

1 Use of amplicon-based sequencing for testing fetal
2 identity and monogenic traits with single circulating
3 trophoblast (SCT) prenatal diagnosis

4 **Short title:** Genotyping fetal trophoblasts in maternal blood

5

6 Xinming Zhuo^{1*}, Qun Wang¹, Vossaert Liesbeth¹, Roseen Salman¹, Adriel Kim³, Ignatia Van den Veyver¹,
7 Amy Breman², Arthur Beaudet^{1*}

8

- 9 1. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX,
10 USA.
11 2. Department of Medical and Molecular Genetics, Indiana University School of Medicine,
12 Indianapolis, IN, USA
13 3. Graduate Program in Diagnostic Genetics, MD Anderson Cancer Center, Houston, TX
14 USA

15 * Email: AB: abeaudet@bcm.edu; XZ: xmzhuo@gmail.com

16

17 **Abstract**

18 A major challenge for cell-based non-invasive prenatal testing (NIPT) is to distinguish individual
19 presumptive fetal cells from maternal cells in female pregnancies. We have sought a rapid,
20 robust, versatile, and low-cost next-generation sequencing method to facilitate this process.
21 Toward this goal, single isolated cells underwent whole genome amplification prior to
22 genotyping. Multiple highly polymorphic genomic regions (including HLA-A and HLA-B) with 10-
23 20 very informative single nucleotide polymorphisms (SNPs) within a 200 bp interval were
24 amplified with a modified method based on other publications. To enhance the power of cell
25 identification, approximately 40 Human Identification SNP (Applied Biosystems) test amplicons
26 were also utilized. This method allowed reliable differentiation of fetal and maternal cells. In fully
27 informative cases, two haplotypes were found within the maternal reads, and fetal cells showed
28 reads with one but not the second maternal haplotype while also showing a novel paternal
29 haplotype absent in the mother. For SNP typing, at least 2 SNPs and 10% of informative SNPs
30 were required to differentiate a fetal cell from a maternal cell. A paternal DNA sample is not
31 required using this method. The assay also successfully detected point mutations causing Tay
32 Sachs disease, cystic fibrosis, and hemoglobinopathies in single lymphoblastoid cells, and
33 monogenic disease-causing mutations in three cell-based NIPT cases. This method could be
34 applicable for any monogenic diagnosis.

35

36 Introduction

37 Since 2015, several influential professional societies, including the International Society for
38 Prenatal Diagnosis (ISPD) and American College of Obstetricians and Gynecologists (ACOG),
39 have stated that noninvasive prenatal testing (NIPT) is an available screening option for all
40 pregnant women. Current NIPT is based on analysis cell-free fetal (cff) DNA, and it has become
41 widely available since its introduction to clinical practice in 2011. In contrast, cell-based NIPT,
42 which relies on the isolation of circulating fetal cells in maternal blood, has been a long-sought
43 alternative to cell-free NIPT and is now approaching commercialization. Currently, the cell-free
44 NIPT approach has the advantage of a faster turnaround time and lower cost. However, the
45 accuracy of cell-free NIPT is impacted by the large amount of maternal DNA in plasma (more
46 than 80% of all circulating DNA) and the highly fragmented nature of this genetic material. Thus,
47 it is only recommended for detection of the common fetal aneuploidies by many professional
48 societies [1, 2]. These drawbacks can be addressed by cell-based NIPT, but it is not yet available
49 as a clinical test. Cell-based NIPT would potentially have a higher positive predictive value
50 compared to cell-free NIPT, since the DNA source is purely fetal or placental in origin without
51 any maternal contamination. [3-5]. Recently, multiple groups have reported successful cases of
52 cell-based NIPT via capturing trophoblast cells [3, 6, 7].

53 The critical step for cell-based NIPT is the recovery of rare fetal cells, such as trophoblasts.
54 As described previously [3, 4], 30-40 mL of blood is collected at 10–16 weeks' gestation,
55 followed by density fractionation or magnetic activated cell sorting (MACS) with anti-
56 trophoblast antibodies to enrich the nucleated cells. Then, the nucleated cells are immunostained
57 to identify trophoblasts that are cytokeratin positive and leukocyte common antigen (CD45)
58 negative. The stained cells are picked individually under fluorescence microscopy with an

59 automatic instrument described previously [3, 4] and subjected to whole genome amplification
60 (WGA), which allows downstream genotyping, and copy number analysis using array
61 Comparative Genomic Hybridization (CGH) or next generation sequencing (NGS).

62 Genotyping is an essential step after isolating the putative fetal cells. Typically, a successful
63 cell-based NIPT would isolate 5-10 cells per 30 mL maternal blood sample. Since the nucleated
64 cell recovery is a complicated multiple-step procedure, and several antibodies are used, there is a
65 chance of picking a maternal cell (~10% in our experience). Whole genome shotgun (WGS)
66 sequencing at low coverage (5-10 million reads per cell) provides good copy number data, but it
67 does not readily distinguish fetal and maternal cells if the fetus is female. Previously, we used
68 short tandem repeat analysis, SNP arrays, or Y-chromosome targeted qPCR to confirm the fetal
69 origin of single cells. However, there are various disadvantages to these approaches, such as
70 inefficiency, ambiguity, high cost, or limited application.

71 In this work, we developed a fast, low-cost, and reliable genotyping assay with amplicon
72 sequencing. We sequenced approximately 90 highly polymorphic SNPs within about 40
73 amplicons. Among these amplicons, four contain multiple common SNPs (see Materials and
74 Methods), which allow for haplotyping of the WGA DNA product. Together, it allows for the
75 effective differentiation of cells containing the fetal genome from cells of maternal origin in most
76 cases. This method could also be easily expanded for the detection of additional disease-
77 associated variants, which would have clinical utility for pregnancies with increased risk for
78 monogenic disorders.

79 **Materials and methods**

80 **Sample collection and preparation**

81 Blood samples were collected from pregnant women from multiple centers under a protocol
82 approved by the Baylor College of Medicine Institutional Review Boards. Approximately 30 mL
83 of blood was collected into anticoagulant EDTA Vacutainer tubes (BD). Fetal cells were
84 enriched with methods described in Breman et al., 2016. Both cyokeratin (CK)-positive putative
85 fetal cells, and CK-negative maternal white blood cells were picked from maternal blood using
86 the CytePicker® equipment (Rarecyte).

87 **Overall strategy**

88 The typical cell-based NIPT workflow yields 3-10 singlet or doublet cells per patient,
89 individually captured in a PCR tube for downstream WGA. The WGA DNA products of those
90 cells must be checked for quality and confirmed as nonmaternal cells before finalizing the
91 interpretation. Thus, a fast, low-cost, and high-throughput genotyping assay is necessary. To
92 meet this need, we designed a single-cell genotyping assay using a modified amplicon
93 sequencing approach for genotyping (Fig 1). The first step is conventional PCR with a pool of
94 amplicons with bridging adaptors, which contains a partial sequence of Illumina i5 and i7
95 adaptors. The second step is adding a dual index with a sequencing adaptor to previous PCR
96 products. This concludes the library construction for the Illumina machine. DNA samples with
97 different indexes were balanced and pooled for sequencing with Illumina Miseq. The sequencing
98 result was demultiplexed with Illumina BaseSpace. The demultiplexed reads were mapped with
99 conventional BWA-MEM. The mapped reads were used for SNP typing and amplicon
100 haplotyping, the details of which are discussed below.

101 **Fig 1. The workflow for amplicon-based genotyping.**

102

103 **Amplicon design**

104 Three groups of amplicons were used to carry out amplicon-seq. The first group consisted of
105 amplicons with multiple common SNPs (>5% prevalence) as suggested in Debeljak et al. [8],
106 including regions in HLA-A, HLA-B, chromosome 7q11 and chromosome 11q22, which have
107 good sequencing coverage in single-cell WGA; all of these amplicons have at least eight
108 common SNPs. The information of these common SNPs can be used for effective haplotyping
109 and identifying the origin of isolated cells. The second group, which consisted of 37 amplicons,
110 contained common SNPs selected from the Human Identification panel (ABI), which covers
111 most chromosomes, including the Y chromosome. They were selected according to sequencing
112 coverage in the single-cell WGA product, which has a high tendency for dropout. The third
113 group consisted of amplicons designed to detect single nucleotide variants associated with
114 certain inherited disease genes of interest, including hemoglobin subunit beta (*HBB*),
115 hexosaminidase A (*HEXA*), and cystic fibrosis transmembrane conductance regulator (*CFTR*)
116 (Table 1). For studies of trophoblasts from specific at-risk cases, we also prepared amplicons for
117 *DHCR7* and *RASPN* (Table 1). The primers of these amplicons were prepared with adaptors
118 compatible for Illumina True-seq HT i5 and i7 adaptors.

119 **Table 1. List of SNPs**

Amplicon_ID	Chr	Amplicon_Start	Amplicon_Stop	Size	Annotation
rs1490413	chr1	4367241	4367415	175	
rs4847034	chr1	105717572	105717689	118	

rs3780962	chr10	17193291	17193386	96	
rs964681	chr10	132698373	132698467	95	
rs1498553	chr11	5708942	5709103	162	
rs901398	chr11	11096160	11096278	119	
rs2269355	chr12	6945833	6946005	173	
rs4530059	chr14	104769098	104769197	100	
rs2016276	chr15	24571747	24571845	99	
rs2342747	chr16	5868655	5868769	115	
rs2292972	chr17	80765753	80765827	75	
rs938283	chr17	77468418	77468592	175	
rs9905977	chr17	2919336	2919452	117	
rs1024116	chr18	75432299	75432467	169	
rs576261	chr19	39559774	39559848	75	
rs12997453	chr2	182413125	182413299	175	
rs1005533	chr20	39487029	39487201	173	
rs445251	chr20	15124851	15125009	159	
rs221956	chr21	43606946	43607048	103	
rs2830795	chr21	28608067	28608212	146	
rs733164	chr22	27816739	27816833	95	
rs1355366	chr3	190806053	190806219	167	
rs4364205	chr3	32417580	32417720	141	
rs6444724	chr3	193207331	193207425	95	
rs1979255	chr4	190318032	190318131	100	

rs159606	chr5	17374818	17374977	160	
rs338882	chr5	178690682	178690774	93	
rs7704770	chr5	159487871	159488033	163	
rs13218440	chr6	12059906	12060004	99	
rs6955448	chr7	4310289	4310423	135	
rs1360288	chr9	128967996	128968115	120	
rs1463729	chr9	126881396	126881493	98	
rs7041158	chr9	27985851	27986020	170	
P256	chrY	8685171	8685289	119	
rs17250845	chrY	8418867	8418960	94	
rs35284970	chrY	2734829	2734921	93	
rs3911	chrY	21733328	21733502	175	
HLA-A	chr6	2911156	2911140	245	Haplotype
HLA-B	chr6	31319491	31319646	156	Haplotype
Chr7q11	chr7	64895160	64895374	215	Haplotype
Chr11q22	chr11	99491336	99491527	192	Haplotype
rs334&rs33930165	chr11	5226980	5227053	74	HBB
rs11393960	chr7	117559481	117559645	165	CFTR
rs75527207&rs74597325	chr7	117587771	117587870	100	CFTR
rs121907954	chr15	72350490	72350617	128	HEXA
rs387906309	chr15	72346528	72346691	164	HEXA
rs147324677	chr15	72346182	72346293	112	HEXA
rs138659167	Chr11	71146795	71146921	127	DHCR7

rs104894299&rs761584017	Chr11	47469333	47469720	388	RASPN
-------------------------	-------	----------	----------	-----	-------

120

121 **Library construction for Illumina**

122 WGA was performed using the PicoPLEX kits (Rubicon/Takara) or, in some cases, the
123 Ampli1 WGA kit (Silicon Biosystems). according to the manufacturer's protocol. Briefly, 100ng
124 of WGA DNA product was used for Amplicon-seq library construction, with two-step PCR. The
125 first step included 20-25 cycles of PCR with NEB Q5HS to amplify amplicons of interest and
126 add designated adaptors. The second step used a previous adaptor sequence as primer binding
127 sites to add Illumina i5 and i7 dual-index adaptors with 10-15 cycles of PCR. The 2-step PCR
128 products were purified with standard AMPure protocol (Beckman) for the 100-300 bp product.
129 PCR products were visualized by gel electrophoresis followed by the Bioanalyzer (Agilent) to
130 check the quality and then quantified by using a KAPA Library Quantification Kit (Kapa
131 Biosystems).

132 **Sequencing with Miseq**

133 The barcoded DNA library was diluted to 2 nM and pooled together for denaturing with the
134 Illumina protocol. An 8-10 pM diluted denature library was mixed with 5-10% PhiX control. The
135 mixed library was loaded onto the Illumina Miseq with 150-cycle v3 kits (Illumina) and
136 sequenced with 2x76 reads and dual index.

137 **Demultiplexing, alignment and variant calling**

138 Sequencing results were demultiplexed by Illumina BaseSpace. The reads were mapped and
139 processed with a shell script (S3 File). In summary, the fastq.gz raw files were aligned with
140 customized reference files (S1 File) by BWA-MEM[9]. Samtools [10, 11] and Bam-readcount

141 follow with a customized R script (S4 File) were used for calling variants within selected
142 intervals (S2 File). The cutoff for calling a variant is at least ten reads of the less frequent allele
143 and 5% of all reads.

144 Variants from each sample were summarized and compared with their paired control with an
145 R script (S6 File). Typically, a cell with more than 2 SNPs and more than 10% of comparable
146 SNPs different from its maternal gDNA control is considered as a likely fetal cell. Otherwise, it
147 will be classified as an uninformative cell.

148 **Haplotype calling**

149 Using the Illumina 2x76 read length necessitated performing stitch overlap for read 1 and
150 read 2 with PEAR (Paired-end read merger) [12] for amplicons shorter than 150 bp. For
151 amplicons larger than 150 bp, we stitched non-overlapping reads with a 15-N padding sequence.
152 Those stitched reads were mapped with BWA-MEM with modified parameters that allow a
153 bigger unmatched gap inside a mapped read. The mapped results were processed with an R script
154 (S5 File) used in the main shell script to extract SNPs and reconstitute a new sequence with
155 CIGAR information. The new concise sequences were tabulated and grouped with sequence
156 similarity according to their Levenshtein distance. We assigned a haplotype to each major group
157 of a concise sequence. Pair-wise haplotype comparison between the maternal gDNA sample and
158 the putative fetal cells were performed with another R script (S7 File). All scripts are hosted and
159 maintained on https://github.com/xmzhuo/NIPT_genotyping.

160

161 Results

162 Target region coverage

163 We compared the coverage of our amplicons for SNPs with gDNA and NIPT cell WGA
164 products (Fig 2). For gDNA, most of the samples have very high coverage, which reflects the
165 distribution of samples with many scorable SNPs. The WGA product of maternal white blood
166 cells (WBCs) shows less scorable SNPs than gDNA, which would be the result of starting with a
167 single diploid genome target in combination with fixation, staining, and amplification during
168 WGA.

169 **Fig 2. Comparison of the coverage of several genomic DNAs and NIPT single cells.** The
170 count of samples with various scorable SNPs was normalized to the total number of samples of
171 each group. (gDNA n=20, WGA n=35). The scorable SNP cutoff is at least 5% minor allele
172 frequency (MAF) and ten reads. Data include NIPT case numbers 946, 977, 982, 983, 984, 988,
173 989, 990, 991, 992, 993, 996, and 998 (clinical data in S8_File table).

174 SNP typing of NIPT WGA products

175 A typical process of SNP typing after BWA-MEM aligner mapping includes variant calling
176 with Samtools and retrieving the read depth with bam-readcount (Fig 3). This step will produce a
177 table containing variant call information of allele fraction and read depth of all SNPs of interest.
178 Information on indels was masked to avoid confusion in later steps. Then, we performed the
179 pair-wise comparison of WGA products with the maternal gDNA. The script will calculate how
180 many SNPs are different between two DNA samples. In an example case with two cells (Fig 3),
181 three SNPs show the difference between the cell (middle panel) and maternal gDNA, which

182 suggests that it is likely to be a fetal cell. The other cell shows identical calls with maternal
183 gDNA, which suggests that it may not be a fetal cell.

184 **Fig 3. The workflow and an example of SNP typing.** Reads for the mutant allele are colored
185 green while reads for the normal allele are colored grey and summed in red. The mother is
186 homozygous for all three SNPs, while cell A is heterozygous for all three SNPs in each case
187 having an allele that the mother does not have. Cell B is homozygous for all SNPs
188 indistinguishable from the mother and is interpreted as maternal or noninformative (fetal with
189 allele drop out for all three SNPs). Data from NIPT case number 1000 (clinical data in S8_File
190 table).

191 We have compared DNA from 156 blood samples, which include gDNA and DNA from
192 WBCs and fetal cells (confirmed by orthogonal methods, such as gender PCR or low coverage
193 WGS), to study the sensitivity and specificity of SNP typing in identifying alleles present in the
194 cell that were not present in the mother and thus being real fetal cells. We used the WBC WGA
195 product from the matched maternal blood as a true negative (Table 2). Typically, a SNP
196 difference of more than 6% can largely avoid a false call in our study without missing real fetal
197 cells. In our application, we used a SNP difference of at least two SNPs and 10% of scorable
198 SNPs for our calling. To date, we found that 68.9% genotyped putative cells from 156 cases are
199 informative (357/518).

200 **Table 2. True positive and False positive value with various differential SNP percentage**

	Maternal cell	female gDNA	Fetal w/o sex	Fetal cell	WBC from different origin
--	---------------	-------------	---------------	------------	---------------------------------

diff SNP %	FPV	TPV	TPV	TPV	TPV
0.01	0.11428571	0.99358974	1	1	0.72435897
0.02	0.11428571	0.99358974	1	1	0.72435897
0.03	0.11428571	0.98076923	1	1	0.72435897
0.04	0.11428571	0.94230769	1	1	0.71153846
0.05	0.11428571	0.91025641	0.85714286	1	0.68589744
0.06	0.08571429	0.86538462	0.85714286	1	0.66025641
0.07	0.08571429	0.82692308	0.85714286	1	0.63461539
0.08	0.08571429	0.80128205	0.71428571	1	0.62820513
0.09	0.08571429	0.75641026	0.71428571	0.85714286	0.60897436
0.1	0.08571429	0.73717949	0.57142857	0.71428571	0.58974359
0.11	0.08571429	0.70512821	0.57142857	0.71428571	0.53205128
0.12	0.05714286	0.64102564	0.42857143	0.71428571	0.51923077
0.13	0.02857143	0.62179487	0.42857143	0.71428571	0.5
0.14	0	0.55769231	0.28571429	0.71428571	0.48076923
0.15	0	0.53205128	0.28571429	0.57142857	0.41025641
0.16	0	0.47435897	0.14285714	0.42857143	0.37820513
0.17	0	0.41025641	0.14285714	0.28571429	0.36538462
0.18	0	0.39102564	0.14285714	0.28571429	0.33974359
0.19	0	0.35897436	0.14285714	0.14285714	0.28205128
0.2	0	0.32692308	0.14285714	0.14285714	0.23717949

202 In NIPT #1000 (Fig 4), we isolated eight putative fetal cells from the mother's blood for
203 WGA. Cells G78, G212, G 227, and G232 all have at least four SNPs which differed from the
204 maternal gDNA; thus, they were confirmed to be fetal in origin. The remaining four cells (G79,
205 G113 and G320) had 0-2 SNPs which differed from the maternal gDNA, and these were
206 considered uninformative and possibly white blood cells accidentally isolated from the mother's
207 blood. Cell G106 has only 2 SNP difference in less than 20 informative SNPs, which was
208 considered as a low-quality sample and uninformative.

209 **Fig 4. Example of SNP typing result of one NIPT case.** The X-axis indicates the number of
210 informative SNPs per sample. The Y-axis is the number of SNPs with non-maternal alleles in a
211 cell. As expected, the maternal gDNA samples gave no alleles not present in the mother. Data
212 from NIPT case number 1000.

213

214 **Haplotyping of NIPT WGA products**

215 The haplotyping of NIPT WGA products potentially has higher power than SNP typing for
216 identifying fetal cells. Since the WGA product typically went through the 14-16 cycles of
217 amplification from trace amounts of input DNA, there is a small chance of introducing new
218 mutations, which would affect the precision of SNP typing at low read-depth. For example, some
219 of the cells from case #1000 (Fig 4) had one low-depth SNP difference from the maternal gDNA.
220 To address this issue, we developed a haplotyping approach for multiple highly polymorphic
221 regions, which contain multiple very common SNPs within the 200 bp amplicon. Thus, we can
222 decrease the impact of random mutations introduced during the WGA process on the final
223 interpretation (i.e., one nucleotide change is less likely to change the classification of a major

224 haplotype group, which is comparable to HLA typing approach). Mutations introducing a
225 random change are not rare, but mutations switching from one allele at a SNP to the other
226 polymorphic allele are much rarer. In addition, the haplotypes of an amplicon can be treated as a
227 permutation of a given number of SNPs, which theoretically generates much more haplotypes
228 than SNP types and has higher power at differentiating two cells. Third, we can estimate the
229 point at which a sequence artifact arose based on the fraction of each minor haplotype group in
230 the total reads for a given amplicon. For example, a high fraction indicates a variant preexisted in
231 the cell, a medium fraction indicates the variant arose during the WGA step, while a low fraction
232 is consistent with an artifact from the final step of amplicon-seq.

233 We performed haplotyping with the following steps (Fig 5). After regular alignment with
234 BWA-MEM with the default setting, we joined the Read 1 and Read 2 with PEAR [12]. The
235 overlapping Read 1 and Read 2 were merged. If the amplicon is longer than two reads joined
236 together, we merged the two reads and padded the gap with a tandem repeat of N. The merged
237 reads were remapped with BWA-MEM again with a lenient setting to tolerate a larger gap. The
238 remapped reads were processed with an R script to extract the selected SNPs, and each read was
239 reconstituted with the concise sequence while preserving the read ID. The concise reads were
240 then tallied and ranked according to frequency (typically, only the top 10 were kept, which
241 usually consist of more than 99.99% of all types of reads). The Levenshtein distances were
242 calculated for these reads, which typically ended up with only one or two major groups to
243 represent the haplotype of this amplicon. To compare the haplotypes of more than two samples,
244 all the top reads of each sample were pooled together, and their distances were calculated, which
245 will determine if these samples share the same read group (haplotype). For example, the maternal
246 gDNA carried a mocked haplotype 1 (TA) and haplotype 2 (GC). A positive fetal cell should be

247 identified to carry at least one new haplotype 3 (GA), which would be TA/GA or GA/GA (result
248 from dropout of haplotype 1 or 2) (Fig 5).

249 **Fig 5. The workflow for haplotyping.** The paired aligned reads were jointed with PEAR and
250 padded with BMAP short read aligner when a gap exists. The new joint reads were remapped
251 with BWA-MEM with a modified configuration. The highly polymorphic sites were extracted
252 for constructing concise haplotypes. To demonstrate the workflow, we present example short
253 reads with TA and GC haplotypes.

254

255 The haplotyping approach can effectively differentiate a candidate cell from maternal gDNA.
256 In the case shown in Fig 6, we have the gDNA from both parents. As described previously, we
257 extracted all 28 SNP sites in the HLA-A amplicon and reconstructed a concise 28 nt sequence for
258 each read. In this case, the top four most frequent read types of potential fetal cells can be
259 grouped into two major groups, with a Levenshtein distance of more than 2 (S1 Fig). The intra-
260 group difference has a distance of less than 0.5, which suggests a difference of only one
261 nucleotide. The difference likely results from artifacts introduced during extensive amplification
262 (WGA then PCR). The same condition was observed in maternal gDNA and paternal gDNA as
263 well. We observed the inheritance pattern of haplotypes when all read types from maternal,
264 paternal, and fetal DNA were plotted together. One fetal haplotype matched with the mother and
265 the other matched with the father. From these haplotype groupings, we concluded that this is a
266 true fetal cell.

267 **Fig 6. Haplotype analysis of a fetal cell compared to both parents.** Each vertical column
268 represents a particular SNP in the haplotype. The left-most four SNPs are informative as the fetal
269 cell has an allele that the mother does not have and indicate that the putative fetal cell is indeed
270 fetal. The 24 SNPs to the right are not informative for the fetal cell as they do not have an allele
271 that the mother does not have. Data from NIPT case number 1000.

272

273 We tested the performance of haplotyping in four amplicons with matched gDNA, WBC, and
274 fetal cells (S2 Fig). For gDNA, all four amplicons performed nicely to distinguish one from the
275 other. For WBC and fetal cells, the performance was not as good, largely owing to the dropout
276 events. However, when four amplicons were combined, we can still distinguish about 50% of all
277 the cells. With combined power of both SNP typing and haplotyping, we can increase the solving
278 rate of differentiating a WBC from around 60% to more than 70% (S3 Fig).

279 **Genotyping for monogenic disease mutations**

280 We also wished to use this method to genotype for monogenic disease mutations. We first
281 evaluated the ability to detect disease variants in single cultured lymphoblasts of known
282 genotypes. Our cell based NIPT provides pure fetal DNA, which allows us to look at monogenic
283 disease mutations. Here, we developed amplicons that contain HEXA c.805C>T, HBB c.19T>A,
284 HBB c.20C>T, and CFTR c.1521_1523delTCT. Corresponding cells carrying certain mutations
285 were obtained from the Coriell Institute and single lymphoblasts were picked from tissue culture
286 and processed for WGA. The cells were isolated and genotyped by methods described herein.
287 We successfully detected the known variants in the WGA products of cells carrying these

288 mutations (Fig 7). The allele dropout rate was 15% and 8% for unfixed and fixed lymphoblasts,
289 respectively.

290 **Fig 7. Testing for point mutations in single lymphoblasts.** From left to right, Tay-Sachs,
291 HEXA c.805C>T het; HBB c.19T>A and c.20C>T compound het; CFTR c.1521_1523delTCT
292 het. The reads are shown in the IGV browser.

293

294 In all of our cell-based NIPT samples studies, there were three families with known
295 monogenic mutations. One was a family where the mother was affected with sickle cell anemia,
296 and the fetus was expected to be heterozygous based on parental information. No amniocentesis
297 or CVS was performed. In a second family, both parents were carriers with the same known
298 pathogenic mutation in *DHCR7*. By amniocentesis, the fetus was heterozygous for the mutation
299 carried by both parents. A third family had a previous affected child with congenital myasthenic
300 syndrome type 11 caused by biallelic, compound heterozygous mutations in *RAPSN*. By CVS,
301 the fetus carried the paternal but not the maternal mutation. For the sickle cell family, ten fetal
302 cells were recovered and five were genotyped. Two cells were heterozygous for the mutation
303 (Fig. 8), while one cell had dropout for the normal allele and one cell had dropout for the mutant
304 allele. For the *DHCR7* family, seven fetal cells were recovered and four were genotyped. One
305 cell was heterozygous for the mutation (Fig. 8), while two cells had dropout for the normal allele
306 and one cell had dropout for the mutant allele. For the *RAPSN* family, four fetal cells were
307 recovered and four were genotyped. In Fig 8C, the mutation in each parent is shown. The G540
308 cell in Fig 8C, shows absence of the maternal mutation but presence of the paternal mutation.
309 All four cells showed absence of the maternal mutation. For the paternal mutation, two cells
310 were heterozygous while one cell had dropout for the normal allele and one cell had dropout for

311 the paternal mutant allele. There is a small probability that the fetus carries the maternal
312 mutation, but there was dropout in all four cells; more likely the fetus does not carry the maternal
313 mutation in agreement with the CVS data. The fetus does carry the paternal mutation.

314 **Fig 8. Genotyping for point mutations in trophoblasts from three cases.** In panel A, the
315 mother is affected and homozygous for the sickle mutation. Fetal trophoblast G54 is
316 heterozygous for the mutation. Reads for the mutant allele are colored green while reads for the
317 normal allele are colored grey and summed in red. In panel B, the mother is heterozygous for a
318 *DHCR7* mutation that is also present in the father. Fetal trophoblast G1286 is also heterozygous
319 for the mutation, although there is biased over-representation of the mutant allele. Reads for the
320 mutant allele are colored blue while reads for the normal allele are colored grey and summed in
321 brown. In panel C, the mother is heterozygous for the pathogenic N88K variant in the *RAPSN*
322 gene, and the father is heterozygous for the V165M variant. Fetal trophoblast G540 is
323 heterozygous for the paternal V165K variant but not for the maternal N88K variant. Allele drop
324 out for the N88K variant cannot be ruled out, and multiple cells must be tested to gain statistical
325 evidence that the fetus has not inherited the N88K variant. All results agreed with data from
326 amniocentesis or CVS. Data from NIPT case numbers 1180, 1492, and 1607.

327 Discussion

328 This single cell genotyping assay can provide an essential step for confirming the fetal origin
329 of cells obtained from cell-based NIPT workflows. Through genotyping, we can reject cells that
330 are of indeterminate or maternal origin and provide metrics for WGA DNA quality using
331 multiple amplicons. Thus, researchers can focus on smaller numbers of cells, which can then be
332 used for a more expensive downstream test or analysis, such as the low-depth NGS and
333 microarray for CNV analysis [3, 4, 6]. We can add more amplicons to cover mutations of interest
334 for detecting recessive or dominant inherited diseases.

335 Although this assay is rapid and reliable, there are still some limitations. First, the dropout
336 rate for individual amplicons is significant, which hinders the power of cell identification and
337 affects the detection of disease-causing mutations. Second, although the cost is relatively low
338 and the running time is short (<24 h), it still takes extra effort to complete and could increase the
339 turnaround time of NIPT. Third, the multiple steps of PCR after WGA are prone to introduce
340 errors that may cause ambiguity at SNP typing, although errors switching from one SNP allele to
341 another or from a mutant to wild-type genotype or vice versa are very rare.

342 There are multiple options for reducing the dropout rate for the improvement of this assay.
343 First, we can increase the size of the panel, which allows more amplicons to compensate for the
344 dropout. Second, we can modify the amplicons to reduce the size of the amplicons or otherwise
345 improve the amplification. Third, improved versions of WGA are being developed which can
346 reduce allele dropout.

347 To minimize the errors introduced by the multiple amplification steps, we can incorporate the
348 genotyping amplification into WGA. We can redesign our genotyping primers to make them

349 compatible with the linear amplification step in WGA, which will result in the enrichment of our
350 regions of interest. The modified WGA product can be used for barcoding directly for low-depth
351 WGS. This approach may eliminate 5-10 cycles of PCR and reduce the chance of dropout.

352 Use of this method for genotyping fetuses at risk for specific monogenic mutations is feasible.
353 Although cell-free NIPT is relatively straight forward for genotyping paternal mutations, it is
354 more complex for determining the maternal contribution to the genotype, although this can be
355 accomplished with more complex analysis [14, 15]. Cell-free NIPT has been used to screen for
356 *de novo* mutations in a panel of genes [13]. Confirmation for *de novo* mutations can be
357 performed by specific reanalysis of maternal plasma if the mother is not mosaic for the mutation.
358 Single cell analysis of circulating trophoblasts can be used to determine the genotype of fetuses
359 at risk for monogenic disorders, although allele dropout must be ruled out by analysis of multiple
360 cells, and a low failure of diagnosis rate is to be expected due to failure to isolate fetal cells
361 and/or to allele drop out for cells isolated.

362 Our genotyping assay has the potential to be used in many clinical research applications. As
363 described previously, it can be used to identify a fetal cell in cell-based NIPT and screen for
364 known disease-causing point mutations. It also has the potential to distinguish between same sex
365 dizygotic twins. Furthermore, this technique would also be adapted for other single cell
366 applications, such as circulating tumor cell analysis. An improved version of this assay may
367 significantly reduce the dropout rate in amplicons and increase the coverage at regions of interest.

368 **Acknowledgment**

369 This work was supported by internal institutional funds at Baylor College of Medicine.

370 We thank the participating patients, study coordinators, and genetic counselors at Baylor
371 College of Medicine, Texas Children's Hospital, and Columbia University Medical Center for
372 recruitment of samples.

373

374 **Conflict of Interest**

375 The Houston authors are faculty and staff at Baylor College of Medicine (BCM), which is a
376 partial owner of a for profit diagnostic company, Baylor Genetics (BG); Houston authors also
377 are employee of or have advisory or lab director roles at BG. ALB is founder and CEO of Luna
378 Genetics, Inc.

379

380 Reference

- 381 1. Committee Opinion No. 640: Cell-Free DNA Screening For Fetal Aneuploidy. *Obstet Gynecol.*
382 2015;126(3):e31-7. doi: 10.1097/AOG.0000000000001051. PubMed PMID: 26287791.
- 383 2. Benn P, Borrell A, Chiu RW, Cuckle H, Dugoff L, Faas B, et al. Position statement from the
384 Chromosome Abnormality Screening Committee on behalf of the Board of the International Society for
385 Prenatal Diagnosis. *Prenat Diagn.* 2015;35(8):725-34. doi: 10.1002/pd.4608. PubMed PMID: 25970088.
- 386 3. Breman AM, Chow JC, U'Ren L, Normand EA, Qdaisat S, Zhao L, et al. Evidence for feasibility of
387 fetal trophoblastic cell-based noninvasive prenatal testing. *Prenat Diagn.* 2016;36(11):1009-19. doi:
388 10.1002/pd.4924. PubMed PMID: 27616633; PubMed Central PMCID: PMC5129580.
- 389 4. Vossaert L, Wang Q, Salman R, Zhuo X, Qu C, Henke D, et al. Reliable detection of
390 subchromosomal deletions and duplications using cell-based noninvasive prenatal testing. *Prenat Diagn.*
391 2018;38(13):1069-78. Epub 2018/11/19. doi: 10.1002/pd.5377. PubMed PMID: 30357877; PubMed
392 Central PMCID: PMC6587831.
- 393 5. Vossaert L, Wang Q, Salman R, McCombs AK, Patel V, Qu C, et al. Validation Studies for Single
394 Circulating Trophoblast Genetic Testing as a Form of Noninvasive Prenatal Diagnosis. *Am J Hum Genet.*
395 2019;105(6):1262-73. Epub 2019/11/27. doi: 10.1016/j.ajhg.2019.11.004. PubMed PMID: 31785788;
396 PubMed Central PMCID: PMC6904821.
- 397 6. Kolvraa S, Singh R, Normand EA, Qdaisat S, van den Veyver IB, Jackson L, et al. Genome-wide
398 copy number analysis on DNA from fetal cells isolated from the blood of pregnant women. *Prenat Diagn.*
399 2016;36(12):1127-34. doi: 10.1002/pd.4948. PubMed PMID: 27761919.
- 400 7. Vestergaard EM, Singh R, Schelde P, Hatt L, Ravn K, Christensen R, et al. On the road to replacing
401 invasive testing with cell-based NIPT: Five clinical cases with aneuploidies, microduplication, unbalanced
402 structural rearrangement, or mosaicism. *Prenat Diagn.* 2017;37(11):1120-4. doi: 10.1002/pd.5150.
403 PubMed PMID: 28881392.

- 404 8. Debeljak M, Freed DN, Welch JA, Haley L, Beierl K, Iglehart BS, et al. Haplotype counting by next-
405 generation sequencing for ultrasensitive human DNA detection. *J Mol Diagn*. 2014;16(5):495-503. doi:
406 10.1016/j.jmoldx.2014.04.003. PubMed PMID: 25132481; PubMed Central PMCID: PMCPMC5691345.
- 407 9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.
408 *Bioinformatics*. 2009;25(14):1754-60. doi: 10.1093/bioinformatics/btp324. PubMed PMID: 19451168;
409 PubMed Central PMCID: PMCPMC2705234.
- 410 10. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and
411 population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-93.
412 doi: 10.1093/bioinformatics/btr509. PubMed PMID: 21903627; PubMed Central PMCID:
413 PMCPMC3198575.
- 414 11. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map
415 format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9. doi: 10.1093/bioinformatics/btp352.
416 PubMed PMID: 19505943; PubMed Central PMCID: PMCPMC2723002.
- 417 12. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd
418 mergeR. *Bioinformatics*. 2014;30(5):614-20. doi: 10.1093/bioinformatics/btt593. PubMed PMID:
419 24142950; PubMed Central PMCID: PMCPMC3933873.
- 420 13. Zhang J, Li J, Saucier JB, Feng Y, Jiang Y, Sinson J, et al. Non-invasive prenatal sequencing for
421 multiple Mendelian monogenic disorders using circulating cell-free fetal DNA. *Nat Med*. 2019;25(3):439-
422 47. doi: 10.1038/s41591-018-0334-x. PubMed PMID: 30692697.
- 423 14. Chang MY, Ahn S, Kim MY, Han JH, Park HR, Seo HK, et al. One-step noninvasive prenatal testing
424 (NIPT) for autosomal recessive homozygous point mutations using digital PCR. *Sci Rep*. 2018;8(1):2877.
425 Epub 2018/02/13. doi: 10.1038/s41598-018-21236-w. PubMed PMID: 29440752; PubMed Central
426 PMCID: PMCPMC5811538.

427 15. Hudecova I, Chiu RW. Non-invasive prenatal diagnosis of thalasseмии using maternal plasma
428 cell free DNA. Best Pract Res Clin Obstet Gynaecol. 2017;39:63-73. Epub 2016/10/26. doi:
429 10.1016/j.bpobgyn.2016.10.016. PubMed PMID: 27887921.

430

431 **Supplement Data**

432 **S1_File Reference sequence fasta file**

433 **S2_File SNVs reference BED**

434 **S3_File Linux shell script and parameters for alignment and analysis**

435 **S4_File R script for genotype calling**

436 **S5_File R script for haplotype calling**

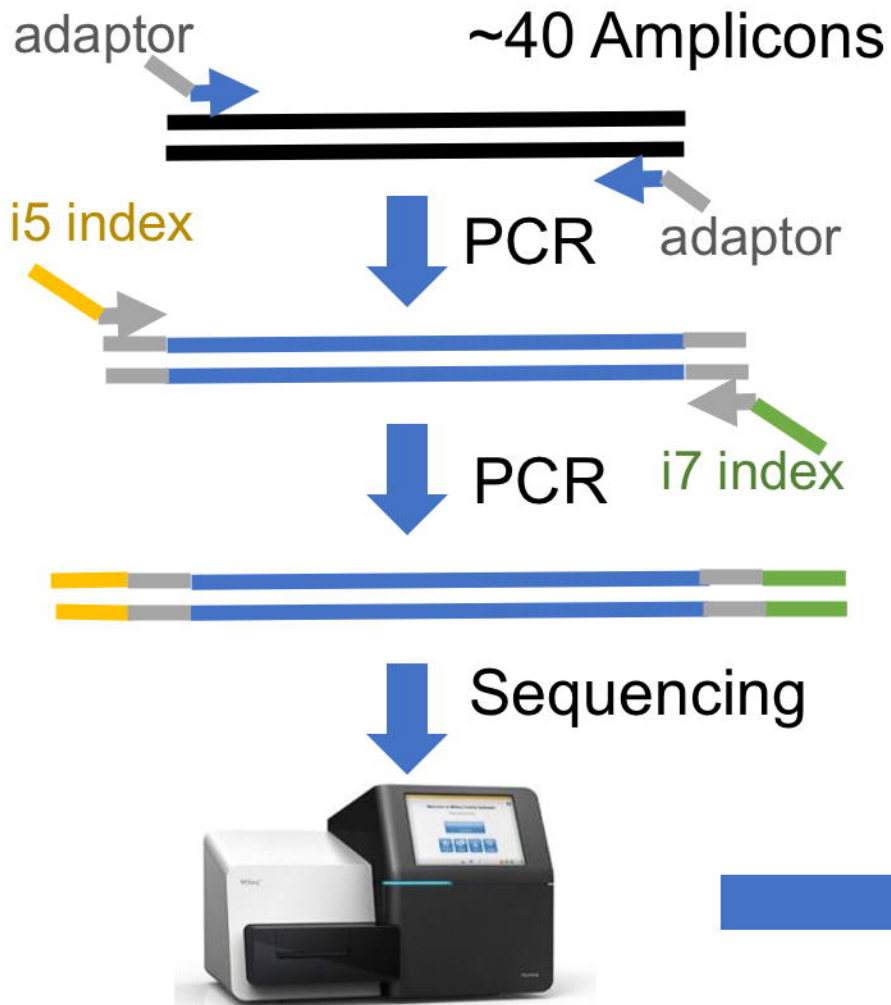
437 **S6_File R script to compare genotype of NIPT cells with parents**

438 **S7_File R script to compare haplotype of NIPT cells with parents**

439 **S8_File Table of Clinical Sample Information**

440

WGA products from single fetal cells

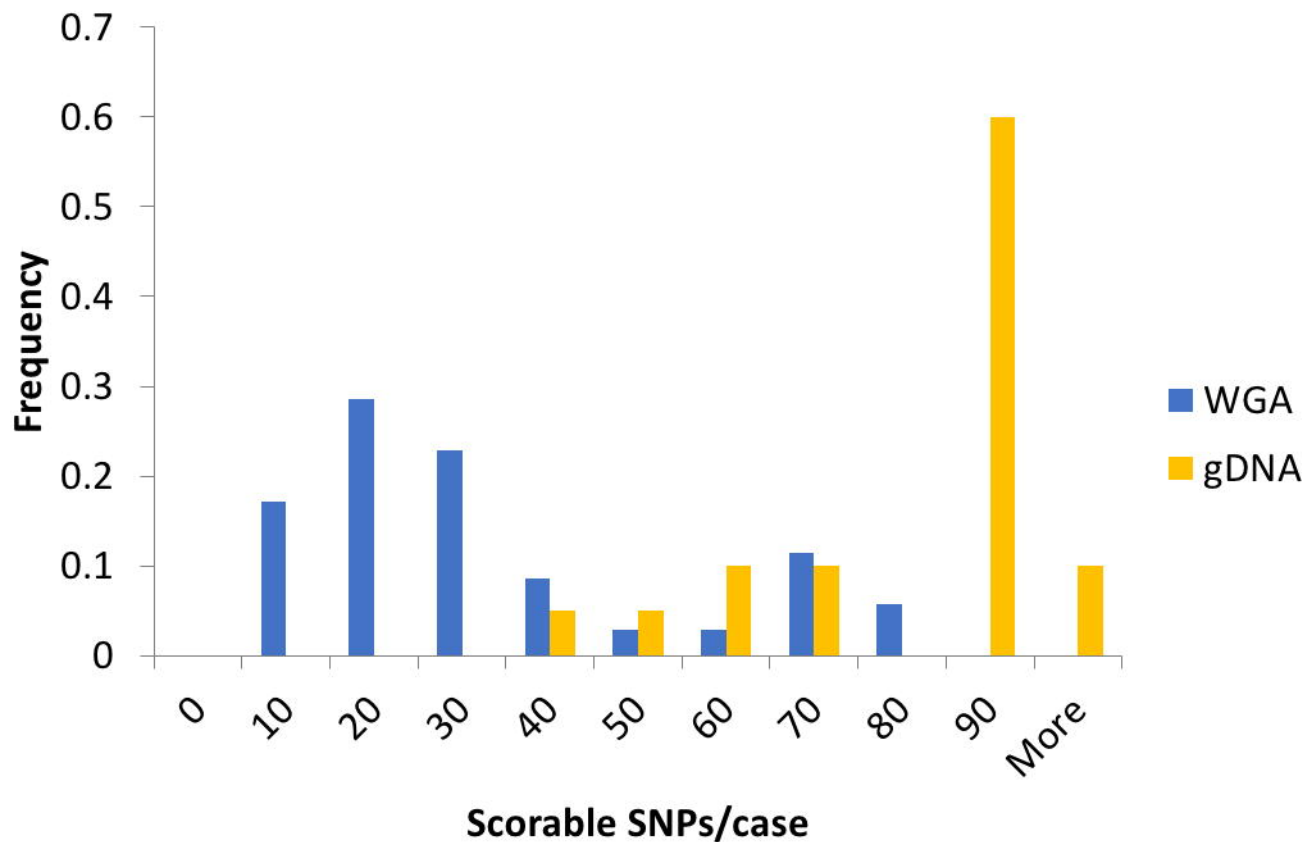


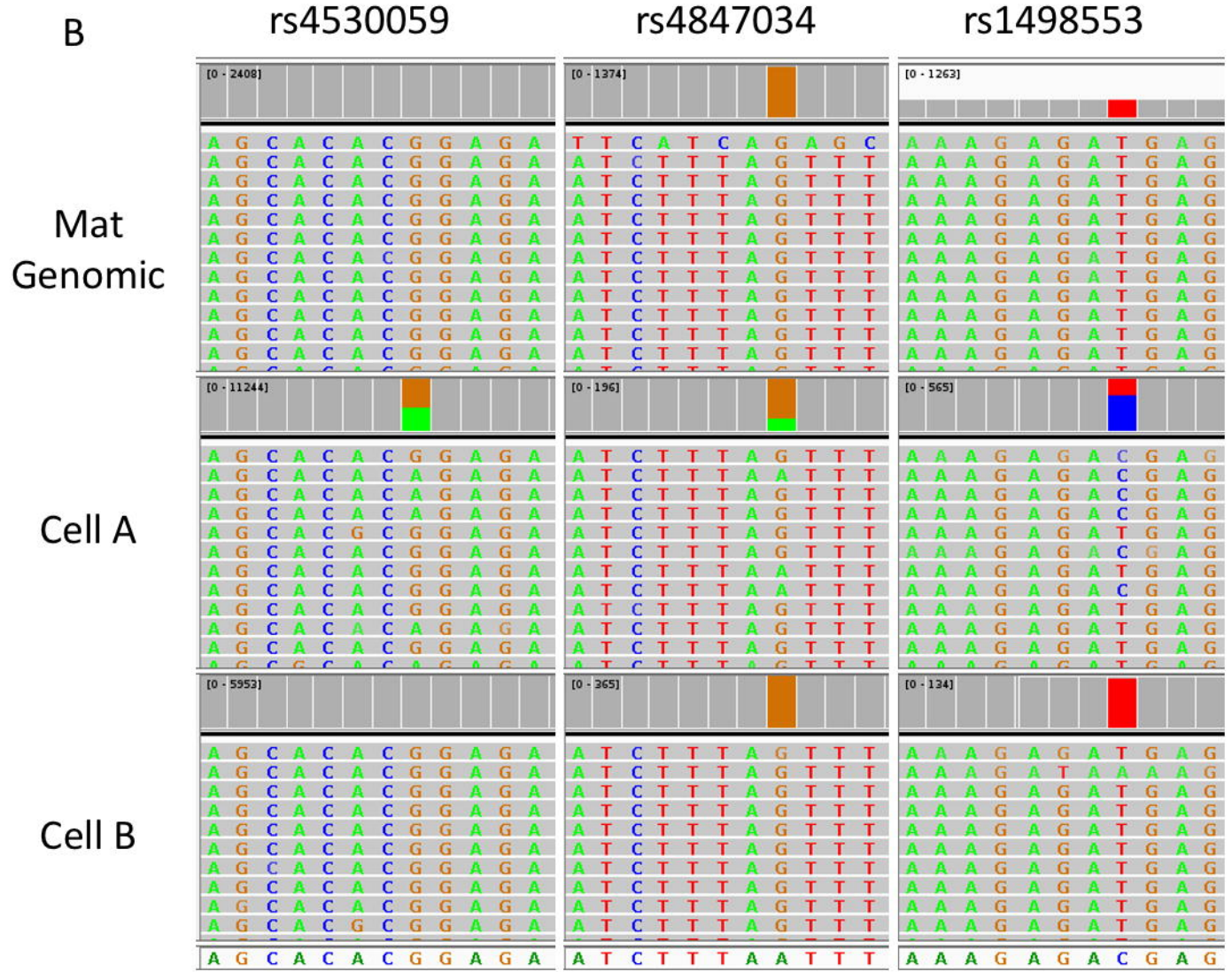
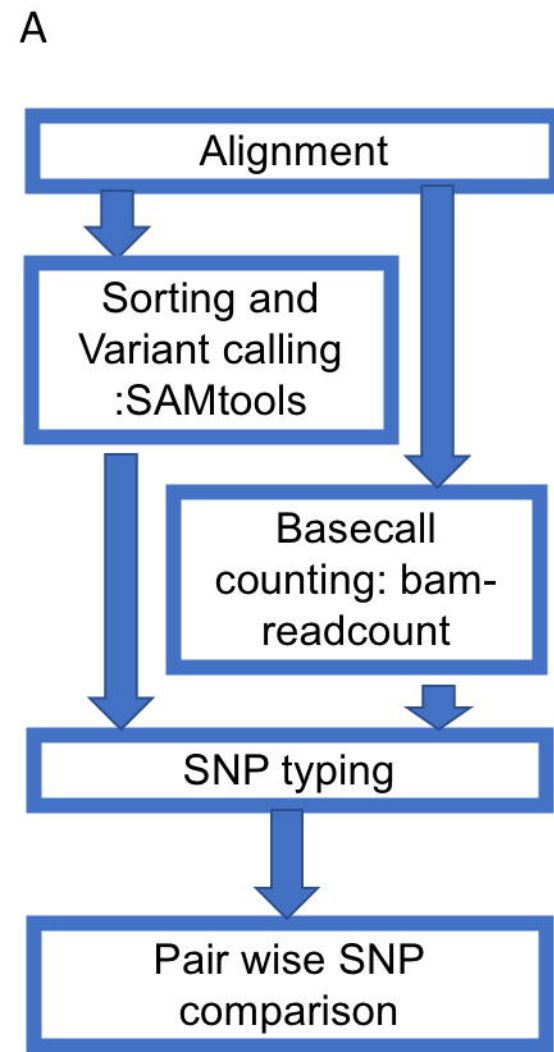
Alignment: BWA-MEM

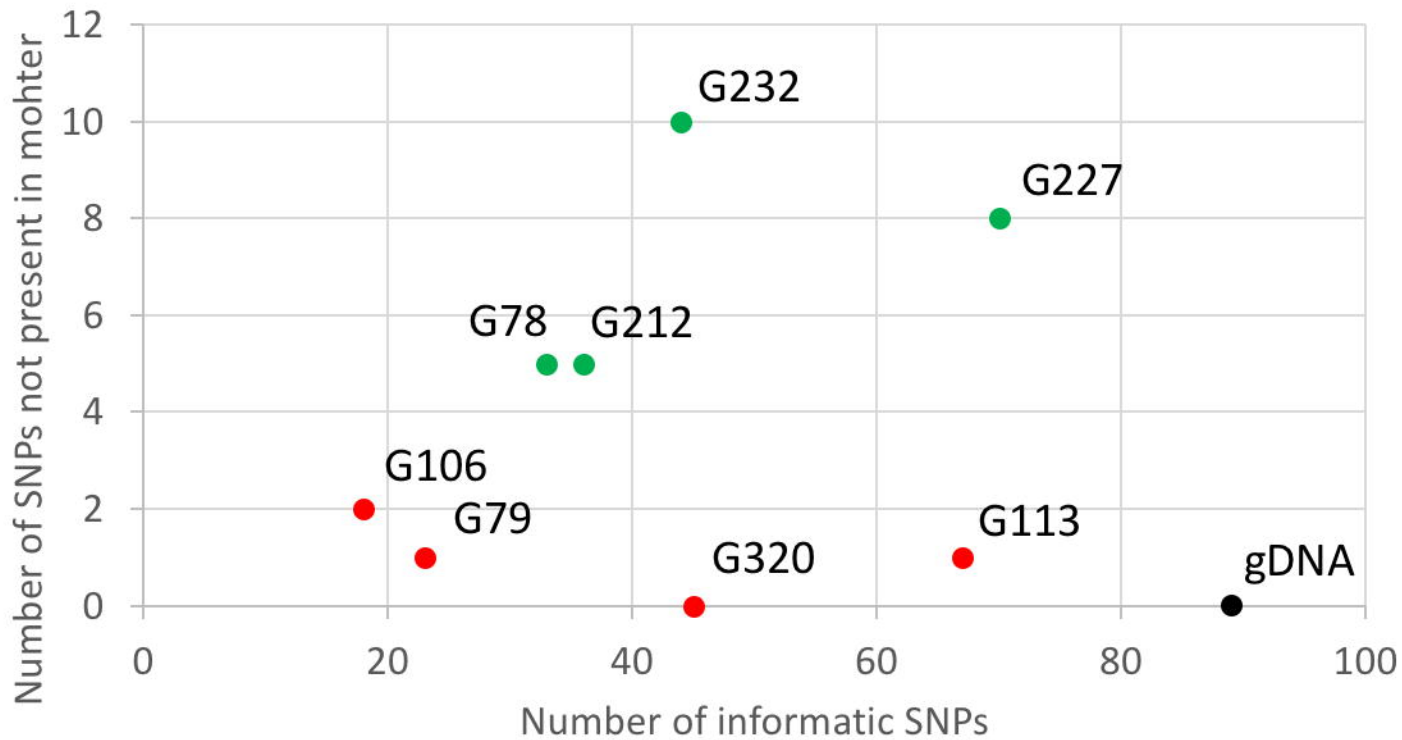
SNP typing

Haplotyping

Genotyping Analysis







● Control ● Fetal ● Mat/uninf

Alignment

Join paired reads

Merge overlapping reads:
PEAR

Padding non-overlapping reads:
BMAP

```
ATATGC
  GCCAAG
>ATATGCCAAG
```

```
CATAT
      CAAGT
>CATATNNCAAGT
```

```
AGATGC
  GCCCAG
>ATATGCCCCAG
```

```
CAGAT
      CCAGT
>CAGATNNCCAGT
```

Merge reads and remapping

Extracting SNP from reads to make concise reads

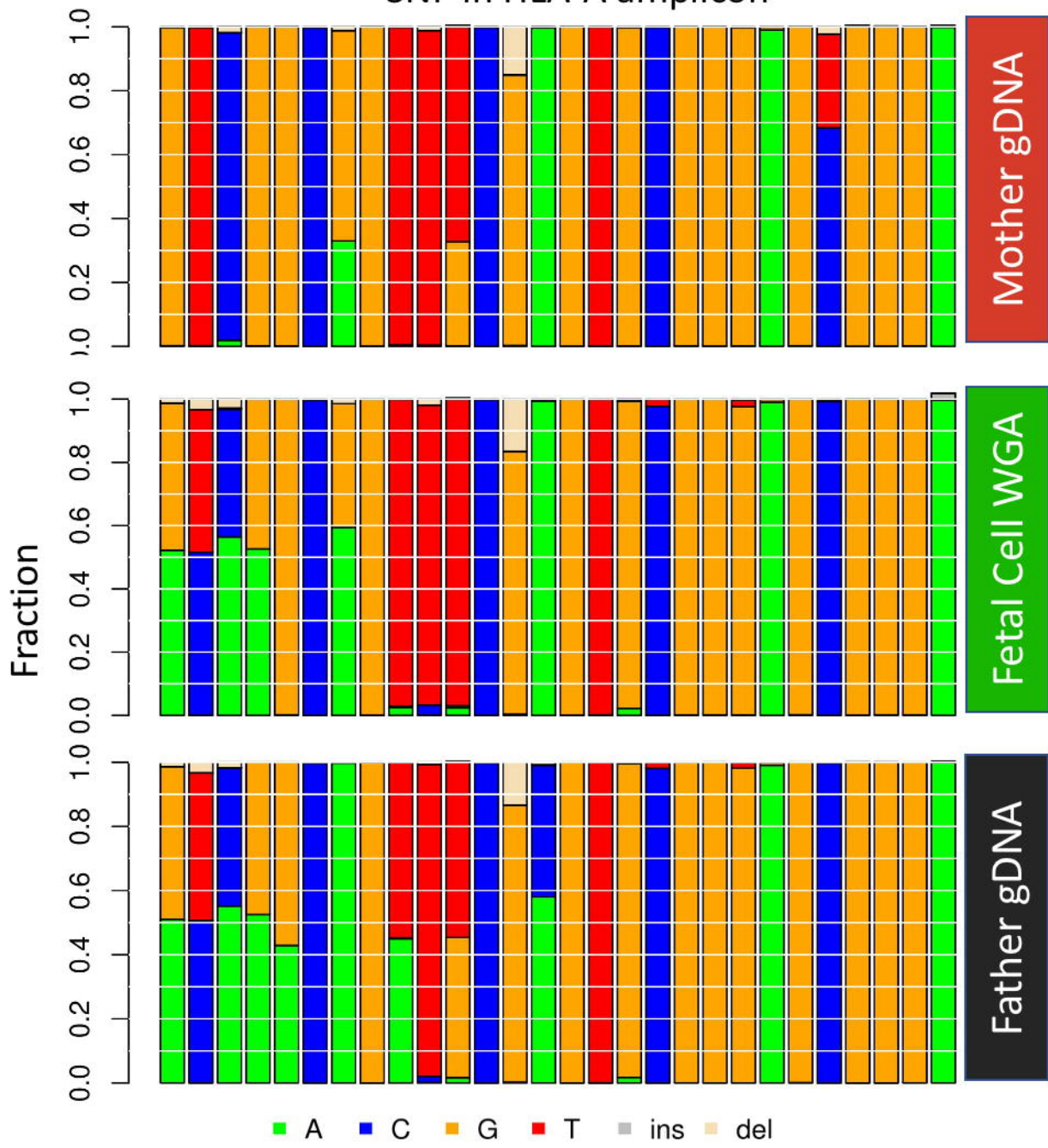
```
ATATGCCAAG    CATATNNCAAGT
> T      A      > T      A
```

```
AGATGCCCAG    CAGATNNCCAGT
> G      C      > G      C
```

Haplotyping and pairwise comparison

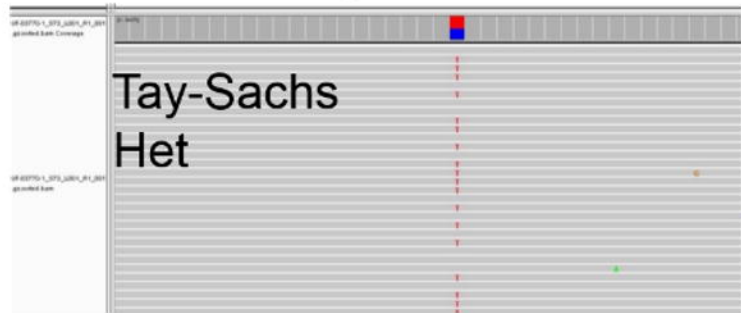
Mother gDNA	Candidate Cell	Call
TA/GC	TA/GC	Mat/Uninf
	TA/TA	Mat/Uninf
	GA/GA	Fetal
	TA/GA	Fetal

SNP in HLA-A amplicon



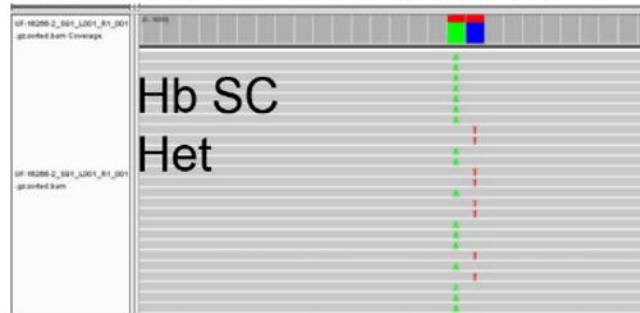
A

HEXA c.805C>T



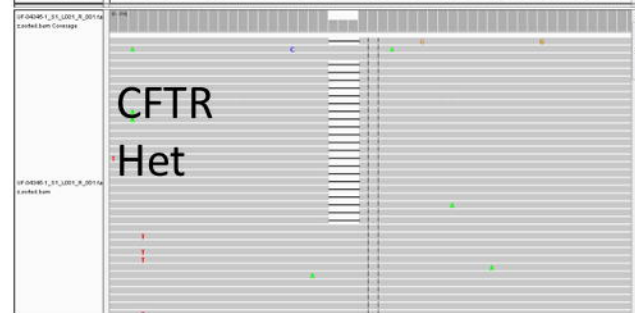
B

HBB c.19T>A and c.20C>T



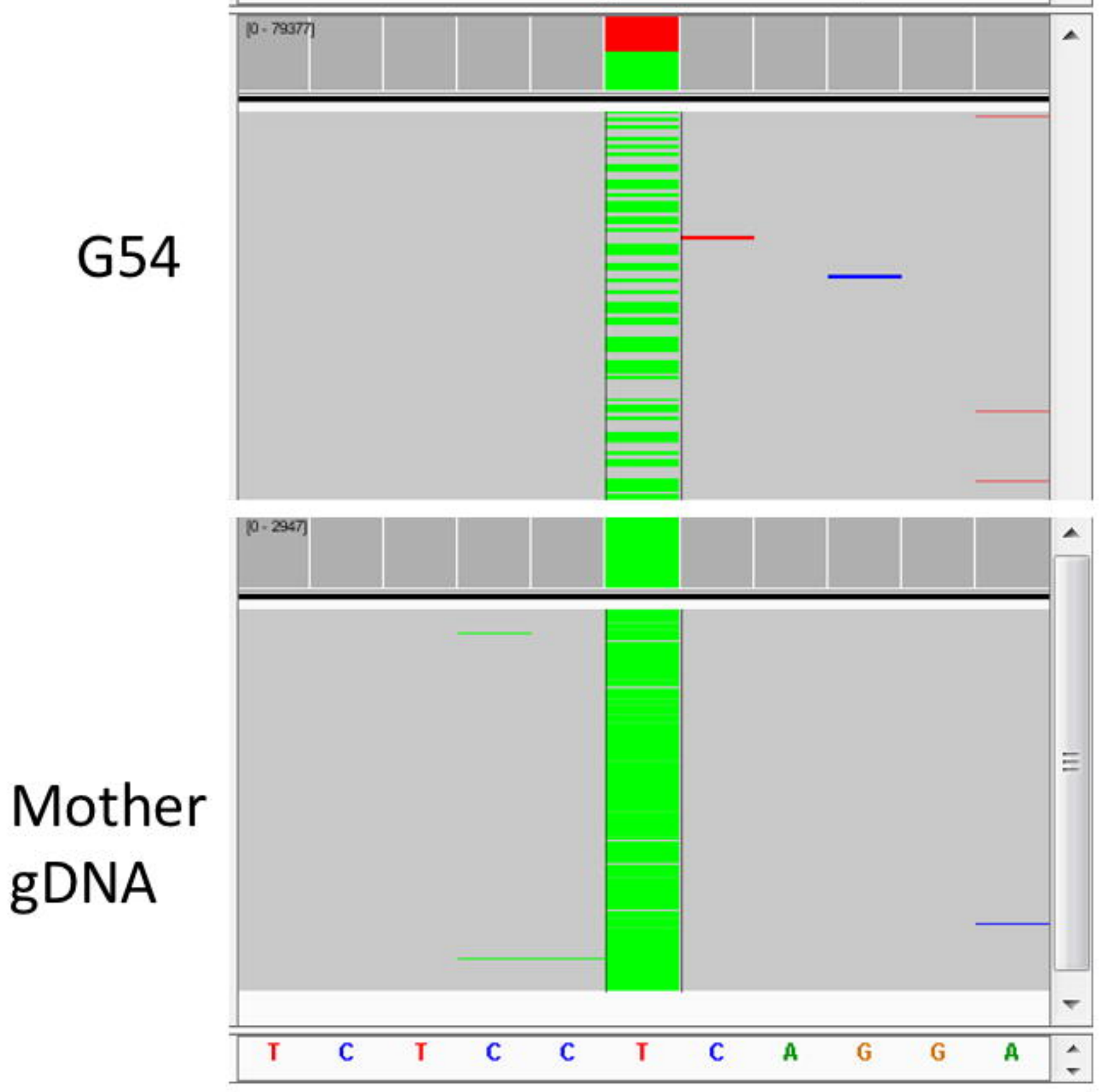
C

CFTR c.1521_1523delTCT



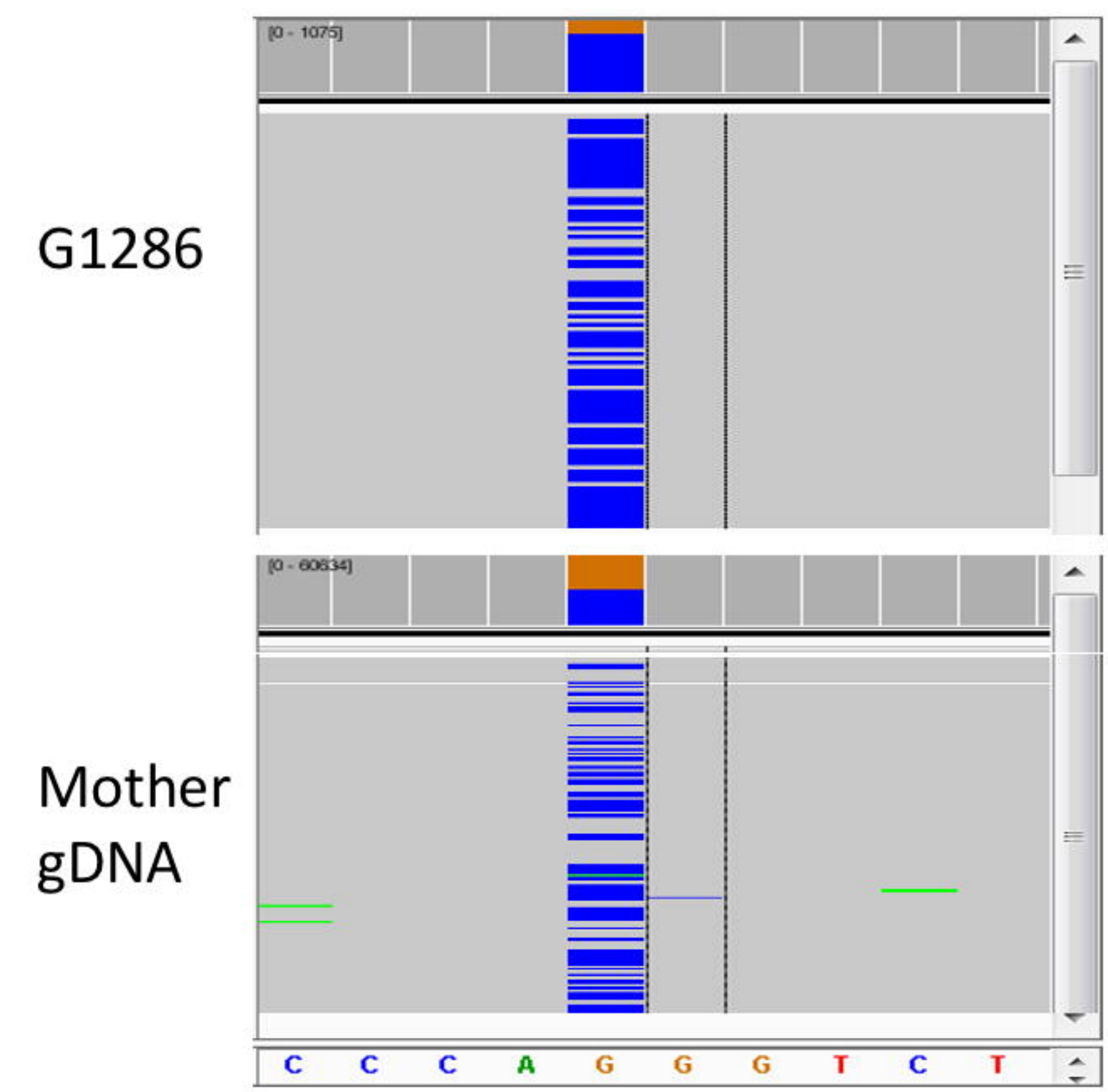
A

NIPT-1180
HBB c.20A>T (p.Glu7Val) rs334



B

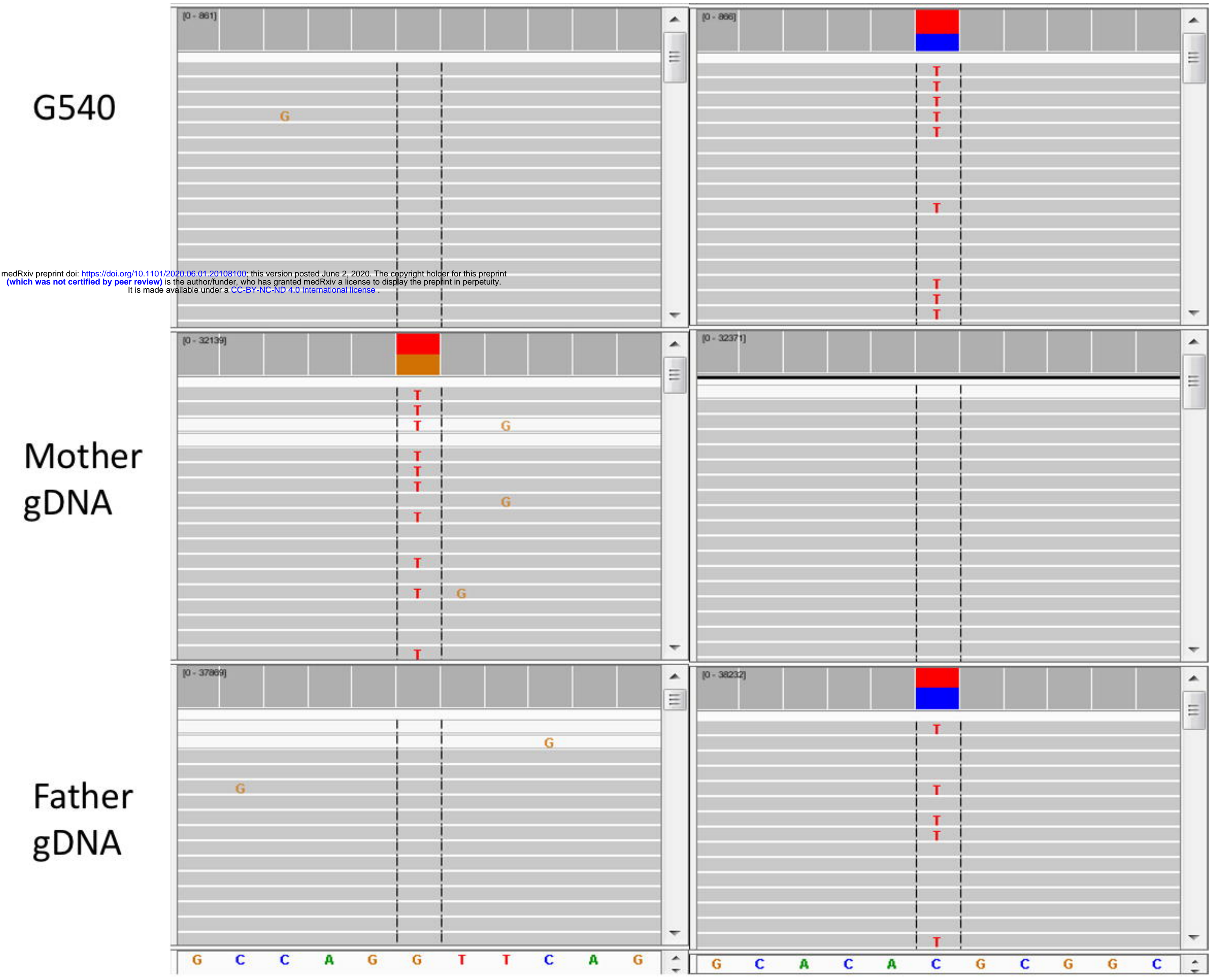
NIPT-1492
DHCR7 c.964-1G>C rs138659167

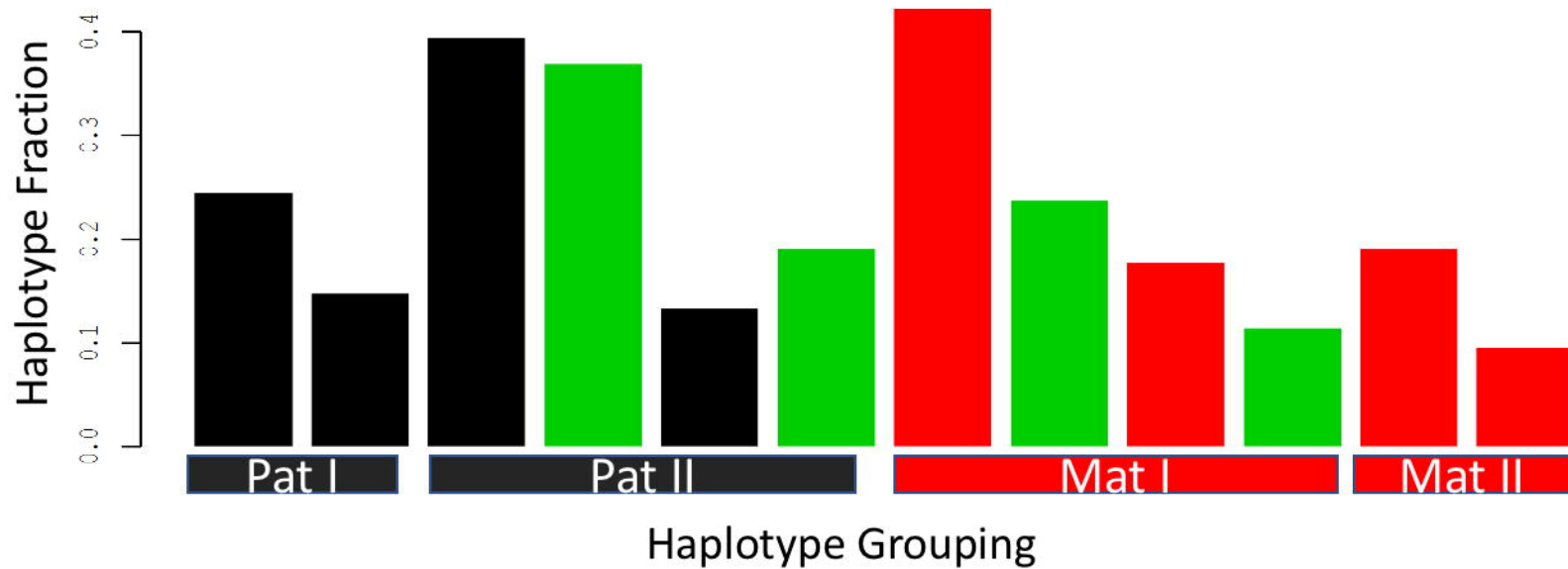
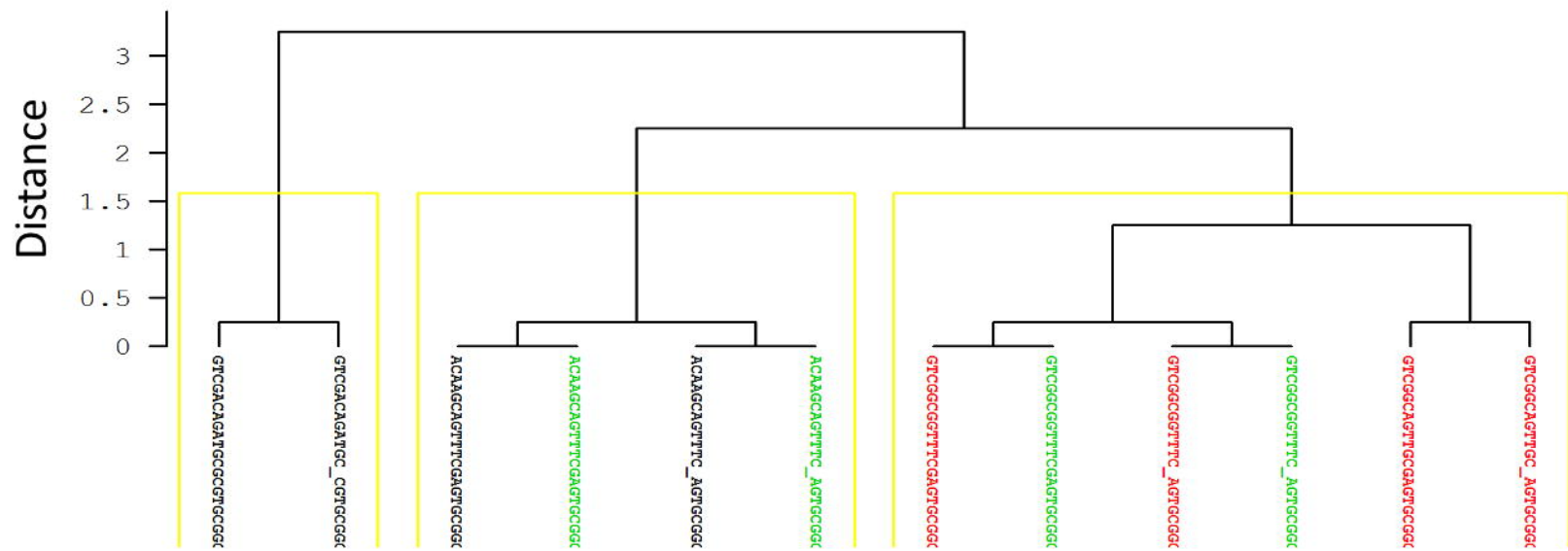


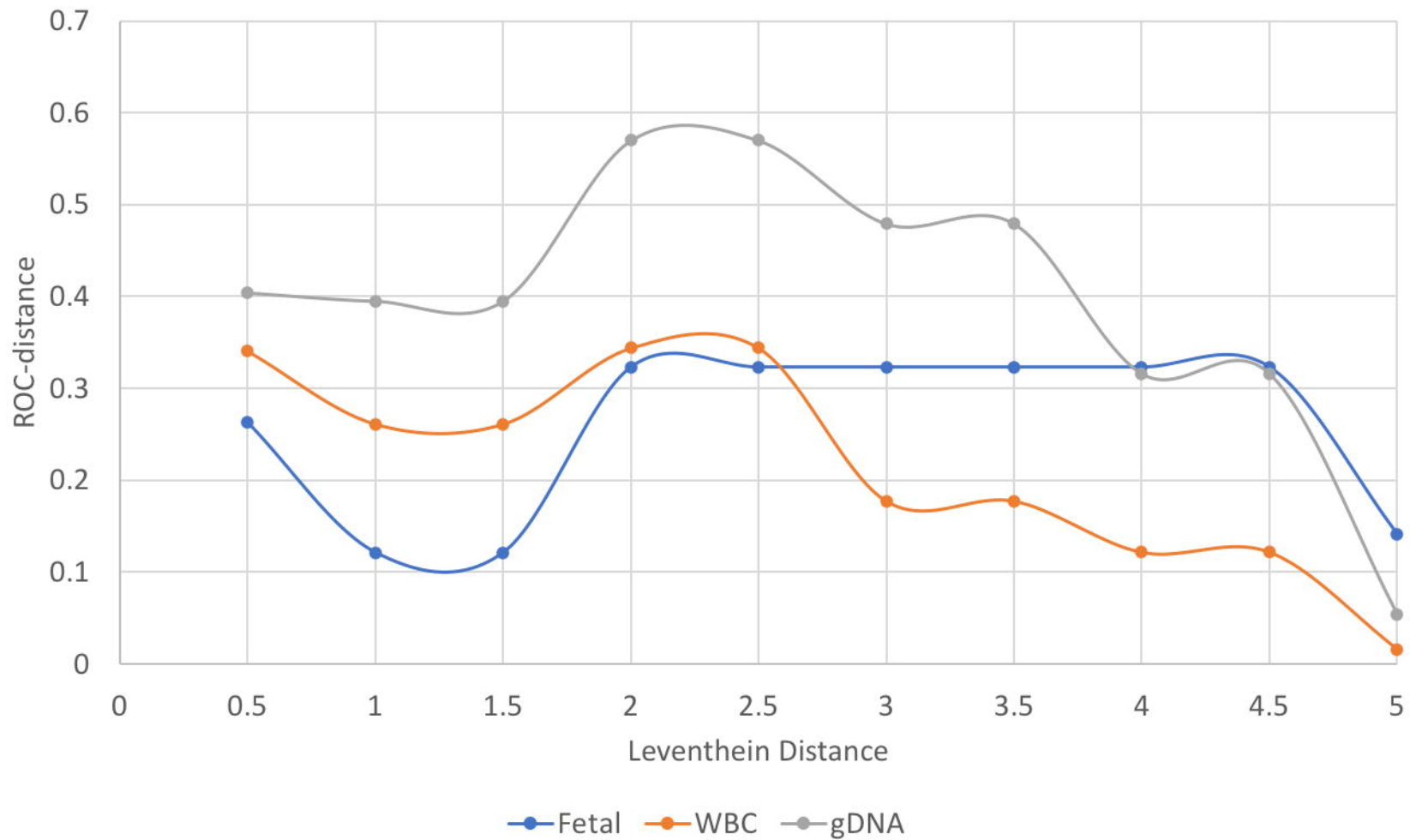
C

RAPSN
c.264C>A (N88K)
rs104894299

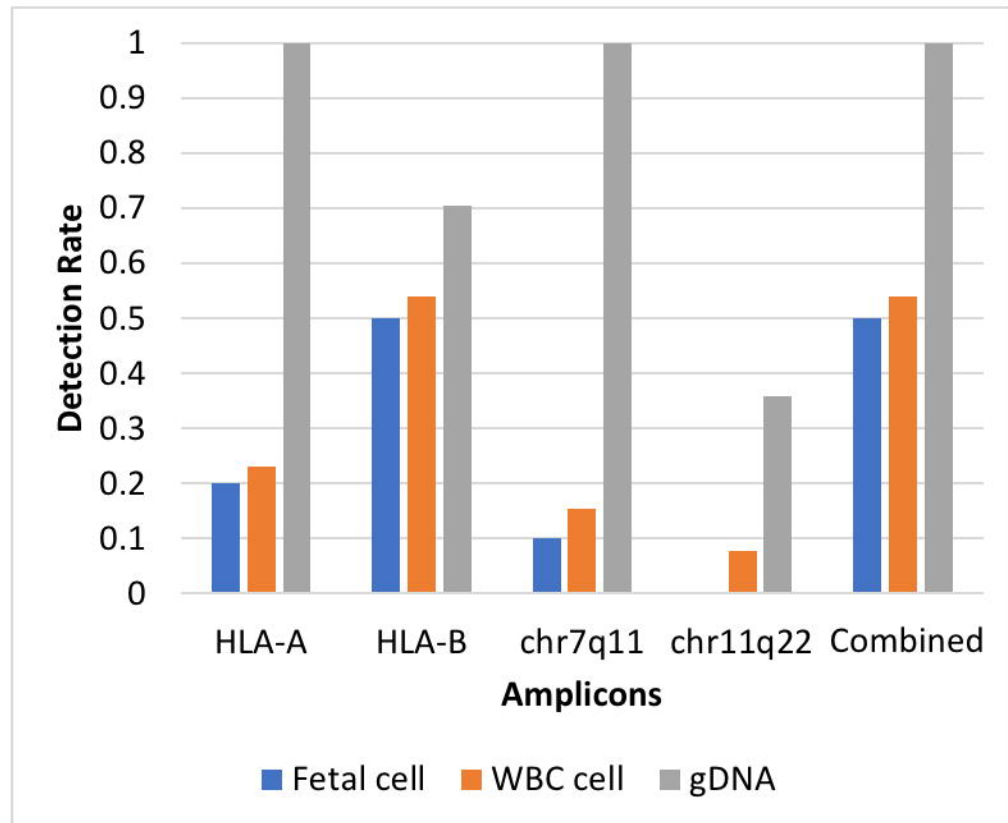
RAPSN
c.493G>A (V165M)
rs761584017







A.



B.

