NIPT technique based on the use of long chimeric DNA reads

Vera Belova
Pirogov Medical University
Daria Plakhina
Genotek Ltd
Sergey Evfratov
Genotek Ltd
Kirill Tsukanov
Genotek Ltd
Gennady Khvorykh
Genotek Ltd
Alexander Rakitko
Genotek Ltd
Alexander Konoplyannikov
Pirogov Medical University
Valery Ilinsky
Genotek Ltd
Denis Rebrikov
Pirogov Medical University
Dmitriy Korostin (d.korostin@gmail.com)
Pirogov Medical University https://orcid.org/0000-0003-1343-2550

Technical advance

Keywords: NIPT, chimeric DNA, cfDNA, NGS, short DNA fragments, fetal fraction, long chimeric reads

Posted Date: February 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-15433/v1

Abstract

Background: Non-invasive prenatal testing for aneuploidy on chromosomes 21, 18 and 13 is actively used in clinical practice around the world. One of the limitations of the wider implementation of this test is the high cost of the analysis itself, as the high throughput sequencing is still relatively expensive. At the same time, there is a trend of increase of the length of reads yielded by sequencers. Since extracellular DNA is short, in the order of 140-160 bp, it is not possible to effectively use long reads.

Results: The authors used high-performance sequencing of cfDNA libraries that went through additional stages of enzymatic fragmentation and random ligation of the resulting products to create long chimeric reads. The authors used a controlled set of samples to analyze a set of cfDNA samples of pregnant women with a high risk of fetus aneuploidy according to the results of the first trimester screening and confirmed by invasive karyotyping of the fetus using laboratory and analytical approaches developed by the authors. They evaluated the sensitivity, specificity, PPV and NPV of the results.

Conclusions: The authors developed a technique for constructing long chimeric reads from short cfDNA fragments and validated the test using a control set of extracellular DNA samples obtained from pregnant women. The obtained sensitivity and specificity parameters of the NIPT developed by the authors corresponded to the approaches proposed earlier (99, 93% and 99.14% for 21 trisomy; 100% and 98.34% for 18 trisomy; 100% and 99.17% for 13 trisomy, respectively).

Background

NIPT studies are carried out early in the pregnancy (starting 9–12 weeks of gestation) to determine the presence of aneuploidy in the fetal genome. This test is done using the freely circulating cfDNA isolated from the mother's blood, part of which is of fetal origin. cffDNA enters the mother's blood as a result of apoptosis of placental trophoblast cells and reaches a proportion of 4–15% by the end of the first trimester of pregnancy. In the mother's blood, cfDNA is in the form of fragments. Fragments of maternal cfDNA are predominantly 166 bp in length, and cffDNA fragments are 143 bp in length. [1]. This distribution of the lengths is associated with nonrandom DNA cutting [2]. NIPT is a method for analysis of the relative number of copies of chromosomes in the extracellular DNA and in most cases it is done using the sWGS approach. The obtained sequencing data of the corresponding chromosome normalized by the remaining chromosome coverages is compared to the results of the analysis of the controlled set of samples with a known karyotype using a Z-test. Reaching threshold values of Z statistics indicates the presence of aneuploidy on the corresponding chromosome in the fetus.

The proportion of fetal DNA is an important parameter, and as such is necessary to assess. If the values are too low, then the NIPT may result in a false negative. The two most common approaches of assessment are analysis of SNPs and cfDNA length distribution. SNP analysis is a deep sequencing (starting at x100) of pre-amplified DNA fragments containing SNPs, according to which the mother and fetus may be differentiated due to alleles inherited from the father. Analysis of several dozen such sites

by counting reads containing paternal alleles determines the proportion of cffDNA. Length distribution analysis uses the results of pair-end mapped reads to obtain cfDNA length distribution data. The cffDNA proportion is estimated by comparing the number of reads of certain lengths, which is related to the difference in the initial length of the fetal and maternal cfDNA.

NGS technologies compete in two areas: the number of molecules processed in a run and the read length, which reaches 200 and 250 bp in PE mode in modern sequencing platforms (MGISEQ-2000, Illumina HiSeq 2500 in Rapid Run Mode). However, at a length as short as 40 bp, 82% of the molecules will be unambiguously mapped to the human genome [3]. Therefore, for applications not related to variant calling, increasing the length of the read does not bring additional benefits. More than 20 years ago methods were developed to combine short DNA or cDNA fragments originating from different molecules into chimeric DNA molecules [4]. Thus, one chimeric molecule contains a set of tags that map to different parts of the genome. Chimeric molecules are extremely useful in studies requiring the determination of numbers of copies, such as differential gene expression studies or CNV searches in the genome.

The goal of this work was to expand the laboratory and bioinformatics stages of NIPT using analysis of tags in chimeric molecules built from cfDNA fragments of pregnant women.

Results

Clinical characteristics

The study included 145 women aged 20 to 48 years. The gestational ages varied from 11 to 25 weeks (15.1 weeks on average). Demographic and clinical characteristics of the patients are presented in Table 1. All subjects had a singleton pregnancy, were not carrying a donor egg, did not have clinically established rearrangements in their genome, did not go through an embryo reduction procedure, were not cancer patients, did not undergo blood transfusions during the last six months before the biomaterial collection procedure, and also did not undergo organ transplantation, including bone marrow or stem cell therapy.

Extraction of cfDNA

We have demonstrated (data not shown) that the efficiency of using the QIAamp DNA Blood Mini Kit to isolate cfDNA from 3-4 ml of blood plasma is comparable with that of the QIAamp Circulating Nucleic Acid Kit, so we used the first kit.

Smash protocol

The library preparation protocol we propose differs from the classical NIPT protocol in three elements:

1. Fragmentation of cfDNA to short fragments
2. Size-selection of the fragmented cfDNA from two sides, leading to a number of fragments with an average length of 40-50 bp remaining in the sample.

3. The random ligation of short cfDNA fragments and thereby the formation of long (more than 300 bp) chimeric DNA molecules to which adapters for NGS are already ligated.

As a result of the sample preparation conditions that we selected and the presence of reagents for sequencing long reads (Rapid Run PE250), we were able to obtain a larger amount of useful diagnostic information from each read compared to the classic NIPT. The average fragment size in the test samples was 43.8 bp, thus, the information content of one direct or reverse read with a length of 250 bases was 5.7 times higher than for the classic single read. Since it was impossible to determine the size of the insertion of a pair of reads by mapping them onto the reference genome, in our case we tried to estimate the initial sizes of sequenced chimeric molecules by fusing them using the FLASH tool. The length distribution of chimeric molecules for one of the typical samples is shown in Figure 2. It can be seen that most of the reads had a length of about 250 bp. This means that it was not possible to fuse the pairs of reads, and the length of the chimeric molecule for them exceeded 460 bp, which correlated with the histogram data from the Bioanalyzer (Figure 1A).

Choice of SNPs

SNPs were selected based on an analysis of 1000 Genomes [12] and gnomAD [13] databases. SNP filtering was performed using the VCFtools software package [14]. Markers in the human genome were selected according to the following criteria:

Insertions, deletions, chromosomes X and Y, mitochondrial DNA and the p-arm of chromosome 6 were not included;

Only diallelic polymorphisms were considered;

Only SNP with identifiers “rs” were considered;

Only alleles with a minimum allele frequency (MAF) of 0.4 were considered;

We considred SNPs with a minimum probability of selection according to Hardy-Weinberg of p 0.00001 (i.e. the selection has almost no effect)

Only unlinked SNPs were considered. They were determined using the sliding window method with the following parameters: correlation coefficient $r2 < 0.5$, sliding window size = 50 SNP, step = 5 SNP.

Next , only those SNPs that were common for samples from European populations were used for further analysis.

Primers were selected for the 610 selected SNPs using the AmpliSeq Designer software (www.ampliseq.com) with an amplicon size of no more than 140 bp. The nucleotide profile around the selected primers was analyzed and those that could have potential amplification problems (hairpins, homopolymers etc.) were removed. Sequential filtering was done using the following criteria:

- absence of homopolymers in the primer (4 or more identical nucleotides in a row);

- no repetitions (eg, “ATATATA ...”);

- absence of a large number of GC at the 3' end;

- GC-composition of the primer ranging from 0.4 to 0.6.

As a result, we designed a panel of primers for 102 SNPs, and ordered them to be synthetized from Thermo Fisher as a custom Ampliseq panel.

Amplifet library preparation

We used the standard approach to library construction from PCR products as the Amplifet library preparation protocol: at the first stage, the amplification of the regions of interest was carried out, and then adapters for sequencing were ligated to the purified amplicons with appropriately modified ends. We did not destroy the primer dimers using FuPa or mixtures containing uracilglycosylase, as they were effectively eliminated by the first washing with Ampure XP.

Smash sequencing results

The sequencing results for a subset of smash libraries are shown in Table 1 of Supplementary files. By filtered fragments we mean subreads from chimeric long reads that have passed the filtration stages described in the corresponding section of Materials and Methods. The total length of the regions in the RepeatMasker track was 1,586,326,530 nucleotides, which is approximately half the human genome, therefore it can be assumed that the total number of fragments per chimeric read before filtering was twice as that after filtering. Chimeric reads were read using the pair-end method for 250 nucleotides in both directions. *Every* pair of 250 bp reads contains 2,37 filtered fragments. It is approximately 4,6 times higher than classic NIPT pair because half of them are filtered during RepeatMasker usage.

Amplifet sample sequencing results

Amplifet library sample sequencing results are shown in Table 2 of Supplementary files. In the control subset of samples, 96 out of 102 systems worked (94%). On average, 5099 reads were obtained for each point. For two samples (za6077, zi2425), the average coverage at the point was less than x50; therefore, they were excluded from further analysis because they provided no reliable assessment of the proportion of fetal DNA. SNPs, according to which the fetus genotype was heterozygous and the mother was homozygous, were considered informative. Thus, the proportion of fetal DNA is defined as the fraction of reads per paternal allele doubled. On average, the proportion of fetal DNA in the set of the control samples was 12%, which is consistent with the results obtained for this range of gestational age [15] (Figure 3). 4 samples were excluded from further analysis as it was not possible to confirm data on gestational dates from the clinic for them (jt4805, ds3710, in5586, gv2133).

Determination of the minimum number of fragments in the sample

We determined how many fragments that passed through the filters is necessary and sufficient to determine aneuploidy. This was done by calculating the sensitivity and specificity of aneuploidy determination on chromosomes 21, 18 and 13 for samples with 1 to 3 million filtered fragments in increments of 0.5 million fragments. The calculation results are shown in Table 2. Figure 4 shows the specificity and sensitivity values calculated for aneuploidy of chromosome 21. Thus, a minimum of 2.5 million filtered fragments per sample is optimal.

Sample filtration

Samples were filtered in accordance with the criteria for the minimum number of filtered fragments defined above, as well as a cutoff value for the minimum fraction of fetal DNA in the sample of 4% [16]. 83 samples were included in the final sampling (normal karyotype – 48 samples, 58%; T21 – 25 samples, 30%; T18 – 6 samples, 7%; T13 – 4 samples, 5%). The list of final samples and their characteristics is presented in Table 3 of Supplementary files.

Sensitivity and specificity assessment for aneuploidy

To assess the sensitivity and specificity of the test, we used a z-score threshold of 3. To increase the sample size in the process of obtaining robust estimates of sensitivity, specificity and AUC, a cross-validation procedure was used, described above in the Materials and methods. The results of the Z-score calculation for the corresponding trisomies are shown in Figure 5, as well as in Table 4 of Supplementary files.

Results of the sensitivity and specificity assessment for aneuploidy are presented in table 3. Our results are consistent with the values obtained by other researchers [17-19].

Calculation of PPV and NPV for aneuploidy

Positive predictive value (PPV) and negative predictive value (NPV) are operational characteristics of the test and take into account not only the sensitivity and specificity metrics of our methodology, but also the disease frequency in the population. Since our test can be used in populations with different trisomy frequencies (for example, in risk groups according to the results of the first trimester screening), we calculated PPV and NPV for the corresponding frequencies of aneuploidy. The results are shown in table 4.

Determination of the sex of the fetus

Determination of the sex of the fetus was carried out using calculations of the normalized coverage of the Y chromosome. As can be seen in Figure 6 and in Table 5 in Supplementary files, male and female fetuses are well divided into groups. If the proportion of fetal DNA exceeds 4% and the proportion of the Y chromosome is not higher than 1%, the fetus is female. It can be seen that the results of the determination of the fetal DNA proportion in male samples for amplifet and the Y chromosome tend to correlate. We predict that further adjustments to the Y chromosome calculation system can closely

correlate with other methods for determining the proportion of fetal DNA. It is also possible that a similar approach of comparing X chromosome and autosome coverages can be used to determine the proportion of fetal DNA in mothers pregnant with females.

Discussion

In this study, we showed the possibility of using long chimeric reads for NIPT in a control set of 83 extracellular DNA samples from pregnant women with different fetal karyotypes. Our results show that the use of the combined approach (smash and amplifet libraries) can make better use of modern massive parallel sequencing technologies that use long reads. This is especially relevant if we consider the trend of increasing read lengths in recent years (for example, single-end 400 bp on MGISEQ-2000 or extra-long reads in nanopore sequencing technology). Our technique can be easily adapted for use with Oxford Nanopore technology. A similar study was conducted by our colleagues for CNV analysis [20]. With a good training set of cfDNA samples from pregnant women, the possibility of NIPT using nanopore sequencers can be promising.

Despite the obvious advantages of using NIPT in clinical practice, the widespread introduction of the test is largely limited by the high cost of its implementation. The technique demonstrated in this publication can significantly reduce the cost of NIPT, thereby making it more accessible to patients. Of course, this requires more extensive studies of the clinical efficacy of our test on larger patient samples.

We also see promise in the technique for estimating the proportion of fetal DNA directly from sequencing data of smash libraries. This will simplify the sample preparation procedure and reduce its cost. Despite the use of additional fragmentation, in which we lose information about the initial length of a particular cfDNA molecule, the ratio of extracellular DNA, as well as its 5 ’and 3’ ends, is preserved in the fetus and mother. Therefore, with an increase in the size of the training set, it will become possible to use methods for analyzing the nucleosome profile of the ends of the fragments [21]. Another approach may be the use of neural networks [22]. This approach is also limited by the size of the training sample set, so, in the future, we plan to conduct additional research on a larger set of samples.

Conclusions

In conclusion, we developed an NIPT and demonstrated the validity of the approach based on long chimeric reads. This technology already seems more economically justified for routine laboratory practice, however there are obvious directions in which our approach should be improved. In particular, it is important to develop alternative methods for assessment of the proportion of fetal DNA.

Methods

Ethics statement

This study was conducted according to the principles expressed in the Declaration of Helsinki. Appropriate institutional review board approval for this study was obtained from the Ethics Committee at Pirogov Russian National Research Medical University (#170 at 18.12.2017). All patients provided written informed consent for the collection of samples, subsequent analysis and publication of .

Sample processing and extraction of cfDNA

Pregnant women referred for invasive diagnostics at the Center for Family Planning and Reproduction in Moscow were recruited from December 2017 to December 2018. Maternal peripheral blood samples were collected into Streck BCT tubes just before obstetric procedures (such as CVS) during the first trimester. Maternal peripheral blood samples were centrifuged at 250 g for 30 min, then plasma was transferred to new 2,0 ml microtubes and centrifuged at 9000 g for 15 minutes. CfDNA was extracted using the QIAamp Blood kit (Qiagen) with an increased volume of lysate transferred to the column (2-4 ml of blood plasma at the beginning) and volume of extraction decreased to 40 uL. QC was performed using the Qubit HS kit measuring on Qubit 2 (ThermoFisher). 2 ng of cfDNA was used in the chimeric DNA library preparation and fetal fraction estimation stages of the assay.

To simplify, we will refer to the two types of libraries used in our method as «smash» and «amplifet» libraries.

Chimeric DNA library preparation (“smash”)

We refer to this type of library as “smash” in honor of the previously developed method [3], which prompted us to develop our methodology.

Fragmentation

2 ng of cfDNA was fragmented using the dsFragmentase kit (NEB): 1.25 uL of Fragmentase reaction Buffer v2; 2.5 uL of dsDNA Fragmentase; 3 uL of 50% PEG-8000 solution (Sigma Aldrich), and MQ water to make up to a total volume of 27 uL . Solution was incubated at 36℃ for 40 minutes.

Double size-selection to obtain 40-50 bp fragments

2x Ampure XP beads (Beckman) were added to 27 uL of fragmented products and incubated for 5 minutes at RT on benchtop. Then the microtube was placed on a magnetic rack until all the beads were concentrated on one side of the microtube. Supernatant with DNA less than 100 bp in length was transferred to a new microtube. Lower size-selection was done using QIAquick Nucleotide Removal Kit (Qiagen) to cut off DNA shorter than 40 bp. DNA was collected to make up a volume of 20 uL.

End-repair

End-repair reaction was done at RT for 180 minutes using a Quick blunting kit (NEB) by adding 2.5 uL of buffer, 2.5 uL dNTP mix and 1 uL of enzyme mix.

Self-ligation

Formation of chimeric DNA molecules was done by adding 4.5 uL of T4 ligation buffer, 0.5 uL of T4 DNA ligase (E320 kit, Sybenzyme), 9 uL of 50% PEG-8000 solution, and 1 uL of 5'-deadenylase (NEB) to the product of the end-repair reaction. Self-ligation was conducted at RT overnight.

A-tailing

4.5 uL of Taq buffer, 2 uL of Taq-pol (PK015L, Evrogen), and 2 uL of dATP (R0181, ThermoFisher) were added to the self-ligation product to make up to a final volume of 47 uL. The solution was incubated for 30 minutes at 65℃ and 2 minutes at 72℃.

Adapter ligation

Oligonucleotides dir_1 (ACACTCTTTCCCTACACGACGCTCTTCCGATCT) and rev_P (P*GATCGGAAGAGCACACGTCTGAACTCCAGTC) were synthesized in Evrogen (Moscow). dir_1 and rev_P were diluted to 5mM and combined in equal volumes, then hybridized in a thermocycler by heating to 95℃ for 5 minutes and slow cooling to RT. To 47 uL of A-tailing product we added: 3 uL of adaper mix; 5,75 uL of ligation buffer; 3 uL of T4 DNA ligase (E320 kit, Sybenzyme); 6.3 uL of 50% PEG-8000 solution; 0.5 uL of 5'-deadenylase. The reaction was incubated in a thermocycler for 100 cycles of the following program: 4℃ for 10 seconds, 16℃ for 30 seconds.

Size-selection

1x volume of MQ water was added to the adapter ligation product. Then x0,4 volumes of x2 concentrated Ampure XP beads was added. Standard cleaning procedure was performed, however, DNA was not eluted from the beads. PCR mix from next step was added directly to the beads with immobilized DNA.

Indexing PCR

In a new micrutube, the following reagents were combined: 1 uL of i5 and 1uL of i7 indexes from the E7600S NEB kit; 5uL of HiFi buffer; 0.5 uL of HiFi pol; 0.75 uL of dNTP's (all KAPA 7958897001). The mixture was added to beads with immobilized DNA and put into a thermocycler with the following program: 95℃ for 2 minutes; 98℃ for 20 sec, 65℃ for 30 sec, 72℃ for 2 minutes (for 15 cycles); 72℃ for 5 minutes.

Cleanup

PCR products were cleaned with x0.5 volume of Ampure XP beads. Elution was done in 20 uL of Low TE.

QC

QC was done using a High Sensitivity Kit for Bioanalyzer 2100 (Agilent). Result were considered optimal if the library peak was in 500-800 bp range and the concentration was more than 4 nM in 200-800 bp

range (picture 1A).

Fetal fraction estimation library preparation (“amplifet”)

We refer to libraries for assessing the proportion of fetal DNA as “amplifet”.

Multiplex PCR

2 ng of cfDNA was added to the first PCR with 20 uL of Amliseq primer mix (ThermoFisher); 8 uL of Phusion buffer; 0.4 uL of Phion U pol (F-555L kit, ThermoFisher); 0.8 uL of dNTP’s (pb006L Evrogen) and MQ water up to 40 uL. The mixture was amplified using the following program: 98℃ for 30 seconds; 98℃ for 10 seconds, 60℃ for 4 minutes, 72℃ for 20 seconds for 27 cycles; 72℃ for 5 minutes.

QC

Length of PCR products was assessed using agarose gel-electrophoresis.

Cleanup

1st PCR product was cleaned with x3 volumes of Ampure XP beads and eluted to 20 uL.

Adapter ligation

To ligate Illumina DIY adapters (same as in the «Adapter ligation» step described above), end-repair and A-tailing reactions were carried out in the same microtube but at a lower temperature. The following reagents were mixed in a new microtube: 10 uL of cleaned amplicons from 1st PCR; 5 uL of Ligase buffer (B302 Sybenzyme); 5 uL of adapter mix; 0.5 uL of T4 DNA ligase (E330 Sybenzyme); 0.5 uL of 5’deadenilase (M0331 NEB); 1 uL of 10 mM ATP (R0441 ThermoFisher); 1 uL of Klenow exo- (m0212L NEB); 1 uL of dATP (R0141 ThermoFisher); 2 uL of T4 PNK (EK0032 ThermoFisher); 12 uL of 50% PEG-8000 solution; 12 uL of MQ water. The mix was incubated at 37℃ for 40 minutes; 10℃ for 10 seconds, 30℃ for 30 seconds (100 cycles).

Cleanup

Ligation product was cleaned with x1.5 volumes of Ampure XP beads and eluted to 30 uL.

Indexing PCR

The following reagents were mixed in a new microtube: 1 uL of i5 and 1uL of i7 indexes from the E7600S NEB kit; 5uL of Phusion buffer; 0.25 uL of Phusion U pol; 0.5 uL of dNTP’s. The mixture was put into a thermocycler and ran through the following program: 98℃ for 30 seconds; 98℃ for 10 sec, 65℃ for 30 sec, 72℃ for 20 seconds (for 14 cycles); 72℃ for 5 minutes.

Cleanup

Cleanup was done using a GeneRead Size Selection Kit (180514 Qiagen). DNA was eluted to 20 uL of Low TE.

QC

QC was done using a High Sensitivity Kit for Bioanalyzer 2100 (Agilent). Results were considered optimal if the library peak was in the 270-280 bp range and the concentration was more than 4 nM in the 270-280 bp range (picture 1B).

Sequencing

NGS was performed using an Illumina HiSeq 2500 instrument with Rapid Run v2 kits designed for 500 cycles (PE250 dual-indexing).

Bioinformatics

Mapping

Raw data in BCL format were converted to FASTQ using bcl2fastq v. 2.20 software. Reads were mapped to the h38 genome in two iterations. Initially, BWA [5] version 0.7.17 with standard settings was used. In this case, smash-read fragments were distributed along the genome depending on where they were mapped. Reads with amplifet libraries were mapped entirely. During the second step, the subreads obtained as a result of the first mapping were extracted and mapped again as separate reads. Additional (supplementary) alignments and mapping onto the minus strand were filtered out (as after the first stage all exported subreads have already been inverted into the plus strand of the reference genome).

Filtering

For each smash library, the following steps were performed:

All BAM files that contain smash data for this sample were downloaded and, if there were several, were combined using samtools [6] version 1.9.

The following reads were filtered out:

imperfectly mapped onto the genome (MAPQ <60)

those that fall into regions of known repeats (RepeatMasker in Genome Browser track)

those that fall into amplicon regions

For each amplifet library, only reads that fall into the amplicons region (off-target reads were depleted) were filtered out.

FLASH tool with the following settings was used to calculate insertion length in smash libraries:

–min-overlap 20 –max-overlap 250 –allow-outies –max-mismatch density 0.20.

Statistics

We ran FetalQuant [7] with the default parameters to estimate the fractional fetal DNA concentration from the SNP data (amplifet target sequencing). The training samples were classified as case/ control based on chromosomal z-scores. We used the R-package NIPTer [8] to perform a variation reduction (peak, GC and chi-squared corrections), match QC and to calculate the z-score.

Filtering

We filtered out all samples in which the estimated fractional fetal DNA concentration was less than 4% or the total number of fragments was less than 2 500 000. We used a threshold of 3 for the z-score in the classification task. To make the estimates of the accuracy characteristics more stable, we implemented a kind of cross-validation procedure. The test dataset was composed of all samples with trisomies (for a certain chromosome) and an equal number of the control samples chosen randomly. The remaining control samples formed the training dataset. For each run we computed sensitivity, specificity and AUC values for the classification task and reported the averaged values over 200 runs.

Determination of the sex of the fetus

To determine the sex, a formula for estimating the fraction of fetal DNA was used based on the analysis of the fraction of fragments of the Y chromosome [9-11].

Abbreviations

cfDNA
cell-free DNA
cffDNA
cell-free fetal DNA
PPV
positive predictive value
NPV
negative predictive value
T13
13th chromosome trisomy
T18
18th chromosome trisomy
T21
21th chromosome trisomy

Declarations

The authors declare that they have no competing interests.

Funding

No funding was obtained for this study

Consent for publication

Not Applicable

Availability of data and materials

All data is available on request to corresponding author.

Authors contributions

VBand DPparticipated in the execution of the experiments, SE and KTs participated in the bioinformatical data analysis, GHv and AR participated in statistical analysis, AK collected samples, VI and DR participated in the study design and critical discussion, DK participated in the study design, sample collection, execution, analysis, interpretation of data and manuscript drafting and editing.

All authors read and approved the final manuscript.

Acknowledgements

We thank the Center for Precision Genome Editing and Genetic Technologies for Biomedicine (Moscow) for the genetic research methods.

We thank Larisa Doronenko for assistance in collecting samples.

We thank Andrey Krivoy for valuable comments during manuscript preparation.

References

1. Wang, T., He, Q., Li, H., Ding, J., Wen, P., Zhang, Q., ... & Mao, Y. (2016). An optimized method for accurate fetal sex prediction and sex chromosome aneuploidy detection in non-invasive prenatal testing. *PloS one*, *11*(7), e0159648.
2. Johansson L. F. et al. Novel algorithms for improved sensitivity in non-invasive prenatal testing //Scientific Reports. – 2017. – Т. 7. – №. 1. – С. 1838.
3. Jiang P. et al. FetalQuant: deducing fractional fetal DNA concentration from massively parallel sequencing of DNA in maternal plasma //Bioinformatics. – 2012. – Т. 28. – №. 22. – С. 2883-2890.
4. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

5. Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, *25*(14), 1754-1760.
6. Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science*, *270*(5235), 484-487.
7. Wang, Z., Andrews, P., Kendall, J., Ma, B., Hakker, I., Rodgers, L., ... & Levy, D. (2016). SMASH, a fragmentation and sequencing method for genomic copy number analysis. *Genome research*, *26*(6), 844-851.
8. Ivanov, M., Baranova, A., Butler, T., Spellman, P., & Mileyko, V. (2015). Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation. *BMC genomics*, *16*(13), S1.
9. Lo, Y. D., Chan, K. A., Sun, H., Chen, E. Z., Jiang, P., Lun, F. M., ... & Chiu, R. W. (2010). Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Science translational medicine*, *2*(61), 61ra91-61ra91.
10. Hudecova, I., Sahota, D., Heung, M. M., Jin, Y., Lee, W. S., Leung, T. Y., ... & Chiu, R. W. (2014). Maternal plasma fetal DNA fractions in pregnancies with low and high risks for fetal chromosomal aneuploidies. *PloS one*, *9*(2), e88484.
11. Xu, X. P., Gan, H. Y., Li, F. X., Tian, Q., Zhang, J., Liang, R. L., ... & Wu, Y. S. (2016). A method to quantify cell-free fetal DNA fraction in maternal plasma using next generation sequencing: its application in non-invasive prenatal chromosomal aneuploidy detection. *PloS one*, *11*(1), e0146997.
12. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes //Nature. – 2012. – T. 491. – №. 7422. – C. 56.
13. Lek M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans //Nature. – 2016. – T. 536. – №. 7616. – C. 285.
14. Danecek P. *et al.* The variant call format and VCFtools //Bioinformatics. – 2011. – T. 27. – №. 15. – C. 2156-2158
15. Wang, E., Batey, A., Struble, C., Musci, T., Song, K., & Oliphant, A. (2013). Gestational age and maternal weight effects on fetal cell-free DNA in maternal plasma. Prenatal diagnosis, 33(7), 662-666.
16. Gil, M. M., Quezada, M. S., Revello, R., Akolekar, R., & Nicolaides, K. H. (2015). Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta- Ultrasound in obstetrics & gynecology, 45(3), 249-266.
17. Bianchi, D. W., Platt, L. D., Goldberg, J. D., Abuhamad, A. Z., Sehnert, A. J., & Rava, R. P. (2012). Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. Obstetrics & Gynecology, 119(5), 890-901.;
18. Norton, M. E., Jacobsson, B., Swamy, G. K., Laurent, L. C., Ranzini, A. C., Brar, H., ... & Cuckle, H. (2015). Cell-free DNA analysis for noninvasive examination of trisomy. New England Journal of Medicine, 372(17), 1589-1597.;
19. McCullough, R. M., Almasri, E. A., Guan, X., Geis, J. A., Hicks, S. C., Mazloom, A. R., ... & Dharajiya, N. (2014). Non-invasive prenatal chromosomal aneuploidy testing-clinical experience: 100,000 clinical samples. PLoS one, 9(10), e109173.

20. Prabakar, R. K., Xu, L., Hicks, J., & Smith, A. D. (2019). SMURF-seq: efficient copy number profiling on long-read sequencers. Genome biology, 20(1), 134.
21. Sun, K., Jiang, P., Wong, A. I., Cheng, Y. K., Cheng, S. H., Zhang, H., ... & Lo, Y. D. (2018). Size-tagged preferred ends in maternal plasma DNA shed light on the production mechanism and show utility in noninvasive prenatal testing. *Proceedings of the National Academy of Sciences*, *115*(22), E5106-E5114.
22. Kim, S. K., Hannum, G., Geis, J., Tynan, J., Hogg, G., Zhao, C., ... & van den Boom, D. (2015). Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts. *Prenatal diagnosis*, *35*(8), 810-815.

Tables

Due to technical limitations, the tables are only available as a download in the supplemental files section.

Figures

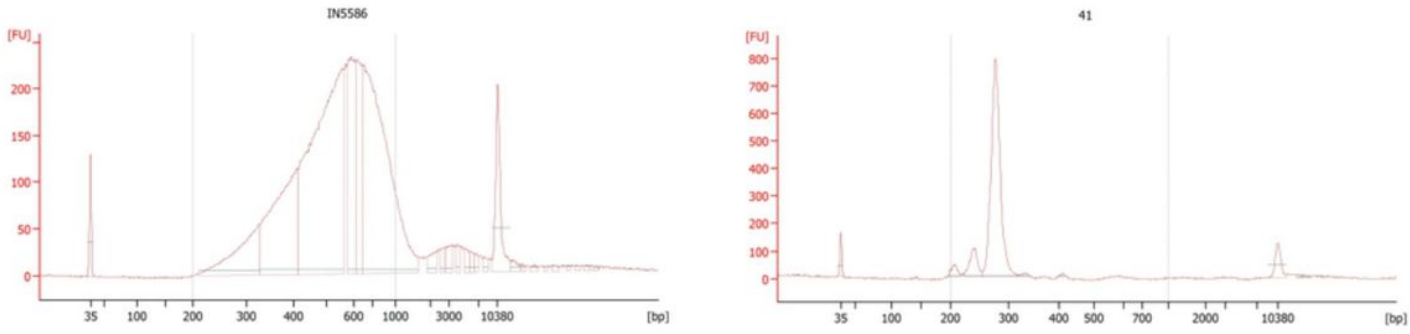

Figure 1

Electrophoregram (Bioanalyzer 2100, Agilent) of chimeric DNA library – «smash» (A) and fetal fraction estimation library – «amplifet» (B).

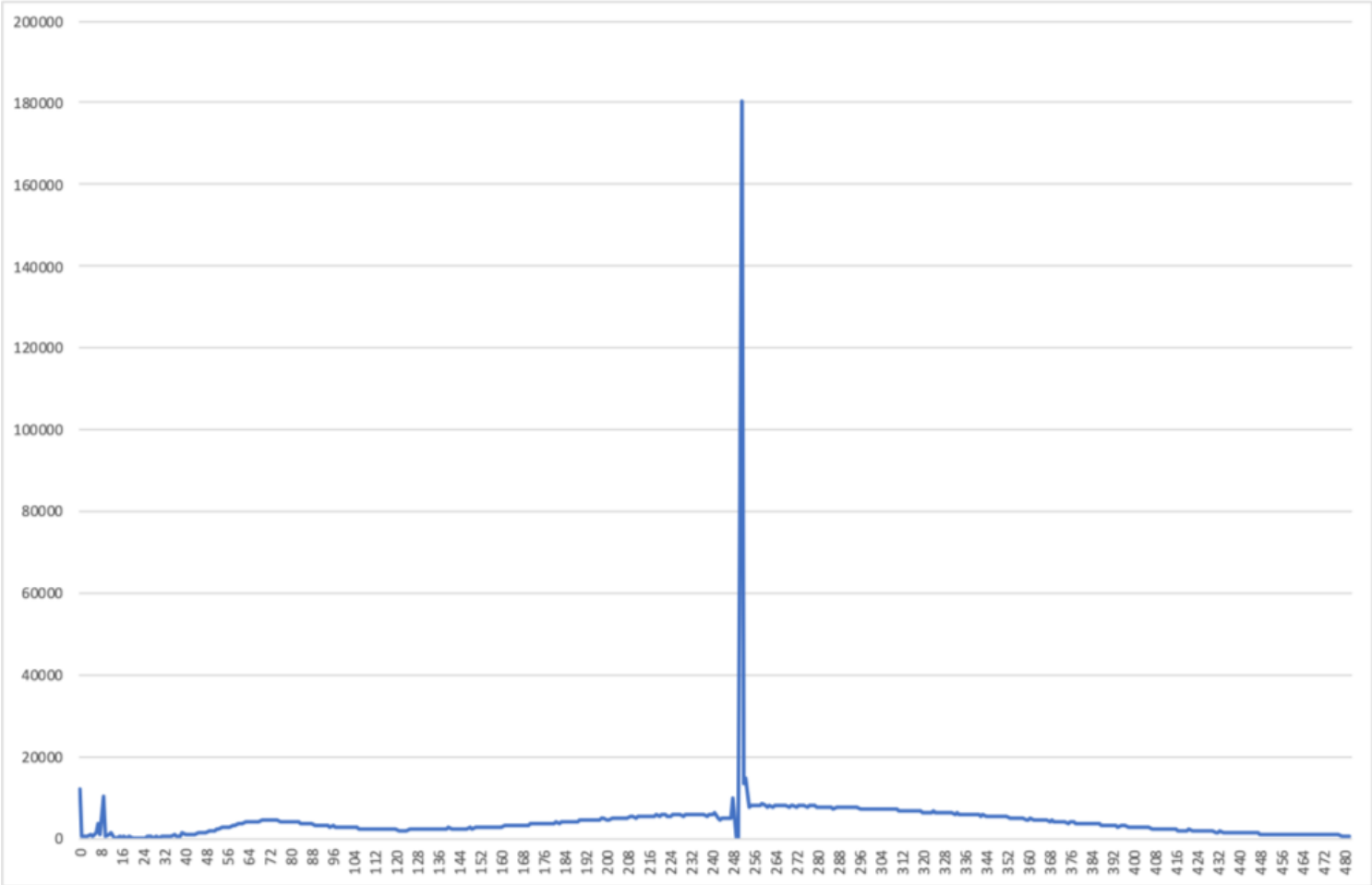

Figure 2

Insertion length distribution for PE reads of the smash library. The X axis is the insertion length in bp, and the Y axis is the number of reads of the corresponding length.

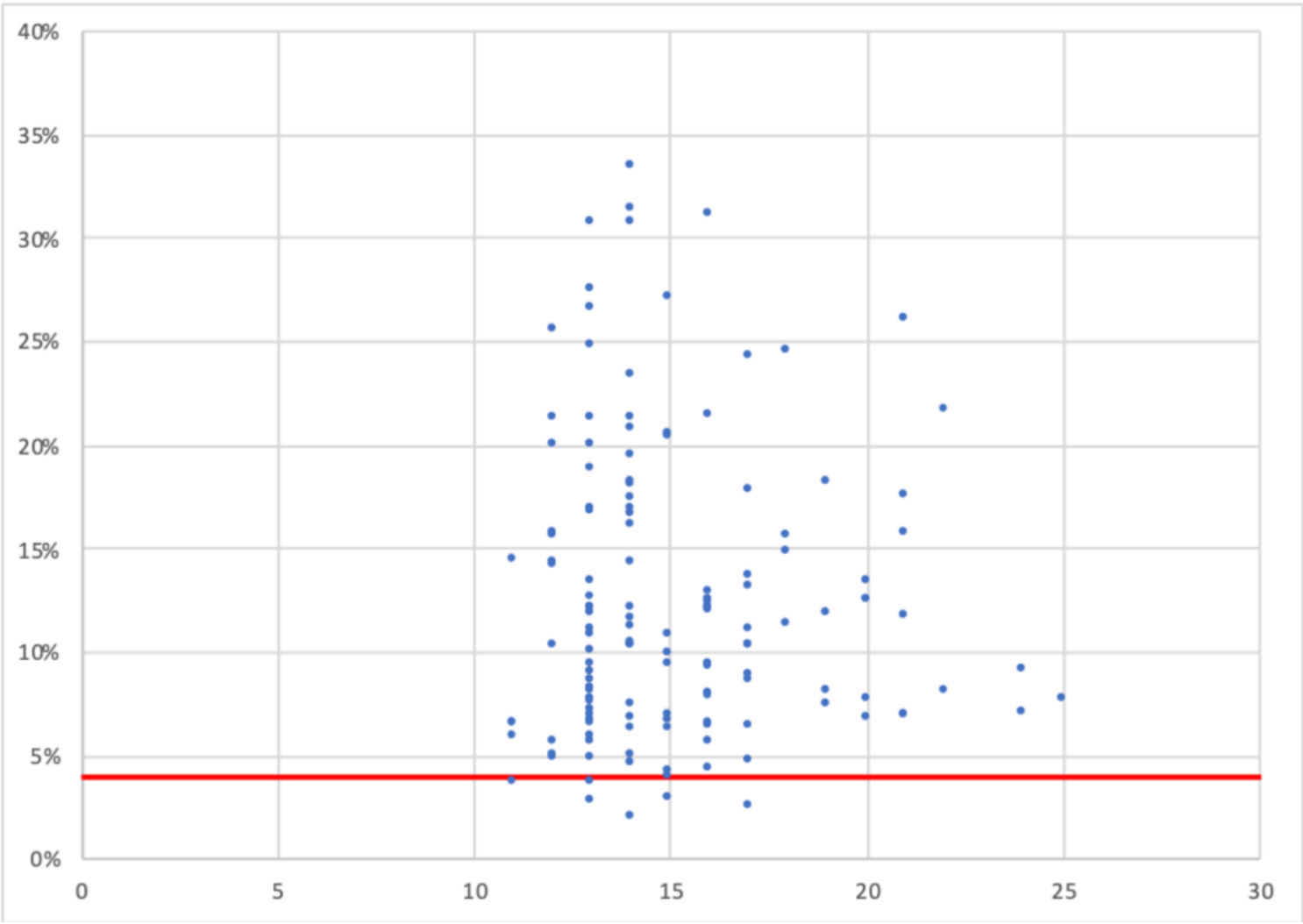

Figure 3

The proportion of fetal DNA (Y axis) versus the gestational age of 139 samples of the control set (X axis), calculated based on the results of amplifet library sequencing. The solid red line is 4% proportion of fetal DNA.

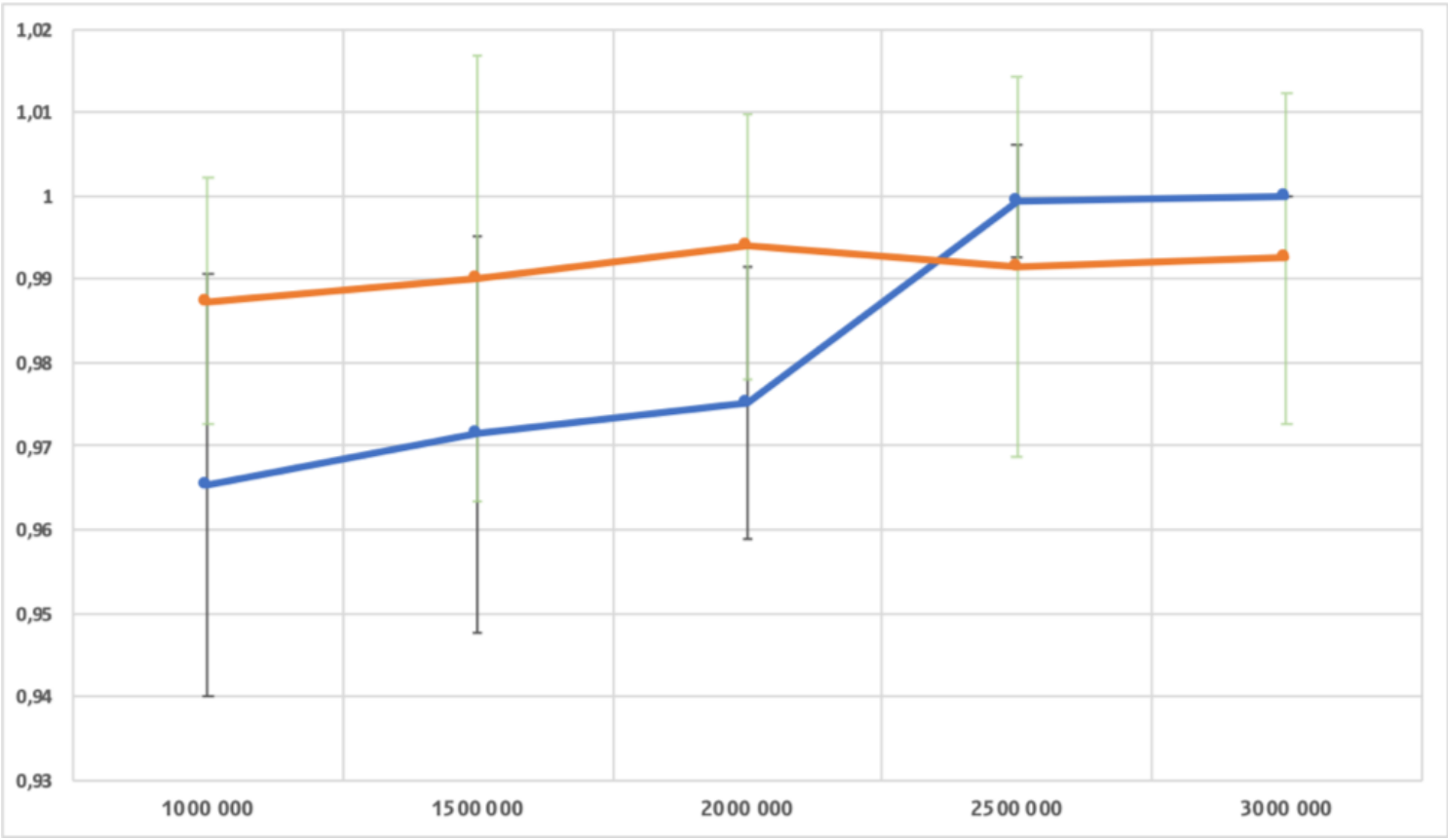

Figure 4

The sensitivity (blue line) and specificity (orange line) values of the determination of aneuploidy on chromosome 21 with regard to the standard deviation depending on the number of filtered fragments per sample.

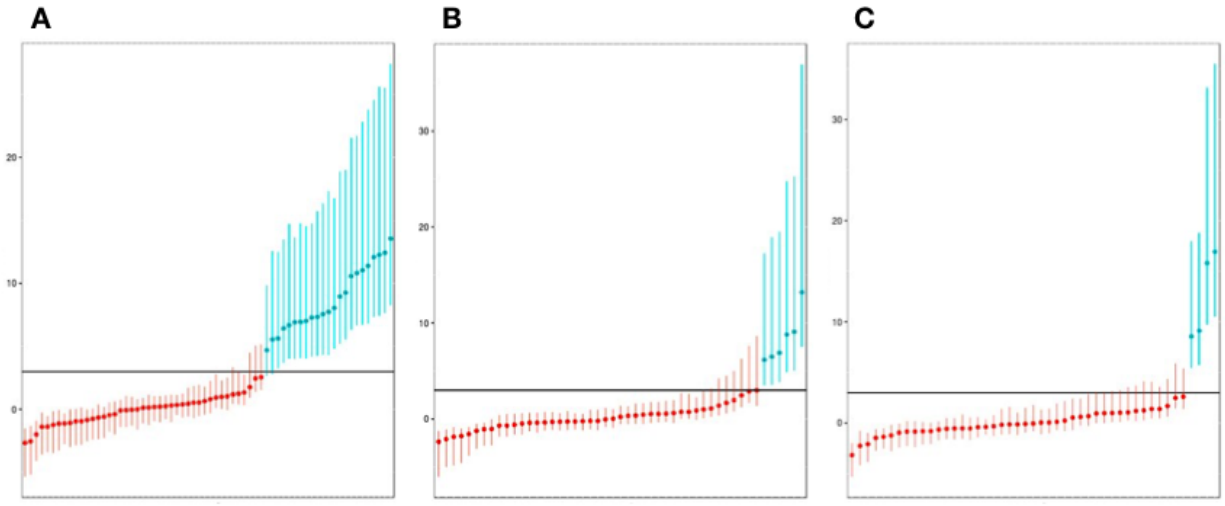

Figure 5

Histograms of Z-score values obtained during the cross-validation of control samples (A – T21, B – T18, C – T13 chromosome). Red indicates samples that do not have the corresponding aneuploidy, green indicates samples with aneuploidy. The maximum and minimum Z values for the sample in the simulations are plotted by error bars. SD values are given in Table 4 of Supplementary files. The black horizontal line in the histograms is Z=3.

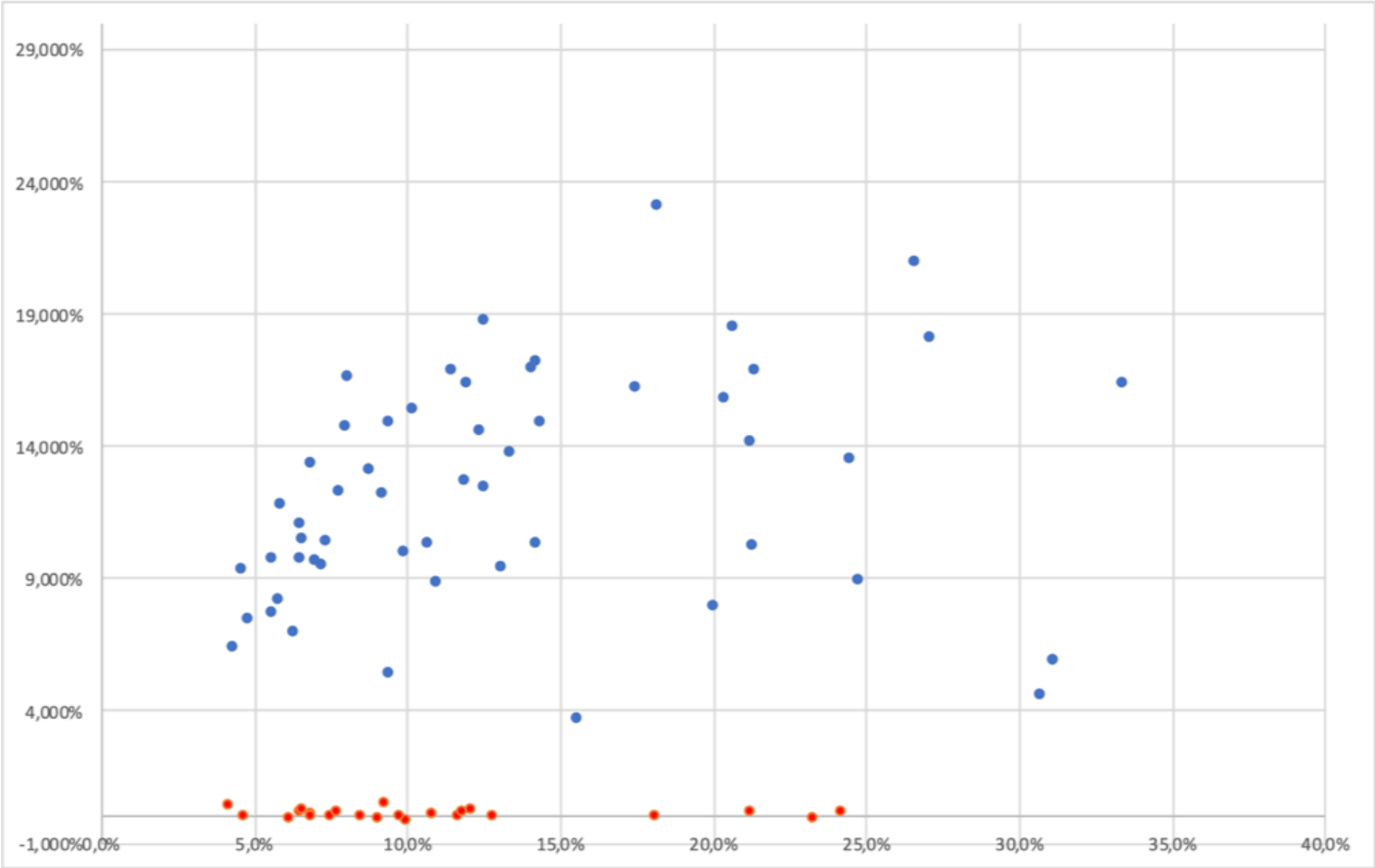

Figure 6

Fetal DNA proportion calculated for amplifet (X axis) and Y chromosome (Y axis). Male samples are indicated in blue, female samples are red.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryfileTable2.xlsx
- SupplementaryfileTable4.xlsx
- SupplementaryfileTable5.xlsx
- Table1.xlsx
- Table2.xlsx
- Table3.xlsx
- Table4.xlsx
- SupplementaryfileTable1.xlsx
- SupplementaryfileTable3.xlsx