is not appropriate for small-size cohorts, probably due to a lack of variability.

**Assessment of fetal aneuploidies**

WisecondorX was used to detect potential fetal aneuploidies and includes three different steps: (i) Conversion from .bam (output of mapping) to .npz (required input format for WisecondorX), (ii) Creation of a reference composed of a subset of samples from the cohort. (iii) Determination of Z-score based on the pre-established reference. To establish reference samples, we selected randomly 100 samples from each cohort and used default method. Performances of WisecondorX were evaluated by comparing the obtained Z-scores to confirmed aneuploidies from our retrospectives cohorts.

**NiPTUNE**

NiPTUNE is composed of seven main independent blocks deployed into an integrated pipeline to perform NIPT analysis. The following paragraphs explain each block composition and purposes, for specific input/output, parameters and specific algorithm details.

**Configurations and input files preprocessing**

This block, composed of two modules, prepares sample files for use by the successive tools of the pipeline and calculates SD. The module triton.py converts the original file format (.bam) to other formats (.pickle, .gcc, .npz), which serve as input for downstream processing in NiPTUNE. The ‘.pickle’ output
Table 1. Main characteristics of ff estimation tools selected for benchmark in the present study.

| ff tools      | Main principle                                                                 | Male specific | Read counting method | GC correction | Input       | Output        | Programming language | Repository link          |
|---------------|--------------------------------------------------------------------------------|---------------|----------------------|---------------|-------------|---------------|----------------------|--------------------------|
| Defrag a      | Based on normalized fraction of reads on chromosome Y                           | Yes           | bin (1Mo)            | Yes           | .bam, .pickle | .csv          | Python               | https://github.com/VUmcCGP/wisecondor/tree/legacy |
| Defrag b      | Based on fraction of reads uniquely mapped to chromosome Y (subset of chromosome Y) | Yes           | bin (1Mo)            | Yes           | .bam, .pickle | .csv          | Python               | https://github.com/VUmcCGP/wisecondor/tree/legacy |
| Seqff         | Based on pretrained bin counts                                                  | No            | bin (50ko)           | Yes           | .sam        | .csv          | R                   | https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1002/pd.4615 |
| iSanefalcon   | Based on the spatial distribution of reads on the estimated nucleosome profiles   | No            | Read start position  | No            | .bam        | .csv          | Python               |                                                                       |

contains aligned reads split in 1 Mb bins (tunable parameter), the ‘.gcc’ output is the result of GC correction at the bin level obtained by applying a locally weighted scatterplot smoothing (lowess) model with respect to the reference genome. Finally, the ‘.npz’ format is a specific file used as input for WisecondorX.

The module despina.py calculates the SD of the sample(s) provided.

Quality control

The module proteus.py performs the PCA on all the samples of the cohort and produce a visual output to check whether there are outlier samples or subpopulations into the cohort. This quality control can be used only if a certain amount of samples is available [26].

Fetal sex prediction

To predict the sex of the fetus the module halimedea.py implements the ‘MagicY’ method combined with the gaussian mixture model.

Reference sets creations

The module larissa.py randomly selects n samples to be used as reference for Defrag and WisecondorX. In the present work, we set n = 100 samples for each cohort. We strongly recommend the use of cohort specific samples to improve precision and reliability of the different tests.

Fetal fraction prediction

Two methods are proposed in NiPTUNE to calculate ff, namely: Defrag_a based on chromosome Y read counts, implemented in the module lamedeida.py and Seqff based on pretrained bin counts in the module neso.py. The core components of Defrag_a and Seqff softwares were integrated in NiPTUNE. We added a script to parse Seqff input to make it more efficient and able to calculate ff on multiple samples in parallel. Sanefalcon is not deployed in the main workflow because our benchmark demonstrated that it performs less efficiently on our samples.

We provide an improved version of this tool to allow users to run it independently of NiPTUNE.

Copy number alteration prediction

We wrapped WisecondorX in the module sao.py. For each sample submitted to this module, a Z-score is calculated on the binned sample with respect to the reference samples. A graphical output is also provided to visually inspect the chromosomes with abnormal counts.

Prepare output

The last module of the pipeline, thalassa.py, collects results from previous modules in a table. Each sample (line) is described by the following columns: quality control (visual output), SD, fetal sex prediction (magicY), ff (2 columns one for each method) and the Z-score for chromosome 13, 18 and 21.

Results

NiPTUNE: a computational pipeline for NIPT in an accurate, integrative and flexible framework

To develop a computational pipeline to perform all steps of NIPT, from sequenced mapped reads to chromosomal aberration prediction, we performed an extensive benchmark and selected the best performing tools. This pipeline is a flexible framework, composed of seven independent modules (Figure 1) that can be used by expert users as stand-alone tool to create personalized pipelines. The first step is quality control performed through PCA analysis. We provide a module that estimates the fetal sex of the fetus with MagicY, a novel method that outperformed existing ones (see Supplementary Information). NiPTUNE provides estimation of ff with two tools, namely Defrag_a and Seqff and chromosomal aberrations are estimated by WisecondorX using a threshold of 5 of the Z-score. Finally, we provide a module that automatically collects all the results of analyzed samples in a table-like format that can be easily processed.
NiPTUNE: an automated pipeline for noninvasive prenatal testing

Figure 1. NiPTUNE workflow. NiPTUNE is composed of seven blocks, each comprising one or multiple modules. The corresponding modules are indicated in white boxes. The name of the associated python script is reported in italic. The last column reports the outcome of each block.

(Supplementary Table 3). Once the NIPT laboratory have calculated PCA on an available cohort, NiPTUNE can run either a single sample or batches of samples, adding this or these new sample(s) to the original cohort without re-running NiPTUNE on already analyzed samples. We validated and benchmarked all steps of the pipeline using two cohorts including a total of 1439 samples with 31 confirmed aneuploidies.

Sequencing quality control

NIPT was performed on two cohorts of pregnant women: Nice (cohort 1) and Marseille (cohort 2). Cohort 1 consists of 377 samples, including 11 fetal aneuploidies. Cohort 2 is composed by 1062 samples, including 20 fetal aneuploidies (Supplementary Table 1). Furthermore, samples from two non-pregnant women were included as controls. We retrospectively used generated sequencing data, post alignment to the GRCh37 reference genome, to validate our approach. We corrected for GC bias applying a loess function with respect to the GRCh37 reference genome. To test the quality of normalized samples, we inspected the distribution of mapped reads onto the genome. Specifically, we divided the genome in bins of 1 Mb and we quantified the number of normalized reads after GC correction. We expect that most of the bins show a count around 1 due to the normalization. To verify that samples do not have aberrant read count distributions, we applied the PCA on binned count of normalized reads for samples belonging to each cohort (Figure 2). We observed that the distribution of points (samples) is coherent in the two cohorts (Figure 2A and B): most
Figure 2. Sequencing quality control. PCA on normalized binned counts of reads for samples belonging to (A) cohort 1 or (B) cohort 2. (C) PCA on joined cohorts. SNPW (sequencing results from nonpregnant women) samples were added as control. (D) PCA on cohort 1 with addition of three test samples, two of them resulted from failed alignments to the reference genome and the third correspond to a maternal aneuploidy. Color code: cohort 1, gold; cohort 2, grey; samples from nonpregnant women (SNPW), black; test samples: failed alignments to the reference genome, red and maternal aneuploidy, green.

of the points are clustered around the centroid of the plot. If cohorts came from two different hospitals, the extraction and sequencing method were identical [23]. We thus joined the two cohorts for cross comparison (Figure 2C). Sequencing results from two nonpregnant women are distributed homogeneously, alike fetal aneuploidies. To test the reliability of the method, we also added three test samples from sequencing results to those from cohort 1. Two of them resulted from failed alignments to the reference genome and the third one corresponded to a maternal aneuploidy. A failed alignment is an erroneous reads mapping resulting in an imbalance of reads distribution on the genome due to an error in the laboratory automatical alignment pipeline. The points corresponding to the three test samples were scattered from the main group (Figure 2D). These analyses support the validity and importance of using PCA as a quality control to test the distribution of read counts after mapping before further analysis. Furthermore, they highlight the contribution of PCA to identify disease-associated genomic maternal abnormalities that could lead to false NIPT interpretation.

Identification of a reliable strategy for fetal gender prediction

Fetal gender prediction is an important step of NIPT pipeline because the gender of the fetus is used by several tools to determine the set of samples to be used as reference. Thus we needed to establish a confident method to predict fetal gender from sequencing data. We tested different tools described in the literature on cohort 1, including 377 samples. Gender could not be confirmed for 28 samples out of 377 because we could not obtain the pregnancy outcome for these fetuses (with no chromosomal abnormality according to NIPT). The analysis was not performed on cohort 2 because gender outcome was not available.

We defined a new calculation that we called 'MagicY'. We selected 7 chromosome Y-specific regions and estimated the proportion of reads belonging to these regions (Supplementary Table 2[24]). A bimodal distribution is observed. Thus, to separate male and female samples, we applied a gaussian mixture model on the distribution of these counts for the entire cohort (Supplementary Figure 1A). This method gave an agreement of 97.1% with gender outcomes, respectively (Supplementary Table 4).

To validate our approach, we compared with other three available methods. First, we used the calculation based on autosomes and chromosome X, proposed by Bayindir et al., as part of their aneuploidy detection pipeline [25]. Beek et al. suggested that Bayindir was not the most reliable tool for ff calculation [10]. Our data agree with theirs, we observed that ff calculation stratifies samples by gender (P-value < 2.2 × 10^{-16} Wilcoxon test; Supplementary Figure 1B). We thus resolved to use this property for gender prediction. We then applied the gaussian mixture model to Bayindir results (Supplementary Figure 1C) and obtained an agreement of 94.3% with gender outcomes (Supplementary Table 4). The suite of tools from WisecondorX
NiPTUNE: an automated pipeline for noninvasive prenatal testing

Figure 3. Fetal fraction trends with respect to gestational age. Fetal fraction (ff) estimated with the four tools, from left to right: Defrag_a, Defrag_b, Seqff, iSanefalcon, shown as violin plots accordingly to gestational age intervals for cohort 1 (A–D) and cohort 2 (E–H). The number of samples is reported above each violin plot. Color code: cohort 1, gold and cohort 2, grey.

predicts fetal gender by calculating the proportion of reads on chromosome Y with respect to the total number of reads for each sample followed by a gaussian mixture model. However the model could not find a threshold to separate male and female populations. Since gender prediction is a mandatory step for WisecondorX, we needed to find another strategy to assess this threshold. Thus we projected the threshold estimated by the mixture model from MagicY counts of y fraction counts obtained by WisecondorX (Supplementary Figure 1D). This procedure gave an agreement with pregnancy outcomes of 87.7% (Supplementary Table 4). This lower agreement could be due to pseudoautosomal regions of X and Y chromosomes leading to a miscount of chromosome Y-specific reads. The Defrag ff estimation tool includes a gender prediction step. Analysis of cohort 1 by Defrag only agreed with 90.3% of gender outcomes (Supplementary Table 4).

In conclusion, these analyses allowed us to develop a novel read count, based on chromosome Y-specific regions, followed by a gaussian mixture model approach to estimate fetal gender with high accuracy called MagicY.

Performances of ff estimation tools with respect to clinical parameters

The ff is a fundamental parameter for reliability of chromosomal anomaly calculation. If aneuploidy is ff-independent per se, its prediction can be affected by a low ff [5]. Despite the importance of ff, a gold standard method for its calculation is not yet established. Here, we used four tools most commonly used in the literature: Defrag_a, Defrag_b, Seqff and Sanefalcon. Characteristics of the ff tools are listed in Table 1. Defrag_a showed the highest performances in previous benchmarks [10, 11]. However, this method is only usable in the case of male fetuses because the ff estimation is chromosome Y-based. Despite Defrag_b as well is only for male fetuses, the two methods are slightly different. Although Defrag_a considers reads mapped on the entire chromosome Y, Defrag_b only uses specific male regions on this chromosome. Both Defrag and Seqff were used with default configuration. Sanefalcon original version of the code was entirely revised in order to automatize and parallelize the multiple steps needed to perform ff calculation from nucleosome profiles estimation for each sample. We provide an improved version of Sanefalcon, called iSanefalcon, which implements the original algorithm with more powerful computation performances and easier configuration. We used iSanefalcon for benchmark purposes in this work after comparison with the original version (see Materials and Methods).

The ff ranges are reported by tool and cohort in Supplementary Table 1. The ranges of ff found by the different tools are consistent between the two cohorts and they are consistent with experimental ranges found by SNP-based techniques [10, 11] by suggesting the reliability of the results.
In contrast, iSanefalcon calculated very different ranges of ff from the two cohorts: from 0 to 7.8 (cohort 1) and 9.1 to 23.3 (cohort 2) suggesting a lower reliability of the results.

Then, we analyzed the trend of ff with respect to clinical parameters (gestational weeks, maternal age and BMI; the latest being available for cohort 1 only). It is well established in the literature that ff increases with gestational weeks [5, 6]. This behavior was confirmed by our analysis on the two cohorts (Figure 3). When ff was calculated by iSanefalcon, no specific trend was observed except for a slight increase at gestational weeks from 28 to 32 for cohort 1 (Figure 3). No particular trends were found for the two other parameters (maternal age and BMI) regardless of the tool (Supplementary Figures 4 and 5). Although no trend was expected for maternal age [5], the correlation between BMI and ff was reported in the literature mainly in large cohorts [5]. Here, the absence of correlation between these two parameters could be explained by the small size of cohort 1, especially considering that cohort 1 includes a few samples [20] of BMI values equal or over 35 (Supplementary Figure 5). As
expected, no stratification by fetal sex was observed with Seqff and iSanefalcon (Supplementary Figure 6).

Benchmark of tools for ff estimation

To compare ff tools performances, we performed pairwise correlations of ff calculated by the four tools and we reported the scatter plots with regression lines in Figure 4 and correlation values in Supplementary Table 5. Defrag_a and Defrag_b do not show a very high correlation. We observed an underestimation of low ff and an over estimation of high ff for Defrag_b compared to Defrag_a. The best correlation was found between Defrag_a and Seqff. However Seqff slightly underestimates ff values with respect to Defrag_a except for extremely high ff. The behavior of Defrag_b was confirmed and accentuated when compared with Seqff, confirming an underestimation of low ff and an overestimation of high ff. As expected, a very poor correlation was observed between iSanefalcon and any other method, due to the very narrow range of ff calculated by this tool on both cohorts.

Although Defrag_b shows stable results for the two cohorts, its behavior regarding extreme ff values can be very misleading in the context of NIPT. Low values of ff can lead to false negative results. Furthermore, NIPT is usually performed between 12 and 16 gestational weeks, period during which a low percentage of ff is expected (Figure 3). iSanefalcon showed both strong cohort-dependent behaviors and very poor correlation with any other tool for ff calculation. Previous benchmarks already compared Sanefalcon with Defrag and Seqff showing better performances with respect to the ones reported here [10, 11]. We are considering two explanations. First, the computational complexity of the original code of Sanefalcon pushed the authors of previous benchmarks to use pretrained reference nucleosome profiles while we recomputed the nucleosome profiles for each of our cohorts, thanks to our new version of the code. These reference profiles were then used to estimate the ff by comparing to the nucleosome track for each sample. To have a reliable estimation of ff, reference profiles should be calculated for a given NIPT protocol since specific tests showed that the nucleosome profiles affect the ff outcome. Second, Sanefalcon uses the reads start positions as input for nucleosome profile/track estimation. In data used for Sanefalcon original publication and previous benchmarks, all single ended reads were short (51 bp) while the median of reads generated in our study is ~170 bp.

Taken together, these results ruled out Defrag_b and iSanefalcon as confident tools for ff calculation.

Computational prediction of chromosomal aberrations

To test for fetal aneuploidies, we used WisecondorX [19], an upgrade of the original version Wisecondor [18]. WisecondorX is more computationally efficient and allows gonosomal copy number variation detection. Briefly, the estimation of chromosomal aberration is performed through a Z-score calculation that represents the ratio between normalized bin counts on chromosomes compared with bin counts from reference samples. A smoothing approach is then used to estimate a global Z-score per chromosome from Z-scores of local bins. The definition of reference samples, from which reference bins are selected, is of fundamental importance for Z-score calculation and reliability of results. Reference samples must not include aberrant samples and should represent the variability of cohorts. Results from WisecondorX on our cohorts are shown in Supplementary Table 6. A Z-score >5 is indicative of a potential chromosomal abnormality. Cohort 1 is composed of 11 aneuploid samples, including five trisomies of chromosome 18 (T18) and six trisomies of chromosome 21 (T21). Cohort 2 contains 20 aneuploidies, among which two T18, sixteen T21, one sample with both T18 and T21 and one T13 (Supplementary Table 1). Among aneuploid samples, two corresponded to dichorionic diamniotic twin pregnancies with one fetus out of two carrying either a T18 or a T21. Importantly, the Z-scores calculation with WisecondorX identified all trisomies in both cohorts, including the ones for twin pregnancies and the double trisomy. Importantly, WisecondorX did not lead to false negative results for chromosomes 13, 18 and 21 (Supplementary Table 6). These results confirm the specificity and the sensibility of WisecondorX approach.

Conclusion

In this study, we have measured and compared the performances several tools used for NIPT analysis showing their strengths and weaknesses. Of note, we performed the first benchmark of fetal fraction tools and the first application of WisecondorX on NIPT sequencing data from Ion Sequencing. We provide a novel method for fetal sex prediction that outperforms existing tools. We also proved for the first time that the PCA analysis is a valid and essential quality control for NIPT samples. This allowed us to develop NiPTUNE that integrates highly performing tools to secure all steps of NIPT, allowing a confident estimation of ff, fetus gender and Z-score.

In conclusion, NiPTUNE is a pipeline for data analysis that allow high accuracy NIPT results. NiPTUNE will reduce both the number of false negatives, leading to undiagnosed aneuploidies and false positives, causing unnecessary invasive testing.

Key Points

• We provide guidelines for NIPT including samples quality control through PCA analysis.
• We developed a novel method for accurate fetus sex prediction.
• We showed that, depending on the tool used to calculate ff, different intervals are obtained.
• NiPTUNE integrates highly performing tools to secure all steps of NIPT, allowing a confident estimation of ff, fetus sex and Z-score.
• We provide the first benchmark of fetal fraction tools and the first application of WisecondorX on NIPT sequencing data from Ion Sequencing.

Authors’ contributions

SB, VD and VPF conceived the study, supervised and interpreted all analysis and wrote the manuscript. VD and JB collected the sequences. DP and MM performed all bioinformatics analysis and developed NiPTUNE. SD participated in data analysis supervision. SB conceived the tool. All authors read and contributed to the final manuscript.

Data and code availability

NiPTUNE, is available in the GitHub repositories: https://github.com/UCA-MSI/NiPTUNE, the code is available from the corresponding author on request. The data underlying this article are available in the article and in its online supplementary material.
Supplementary data

Supplementary data are available online at Briefings in Bioinformatics.

Acknowledgments

We acknowledge the Centres Pluridisciplinaires de Diagnostics Prenatal (CPDPN) from CHU of Nice and Timone Hospital (Marseille) for providing files of pregnant women. We thank Dr Marco Lorenzi for insightful discussions.

Conflict of Interest

The authors declare that they have no competing interests.

Funding

This work was supported by the French government, through the UCA JEDI Investments in the Future project managed by the National Research Agency (ANR) under reference number ANR-15-IDE-01.

References

1. Tamminga S, van Maarle M, Henneman L, et al. Maternal plasma DNA and RNA sequencing for prenatal testing. Ado Clin Chem 2016;74:63–102.
2. Chiu RWK, Chan KCA, Gao Y, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. PNAS 2008;105:20458–63.
3. Lo YMD. Noninvasive prenatal testing using massively parallel sequencing of maternal plasma DNA: from molecular karyotyping to fetal whole-genome sequencing. Reprod Biomed Online 2013;27:593–8.
4. Lo YMD, Corbetta N, Chamberlain PF, et al. Presence of fetal DNA in maternal plasma and serum. The Lancet 1997;350:485–7.
5. Ashoor G, Syngelaki A, Poon LCY, et al. Fetal fraction in maternal plasma cell-free DNA at 11-13 weeks' gestation: relation to maternal and fetal characteristics. Ultrasound Obstet Gynecol 2014;43:26–32.
6. Palomaki GE, Kloza EM, Lambert-Messerlian GM, et al. DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. Genet Med 2011;13:913–20.
7. Chan RWY, Jiang P, Peng X, et al. Plasma DNA aberrations in systemic lupus erythematosus revealed by genomic and methylation sequencing. PNAS 2014;111:E5302–11.
8. Schuring-Blom H, Lichtenbelt K, van Galen K, et al. Maternal vitamin B12 deficiency and abnormal cell-free DNA results in pregnancy: maternal vitamin B12 deficiency and abnormal NIPT outcome. Prenat Diagn 2016;36:790–3.
9. Hui L, Bianchi DW. Fetal fraction and noninvasive prenatal testing: what clinicians need to know. Prenat Diagn 2020;40:155–63.
10. van Beek DM, Straver R, Weiss MM, et al. Comparing methods for fetal fraction determination and quality control of NIPT samples. Prenat Diagn 2017;37:769–73.
11. Hestand MS, Bessem M, van Rijn P, et al. Fetal fraction evaluation in non-invasive prenatal screening (NIPS). Eur J Hum Genet 2018.
12. Peng XL, Jiang P. Bioinformatics approaches for Fetal DNA fraction estimation in noninvasive prenatal testing. Int J Mol Sci 2017;18:453.
13. Hudecova I, Sahota D, Heung MMS, et al. Maternal plasma fetal DNA fractions in pregnancies with low and high risks for fetal chromosomal aneuploidies. PLoS One 2014;9.
14. Kim SK, Hannum G, Geis J, et al. Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts: determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts. Prenat Diagn 2015;35:810–5.
15. Straver R, Oudejans CBM, Sistermans EA, et al. Calculating the fetal fraction for noninvasive prenatal testing based on genome-wide nucleosome profiles. Prenat Diagn 2016;36:614–21.
16. Talevich E, Shain AH, Botton T, et al. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput Biol 2016;12:e1004873.
17. Klambauer G, Schwarzbauer K, Mayr A, et al. cnMOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with low false discovery rate. Nucl Acids Res 2012;40:e69–9.
18. Straver R, Sistermans EA, Holstege H, et al. WISECONDOR: detection of fetal aberrations from shallow sequencing maternal plasma based on a within-sample comparison scheme. Nucl Acids Res 2014;42:e31–1.
19. Raman I, Dheedene A, De Smet M, et al. WisecondorX: improved copy number detection for routine shallow whole-genome sequencing. Nucl Acids Res 2019.
20. Boeva V, Zinovyev A, Bleakley K, et al. Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. Bioinformatics 2011;27:268–9.
21. Scheinin I, Sie D, Bengtsson H, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. Genome Res 2014;24:2022–32.
22. Xi R, Lee S, Xia Y, et al. Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants. Nucl Acids Res 2016;44:6274–86.
23. El Khattabi LA, Brun S, Gueguen P, et al. Performance of semiconductor sequencing platform for non-invasive prenatal genetic screening for fetal aneuploidy: results from a multicenter prospective cohort study in a clinical setting. Ultrasound Obstet Gynecol 2019;54:246–54.
24. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 2003;423:825–37.
25. Bayindir B, Dehaspe L, Brison N, et al. Noninvasive prenatal testing using a novel analysis pipeline to screen for all autosomal fetal aneuploidies improves pregnancy management. Eur J Hum Genet 2015;23:1286–93.
26. Mundfrom DJ, Shaw DG, Ke TL. Minimum sample size recommendations for conducting factor analyses. Int J Testing 2005;5:159–68.