Journal of Human Genetics (2018) 63:1129–1137
https://doi.org/10.1038/s10038-018-0489-9

ARTICLE

Identification of a de novo fetal variant in osteogenesis imperfecta by targeted sequencing-based noninvasive prenatal testing

Xiuju Yin[1] · Yang Du[2] · Han Zhang[2] · Zhandong Wang[2] · Juan Wang[2] · Xinxin Fu[2] · Yaoyao Cui[2] · Chongjian Chen[2] · Junbin Liang[2] · Zhaoling Xuan[2] · Xiaohong Zhang[1]

Received: 26 February 2018 / Revised: 14 June 2018 / Accepted: 26 June 2018 / Published online: 21 August 2018

Abstract

Noninvasive prenatal testing (NIPT), which involves analysis of circulating cell-free fetal DNA (cffDNA) from maternal plasma, is highly effective for detecting feto-placental chromosome aneuploidy. However, recent studies suggested that coverage-based shallow-depth NIPT cannot accurately detect smaller single or multi-loci genetic variants. To assess the fetal genotype of any locus using maternal plasma, we developed a novel genotyping algorithm named pseudo tetraploid genotyping (PTG). We performed paired-end captured sequencing of the plasma cell-free DNA (cfDNA), in which case a phenotypically healthy woman is suspected to be carrying a fetus with genetic defect. After a series of independent filtering of 111,407 SNPs, we found one variant in *COL1A1* graded with high pathogenic potential which might cause osteogenesis imperfecta (OI). Then, we verified this mutation by Sanger sequencing of fetal and parental blood cells. In addition, we evaluated the accuracy and detection rate of the PTG algorithm through direct sequencing of the genomic DNA from maternal and fetal blood cells. Collectively, our study developed an intuitive and cost-effective method for the noninvasive detection of pathogenic mutations, and successfully identified a de novo variant in *COL1A1* (c.2596 G > A, p.Gly866Ser) in the fetus implicated in OI.

Introduction

Noninvasive prenatal testing (NIPT) of cell-free fetal DNA (cffDNA) is commonly used in clinical diagnostics for Rh-D blood typing, paternity testing, as well as for the testing of fetal chromosomal aneuploidies like trisomy 13, trisomy 18, and trisomy 21 [1–3]. An earlier study discovered that a fraction of cffDNA is detectable in maternal peripheral blood, which initiated the clinical use of noninvasive prenatal testing (NIPT) [4]. However, several technical limitations and lack of robust analytical methods currently limit the broader application of NIPT.

Noninvasive diagnosis in the fetus relies on distinguishing or quantifying fetus-specific DNA from the mixed cfDNA in the peripheral blood. The biological characteristics of circulating cffDNA have been investigated using polymerase chain reaction (PCR) [5–8]. In addition, cffDNA concentrations were successfully quantified from maternal plasma circulating cell-free DNA (cfDNA) by exploring DNA-methylation patterns [9], placenta-specific mRNA expression [10], Y chromosomes coverage [4], as well as genotype information of single-nucleotide polymorphism (SNP) [11]. With improved technology and reduced cost for higher sequencing depth, the scope of next-generation sequencing (NGS)-based NIPT of maternal plasma DNA is currently expanding to identify fetal copy-number variations (CNVs) and monogenic mutation [12–21].

Deep sequencing of maternal plasma DNA through NGS has recently been shown to be feasible for detecting CNVs as well as single-gene alterations. Lo et al. [18] recently

These authors contributed equally: Xiuju Yin and Yang Du.

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s10038-018-0489-9) contains supplementary material, which is available to authorized users.

✉ Zhaoling Xuan
zhlxuan@annoroad.com

✉ Xiaohong Zhang
happyxh713@sina.com

[1] Department of Obstetrics and Gynecology, Peking University People's Hospital, Beijing 100044, China

[2] Annoroad Gene Technology Co., Ltd, Beijing 100176, China

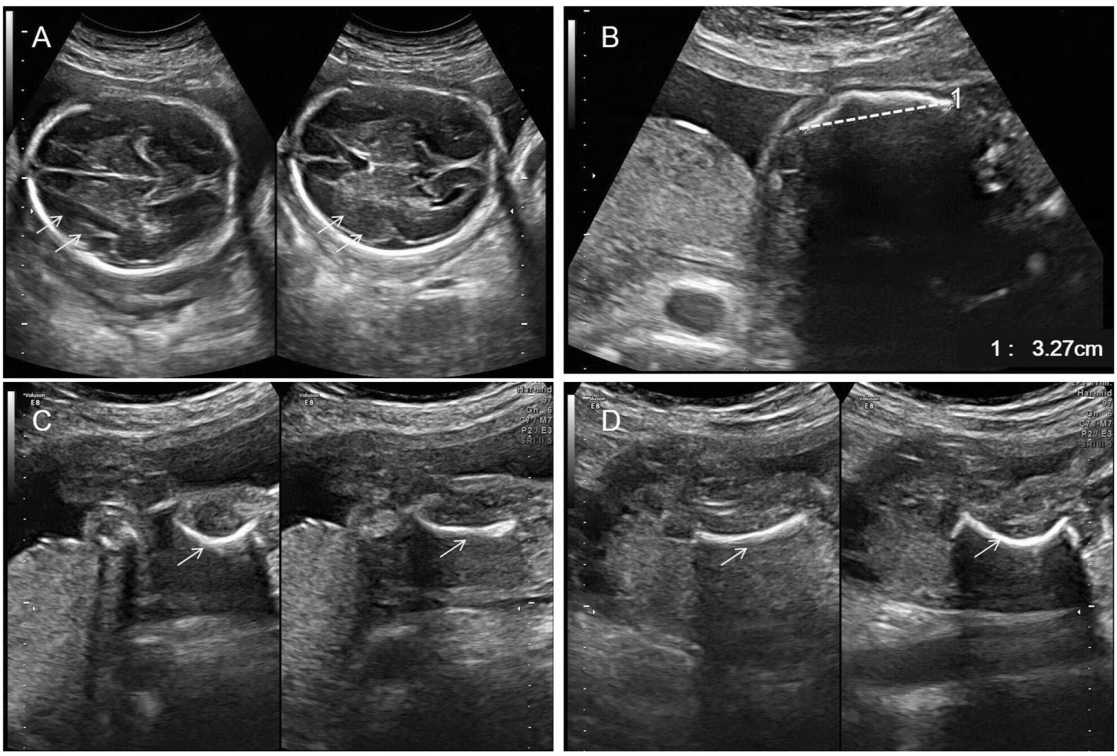


Fig. 1 Real-time graphs of B-mode ultrasonography of the fetus at 26 weeks of gestation. **A** The yellow arrow indicates the skull region. Echo in this region is reduced, which suggests that the fetal skull is too thin. **B** The dashed line indicates the shortened bilateral femur, which was only 3.27 cm. **C**, **D** The curved tibia and fibula. Yellow arrows indicate the corresponding bone defects. Color figure online

developed an algorithm named "relative haplotype dosage analysis" (RHDO) to make inference of the local fetal haplotype based on parental haplotypes through linkage analysis, whereby one of the parents is homozygous while the other parent is heterozygous. Kitzman et al. [21] determined the maternal haplotype through large fragment cloning and sequencing and detected paternal-specific alleles through a site-by-site strategy (SBSS). Different reads matching the same paternal-specific allele were counted as evidence of transition from A to B. Fan et al. [19] deduced fetal genomic information through counting the paternal-specific haplotypes following whole-genome sequencing of plasma cfDNA. The prediction of fetal genomic information was mainly based on the allelic imbalance principle.

In order to obtain parental haplotypes, Chen et al. [20] developed a combined strategy of trio and unrelated individuals. The fetal haplotype was noninvasively recovered using a hidden Markov model (HMM) and Viterbi algorithm, with 95.37, 98.57, and 98.45% of the maternal autosomal alleles, paternal autosomal alleles, and alleles from the chromosome X recovered in fetal haplotype. However, all present studies were based on bi-parental haplotypes with a higher cost, a fact that renders clinical application highly problematic.

Obtaining haplotype information using whole-genome sequencing remains expensive and requires complex analytical procedures. As demonstrated by Liao et al. [22], target region capture sequencing (TRCS) achieved deep sequencing of specific regions at reasonable cost. Here, we developed a novel algorithm, pseudo tetraploid genotyping (PTG), which directly deduces the combined genotype of mother and child. To confirm these results, we evaluated the PTG through exome sequencing of fetal and maternal blood cells. Using our PTG algorithm, we deduced the genotype of a miscarried fetus with osteogenesis imperfecta (OI) disease from maternal plasma and a de novo mutation in *COL1A1* was identified. The study suggested our newly developed algorithm is able to correctly infer fetal genotype from maternal plasma.

Materials and methods

Sample collection and DNA extraction

The pregnant mother was recruited with informed consent and approval of the Ethics Committee of the Peking University People's Hospital. The fetus was suspected to have OI according to echograph imaging, as shown in Fig. 1.

After obtaining the fully informed consent from the patient, we collected mother's peripheral blood as well as fetal umbilical cord blood through umbilical vein puncture at week 30.

Maternal peripheral blood (5 mL) was collected and centrifuged for 10 min at 4 °C at 1,600 × *g*. The blood cell portion was centrifuged again at 2,500 × *g* for 10 min and the plasma portion at 16,000 × *g* for 10 min, the blood cells portion and plasma samples were immediately stored at −80 °C until further processing [23].

Total genomic DNA from the maternal blood cells and fetal blood cells were isolated using the Amp Genomic DNA Kit (TIANGEN, China). For analysis, 1 mL plasma was extracted using the QIAamp Blood Kit (Qiagen, Germany), and we obtained ~10 ng DNA for library preparation, measured by Qubit 3.0 (Thermo Fisher Scientific).

Library preparation and DNA sequencing

The gDNA samples from maternal blood cells and fetal blood cells were fragmented by using the Covaris S220 (Covaris, Woburn, MA, USA) according to the manufacturer's protocol. Fragmented gDNA and plasma DNA were then used to construct sequencing libraries. Standard protocol was followed including blunt-ending, "A" tailing, and adapter ligation. A total of 17 cycles of PCR were performed to amplify the adapter-ligated DNA fragments. Sequencing ready DNA templates were subsequently hybridized to the SeqCap EZ Probes oligo pool (Roche Nimblegen, USA), with a customary designed 1-M target capture panel covering genetic hotspots of various known diseases. We evaluated the amplified libraries using Agilent 2100 Bioanalyzer (Agilent, USA) and by quantitative PCR. The three libraries with different barcodes were pooled and sequenced (paired-end 2 × 100 bp) using an Illumina HiSeq 2500 sequencer (Illumina, USA) following standard instructions [24].

To eliminate low quality reads, reads with more than 5% N or with at least 50% of all bases' quality not larger than 30 were filtered out from the raw data with our in-house scripts, and sequencing adaptors were removed. Using the Burrows-Wheeler alignment (BWA) tool [25], remaining reads were aligned to human genome reference sequences (Hg19, NCBI build 37) with fixed parameters (aln -o1-e63 -i15 -L -l31 -k1 -t6). To reduce the bias of PCR amplification, alignment files were further processed with SAMtools [26].

Pseudo tetraploid genotyping

To deduce the fetal genotype from the maternal plasma, we developed a novel algorithm, pseudo tetraploid genotyping (PTG), which jointly considers the combined genotype of both the mother and the fetus. The probability of such a pseudo tetraploid genotype can be calculated by multiplying the probability of any two genotypes or one to itself. However, out of the nine possible combinations, only seven ("AAAA", "AAAB", "ABAA", "ABAB", "ABBB", "BBAB" and "BBBB") are possible for a mother–child genotype mixture one can recover from maternal plasma used for NIPT. Here, "A" stands for the reference allele, and "B" the most frequent alternative allele other than "A". The first two letters of each four-letter set represent the maternal genotype, while the last two letters represent the fetal genotype. Among the seven combinations, only four mother–child genotype combinations ("AAAB", "ABAA", "ABBB", and "BBAB") can theoretically be used to calculate the relative concentration of cffDNA, which is also known as fetal concentration (FC).

The algorithm borrows information from the 1000 Genome Project to empirically learn the population-wide minor allele frequency (MAF), θ, and the prior probabilities of the three genotypes (AA, AB, BB) for any given point mutation.

$$P(\mathrm{AA}) = (1-\theta)^2; P(\mathrm{AB}) = \theta(1-\theta); P(\mathrm{BB}) = \theta^2.$$

We considered the following relationship between FC and the observed MAF and utilized an expectation maximization (EM) algorithm to obtain the optimal FC estimate and the most likely pseudo tetraploid genotype at each mutation point. We hypothesized that the probability density of the observed alternative allele coverage at a given site j follow a negative binomial distribution, $NB_{gj}(C_{aj} + C_{bj}, f_g(FC))$, with the size parameter equal to the total coverage at that site and the probability parameter given by a function of FC for each of the seven pseudo tetraploid genotypes,

$$f_{\mathrm{aaaa}} = \varepsilon, f_{\mathrm{aaab}} = \tfrac{\mathrm{FC}}{2}, f_{\mathrm{abaa}} = 0.5 - \mathrm{FC}$$
$$f_{\mathrm{abab}} = 0.5, f_{\mathrm{abbb}} = 0.5 + \mathrm{FC}, f_{\mathrm{bbab}} = 1 - \tfrac{\mathrm{FC}}{2}, f_{\mathrm{bbbb}} = 1 - \varepsilon$$

where ε is a numeric constant close to zero to represent an infinitesimal effect. The pseudo tetraploid genotype G_j with the highest density is assigned to that site. With the updated genotypes at all positions, a new estimation of the FC can be made using the four mother–child genotype combinations ("AAAB", "ABAA", "ABBB", and "BBAB").

$$\mathrm{FC}_{\mathrm{aaab}} = 2 \times \underset{G_j=\mathrm{aaab}}{\mathrm{median}}\left(\frac{C_{bj}}{C_{bj}+C_{aj}}\right),$$
$$\mathrm{FC}_{\mathrm{abaa}} = 1 - 2 \times \underset{G_j=\mathrm{abaa}}{\mathrm{median}}\left(\frac{C_{bj}}{C_{bj}+C_{aj}}\right)$$
$$\mathrm{FC}_{\mathrm{abbb}} = 1 - 2 \times \underset{G_j=\mathrm{abbb}}{\mathrm{median}}\left(\frac{C_{aj}}{C_{bj}+C_{aj}}\right),$$
$$\mathrm{FC}_{\mathrm{bbab}} = 2 \times \underset{G_j=\mathrm{aaab}}{\mathrm{median}}\left(\frac{C_{aj}}{C_{bj}+C_{aj}}\right)$$

The EM process stops when the estimated FC converge to some constant or a maximum number of iterations is reached, together with updated pseudo tetraploid genotype assigned to each site to represent mother–child genotype combination.

Sanger sequencing

To validate the genotype of the potentially pathogenic candidate detected in NGS via PTG, target amplicon was designed to be of 534 bp in length using two primers, 7237F (5′-CCCAACCTAGAGCAGTGGAC-3′) and 7771R (5′CTCCCCCAGGTAGTGGAAAC-3′). The PCR system was set up as follows: 25 µL of total reaction volume, 20 ng of genomic DNA, 2 µL of dNTP (10 mM), 0.3 µL of each primer (10 pmol), 2.5 µL 10 × reaction buffer, 0.3 µL DNA Taq polymerase (5 U/µL), and 18.6 µL PCR-grade H_2O. The PCR reaction was programmed as follows: 95 °C for 2 min, 95 °C for 10 s, annealing temperature 60 °C for 30 s, extension at 70 °C for 1 min, 30 cycles from step 2, 70 °C for 10 min; and hold at 4 °C. The reaction was performed using a Veriti Thermal Cycler (Thermo Fisher Scientific, USA). PCR products were tested using 1% agarose gel electrophoresis and purified using a QIAquick PCR Purification Kit (Qiagen, Germany). The amplicon was sequenced using a 3730xl DNA analyzer (Thermo Fisher Scientific, USA).

Table 1 Obstetric examinations of the affected gestation in chronological order

First trimester	
No fever, rash, and other related disease	
No access to toxin, radiation, or related pathogen	
Second trimester	
13th week	Negative for TORCH virus test, including Toxoplasma, rubella virus, cytomegalovirus, and herpes simplex virus
14th week[a]	Nuchal translucency (NT) measurement 0.14 cm
17th week[b]	Serum screen for 21 trisomy and 18 trisomy
	Both risks are lower than 1:50,000
26th week	Thin fetal skull with reduced ECHO
	Bilateral femur (3.27 cm), curved tibia and fibula
Third trimester	
30th week	Characterized to be fetal IO, suggestion for elective termination
	Negative TORCH tests
	Karyotype: 46, XY

[a]At 14 weeks gestation, the nuchal translucency (NT) value was 0.14 cm which is considered of low risk of chromosomal abnormality

[b]At 17 weeks gestation, the serum screen suggested that the fetus had very low risk of trisomy 21 or trisomy 18

Results

Case description

The patient is a 28-year-old female with a history of three miscarriages. She scheduled a series of prenatal examinations during the 2nd trimester, including a TORCH test at the 13th week, nuchal translucency (NT) test at week 14, serum screen at week 17, and a B-mode echography at week 26 of gestational age (Table 1).

B-mode echography of the fetus demonstrated developmental defects and structural anomalies of several fetal bones, including thin skull, curved tibia and fibula, and shortened bilateral femur (Fig. 1), suggesting high possibility of congenital defects of OI or thanatophoric dysplasia. Following confirmation through prenatal diagnostic testing at 30 weeks of gestation, the mother decided to receive an induced elective termination. Following elective termination, the male fetus exhibited clear signs of OI, including fractured tibia and fibula of the bilateral lower limbs, as well as softened skull. The karyotype of the fetus was 46, XY with no visible chromosomal abnormalities (Supplementary Figure 1).

Pseudo tetraploid genotyping

Using the maternal plasma sample, we mapped ~98-M clean reads to the target regions, resulting in a 97.26% target coverage and a mean sequencing depth of 118×. In addition, we mapped 69-M clean reads to the flanking regions, with a 95.60% coverage and a mean sequencing depth of 59 fold. After joint genotypes calling with PTG, 111,407 loci were detected (Fig. 2B), final fetal fraction was estimated to be 11.59701%. Noting the fact that both biological parents are phenotypically healthy, the disease-related gene most likely carries two possible genotypes in the maternal-fetal combined genotypes: AAAB or ABBB (first two letters for the mother and the second two for the fetus), considering the disease-causing variant is either a de novo autosomal dominant mutation or an autosomal recessive mutation. As a result, 24,416 loci with genotypes either AAAB or ABBB were selected for further analysis. We annotated the resulting 24,416 candidate SNPs with ANNOVAR [27] and sorted the SNPs with a series of filtrating including mutant alleles, variant location, dbSNP annotation [28], population frequency, and variant impact predictions. Only seven SNPs were left after the filtering. Finally, one candidate mutation located in the protein-coding region of *COL1A1* (c.2596 G > A, p.Gly866Ser) was considered to be the main triggering mutation for the symptoms. The affected gene has been frequently reported as the genetic cause of OI [29, 30].

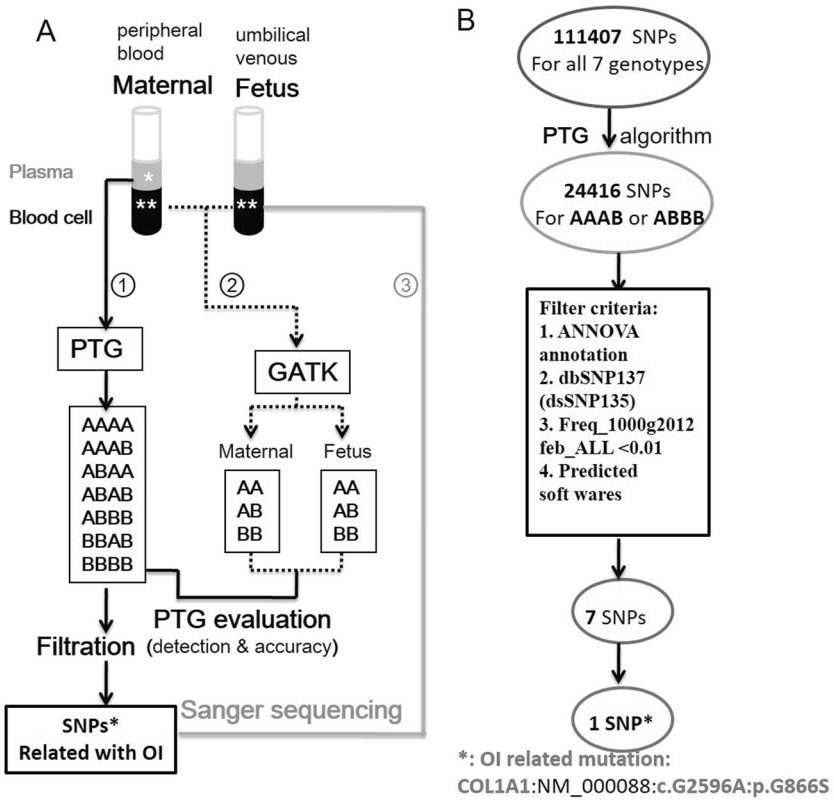


Fig. 2 Work flow of the pseudo tetraploid genotyping (PTG) algorithm. **A** (1) Sequencing of the maternal plasma and joint genotype calling with PTG; (2) Sequencing of maternal blood cells and fetal blood cells in umbilical venous and individual genotype calling with GATK; (3) Sanger sequencing of fetal blood cell DNA was performed for validation. **B** 111,407 variants were jointly called by PTG on the maternal plasma sample. After genotype selection of "AAAB" or "ABBB", 24,416 loci remained. After extensive filtering, 7 SNPs were identified. The filtering rules are described as follows: (1) Located in UTRs, exon regions, and splicing regions. (2) Excluding mutations in dbSNP (Build135) but not flagged as clinically associated in dbSNP (Build137). (3) Population frequency from 1000 Genome < 0.01; AVSIFT < 0.05; SIFT > 0.9; PhyloP > 0.95; PolyPhen2 > 0.15

PTG evaluation

In order to evaluate the overall accuracy of PTG, we used the combined genotype made up from the GATK called variants of maternal and fetal blood cells as truth [31]. The accuracy rate was defined as the ratio of matched genotypes in the total GATK derived genotypes. In comparison, the detection rate was defined as the ratio of matched genotypes in the total PTG derived genotypes. In our specific OI case, "AAAB" and "ABBB" are two possible combinations of pathogenic genotypes. Taking "AAAB" as an example, combined genotypes of 10,157 loci are consistent between GATK and PTG, whereby combined individual GATK calls resulted in 11,965 loci, and PTG jointly identified 15,811 loci for the mother-fetus combination. Therefore, "AAAB" has an accuracy rate of 84.89% (10,157/11,965) and a detection rate of 64.24 % (10,157/15,811) (Fig. 3A). Similarly, "ABBB" has an accuracy rate of 54.77% and a detection rate of 63.02%. A detailed summary of all seven genotypes is shown in Fig. 3B.

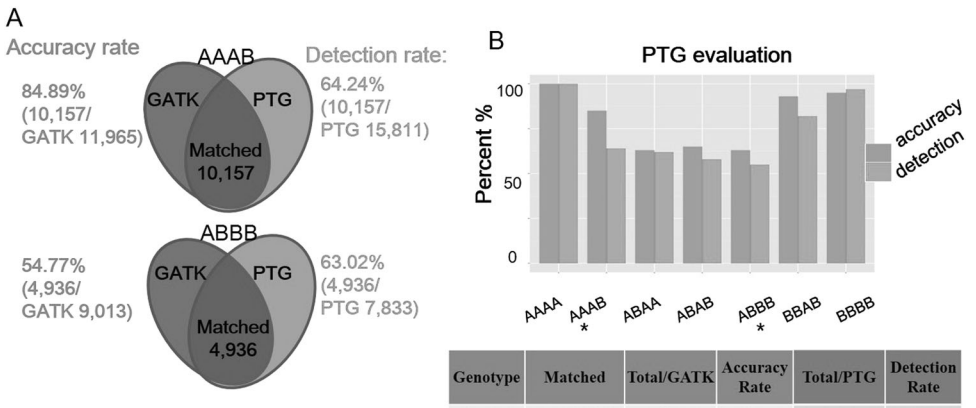
Verification of *COL1A1* mutation by Sanger sequencing

To confirm the mutation in *COL1A1* (c.2596G>A, p.Gly866Ser), we performed Sanger sequencing using parental and fetal blood cells (Fig. 4). We amplified the genomic region including c.G2596A using primers7237F and 7771R, yielding a 534-bp amplicon in the three samples. However, only the fetal blood cell group harbored the 5′-G/A-3′ mutation (7237F direction) or 3′-C/T-5′ mutation (Reverse direction, 7771R) (Fig. 4). The mutation in *COL1A1* is a de novo mutation unidentifiable in the genome of either parents. Together, the Sanger sequencing results confirmed that the de novo fetal mutation was correctly identified by PTG.

Discussion

In the current work, we developed an intuitive method (PTG) to noninvasively detect fetal single-gene defects using only maternal plasma. Using this method, we examined the possibility of noninvasive prenatal diagnosis of monogenic diseases.

The overall genotyping performance of PTG was evaluated using genotype calls made individually with maternal and fetal blood cells. When looking at different genotype combinations separately, the algorithm clearly performs better when the maternal allele is homozygous, in which case an overall accuracy rate of 84.89% was achieved, whereas in the cases of maternal heterozygous groups lower accuracy of 54.77% was obtained. Thus, our proposed method performs well for cases of autosomal dominant



Genotype	Matched	Total/GATK	Accuracy Rate	Total/PTG	Detection Rate
AAAA	89,078,219	89,089,116	99.99%	89,080,275	99.99%
*AAAB	10,157	11,965	84.89%	15,811	64.24%
ABAA	9,968	15,788	63.14%	16,019	62.23%
ABAB	16,203	25,075	64.62%	27,914	58.05%
*ABBB	4,936	7,833	63.02%	9,013	54.77%
BBAB	7,274	7849	92.67%	8,920	81.54%
BBBB	28,284	29,668	95.34%	29,105	97.18%

Fig. 3 Evaluation of the pseudo tetraploid genotyping (PTG) algorithm. **A** Detection rate and accuracy rate of "AAAB" and "ABBB". For "AAAB", we identified 10,157 matched genotypes, 11,965 combined individual calls made with GATK on the two blood cell samples, and 15,811 joint calls of the plasma sample with PTG. The accuracy rate was 84.89% and the detection rate was 64.24%. Similarly, the accuracy rate for "ABBB" was 54.77% and the detection rate was 63.02%. **B** The detailed summary of the detection rate and accuracy rate of all seven genotypes. * The two pathogenic candidate genotypes combinations considering a healthy parents–affected child scenario

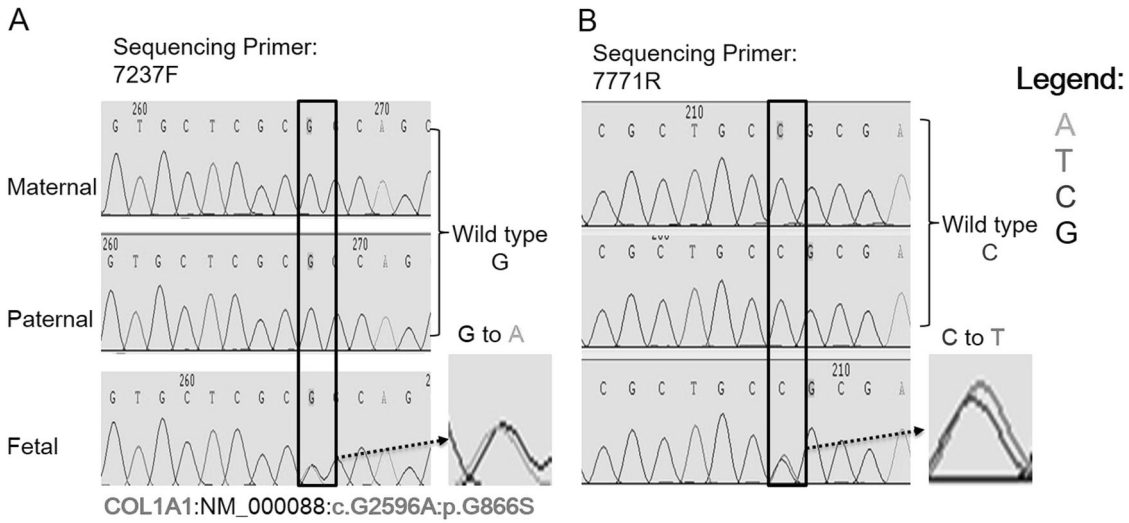


Fig. 4 Mutation validation by Sanger sequencing *COL1A1* (c.2596 G > A, p.Gly866Ser) was validated by Sanger sequencing in fetal and parental blood cells. **A** When sequencing with forward primer, only the G/A mutation was detected in the fetal blood cells. **B** When sequencing with the reverse primer, the C/T mutation was also identified. Green, red, blue, and black curves represent A, T, C, and G, respectively. Color figure online

diseases inherited from the father or pathogenic mutation which occurred de novo.

On the other hand, drawback is being speculated in the case of autosomal recessive diseases, in which case the lower genotyping accuracy renders it difficult to differentiate carrier (ABAB) and affected (ABBB) if the mother is a heterozygous carrier. Since the method assumes strong equilibrium of 1:1 ratio for the allele coverage of heterozygous site, significant deviation from MAF of 0.5 is considered due to the contribution of homozygous genotype at the same locus from the fetus. This issue becomes even more problematic for cases where the fetal fraction is low,

rendering the expected deviation from 0.5 even smaller, and the true signal becomes difficult to be distinguished from random variation form 0.5.

Such a phenomenon of unexpected deviation from equilibrium is commonly observed in short read sequencing and is likely due to a selection bias occurring during library preparation and also the not yet sufficient sequencing depth. Not directly related to our problem in sequencing cfDNA, there is a known problem commonly referred to as allele drop out (ADO) occurs in single cell genomic [32, 33]. There, extreme ADO can result in complete failure to amplify one of the two alleles at a heterozygous locus. ADO can affect up to 40% of amplifications in extreme circumstances and result in misdiagnose of diagnostic errors preimplantation genetic diagnosis (PGD) of single-gene disorders [34]. A report by Zhong et al. [34] also indicated that alleles drop out due to amplification bias can result in the low detection yield for SNVs. In addition, DNA damage could also contribute to such imbalance [35, 36]. A report by Chen et al. showed that mutagenic damage represents the dominating factor for the erroneous identification of variants with lower frequency (1 to 5%). Such damage induced sequencing error was found in widely used databases, such as the 1000 Genomes Project [35].

Osteogenesis imperfecta (OI) is a genetically heterogeneous group of connective tissue disorders characterized by increased bone fragility, low bone mass, short stature, and other connective tissue manifestations [37, 38]. OI was originally classified into four types according to clinical and histological manifestations, while the perinatally lethal form (type II) is the most severe. OI classification has now been expanded to 15 different types based on the gene affected and severity of OI phenotypes. Previous studies have shown that the primary cause of OI is mutations in the *COL1A1* and *COL1A2* genes, which edcode procollagen type I α1 and α2 chains, respectively [39–43]. Also, one-third of the mutations that cause glycine substitutions in *COL1A1* result in lethal subtype of OI [37, 44]. Here, we report a de novo variant of *COL1A1* gene (c.2596 G > A, p.Gly866Ser) and we consider the variant as the potential cause for the symptoms observed in the fetus. This de novo variant we discovered, and the complete genotype–phenotype reports provide a yet another unique case for future clinical research in the pathogenesis of OI.

The discovery of fetal cell-free DNA in maternal plasma resulted in the invention and development of noninvasive prenatal testing. Furthermore, noninvasive diagnosis of single-gene disorders has been reported for the detection of several monogenic diseases in clinical research [45–52]. Nevertheless, to our knowledge, detection of de novo OI mutations in *COL1A1* gene with noninvasive prenatal diagnosis (NIPD) has not been reported previously. Our computational method should greatly assist noninvasive diagnostic procedures in identifying novel mutations, without sequencing of the parents, and also enabling the early detection of similar de novo pathogenic mutation comparing to current echography-based screening. Importantly, our proposed method is highly cost-efficient, as it requires only a single round of plasma screening. In comparison, other approaches rely on haplotyping [16, 20] whereby haplotype blocks must be constructed locally around the known pathogenic mutation derived from the proband. In addition, expanding the panel by adding evenly distributed target regions along the genome should readily unify the standard NIPT for common fetal aneuploidy with noninvasive screening of monogenic disease.

Acknowledgements We thank the patients and their families for their participation in this study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Dondorp W, de Wert G, Bombard Y, Bianchi DW, Bergmann C, Borry P, et al. Non-invasive prenatal testing for aneuploidy and beyond: challenges of responsible innovation in prenatal screening. Eur J Hum Genet. 2015;23:1592.
2. Hayden EC. Prenatal-screening companies expand scope of DNA tests. Nature. 2014;507:19.
3. Agarwal A, Sayres LC, Cho MK, Cook-Deegan R, Chandrasekharan S. Commercial landscape of noninvasive prenatal testing in the United States. Prenat Diagn. 2013;33:521–31.
4. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, et al. Presence of fetal DNA in maternal plasma and serum. Lancet. 1997;350:485–7.
5. Lo YM, Zhang J, Leung TN, Lau TK, Chang AM, Hjelm NM. Rapid clearance of fetal DNA from maternal plasma. Am J Hum Genet. 1999;64:218–24.
6. Angert RM, LeShane ES, Lo YM, Chan LY, Delli-Bovi LC, Bianchi DW. Fetal cell-free plasma DNA concentrations in maternal blood are stable 24 h after collection: analysis of first- and third-trimester samples. Clin Chem. 2003;49:195–8.
7. Chan KC, Zhang J, Hui AB, Wong N, Lau TK, Leung TN, et al. Size distributions of maternal and fetal DNA in maternal plasma. Clin Chem. 2004;50:88–92.
8. Li Y, Zimmermann B, Rusterholz C, Kang A, Holzgreve W, Hahn S. Size separation of circulatory DNA in maternal plasma permits ready detection of fetal DNA polymorphisms. Clin Chem. 2004;50:1002–11.
9. Tong YK, Ding C, Chiu RW, Gerovassili A, Chim SS, Leung TY, et al. Noninvasive prenatal detection of fetal trisomy 18 by epigenetic allelic ratio analysis in maternal plasma. Theor Empir Consid Clin Chem. 2006;52:2194–202.
10. Lo YM, Tsui NB, Chiu RW, Lau TK, Leung TN, Heung MM, et al. Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection. Nat Med. 2007;13:218–23.
11. Finning KM, Martin PG, Soothill PW, Avent ND. Prediction of fetal D status from maternal plasma: introduction of a new

noninvasive fetal RHD genotyping service. Transfusion. 2002;42:1079–85.
12. Phylipsen M, Yamsri S, Treffers EE, Jansen DT, Kanhai WA, Boon EM, et al. Non-invasive prenatal diagnosis of beta-thalassemia and sickle-cell disease using pyrophosphorolysis-activated polymerization and melting curve analysis. Prenat Diagn. 2012;32:578–87.
13. Lam KW, Jiang P, Liao GJ, Chan KC, Leung TY, Chiu RW, et al. Noninvasive prenatal diagnosis of monogenic diseases by targeted massively parallel sequencing of maternal plasma: application to beta-thalassemia. Clin Chem. 2012;58:1467–75.
14. Bustamante-Aragones A, Rodriguez de Alba M, Perlado S, Trujillo-Tiebas MJ, Arranz JP, Diaz-Recasens J, et al. Non-invasive prenatal diagnosis of single-gene disorders from maternal blood. Gene. 2012;504:144–9.
15. Lench N, Barrett A, Fielding S, McKay F, Hill M, Jenkins L, et al. The clinical implementation of non-invasive prenatal diagnosis for single-gene disorders: challenges and progress made. Prenat Diagn. 2013;33:555–62.
16. Papasavva T, van Ijcken WF, Kockx CE, van den Hout MC, Kountouris P, Kythreotis L, et al. Next generation sequencing of SNPs for non-invasive prenatal diagnosis: challenges and feasibility as illustrated by an application to beta-thalassaemia. Eur J Hum Genet. 2013;21:1403–10.
17. Ge H, Huang X, Li X, Chen S, Zheng J, Jiang H, et al. Non-invasive prenatal detection for pathogenic CNVs: the application in alpha-thalassemia. PLOS ONE. 2013;8:e67464.
18. Lo YM, Chan KC, Sun H, Chen EZ, Jiang P, Lun FM, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med. 2010;2:61ra91.
19. Fan HC, Gu W, Wang J, Blumenfeld YJ, El-Sayed YY, Quake SR, et al. Non-invasive prenatal measurement of the fetal genome. Nature. 2012;487:320–4.
20. Chen S, Ge H, Wang X, Pan X, Yao X, Li X, et al. Haplotype-assisted accurate non-invasive fetal whole genome recovery through maternal plasma sequencing. Genome Med. 2013;5:18.
21. Kitzman JO, Snyder MW, Ventura M, Lewis AP, Qiu R, Simmons LE, et al. Noninvasive whole-genome sequencing of a human fetus. Sci Transl Med. 2012;4:137ra176.
22. Liao GJ, Lun FM, Zheng YW, Chan KC, Leung TY, Lau TK, et al. Targeted massively parallel sequencing of maternal plasma DNA permits efficient and unbiased detection of fetal alleles. Clin Chem. 2011;57:92–101.
23. Chan KC, Ding C, Gerovassili A, Yeung SW, Chiu RW, Leung TN, et al. Hypermethylated RASSF1A in maternal plasma: A universal fetal DNA marker that improves the reliability of noninvasive prenatal diagnosis. Clin Chem. 2006;52:2211–8.
24. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012:251364.
25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. Genome project data processing S: the sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
27. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
28. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29:308–11.
29. Maasalu K, Nikopensius T, Koks S, Noukas M, Kals M, Prans E, et al. Whole-exome sequencing identifies de novo mutation in the *COL1A1* gene to underlie the severe osteogenesis imperfecta. Hum Genom. 2015;9:6.
30. Cho SY, Lee JH, Ki CS, Chang MS, Jin DK, Han HS. Osteogenesis imperfecta Type I caused by a novel mutation in the start codon of the COL1A1 gene in a Korean family. Ann Clin Lab Sci. 2015;45:100–5.
31. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
32. Seo SB, King JL, Warshauer DH, Davis CP, Ge J, Budowle B. Single nucleotide polymorphism typing with massively parallel sequencing for human identification. Int J Leg Med. 2013;127:1079–86.
33. Beal MA, Glenn TC, Somers CM. Whole genome sequencing for quantifying germline mutation frequency in humans and model species: cautious optimism. Mutat Res. 2012;750:96–106.
34. Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science. 2012;338:1622–6.
35. Chen L, Liu P, Evans TC Jr, Ettwiller LM. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. Science . 2017;355:752–6.
36. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, et al. Integrated digital error suppression for improved detection of circulating tumor DNA. Nat Biotechnol. 2016;34:547–55.
37. Van Dijk FS. Sillence DO.Osteogenesis imperfecta: clinical diagnosis, nomenclature and severity assessment. Am J Med Genet A. 2014;164A:1470–81.
38. Forlino A, Cabral WA, Barnes AM, Marini JC. New perspectives on osteogenesis imperfecta. Nat Rev Endocrinol. 2011;7:540–57.
39. Zhytnik L, Maasalu K, Reimann E, Prans E, Koks S, Martson A. Mutational analysis of *COL1A1* and *COL1A2* genes among Estonian osteogenesis imperfecta patients. Hum Genom. 2017;11:19.
40. Rolvien T, Kornak U, Sturznickel J, Schinke T, Amling M, Mundlos S, et al. A novel *COL1A2* C-propeptide cleavage site mutation causing high bone mass osteogenesis imperfecta with a regional distribution pattern. Osteoporos Int. 2018;29:243–6.
41. Andersson K, Dahllof G, Lindahl K, Kindmark A, Grigelioniene G, Astrom E, et al. Mutations in *COL1A1* and *COL1A2* and dental aberrations in children and adolescents with osteogenesis imperfecta - a retrospective cohort study. PLOS ONE. 2017;12:e0176466.
42. Kantaputra PN, Sirirungruangsarn Y, Intachai W, Ngamphiw C, Tongsima S, Dejkhamron P. Osteogenesis imperfecta with ectopic mineralizations in dentin and cementum and a *COL1A2* mutation. J Hum Genet. 2018;63:811–820.
43. Monti E, Mottes M, Fraschini P, Brunelli P, Forlino A, Venturi G, et al. Current and emerging treatments for the management of osteogenesis imperfecta. Ther Clin Risk Manag. 2010;6:367–81.
44. Marini JC, Forlino A, Cabral WA, Barnes AM, San Antonio JD, Milgrom S, et al. Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. Hum Mutat. 2007;28:209–21.
45. Ma D, Ge H, Li X, Jiang T, Chen F, Zhang Y, et al. Haplotype-based approach for noninvasive prenatal diagnosis of congenital adrenal hyperplasia by maternal plasma DNA sequencing. Gene. 2014;544:252–8.
46. Meng M, Li X, Ge H, Chen F, Han M, Zhang Y, et al. Non-invasive prenatal testing for autosomal recessive conditions by maternal plasma sequencing in a case of congenital deafness. Genet Med. 2014;16:972–6.

47. New MI, Tong YK, Yuen T, Jiang P, Pina C, Chan KC, et al. Noninvasive prenatal diagnosis of congenital adrenal hyperplasia using cell-free fetal DNA in maternal plasma. J Clin Endocrinol Metab. 2014;99:E1022–1030.
48. Xu Y, Li X, Ge HJ, Xiao B, Zhang YY, Ying XM, et al. Haplotype-based approach for noninvasive prenatal tests of Duchenne muscular dystrophy using cell-free fetal DNA in maternal plasma. Genet Med. 2015;17:889–96.
49. Yoo SK, Lim BC, Byeun J, Hwang H, Kim KJ, Hwang YS, et al. Noninvasive prenatal diagnosis of duchenne muscular dystrophy: comprehensive genetic diagnosis in carrier, proband, and fetus. Clin Chem. 2015;61:829–37.
50. Xiong L, Barrett AN, Hua R, Tan TZ, Ho SS, Chan JK, et al. Non-invasive prenatal diagnostic testing for beta-thalassaemia using cell-free fetal DNA and next generation sequencing. Prenat Diagn. 2015;35:258–65.
51. Xiong L, Barrett AN, Hua R, Ho S, Jun L, Chan K, et al. Non-invasive prenatal testing for fetal inheritance of maternal beta-thalassaemia mutations using targetted sequencing and relative mutation dosage: a feasibility study. BJOG. 2018; 125:461–8.
52. Tanner L, Vainio P, Sandell M, Laine J. Novel *COL1A1* mutation c.3290G>T associated with severe form of osteogenesis imperfecta in a fetus. Pediatr Dev Pathol. 2017;20:455–9.