

ARTICLE

# Next generation sequencing of SNPs for non-invasive prenatal diagnosis: challenges and feasibility as illustrated by an application to $\beta$ -thalassaemia

Thessalia Papasavva<sup>\*1</sup>, Wilfred FJ van IJcken<sup>2</sup>, Christel EM Kockx<sup>2</sup>, Mirjam CGN van den Hout<sup>2</sup>, Petros Kountouris<sup>1</sup>, Loukas Kythreotis<sup>3</sup>, Eleni Kalogirou<sup>3</sup>, Frank G Grosveld<sup>2</sup> and Marina Kleanthous<sup>1</sup>

$\beta$ -Thalassaemia is one of the most common autosomal recessive single-gene disorder worldwide, with a carrier frequency of 12% in Cyprus. Prenatal tests for at risk pregnancies use invasive methods and development of a non-invasive prenatal diagnostic (NIPD) method is of paramount importance to prevent unnecessary risks inherent to invasive methods. Here, we describe such a method by assessing a modified version of next generation sequencing (NGS) using the Illumina platform, called 'targeted sequencing', based on the detection of paternally inherited fetal alleles in maternal plasma. We selected four single-nucleotide polymorphisms (SNPs) located in the  $\beta$ -globin locus with a high degree of heterozygosity in the Cypriot population. Spiked genomic samples were used to determine the specificity of the platform. We could detect the minor alleles in the expected ratio, showing the specificity of the platform. We then developed a multiplexed format for the selected SNPs and analysed ten maternal plasma samples from pregnancies at risk. The presence or absence of the paternal mutant allele was correctly determined in 27 out of 34 samples analysed. With haplotype analysis, NIPD was possible on eight out of ten families. This is the first study carried out for the NIPD of  $\beta$ -thalassaemia using targeted NGS and haplotype analysis. Preliminary results show that NGS is effective in detecting paternally inherited alleles in the maternal plasma.

*European Journal of Human Genetics* (2013) 21, 1403–1410; doi:10.1038/ejhg.2013.47; published online 10 April 2013

**Keywords:** NIPD;  $\beta$ -thalassaemia; SNPs; next generation sequencing; targeted sequencing; maternal plasma

## INTRODUCTION

$\beta$ -Thalassaemia is one of the most common autosomal recessive single-gene disorders in Cyprus, where about 12% of the population are carriers.<sup>1</sup> Currently, fetal genetic material for prenatal diagnosis is sampled by invasive procedures, which are associated with a significant risk of induced abortion.<sup>2,3</sup> The discovery that during pregnancy there is a median of 10% of cell-free fetal DNA in the maternal circulation opened up new avenues in the diagnostics of fetal genetics.<sup>4–6</sup> However, this poses a technical challenge as fetal DNA represents a minor population in maternal plasma, exacerbated by the fragmentation of fetal DNA.<sup>7,8</sup>

As the advent of next generation sequencing (NGS), a lot of effort has been concentrated on exploiting this technology for the measurement of aneuploidies in maternal plasma.<sup>9–14</sup> Using this technology, non-invasive prenatal diagnostic (NIPD) for trisomy 21, 18 and 13 has recently reached the clinical setting by analysing the relative amount of chromosomes in circulating cell-free DNA from maternal plasma.<sup>15,16</sup>

However, approaches permitting reliable detection of single-gene mutations or single-nucleotide polymorphisms (SNPs) using cell-free fetal DNA in maternal plasma are still under development. This is considerably more difficult because it concerns fetal genetic changes that differ only slightly from the maternal genome. Recently, a number of different strategies have been investigated to meet the

challenges for the non-invasive detection of  $\beta$ -thalassaemia using maternal plasma. Allele-specific real-time PCR is one of the first approaches that have been used to exclude paternal mutations in the maternal circulation.<sup>17</sup> Preferential detection of fetal alleles was achieved through initial enrichment of fetal DNA,<sup>8,18</sup> while others enhanced the production of the mutated fetal allele by employing either peptide nucleic acid probes<sup>19</sup> or COLD PCR.<sup>20</sup> In the specific case of  $\beta$ -thalassaemia, MALDI-TOF mass spectrometry has been also investigated.<sup>8</sup>

Moreover, Lun *et al*<sup>21</sup> employed digital size selection to investigate the relative mutation dosage for NIPD of  $\beta$ -thalassaemia. Our group employed the APEX/thalassochip approach, based on the detection of polymorphic SNPs, in order to successfully identify the paternally inherited allele of the fetus in the maternal plasma<sup>22,23</sup> while, more recently, Phylipsen *et al*<sup>24</sup> employed pyrophosphorolysis activated polymerisation analysis using polymorphic SNPs to detect the paternal allele in maternal plasma of  $\beta$ -thalassaemia carriers. However, more SNPs need to be included in the study to link the paternal allele with minimal risk of misdiagnosis. Hence, despite the advances in technology, NIPD of  $\beta$ -thalassaemia has yet to reach clinical practice.<sup>8,18,19,22,23,25</sup>

Our approach is based on the detection of the paternally inherited alleles as maternal alleles cannot be differentiated from fetal ones. Therefore, NIPD is possible only in those cases where the fetus

<sup>1</sup>Molecular Genetics Thalassaemia Department, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; <sup>2</sup>Department of Cell Biology and Erasmus Center for Biomics, Erasmus Medical Centre (EMC), Rotterdam, The Netherlands; <sup>3</sup>Thalassaemia Center, Makarios III Hospital, Nicosia, Cyprus

\*Correspondence: Dr T Papasavva, Molecular Genetics Thalassaemia Department, The Cyprus Institute of Neurology and Genetics, 6, International Airport Avenue, Agios Dometios, Nicosia 1683, Cyprus. Tel: +357 22 392 664; Fax: +357 22 392 615; E-mail: thesalia@cing.ac.cy

Received 13 June 2012; revised 18 February 2013; accepted 20 February 2013; published online 10 April 2013

inherits the normal allele of the father. The aim of this study is to assess the analytical power and specificity of a modified version of NGS using the Illumina platform, called 'targeted sequencing', for the reliable detection of paternally inherited SNPs in the maternal plasma of at risk pregnancies for  $\beta$ -thalassaemia. Moreover, this study aims to use this platform for the development of a fast and cost-effective non-invasive diagnostic assay for  $\beta$ -thalassaemia. The principle of our targeted sequencing is to selectively amplify or capture targeted regions of a DNA before sequencing. For this purpose, we have developed a method of integrating selectively amplified targeted DNA regions and NGS. The proof of principle results presented in this study show that the detection of paternally inherited SNPs in the maternal plasma is possible and reliable using targeted sequencing.

## MATERIALS AND METHODS

### Sample collection and processing

The study was approved by the Cyprus National Bioethics Committee and all subjects gave informed consent. Blood samples, as well as corresponding chorionic villi samples (CVS), were collected from ten families (including parents and grandparents) at risk for  $\beta$ -thalassaemia in their newborns. Approximately 9 ml maternal blood samples were collected into EDTA-containing tubes between the 10th and 11th week of gestation and before chorionic villus sampling. Plasma was separated from cells by centrifugation at low speed, 2500 g for 10 min without braking. It was transferred to micro-centrifuge tubes and subjected to a second centrifugation step at 16 000 g for 40 min to remove any residual cells. The two centrifugation steps were performed within 4–8 h of collection.

### DNA extraction

Cell-free DNA was extracted from 1 ml of maternal plasma using QIAamp Circulating Nucleic Acid Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions. Genomic DNA was extracted using the Puregene Blood Core Kit C (Qiagen Sciences, Germantown, MD, USA).

### Selection of SNPs and primer design

Four SNPs, rs3834466\_IIe, rs968857\_3' $\psi$  $\beta$ , rs10768683\_AvaII and rs7480526\_II74, located on the  $\beta$ -globin gene cluster showing high heterozygosity in the Cypriot population were selected for analysis (Figure 1). SNPs rs968857\_3' $\psi$  $\beta$  and rs7480526\_II74 were selected after a genotyping analysis carried out in a previous study,<sup>26</sup> whereas SNPs rs10768683\_AvaII and rs3834466\_IIe are routinely used in Cyprus for prenatal diagnosis.<sup>27</sup> Two of the SNPs are located within the  $\beta$ -globin gene and the other two SNPs 5' to the  $\delta$ -globin gene, therefore, any recombination events would be instantly recognisable.

Primer sequences were designed using the reference sequence of haemoglobin gene locus with accession number NG\_000007 from the NCBI database ([http://www.ncbi.nlm.nih.gov/nucore/NG\\_000007](http://www.ncbi.nlm.nih.gov/nucore/NG_000007)). The largest PCR-fragment generated was no more than 268 bp (Supplementary Table 1).

### PCR amplification

Targeted PCR was performed for the selected SNPs with the corresponding primers. Each reaction was carried out with 5 ng of genomic DNA or 1–5 ng of maternal plasma DNA in a total reaction volume of 25  $\mu$ l that contained 1  $\times$  PCR buffer with 1.5 mM MgCl<sub>2</sub>, 200  $\mu$ M dNTPs and one unit of AmpliTaq Gold DNA polymerase (Applied Biosystems by Roche Molecule Systems Inc., Branchburg, NJ, USA). The amplification procedure consisted of an initial

denaturation step at 95 °C for 10 min, followed by 45 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s, extension at 72 °C for 30 s, followed by a final extension step at 72 °C for 7 min. The PCR products were purified using the MinElute PCR purification kit, Qiagen, according to the manufacturer's instructions.

We used a 'spiking' approach with genomic DNA to determinate the specificity of the assay for each of the four SNPs. We used 100% genomic DNA having a homozygous genotype for the SNP and a sample of 90% homozygous DNA spiked with 10% DNA having a heterozygous genotype. For the confirmation of the reproducibility, four replicates were performed for each sample per SNP.

### Nested PCR

New genomic primers for the nested PCR were designed using NCBI primer-Blast software so that the SNP of interest was within 36 bp from the nested primer. All genomic primers contained the sequence for the Illumina adaptors and the reverse primers also contained a 6 bp Illumina index barcode. For rs3834466\_IIe, rs968857\_3' $\psi$  $\beta$  and rs10768683\_AvaII the Illumina index barcodes 1, 2, 4, 5, 6, 7, 8, 9, 10 and 12 were used. For rs7480526\_II74, the Illumina index barcodes 2, 4, 5, 6, 7, 8, 9, 12, 13 and 14 were used (Supplementary Table 1).

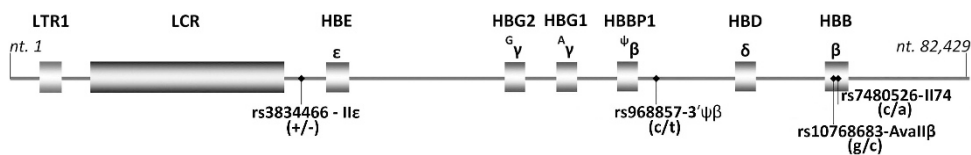
One microliter of the targeted PCR products was loaded on an Agilent Technologies 2100 Bioanalyzer (Santa Clara, CA, USA) using a DNA 1000 assay to determine the concentration and to check for quality. Nested PCR was performed to introduce the Illumina adaptors and the index barcode to the PCR products. Each reaction was carried out with 0.5 ng of targeted PCR products in a total reaction volume of 50  $\mu$ l that contained 1  $\times$  phusion HF buffer, 200  $\mu$ M dNTPs and one unit Phusion DNA Polymerase (Finnzymes part of Thermo Fisher Scientific, Waltham, MA, USA). The nested PCR consisted of an initial denaturation step at 98 °C for 30 s followed by 18 cycles of denaturation at 98 °C for 10 s, annealing at 57.4 °C for 30 s, extension at 72 °C for 30 s, followed by a final extension step at 72 °C for 5 m. The PCR was performed in a Biometra TProfessional (Goettingen, Germany). The PCR fragments were purified using AMPure XP beads according to the manufacturer's instructions. The products were eluted in 15  $\mu$ l of elution buffer. One microliter of the product was loaded on an Agilent Technologies 2100 Bioanalyzer using a DNA 1000 assay to determine the quality and the quantity of the genomic library.

### Bridge amplification and sequencing-by-synthesis

Cluster generation was performed according to the Illumina TruSeq SR Cluster kit v2 (cBot) Reagents Preparation Guide ([www.illumina.com](http://www.illumina.com)). Briefly, 40 PCR libraries were pooled together to get a stock of 10 nM. One microliter of the 10 nM stock was denaturated with NaOH, diluted to 6.5 pM and hybridised onto the flowcell. The hybridised products were sequentially amplified, linearised and end-blocked according to the Illumina Single Read Multiplex Sequencing user guide. After hybridisation of the sequencing primer, sequencing-by-synthesis was performed using the HiSeq 2000 with a 36-cycle protocol. The sequenced fragments were denaturated with NaOH using the HiSeq 2000 and the index-primer was hybridised onto the fragments. The index was sequenced with a 7-cycle protocol.

### Sequence analysis

We used NARWHAL<sup>28</sup> to demultiplex the raw data from the sequencer into FASTQ files per sample. For the spiked genomic samples, we did not perform any alignment, but we filtered the data to keep only reads in which all bases have at least a base-calling phred score of 10 and a mean phred score over the whole read of at least 30. We directly counted the number of times the different



**Figure 1** Four SNPs located on the  $\beta$ -globin gene cluster used in targeted NGS analysis.

read sequences occurred. For the maternal plasma samples, we performed demultiplexing and BWA<sup>29</sup> alignment against the human reference genome (version hg18) with NARWHAL and counted the occurrence of the reads for both alleles in the BWA aligned SAM files using standard Unix tools *grep* and *sed*. All subsequent calculations were performed in Microsoft Excel.

## RESULTS

### Specificity of NGS on spiked samples

In order to develop a reliable method, we selected four SNPs in the  $\beta$ -globin locus, that are common in the Cypriot population, rs3834466\_IIe, rs968857\_3' $\psi$  $\beta$ , rs10768683\_AvaII and rs7480526\_II74, each showing high heterozygosity (Figure 1 and Materials and Methods). Next, we designed primers to specifically amplify the DNA containing these SNPs. Homozygous DNA was amplified directly or after spiking with heterozygous containing DNA, followed by Illumina sequencing and more than 1.5 million reads per sample were obtained (Table 1). In all homozygous samples for all SNPs (100%), the sequence reads match the expected homozygous sequence. In the spiked samples (90% + 10%), the two different sequences present were accurately detected and differentiated in the expected (5%) allele concentration. We noted a somewhat higher concentrations for SNP rs7480526\_II74 at around 8.2%. For this SNP, we also saw reads with an additional variant 7 bp before the SNP position. These variant reads have not been taken into account. A negligible variation was observed between replicates. We conclude that the analytical power of the platform is sufficiently high and specific to be tested on maternal plasma analysis.

### Family study

The maternal and paternal genotypes for 101 families at risk were previously determined by MALDI-TOF MS for a panel of 49 SNPs.<sup>26</sup> SNPs were considered informative when the mother was homozygous and the father was heterozygous for the same SNP. Ten families having at least two informative SNPs of the four analysed above were selected for the NGS analysis. SNPs where the mother was homozygous for one allele and the father was homozygous for the other allele were also included for confirmation of the paternal allele, whereas SNPs where both parents are homozygous were used for the determination of the sequencing error rate. SNPs where the mother was heterozygous were not used for the deduction of the paternal allele. The corresponding CVS material was also typed for the selected SNPs for the confirmation of the maternal plasma result.

### Efficiency of NGS on maternal plasma

To detect the paternally inherited allele in the maternal plasma, we analysed the selected couples at risk of carrying a  $\beta$ -thalassaemia child for the corresponding informative SNPs. Each sample was analysed in triplicate. About two million reads were generated per sample replicate and the detailed results are shown in Table 2. The exact number of reads was determined for all 40 samples for both alleles (ten samples of four SNPs). The proportion of DNA molecules in the maternal plasma sample that originated from the fetus was determined based on the fractional fetal DNA ( $f$ ) in the maternal plasma, which fetal is given by:

$$f = \frac{2p}{p+q}$$

where  $p$  is the number of sequenced reads of the fetal specific allele and  $q$  is the read count of the other allele, which is shared by the maternal and fetal genomes.<sup>25</sup>  $f$  was calculated from the sequencing data for each SNP and replicate (Table 2). Theoretically, a non-present

fetal allele gives a calculated fetal fraction of 0, whereas the fetal fraction DNA in the maternal circulation can be as low as 3% with an average of 10%.<sup>6,7</sup> We used a cutoff of 2.5%, which is somewhat below the minimum value. We consider the fetal allele detected if  $f$  is larger than 2.5% for at least two out of three replicates.

All maternal genotypes determined from the sequence data were correct for the four SNP's. Six sites were not used for the deduction of the paternal allele as the mother was heterozygous for those. The results for the fetal genotypes were compared with the results of the CVS analysis previously performed for prenatal diagnosis. From 34 samples analysed, concordance with CVS was observed in 27 cases, where we positively detected and differentiated the paternal allele in the maternal plasma in nine of the cases and in 18 cases no other measurable allele was observed (negative detection of the paternal allele as expected). However, we also observed four false-positive and three false-negative results. SNP rs3834466\_IIe was analysed for nine samples as follows: for sample five the mother is heterozygous for the SNP and, therefore, the paternally inherited allele in the maternal background could not be discriminated. Six out of nine samples showed concordance with the CVS result, where we correctly detected the paternally inherited allele (true-positive detection) in three of the cases, whereas in three cases we correctly did not detect the paternal allele (true-negative detection). However, two false negatives and one false positive were observed. For SNP rs968857-3' $\psi$  $\beta$ , concordance with the CVS was observed in seven out of ten analysed samples with positive detection of the paternally inherited allele in five cases and negative detection in two cases. Two false positives and one false negative were observed. SNP rs10768683\_AvaII $\beta$  was analysed for eight samples showing correct negative detection of the paternally inherited allele in seven cases, whereas in one case a single false positive detection was found, although this analysis appears suspect due to the very high positive score in one of the samples of the triplicate. For SNP rs7480526\_II74, seven out of ten samples were analysed showing concordance with CVS in all samples analysed. We observed true-positive detection of the paternal allele in a single sample and true-negative detection in six of the cases.

### Non-invasive fetal haplotyping

To investigate the feasibility of SNPs for the NIPD of  $\beta$ -thalassaemia analysed by Illumina sequencing of the maternal plasma, haplotype analysis was performed. The paternal haplotype was determined from previous family studies in our lab for prenatal diagnostic purposes. Based on the results obtained from NGS of maternal plasma, the haplotypes of the fetus were generated and the alleles of the fetus were correctly linked to the paternal normal or  $\beta$ -thal allele for eight out of ten families (Table 3). The haplotypes were inferred if two out of three or three out of four SNPs had the expected result and given that paternal alleles could be differentiated. More specifically, for families 3, 5, 8 and 10 the paternally inherited allele of the fetus was correctly linked based on the result of all SNPs analysed. For families 1, 6, 7 and 9, the fetal allele was correctly linked to the paternal, even though one of the SNPs analysed gave incorrect result. In these cases, the information obtained from the other SNPs was sufficient to link the fetal allele with the paternal one. However, for families 2 and 4, the fetal allele could not be linked as either more than one SNP showed an unexpected result (family 2) or there was insufficient information from the analysed SNPs (family 4). In these cases, the NIPD was inconclusive.

For families 1, 3, 6, 7 and 10, the fetal allele was correctly linked to the paternal  $\beta$ -thal allele and, consequently, the inheritance of the mutated paternal allele indicated that NIPD for these cases is  $\beta$ -thal

**Table 1 Detailed results of targeted sequencing on spiked genomic samples with Illumina platform**

SNP	Read	Allele	Number of reads in												
			Index 1	Index 2	Index 4	Index 5	Index 6	Index 7	Index 9	Index 12	100% T/T	90% T/T + 10% T/-	100% C/C	90% C/C + 10% T/C	
rs3834466_11e (T/-)	5'-GCATTAAACATTTGCCTTAAAGGTGGTGACAGTT-3' <sup>28</sup> GA	T (reference) Minor allele (%)	2280896	1989313	2075229	1714959	1734372	1992984	1816358	2200381					
	5'-GCATTAAACATTTGCCTTAAAGGTGGTGACAGT_GAC-3'		11	7	21	49	75563	84364	80019	111877					
			Mean	0.00048226	0.00035188	0.00101193	0.00285713	4.17490131	4.06113949	4.21957237					
			SD	0.0011758	0.00115666	0.00115666	0.00115666	4.32042391	4.32042391	4.32042391					
			100% C/C												
rs968857_3 <sup>ψ</sup> β (C/T)	5'-GTGAATAAATGCATGACACATGCTTGGTGAATAATC-3'	C (reference) Minor allele (%)	1972133	1582060	2502668	2118741	1615931	1551738	1452447	1659819					
	5'-GTGAATAAATGCATGACACATGCTTGGTGAATAATC-3'		1758	1226	2079	1573	104641	98015	96636	110047					
			Mean	0.08906267	0.07743389	0.0830024	0.07418713	6.08175653	5.94119241	6.23827129					
			SD	0.08092132	0.08092132	0.08092132	0.08092132	6.1197589	6.1197589	6.1197589					
			100% G/G												
rs10768683_Ava (G/C)	5'-CCATAGAAAAAGGGGAAAAAACAATCAAGGGTC-3'	G (reference) Minor allele (%)	1666708	2058647	2196255	2034668	1774046	1873629	1781723	1993217					
	5'-CCATAGAAAAAGGGGAAAAAACAATCAAGGGTC-3'		207	76	106	99	114245	118662	116282	129296					
			Mean	0.01241815	0.00369161	0.00482616	0.00486542	6.05017977	5.95605762	6.12653813					
			SD	0.00645034	0.00645034	0.00645034	0.00645034	6.05610564	6.05610564	6.05610564					
			index 2	index 4	index 5	index 6	index 7	index 12	index GACGAA	index TCGTCT					
			100% C/C												
rs7480526_1174 (C/A)	5'-TCCCATTCTAAACTGTACCCCTGTACTTCCCCCTT-3'	C (reference) Minor allele (%)	1738020	2099298	1819746	1757466	1185112	1139447	979076	1081119					
	5'-TCCCATTCTAAACTGTACCCCTGTACTTCCCCCTT-3'		1154	1848	1478	1893	104165	103516	90399	95544					
			Mean	0.06635334	0.087952	0.08115421	0.10759601	8.07933439	8.32816423	8.452652					
			SD	0.08576389	0.08576389	0.08576389	0.08576389	8.24501565	8.24501565	8.24501565					

<sup>a</sup>Highlighted nucleotide denotes the SNP of interest.

**Table 2 Detailed results of targeted sequencing on the maternal plasma analysis with Illumina platform and comparison with CVS**

Sample Replicate	no	rs3834466_c.11e (-t)				rs568857_3'UTR (ct)				rs10766683_AvalIII (g/c)				rs7480526_1174 (c/a)						
		Maternal allele	Fetal fraction (%) <sup>a</sup>	Fetal genotype (mat. plasma)	Predicted genotype (mat. plasma)	Maternal allele	Fetal fraction (%) <sup>a</sup>	Fetal genotype (mat. plasma)	Predicted genotype (mat. plasma)	Maternal allele	Fetal fraction (%) <sup>a</sup>	Fetal genotype (mat. plasma)	Predicted genotype (mat. plasma)	Maternal allele	Fetal fraction (%) <sup>a</sup>	Fetal genotype (mat. plasma)	Predicted genotype (mat. plasma)			
1	1	2073484	896	0.09	mo/t	mo/t	1548746	10277	c/c	1897149	499	0.05	mo/g	g/g	1045459	1385449	113.99	mo/c	NA	NA
2	2	1957152	56566	5.62	Yes	1.32 mo/c	1295375	7582	1.16 No	1694464	543	0.06	No	g/g	1329232	1214239	95.48	NA	NA	NA
3	3	1819521	63366	6.73		1.11	1387503	7757		1884040	563	0.06			1351923	815548	75.25			
4	4	2357978	1100	0.09	mo/t	5.82 mo/c	1160722	34820	ct	1825846	569	0.05	mo/g	g/g	1298378	1241940	97.78	mo/c	NA	NA
5	5	1750844	12450	1.41	No	7.75 Yes	1051814	42483	ct	1910171	36498	3.75	Yes	g/g	969808	1519854	122.09	NA	NA	NA
6	6	1728292	666	0.08		3.64	1324234	24515		1666173	120973	13.54			1436671	2150903	119.91			
7	7	2240645	65379	5.67	mo/-	4.76 mo/t	1703340	41536	ct	2208515	811	0.07	mo/g	g/g	1872489	16531	1.75	mo/c	c/c	c/c
8	8	1944017	22796	2.32	Yes	9.76 Yes	1872737	96027	ct	1894954	680	0.07	No	g/g	2254514	16372	1.44	no		
9	9	2296009	70803	5.98		3.21	1615333	26390		1971572	747	0.08		g/g	1963487	12282	1.24			
10	10	2219553	1782	0.16	mo/t	1.42 mo/c	1429699	10218	ct	2095634	688	0.07	mo/g	g/g	341939	1511662	163.11	mo/c	NA	NA
11	11	1960895	1636	0.17	No	6.99 Yes	1037324	37577		1832665	698	0.08	No	g/g	644981	1342593	135.10	NA	NA	NA
12	12	2103371	1491	0.14		8.14	1335824	56647		1865159	696	0.07		g/g	651981	965476	119.38			
13	13	1053614	89692	91.95	mo/-t	5.20 mo/c	1344138	35877	ct	2104022	826	0.08	mo/g	g/g	2321198	22757	1.94	mo/c	c/c	c/c
14	14	936608	878892	96.82	NA	2.17 Yes	1236671	13578	ct	1796659	777	0.09	No	g/g	2083631	17684	1.68	No		
15	15	854173	1171199	115.65		4.18	1060075	22627		1883374	798	0.08		g/g	1774714	12425	1.39			
16	16	2171601	3513	0.32	mo/-	3.50 mo/t	1349070	24017	ct	1045257	696087	79.95	mo/g	g/g	2104022	19252	5.53	mo/c	c/a	c/a
17	17	2200687	3182	0.29	No	4.55 Yes	1053813	24517	ct	942509	909284	98.21	NA	NA	851481	14683	3.39	mo/a	Yes	Yes
18	18	1937523	2728	0.28		2.65	1304523	17539		950240	893464	96.92		g/g	701110	14476	4.05			
19	19	2026011	1100	0.11	mo/t	6.54 mo/c	1208208	40848	ct	1934519	611	0.05	mo/g	g/g	2093685	19675	1.86	mo/c	c/c	c/c
20	20	2190300	965	0.09	No	3.70 Yes	1124971	21212	ct	2060072	759	0.08	No	g/g	2158817	16864	1.55	No		
21	21	2038883	903	0.09		6.62	959038	32842		1713765	543	0.06		g/g	1884662	12393	1.31			
22	22	2033758	30848	2.99	mo/t	7.07 mo/c	1260241	46185	ct	1652465	427	0.05	mo/g	g/g	2102837	22228	2.09	mo/c	c/c	c/c
23	23	2235247	36166	3.20	Yes	7.96 Yes	1094286	46367	ct	1675197	503	0.06	No	g/g	2294171	20870	1.80	No		
24	24	2158341	591	0.05		8.48	1369524	60658		1546598	409	0.05		g/g	1839280	14216	1.53			
25	25	2051763	36363	3.48	mo/t	1.73 mo/c	1321919	11531	ct	1792980	719	0.08	mo/g	g/g	793248	9537	2.38	mo/a	a/a	a/a
26	26	1944390	1179	0.12	Yes	1.50 No	1241992	9361	ct	1870102	784	0.08	No	g/g	871091	9846	2.24	No		
27	27	2298786	83406	7.00		1.39	1106279	7763		1762583	711	0.08		g/g	598521	6544	2.16			
28	28	2482264	3305	0.11	mo/t	1.56 mo/c	1569485	12349	ct	993793	9080	90.80	mo/g	g/g	2277314	37611	3.25	mo/c	c/c	c/c
29	29	1886618	939	0.10	No	1.39 No	1430534	10039		860463	780507	95.71	NA	NA	2670000	21045	2.01	No		
30	30	2323991	1027	0.09		1.25	1272596	8017		1160267	834042	83.64			1754158	15404	1.74			

Abbreviation: CVS, chorionic villus sampling; NA, not applicable.

<sup>a</sup>*n* = 2\*Fetal/(mat + fetal).

<sup>b</sup>Highlighted genotypes indicate discordance of maternal plasma analysis with CVS analysis.

<sup>c</sup>Mother heterozygous, father heterozygous for the SNP.

<sup>d</sup>Fetal allele is detected if *f* ≥ 2.5 for at least two out of three replicates.

**Table 3 Haplotype analysis and NIPD of the ten families for the four SNPs**

Family no.	Haplotype	Mother				Father				Maternal Plasma				CVS				PND diagnosis		
		rs3834	rs9688	rs10768	rs748052	rs38344	rs9688	rs107686	rs748052	Allele	rs38344	rs9688	rs107686	rs748052	Allele	rs3834	rs9688		rs10768	rs7480
		466_11c	57_3'v/jf	683_Aval/jf	6_1174	66_11c	57_3'v/jf	3_Aval/jf	6_1174	phase <sup>a</sup>	83_Aval/jf	26_1174	phase <sup>a</sup>	NIPD diagnosis	466_11c	57_3'v/jf	683_Aval/jf	526_1174	phase	
1	1	t	c	g	ND	t	c	g	ND	t	g	ND	$\beta$ -thal	t	c	g	normal	$\beta$ -thal trait	$\beta$ -thal trait	
	2	t	c	g	ND	—	t	c	ND	—	g	$\beta$ -thal	Normal	—	c	g	$\beta$ -thal	normal	$\beta$ -thal trait	
2	1	t	c	g	ND	t	c	g	ND	t	g	ND	$\beta$ -thal	t	c	g	$\beta$ -thal	$\beta$ -thal	$\beta$ -thal major	
	2	t	c	g	ND	—	t	c	ND	—	g	?	Normal	—	c	g	$\beta$ -thal	$\beta$ -thal	$\beta$ -thal major	
3	1	—	t	g	ND	t	c	g	c	—	g	ND	$\beta$ -thal	—	t	g	$\beta$ -thal	$\beta$ -thal	$\beta$ -thal major	
	2	—	t	g	ND	t	c	g	a	—	g	c	Normal	—	c	g	$\beta$ -thal	$\beta$ -thal	$\beta$ -thal major	
4	1	t	c	g	ND	t	c	g	ND	t	g	ND	$\beta$ -thal	t	c	g	Normal	Normal	Normal	
	2	t	c	g	ND	—	t	c	ND	—	g	?	Normal	—	t	g	Normal	Normal	Normal	
5	1	c	c	g	ND	c	c	g	a	c	g	ND	$\beta$ -thal	t	c	g	normal	normal	normal	
	2	c	c	g	ND	c	c	g	c	c	g	Normal	Normal	—	t	g	normal	normal	normal	
6	1	—	t	g	ND	t	c	g	a	—	a	ND	$\beta$ -thal	—	t	a	Normal	$\beta$ -thal	$\beta$ -thal trait	
	2	—	t	g	ND	t	c	g	a	—	a	ND	$\beta$ -thal	—	c	a	Normal	$\beta$ -thal	$\beta$ -thal trait	
7	1	t	c	g	ND	t	c	g	c	c	g	ND	$\beta$ -thal	t	c	g	Normal	Normal	$\beta$ -thal trait	
	2	t	c	g	ND	—	t	c	a	—	g	ND	$\beta$ -thal	—	c	g	Normal	$\beta$ -thal	$\beta$ -thal trait	
8	1	t	c	g	ND	t	c	g	c	c	g	ND	$\beta$ -thal	t	c	g	Normal	Normal	Normal	
	2	t	c	g	ND	—	t	c	c	—	g	Normal	Normal	—	t	g	Normal	Normal	Normal	
9	1	t	c	g	ND	t	c	g	a	a	g	ND	$\beta$ -thal	t	c	g	Normal	Normal	Normal	
	2	t	c	g	ND	—	t	c	a	—	g	Normal	Normal	—	t	g	Normal	Normal	Normal	
10	1	t	c	g	ND	t	c	g	c	c	g	ND	$\beta$ -thal	t	c	g	$\beta$ -thal	$\beta$ -thal	$\beta$ -thal major	
	2	t	c	g	ND	—	t	c	c	—	g	Normal	Normal	—	c	c	$\beta$ -thal	$\beta$ -thal	$\beta$ -thal major	

Abbreviation: CVS, chorionic villus sample; ND, non-determined; NIPD, non-invasive prenatal diagnosis; PND, prenatal diagnosis with CVS.  
<sup>a</sup>Allele phase was determined if two out of three or three out of four, and if the paternal alleles could be differentiated.

trait or major. Therefore, direct invasive prenatal diagnosis in a fetal sample was recommended to confirm the diagnosis of the fetus. For families 5 and 9, the fetal allele is correctly linked to the normal allele of the father, indicating that NIPD is normal or  $\beta$ -thal trait and, therefore, invasive prenatal procedures may be avoided. We concluded that more than four SNPs and more replicates are needed to develop a reliable assay for haplotype analysis. The above analysis was only used as a model in order to demonstrate the effective linkage of the fetal allele to the paternal based on SNPs. In a diagnostic setting, where more SNPs and more replicates will be included per family, a diagnostic algorithm would have to be included in order to derive the final haplotypes and, in turn, the final diagnosis.

## DISCUSSION

Most NIPD studies carried out on  $\beta$ -thalassaemia were based on detection of the paternally inherited mutation and, therefore, limited to the couples sharing different mutations.<sup>17–20</sup> In view of this limitation, we previously showed that the detection of the paternally inherited SNPs is feasible for the NIPD of  $\beta$ -thalassaemia.<sup>22,23</sup> SNPs can be used regardless of the mutation of the carrier couples, they provide positive detection of the paternal allele, normal or mutant, the result can be confirmed with more than one SNP and, importantly, the more SNPs used the less diagnostic risk.

In this study, we took advantage of the analytical power of NGS Illumina platform to reliably detect and quantify all the sequences present in a sample to detect the paternally inherited SNPs in the maternal plasma of  $\beta$ -thal carriers.

The specificity of the platform to detect and differentiate the minor allele present in the overwhelming background of the maternal allele was confirmed using spiked genomic samples. This demonstrates the extreme precision and analytical power of the method, as the correct result in all samples was obtained with insignificant variation between the replicates.

The reliability of the method was assessed with a preliminary analysis of maternal plasma samples. The current study showed that the detection of paternally inherited allele in the maternal plasma is possible with the use of targeted sequencing and SNPs.

Haplotype analysis based on the sequence results showed diagnosis was possible for eight out of ten families. The importance of having a high number of SNPs for each family was illustrated in all cases. In four of the cases, the paternal inheritance of the fetus was correctly deduced based on all the SNPs analysed. However, in four cases where one SNP gave unexpected result, the paternally inherited allele was correctly linked based on the information given by the other three SNPs with an insignificant risk of misdiagnosis. In two cases where NIPD was inconclusive, the fetal allele could not be linked to the paternal one as more than one SNP gave a deviating result or there was lack of adequate information from the analysed SNPs. It is important to emphasise that no misdiagnosis was made even in the cases where some SNPs gave an incorrect result. Including a higher number of SNPs will further increase the statistical power to differentiate the maternal from the paternal allele through haplotype analysis with a higher level of accuracy and for a greater proportion of carrier couples. Such an increase can be easily incorporated in the current Illumina technology platform with only a small increase in costs.

However, further studies are needed in order to improve the reliability of the used assay to eliminate the false positives and negatives. Possible reasons for the observed false negatives are the inefficiency in the isolation of fetal material in the maternal plasma and the inefficiency of the DNA amplification due to its very small

size and quantity. In some cases, the percentage of the minor paternal allele, as measured with the targeted sequencing approach, might be smaller than the actual percentage causing changes in the ratio of the alleles and, in turn, creating problems in their analysis. This depends on the position of the SNP on the amplicon in terms of where the fetal fragment is cut, as well as on the size of the fetal fragment. Fetal-derived DNA molecules are  $<300$  bp<sup>7</sup> showing a prominence at 143 bp.<sup>25</sup> The amplicon sizes of our fragments were between 170 bp and 268 bp and therefore, fragments longer than 146 bp, in combination with the position of the SNP on the amplicon, are expected to result in the amplification of only a fraction of the fetal molecules present in the maternal plasma resulting in false-negative results.

Cross contamination between samples that might have arisen during the extraction process could be a possible cause of false positives, but highly unlikely here as it would have been observed in other SNPs for the same samples. False-positive results and erroneous base calls have been observed by other teams that have used different platforms for the analysis of cell-free fetal DNA but also the technology of NGS.<sup>9,10,15,16</sup> Palomaki *et al*<sup>16</sup> and Chiu *et al*<sup>10</sup> reported false-positive rates of 1.4% and 2.1%, respectively, which were improved in a subsequent study.<sup>9</sup> False-positive SNP calls could be derived from erroneous alignments of short reads.<sup>30</sup> Therefore, different bioinformatics parameters, as well as improvements, in statistical analysis have been investigated by these teams in order to eliminate the erroneous calls and decrease the observed false-positive rate. In order to accomplish that larger scale studies would need to be performed. Moreover, the PCR amplification used in this study might also be a cause of increased error rates by amplifying nonspecific fragments due to high number of cycles used, although the nested PCR partially eliminates unspecific amplified fragments from the first PCR reaction.

We have also observed some variations in the reads of sequenced fragments from sample to sample and from replicate to replicate. It is unclear at this point whether this stems from the quality of the sample, PCR artefacts because of the high number of amplification cycles or during sequencing library preparation or cluster generation.

To avoid the observed discrepancies and to improve the diagnostic efficiency, it is suggested to include a higher number of maternal samples, as well as more SNPs and more replicates per sample with improved conditions for each SNP. This will also aid to derive statistical cutoff values specific for the data of each SNP as opposed to fixed cutoff values. Furthermore, it is suggested to have fragments around 143 bp in size in order to capture and amplify the maximum of the fetal molecules present in the maternal plasma. Furthermore, it is recommended to use less PCR cycles in order to avoid the introduction of erroneous bases. Finally, in future experiments free DNA isolated from plasma of non-pregnant women negative for the SNPs can also be included for a better evaluation of the false positives.

In this study, we have outlined and demonstrated the analytical power of NGS and targeted sequencing for the analysis of fetal DNA sequences in the maternal plasma. The accuracy and precision demonstrated allowed us to detect and differentiate the paternally inherited allele of the fetus in the overwhelming background of the maternal alleles. The linkage of paternally inherited allele was possible based on haplotype analysis, provided that other family members or a previously born child were available for testing. The group of Lam<sup>31</sup> employed the relative haplotype approach to deduce the paternally inherited allele without the need of other family members. However, their procedure is complicated and costly. Although the NGS method is a complex method, it has been demonstrated that it can be used to

detect maternal mutations with accuracy and precision.<sup>25,31–33</sup> Moreover, the targeted sequencing protocol using short 36 bp reads is a faster and more cost-effective than whole genome amplification used by Lo *et al*<sup>25</sup> for NIPD of  $\beta$ -thalassaemia. However, in order to implement our targeted SNP sequencing assay in diagnostics the method has to be simplified. The continuous announcements of smaller, more personalised NGS platforms promise new low-cost, rapid and less complicated sequencing accessible to more laboratory settings.

Directions for future development would be to improve the diagnostic efficiency, precision, accuracy and reliability by eliminating false positives and false-negative results. Finally, as this approach applies only to the 50% of the cases where the fetus inherits the normal allele of the father, in the future one should look for other approaches using this technology to also deduce the maternally inherited allele.

### CONFLICT OF INTEREST

The authors declare no conflict of interest

### ACKNOWLEDGEMENTS

We thank Mrs Elena Kyriacou for her secretarial assistance.

- Kyrri EK, Loizidou D, Ioannou C, Makariou C, Kythreotis LM, Phylactides PK *et al*: The changing epidemiology of  $\beta$ -thalassaemia in the Greek-Cypriot population. *Hemoglobin* 2013 (in press).
- Alfirevic Z, Sundberg K, Brigham S: Amniocentesis and chorionic villus sampling for prenatal diagnosis. *Cochrane Database Syst Rev* 2003; CD003252.
- Tabor A, Philip J, Madsen M, Bang J, Obel EB, Norgaard-Pedersen B: Randomised controlled trial of genetic amniocentesis in 4606 low-risk women. *Lancet* 1986; **1**: 1287–1293.
- Lo YM, Corbetta N, Chamberlain PF *et al*: Presence of fetal DNA in maternal plasma and serum. *Lancet* 1997; **350**: 485–487.
- Lun FM, Chiu RW, Allen Chan KC, Yeung Leung T, Kin Lau T, Dennis LoYM: Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. *Clin Chem* 2008; **54**: 1664–1672.
- Lo YM, Tein MS, Lau TK *et al*: Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. *Am J Hum Genet* 1998; **62**: 768–775.
- Chan KC, Zhang J, Hui AB *et al*: Size distributions of maternal and fetal DNA in maternal plasma. *Clin Chem* 2004; **50**: 88–92.
- Li Y, Di Naro E, Vitucci A *et al*: Size fractionation of cell-free DNA in maternal plasma improves the detection of a paternally inherited beta-thalassaemia point mutation by MALDI-TOF mass spectrometry. *Fetal Diagn Ther* 2009; **25**: 246–249.
- Chen EZ, Chiu RW, Sun H *et al*: Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. *PLoS One* 2011; **6**: e21791.
- Chiu RW, Akolekar R, Zheng YW *et al*: Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. *BMJ* 2011; **342**: c7401.
- Chiu RW, Chan KC, Gao Y *et al*: Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc Natl Acad Sci USA* 2008; **105**: 20458–20463.
- Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR: Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc Natl Acad Sci USA* 2008; **105**: 16266–16271.
- Fan HC, Quake SR: Sensitivity of noninvasive prenatal detection of fetal aneuploidy from maternal plasma using shotgun sequencing is limited only by counting statistics. *PLoS One* 2010; **5**: e10439.
- Chiu RW, Sun H, Akolekar R *et al*: Maternal plasma DNA analysis with massively parallel sequencing by ligation for noninvasive prenatal diagnosis of trisomy 21. *Clinical Chem* 2010; **56**: 459–463.
- Palomaki GE, Deciu C, Kloza EM *et al*: DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study. *Genet Med* 2012; **14**: 296–305.
- 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**: 913–920.
- Chiu RW, Lau TK, Leung TN, Chow KC, Chui DH, Lo YM: Prenatal exclusion of beta thalassaemia major by examination of maternal plasma. *Lancet* 2002; **360**: 998–1000.
- Li Y, Di Naro E, Vitucci A, Zimmermann B, Holzgreve W, Hahn S: Detection of paternally inherited fetal point mutations for beta-thalassaemia using size-fractionated cell-free DNA in maternal plasma. *JAMA* 2005; **293**: 843–849.
- Galbiati S, Foglieni B, Travi M *et al*: Peptide-nucleic acid-mediated enriched polymerase chain reaction as a key point for non-invasive prenatal diagnosis of beta-thalassaemia. *Haematologica* 2008; **93**: 610–614.
- Galbiati S, Brisci A, Lalatta F *et al*: Full COLD-PCR protocol for noninvasive prenatal diagnosis of genetic diseases. *Clin Chem* 2011; **57**: 136–138.
- Lun FM, Tsui NB, Chan KC *et al*: Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. *Proc Natl Acad Sci USA* 2008; **105**: 19920–19925.
- Papisavva T, Kalakoutis G, Kalikas I *et al*: Noninvasive prenatal diagnostic assay for the detection of beta-thalassaemia. *Ann N Y Acad Sci* 2006; **1075**: 148–153.
- Papisavva T, Kalikas I, Kyrri A, Kleantous M: Arrayed primer extension for the noninvasive prenatal diagnosis of beta-thalassaemia based on detection of single nucleotide polymorphisms. *Ann N Y Acad Sci* 2008; **1137**: 302–308.
- Phylipsen M, Yamsri S, Treffers EE *et al*: Non-invasive prenatal diagnosis of beta-thalassaemia and sickle-cell disease using pyrophosphorolysis-activated polymerization and melting curve analysis. *Prenatal Diagnosis* 2012; **32**: 578–587.
- Lo YM, Chan KC, Sun H *et al*: Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med* 2010; **2**: 61ra91.
- Papisavva TE, Lederer CW, Traeger-Synodinos J *et al*: A minimal set of SNPs for the noninvasive prenatal diagnosis of beta-thalassaemia. *Ann Hum Genet* 2013; **77**: 115–124.
- Old JM, Petrou M, Modell B, Weatherall DJ: Feasibility of antenatal diagnosis of beta thalassaemia by DNA polymorphisms in Asian Indian and Cypriot populations. *Br J Haematol* 1984; **57**: 255–263.
- Brouwer RW, van den Hout MC, Grosveld FG, van Ijcken WF: NARWHAL, a primary analysis pipeline for NGS data. *Bioinformatics* 2012; **28**: 284–285.
- Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–1760.
- Dohm JC, Lottaz C, Borodina T, Himmelbauer H: Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acid Res* 2008; **36**: e105.
- Lam KW, Jiang P, Liao GJ *et al*: Noninvasive prenatal diagnosis of monogenic diseases by targeted massively parallel sequencing of maternal plasma: application to beta thalassaemia. *Clin Chem* 2012; **58**: 1467–1475.
- Fan HC, Gu W, Wang J, Blumenfeld YJ, El-Sayed YY, Quake SR: Non-invasive prenatal measurement of the fetal genome. *Nature* 2012; **487**: 320–324.
- Kitzman JO, Snyder MW, Ventura M *et al*: Noninvasive whole-genome sequencing of a human fetus. *Sci Transl Med* 2012; **4**: 137ra176.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)