

1 **Title**

2 **Non-Invasive Prenatal Diagnosis of Single Gene Disorders with enhanced Relative**  
3 **Haplotype Dosage Analysis for diagnostic implementation**

4 NIPD of single gene disorders with eRHDO

5

6 **Author list & Affiliations**

7

8 **Mathilde Pacault,<sup>1,2,\*</sup> Camille Verebi,<sup>2</sup> Magali Champion,<sup>3</sup> Lucie Orhant,<sup>2</sup> Alexandre Perrier,<sup>2</sup>**  
9 **Emmanuelle Girodon,<sup>2</sup> France Leturcq,<sup>2</sup> Dominique Vidaud,<sup>2</sup> Claude Férec,<sup>1</sup> Thierry**  
10 **Bienvenu,<sup>2</sup> Romain Daveau,<sup>4</sup> Juliette Nectoux,<sup>2,\*</sup>**

11

12 <sup>1</sup>Laboratoire de Génétique Moléculaire et Histocompatibilité, CHU, 29200 Brest

13 <sup>2</sup>Service de Médecine Génomique des maladies de système et d'organe, APHP.Centre -  
14 Université Paris Cité, Hôpital Cochin, 75014 Paris

15 <sup>3</sup>Université de Paris, CNRS, MAP5 UMR8145, 75006 Paris

16 <sup>4</sup>MOABI, Plateforme bio-informatique AP-HP, Département I&D, DSI, 75012 Paris

17 \*Correspondance : [juliette.nectoux@aphp.fr](mailto:juliette.nectoux@aphp.fr), [mathilde.pacault@gmail.com](mailto:mathilde.pacault@gmail.com)

18

19

20

## 21 **Abstract**

22 Non-invasive prenatal diagnosis for single-gene disorders (SGD-NIPD) has been widely adopted  
23 by patients, but is mostly limited to the exclusion of paternal or de novo mutations. Indeed, it is  
24 still difficult to infer the inheritance of maternal allele from cell free DNA (cfDNA) analysis. Based  
25 on the study of maternal haplotypes imbalance in cfDNA, relative haplotype dosage (RHDO) was  
26 developed to address this challenge. Although RHDO has proven to be reliable, robust control of  
27 statistical error and explicit delineation of critical parameters for assessing the quality of analysis  
28 have not been fully considered yet.

29 Here we propose a universal and adaptable enhanced-RHDO procedure (eRHDO) through an  
30 automated bioinformatics pipeline with a didactical visualization of results that aims to be applied  
31 for any SGD-NIPD in routine care. A training cohort of 43 families carriers for *CFTR*, *NF1*, *DMD*,  
32 or *F8* mutations allowed the characterization and optimal setting of several adjustable data  
33 variables, such as minimal sequencing depth and type 1 and type 2 statistical errors, as well as  
34 the quality assessment for intermediate steps and final result through block score and  
35 concordance score. Validation was successfully carried out on 56 pregnancies of the test cohort.  
36 Finally, computer simulations were used to estimate the effect of fetal-fraction, sequencing depth  
37 and number of informative SNPs on the quality of results. Our workflow proved to be robust, as  
38 we obtained 94.9% conclusive and correctly inferred fetal genotypes, without any false negative  
39 or false positive result.

40 By standardizing data generation and analysis, we fully describe a turnkey protocol for laboratories  
41 wishing to offer eRHDO-based non-invasive prenatal diagnosis for single-gene disorders as an  
42 alternative to conventional prenatal diagnosis.

43

## 44 Introduction

45 The identification of fetal DNA in maternal blood by Dennis Lo in 1997 has dramatically changed  
46 the landscape of prenatal diagnosis (PND).<sup>1</sup> The major impact to date is probably the widespread  
47 use of cell-free DNA (cfDNA) testing for frequent fetal aneuploidies and, more recently,  
48 subchromosomal abnormalities,<sup>2-5</sup> which was largely adopted by clinicians and pregnant women,<sup>6</sup>  
49 and drastically reduced invasive sampling rate for this indication over a few years.<sup>7,8</sup> Detection of  
50 fetal-specific sequences that are absent or different from maternal genome is also routinely  
51 performed, in the context of fetal sex determination,<sup>9,10</sup> fetal RhD genotyping<sup>11,12</sup> and exclusion of  
52 paternally-inherited or *de novo* mutations for fetuses at risk of single-gene disorders (SGD).<sup>13-15</sup>  
53 However, determination of fetal status towards maternal pathogenic variants remains more  
54 challenging, as haploidentical maternal and fetal sequences cannot be easily distinguished.

55 Whereas detection of paternally inherited or *de novo* variants relies on a qualitative  
56 presence/absence approach, most workflows described to date for maternal inheritance  
57 determination rely on the fine quantification of each maternal haplotype, called relative haplotype  
58 dosage analysis (RHDO), initially described by Lo et al.<sup>16</sup> In this approach, based on Massively  
59 Parallel Sequencing (MPS), parental, proband and cfDNA are sequenced in parallel at multiple  
60 polymorphic positions along the locus of interest. The two haplotypes of each parent are phased  
61 by comparison with the proband's genotype. Paternal inheritance is then determined based on the  
62 qualitative detection of polymorphisms where the two paternal alleles can be distinguished, in  
63 other words where the father is heterozygous and the mother is homozygous, and both paternal  
64 alleles have been identified as linked to the at-risk or the normal haplotype. On the contrary,  
65 maternal inheritance is based on the calculation of an allelic ratio at positions where the two  
66 maternal alleles can be distinguished, that is to say where the mother is heterozygous and the  
67 father is homozygous. The detection of a haplotypic imbalance would reflect the fetal contribution,  
68 from which we may deduce the maternal haplotype that was transmitted to the fetus, normal or at-

69 risk. By relying on indirect diagnosis, RHDO theoretically allows diagnosis for every family,  
70 regardless of the inheritance pattern or the type of molecular anomaly and has been successfully  
71 applied to several SGD.<sup>17-21</sup> However, to our knowledge, clinical implementation in public health  
72 service laboratories around the world remains sparse, probably because quality controls and  
73 decisional thresholds remain complex to define.<sup>22,23</sup>

74 In this report, we describe our RHDO-based NIPD workflow, that we call *enhanced*-RHDO  
75 (eRHDO), where we explain in detail our informatics and statistical analyses with precise risks  
76 control and present a straightforward and comprehensive result visualization. Using computer  
77 simulations, we discuss the impact of the different biological and analytical parameters on the  
78 quality of the result, and propose thresholds and objective quality scores for easy implementation  
79 in a diagnosis laboratory.

## 80 **Materials and Methods**

### 81 **Participants and sample processing**

82 Pregnant women who are known carriers of cystic fibrosis (*CFTR* gene), neurofibromatosis type  
83 1 (*NF1* gene), Duchenne/Becker muscular dystrophy (*DMD* gene) or hemophilia (*F8* and *F9*  
84 genes) were recruited nationwide through the DANNI and NID studies, which were ethically  
85 approved by the French "Comité Consultatif sur le Traitement de l'Information en matière de  
86 Recherche dans le domaine de la Santé" (ref. 13.386) and "Comité de Protection des Personnes"  
87 (ref. 2014-janvier-13465 and 29BRC18.0055). We separated our cohort in two groups. The first  
88 group, called "training cohort", included families for whom only parental, fetus and cfDNA were  
89 available, meaning parental haplotypes were determined using the current pregnancy's fetus  
90 information. Blood samples from women in this group were initially used to validate the efficiency,  
91 accuracy and multiplexing capacity of our method. When genomic DNA (gDNA) from a first child

92 or close relative was available, pregnant women were included in the group we called “test cohort”.

93 For these patients, we used the proband’s gDNA to identify each parental haplotypes, allowing

94 these families to be tested in a real-life-like setting.

95 The DNA samples needed for each family included the cfDNA extracted from maternal plasma,

96 the maternal and paternal gDNA extracted from leukocytes, the proband gDNA when available

97 and the fetal gDNA from invasive sampling to confirm the NIPD result (Figure 1). All samples from

98 the same family were processed simultaneously and pooled together prior to targeted capture

99 enrichment and MPS. Details on DNA extraction and sample processing for MPS libraries

100 preparations are reported in the supporting information. After targeted MPS, sequencing data from

101 each DNA sample were used to perform eRHDO analysis and determine fetal inheritance at the

102 locus of interest. The outcomes were set as haplotype A, B, C or D for the training cohort, as the

103 aim was to refine analysis thresholds and determine whether the fetal haplotype was correctly

104 inferred, regardless of its risk of SGD. For the test cohort, we studied only the gene involved in

105 the familial SGD, and outcomes were set as "HapI" or "HapII" for maternal at-risk or non-at-risk

106 haplotype, respectively; and "HapIII" or "HapIV" for paternal at-risk or non-at-risk haplotype,

107 respectively.

108

109 **Figure 1. Workflow and processing steps of samples obtained from training and test cohorts. Training cohort**

110 **involved couples at risk for SGD who were offered invasive prenatal testing and for whom parental haplotypes**

111 **were determined using the current pregnancy’s fetus genomic DNA. Test cohort involved pregnant women at**

112 **risk for SGD who were offered invasive prenatal testing and for whom a proband’s genomic DNA was available**

113 **for haplotype reconstruction. Study of training cohort allowed to evaluate and set up the adjustable data**

114 **variables. Then, the best combination of variables was used for test cohort exploration. Finally, the diagnostic**

115 **performance of our SGD-NIPD workflow with eRHDO was evaluated and compared to gold standard as defined**

116 **by fetal status obtained through invasive sampling. Quality controls checks are described to minimize mistakes**

117 **and to help with technical validation and biological interpretation of the result.**

## 118 **SGD-NIPD Workflow: processing steps and terminology**

119 After sequencing, a family-specific pile-up was generated, counting the number of reads for each  
120 base, from parental at-risk and non-at-risk haplotypes, at every SNP targeted in our capture panel.

121 The two maternal haplotypes are HapI and HapII, HapI being the maternal at-risk haplotype and  
122 HapII the maternal non-at-risk haplotype; while the two paternal haplotypes are HapIII and HapIV,  
123 HapIII being the paternal at-risk haplotype and HapIV the paternal non-at-risk haplotype.

124 Each SNP of the pile-up was categorized based on parental inheritance, as initially described by  
125 Lo et al.<sup>16</sup> Briefly, SNP1 are defined as those for which the father and mother are both homozygous  
126 for different alleles (mother AA and father BB). SNP2 are those in which the father and mother are  
127 both homozygous for the same allele (mother AA and father AA). SNP3 are those for which the  
128 father is heterozygous and the mother is homozygous (mother AA and father AB). Finally, SNP4  
129 are those for which the father is homozygous and the mother is heterozygous (mother AB and  
130 father AA). SNP4 can be subclassified as SNP4a and SNP4b. SNP4a are those for which the  
131 paternal allele is the same as the maternal allele from HapI, while SNP4b are those for which the  
132 paternal allele is the same as the maternal allele from HapII. The other genomic positions of the  
133 pileup are categorized as WARNING\_1 or WARNING\_2 (Tables S1 and S2).

134 SNP from each category were then analyzed in maternal plasma through SGD-NIPD investigation  
135 to provide simultaneously i/ sequencing error rate determination (SNP2 analysis), ii/ fetal fraction  
136 evaluation (SNP1 analysis or uncategorized SNP analysis), iii/ qualitative paternal haplotype  
137 transmission determination (SNP3 analysis) and iv/ quantitative maternal haplotype transmission  
138 determination (SPRT-based eRHDO SNP4 analysis). After technical validation of the result,  
139 biological interpretation and medical conclusion can be conducted.

140 The SGD-NIPD workflow is reported in Figure 1.

## 141 Data analysis

142 Using simulated cfDNA sequencing data and data from the training cohort, we defined adjustable  
143 data variables, namely consideration of PCR duplicates, thresholds used to discriminate a  
144 heterozygous position in gDNA and minimal sequencing depth for calling SNP in gDNA (NDP)  
145 and cfDNA (PDP). Minimal interSNP distance for eRHDO analysis, as well as type I and type II  
146 error risks associated with SPRT haplotype block delimitation were defined as adjustable output  
147 data variables. Afterwards, we tested different combinations for input and output data variables  
148 and examined objectively their impact on the quality scores in order to select the best combination  
149 of input and output data variables for test cohort exploration. Finally, the diagnostic performance  
150 of our noninvasive test was settled from the test cohort analysis.

151 Details of the bioinformatics pipeline for sequencing data analysis are reported in Supplemental  
152 Methods. We generated a graphical visualization of the results that include preanalytical and  
153 analytical information, such as input data (inheritance mode, proband gender for X-linked  
154 disorders), fetal fraction and quality parameters (position of the parental variant, total number of  
155 SNPs called for each analysis and distance between tested SNPs). For eRHDO analysis, the  
156 repartition of SNP4a and SN4b along the locus, the mean count of SNPs per haplotype block, the  
157 haplotype block count, the concordance between forward- and reverse-defined blocks ( $pc$ ), as  
158 well as the proportion of positions where the SPRT remained non-conclusive ( $nc$ ) are specified.  
159 SNPs were arbitrarily colored in green for the wild-type linked-haplotype, in red for the mutant-  
160 linked haplotype, and in gray for non-conclusive results. This whole analysis pipeline was created  
161 and refined using our training cohort. Its reliability and robustness, as well as pertinence of the  
162 chosen thresholds were tried on our test cohort. In this research study, all results were compared  
163 to the fetal genotype obtained on invasive sampling.

## 164 Quality parameters for eRHDO analysis

165 We implemented quality controls tags as well as two eRHDO-specific quality scores, that we  
166 named concordance score  $S_c$  and block score  $S_b$ , with the aim of reflecting forward and reverse  
167 concordance, haplotype classification error rate as well as the ease to conclude in favor of the  
168 over-representation of one of the maternal haplotype,

169 The concordance score  $S_c$  is defined as the difference between the concordance proportion  
170 between forward and reverse analyses  $pc$  and the proportion of SNPs that remained non-  
171 conclusive  $nc$ . To strongly discriminate against misclassified haplotype blocks (low  $pc$ ) and low-  
172 quality results (high  $nc$ ), which would reflect a low quality result to be interpreted with caution,  $S_c$   
173 is calculated as follows:  $S_c = -\log_{100}(1 - pc) - nc$ . A  $S_c$  closer to 1 reflects a more robust result.

174 The block score  $S_b$  is obtained by normalizing the averaged number of SNPs per haplotype block  
175 by the total number of SNPs tested on the locus, expressed in a logarithmic scale for more  
176 discrimination:  $S_b = -\log_{100}\left(\frac{\text{SNPs per block}}{\#SNPs}\right)$ . Shorter haplotype blocks means that fewer positions  
177 were tested before reaching statistical significance. In other words, a block score close to 1 reflects  
178 a bigger imbalance between both maternal alleles, therefore a higher confidence in the fetal  
179 haplotype that is called.

180

## 181 Statistical analysis

182 All statistical analyses were performed using R. To measure the effect of the adjustable data  
183 variables on the results quality, we performed Wilcoxon tests and Friedman tests for variables with  
184 two (heterozygous position in gDNA) or more than two groups (minimal sequencing depth and  
185 minimal interSNP distance). In addition, we represented the associations using boxplots.



186 The effect of all quality-influencing factors (fetal fraction, number of SNPs and sequencing depth)  
187 and their combinations on the results quality was quantified using Sobol sensitivity indices, ranging  
188 between 0 (minor impact) and 100 (major impact). P-values for ANOVA tests were also computed  
189 to verify their significance. All p-values were adjusted for multiple comparisons using the  
190 Bonferroni correction.

## 191 **Results**

### 192 **Patients information**

193 The training cohort included 43 pregnancies, divided between 27 pregnancies at 25% risk of cystic  
194 fibrosis, 9 pregnancies at 50% risk of maternal transmission of neurofibromatosis type 1, 3 and 4  
195 pregnancies involving a male fetus at 50% risk Duchenne/Becker Muscular Dystrophy and  
196 Hemophilia, respectively. Gestational age at sampling ranged from 6 to 29 weeks of gestation  
197 (mean, 12 weeks).

198 The test cohort included 56 pregnancies from 45 families, divided between 25 pregnancies at 25%  
199 risk of cystic fibrosis (paternal and maternal haplotypes transmission were non-invasively  
200 determined in an autosomal recessive pattern if both parents carried the same mutation, or with  
201 autosomal paternal + maternal dominant patterns if parents carried different mutations), 11  
202 pregnancies at 50% risk of maternal transmission of neurofibromatosis type 1 (maternal haplotype  
203 transmission was non-invasively determined in an autosomal dominant pattern), and 20  
204 pregnancies involving a male fetus at 50% risk of Duchenne/Becker Muscular Dystrophy (maternal  
205 haplotype transmission was non-invasively determined in an recessive X-linked pattern).  
206 Gestational age at sampling ranged from 6 to 17 weeks of gestation (mean, 11 weeks).

207 Patients' information is summarized in Table S3.

## 208 **Optimal setting of the adjustable data variables**

209 To refine our pipeline, the impact of several parameters on the quality of the result was tested  
210 using data from the training cohort.

### 211 **Effect of PCR duplicates**

212 As cfDNA has been largely shown to be non-randomly fragmented,<sup>24,25</sup> two distinct fragments may  
213 be misinterpreted as PCR duplicates and eliminated on bioinformatics analyses, resulting in a loss  
214 of information and a risk of bias in the interpretation. To alleviate this risk, quality parameters of  
215 analyses were compared with or without PCR duplicates. As we obtained no difference in the fetal  
216 fraction, we settled on retaining the PCR duplicates removal in the final bioinformatics pipeline  
217 (data not shown).

### 218 **Effect of the thresholds used to specify a heterozygous position in gDNA**

219 We investigated two alternatives to define heterozygous SNPs in gDNA. In the extensive  
220 definition, heterozygous SNPs in gDNA are defined by an allelic frequency of the alternative allele  
221 (AF) comprised between 0.15 and 0.85, while homozygous positions in gDNA are defined by an  
222 AF <0.15 or >0.85. In contrast, the restrictive definition corresponds to heterozygous SNPs in  
223 gDNA defined by an AF comprised between 0.35 and 0.65 while homozygous positions' AF in  
224 gDNA remain unchanged at <0.15 and >0.85. An extensive definition of a heterozygous position  
225 could retain more informative positions than the restrictive definition but is likely to introduce SNP  
226 categorization errors due to misspecification of SNP positions associated with biased allele  
227 frequency. We explored the effect of extensive versus restrictive definition of heterozygous SNPs  
228 in gDNA on the results by examining their impact on both scores (Figure S1). As expected, the  
229 use of the restrictive definition leads to a reduction in the number of SNP4 that can be used for  
230 the eRHDO analysis of the training cohort, which results in a lower  $\delta_b$ . On the other hand, this

231 restrictive definition is associated with a significantly increased  $\mathcal{S}_c$ , reflecting a marked  
232 improvement in the quality of the results (Table S4). Thus, we decided to opt for the restrictive  
233 definition.

#### 234 **Effect of the minimal sequencing depth**

235 As a lower minimal sequencing depth threshold may retain information at more positions but is  
236 likely less precise in allelic ratio determination, we explored different minimal thresholds for  
237 sequencing depth for our gDNA and cfDNA analyses, denoted by gDNA sequencing depth (NDP)  
238 and cfDNA minimal sequencing depth (PDP), tested at 8x, 15x and 30x, and 15x, 30x, 45x and  
239 60x, respectively. We examined objectively their impact on  $\mathcal{S}_c$  and  $\mathcal{S}_b$ . At PDP fixed, we found no  
240 difference in our scores with higher NDP (Figure S2). Similarly, PDP variation did not impact the  
241  $\mathcal{S}_c$  (Figure S3), indicating no impact on the concordance of haplotype blocks classification between  
242 forward and reverse orientations. However, we found significant differences in terms of  $\mathcal{S}_b$  ( $p$ -value  
243 for Friedman test of  $4.8 \times e^{-24}$ ; Table S5).  $\mathcal{S}_b$  value rises at low PDP, which reflects a better quality  
244 of analysis: the smaller the minimal sequencing depth for calling a SNP4 in cfDNA, the greater the  
245 number of SNP4 taken into consideration in the eRHDO analysis, with potentially a greater the  
246 number of blocks. Therefore, we fixed the minimal sequencing depth for calling SNP in cfDNA at  
247 15x.

#### 248 **Effect of the threshold for minimal interSNP distance**

249 Previous reports used a threshold for minimal interSNP distance of 200 base pair (bp) to minimize  
250 the risk of linkage between two SNPs and a resulting bias that would hinder allelic dosage analysis,  
251 but may also discard positions that would be informative.<sup>18,19</sup> In order to evaluate this risk, we  
252 tested the impact of this threshold at 50 bp, 100 bp and 200 bp. The interSNP distance significantly  
253 impacted  $\mathcal{S}_b$ , with a  $p$ -value for Friedman test of  $3.6e^{-46}$ , suggesting an increased distance  
254 between two SNPs might result in a loss of information by lowering the number of analyzed events

255 (Figure S4 and Table S6). To maximize  $\mathcal{S}_b$ , we settled for a minimal distance of 50 bp between  
256 two SNPs to be considered.

## 257 **Effect of the statistical risks**

258 When implementing the RHDO method, the statistical risks associated with the SPRT control the  
259 proportion of tolerated misclassification errors, thus having a major impact on both  $\mathcal{S}_c$  and  $\mathcal{S}_b$ .  
260 Permissive risks, in other words risks closer to 1, make easier a conclusion of the SPRT test,  
261 which obviously leads to smaller haplotype blocks and better associated  $\mathcal{S}_b$  (Figure S5), but they  
262 have a variable impact on  $\mathcal{S}_c$  (Figure S6). For low values of the parameters fetal fraction, number  
263 of SNPs and depth of sequencing, obtaining conclusive results is rather challenging and higher  
264 values of the statistical risks, associated with smaller blocks, helps in that sense (Figure S7).  
265 However, in reasonable situations, the misclassification errors (Figure S8) induced by permissive  
266 statistical risks increases drastically. Therefore, we decided to set the two risks to 1/1000, with  
267 the aim of reducing at most the erroneous conclusions. On note that our pipeline allows to adjust  
268 each risk to other values if deemed necessary.

269 As a conclusion, the investigation of the training cohort enabled to test different combinations for  
270 input and output data variables, to objectively examine their impact on the quality scores and to  
271 pinpoint important quality control checkpoints. In fine, the optimal combination of input and output  
272 data variables for test cohort exploration corresponded to restrictive definition of heterozygous  
273 SNP, minimal NDP=8x, minimal PDP=15x, minimal interSNP distance=50bp, and type I and type  
274 II error risks associated with SPRT =1/1000.

275

## 276 **Quality assessment**

277 Quality control points have been implemented at all stages of the workflow, to minimize the risk of  
278 error, but also to allow the users to assess the robustness and confidence they can have in their  
279 technical result.

## 280 **Quality control tags associated with identity check**

281 After at-risk and non-at-risk haplotype determination, SNPs are split into the previously described  
282 categories depending on the parental genotypes' combination<sup>16</sup>. We adapted SNP categorization  
283 by considering the proband's relationship to the couple and his status towards the familial SGD  
284 for subclassification, as well as their sex for X-linked transmission. Indeed, possible genotype  
285 combinations and decisional thresholds may differ whether the individual tested for parental  
286 haplotyping is a previous child or a relative of the couple. For quality purposes, we also included  
287 two "WARNING" categories, as consanguinity or undisclosed false paternity may hinder  
288 interpretation. Previously called SNP5, where both parents are heterozygous and the at-risk  
289 haplotypes cannot be inferred, were named "WARNING\_2", for "uninformative genotype  
290 combination". Impossible genotypes combinations were called "WARNING\_1", for example  
291 parents "AA" and "BB" with a first child as a proband genotyped "AA". This second category may  
292 be informative of a sample swap, a wrong proband relationship assignment, or low-quality  
293 sequencing (Tables S1 and S2).

## 294 **Quality control tags associated with DNA capture and sequencing**

295 Important allele drop-out (ADO) may result in erroneous determination of fetal fraction, lower  
296 confidence in paternal inheritance or hinder maternal haplotypes ratio determination. We therefore  
297 return an evaluation of the proportion of ADO using SNP1 (fetal genotype AB and maternal  
298 genotype AA), as the proportion of fetal specific B allele not detected in maternal plasma (AF of B

299 allele in maternal plasma = 0). Although this parameter does not block subsequent SGD-NIPD  
300 analyses, a significant ADO should result in a cautious interpretation of the result.

301 Sequencing error rate for each family was estimated using SNP2 density plots, corresponding to  
302 positions where both parents are homozygous for the same allele: the fetus is expected to be  
303 homozygous as well and any other allele found at the locus can be considered a sequencing error.  
304 Here again, a significant sequencing error rate associated to one or more samples should result  
305 in a cautious interpretation of the result of the SGD-NIPD.

## 306 **Test cohort results**

307 Patients' information and eRHDO-based SGD-NIPD results are presented in Table S3.

308 Sequencing error rate was consistently low for all our families, ranging from 0.3% to 1.4% (mean,  
309 0.68%, data not shown), consistent with the sequencing error profile of the platform.<sup>26</sup> Although  
310 we chose not to set a fixed threshold for interpretation, an increased sequencing error rate should  
311 alert on the risk of low-quality result secondary to degraded DNA or a technical issue during DNA  
312 extraction or sample processing and sequencing.

313 Fetal fraction estimated using the distribution of all sequenced positions (minor allele frequency,  
314 MAF), as described in supplemental data, ranged from 2.9 to 19.3% (mean 9.3%). Three major  
315 discrepancies were found between MAF-based determination and SNP1-based determination,  
316 related to an abnormal distribution of the fetal-specific SNP1 allele frequency in maternal plasma  
317 caused by major ADO in these three samples (families 2\_29, 2\_33, 2\_40).

318 An example of our result visualization is displayed in Figure 2.

319 **Figure 2. Example of a graphic result for autosomal inheritance in a family at risk of transmitting Cystic Fibrosis**  
320 **(family 2), subdivided in paternal (A, B) and maternal (C) inheritance. In this example, SNP3A/D/E/H correspond**  
321 **to SNP3 located on the “non at risk” paternal haplotype (A), whereas SNP3B/C/F/G correspond to SNP3 located**  
322 **on the “at risk” paternal haplotype (B). “Non at risk” and “at risk” haplotypes are represented in green and red,**

323 respectively. (a) For each result, input data are summarized as follow : family ID, tested locus and parental  
324 inheritance, gDNA/cfDNA minimal sequencing depth, transmission mode (ARI = autosomal recessive  
325 inheritance, ADI = autosomal dominant inheritance, RXI = recessive X-linked inheritance) and number of SNP3.  
326 (b) Qualitative detection of fetal-specific SNPs at an allelic frequency (VAF) higher than our threshold visualized  
327 in (c) (AU = arbitrary units). SNP3A/D/E/H from the “non at risk haplotype” were not detected in cfDNA (top),  
328 whereas the majority of SNP3B/C/F/G from the “at risk” haplotype were qualitatively detected above the  
329 background threshold (bottom). (d) Fetal genotype from sequencing of fetal gDNA obtained through invasive  
330 sampling (e) Representation of inter-SNP distance. Each peaks indicates a longer genomic distance between  
331 two consecutive tested positions. (f) Location of the parental mutation, symbolized by a vertical dashed line.  
332 (g) Input data and quality parameters of the SPRT analysis, namely number of SNP4, mean number of SNPs  $4\alpha$   
333 or  $4\beta$  per block (na and nb), concordance between conclusive haplotype blocks in forward and reverse  
334 orientation (pc) and proportion of non conclusive haplotype blocks (in grey color) among all blocks (nc), and  
335 type I and type II statistical error used in SPRT test (a/b). (h) and (i) SPRT analysis in forward and reverse  
336 orientation, respectively. (j) Visualization of haplotype blocks, divided in  $4\alpha$  (top) and  $4\beta$  (bottom) analyses,  
337 with number of conclusive blocks. (k) Distribution of SNPs  $4\alpha$  (blue) and  $4\beta$  (purple) along the genomic region.

338 Paternal inheritance was determined only for families at risk of recessive cystic fibrosis (CF). All  
339 results returned conclusive and concordant with the fetal genotype obtained on invasive sampling  
340 except one. In family 2\_12, a recombination occurred close to the paternal variant location and  
341 the transmitted paternal haplotype at this position could not be inferred. Although the mother was  
342 a carrier of a different pathogenic variant, as this position was not targeted by our capture panel,  
343 the paternal variant could not be qualitatively detected by direct visualization of the paternal-  
344 specific allele on the bam sequence files.

345 Considering maternal inheritance, 59 analyses were performed on 56 samples, as one family  
346 (family 2\_26) was included three times during three successive pregnancies in our study and the  
347 cfDNA of one given pregnancy was analyzed twice using either PND's gDNA for parental  
348 haplotype inference. 56/59 samples returned conclusive (94.9%) and concordant with the fetal  
349 genotype obtained on invasive sampling. Among these, 43/59 results (72.9%) returned conclusive  
350 with high quality results. All these high quality results displayed a  $\delta_c$  and a  $\delta_b$  higher than 0.7

351 (Figure 3A). 13/59 (22%) returned conclusive and concordant with the fetal genotype but with  
352 slightly lower quality results. As suggested by our computer simulations, these low-quality results  
353 probably happened because of a small number of tested events, either because of low cfDNA  
354 sequencing depth, consistently lower than 100x (meaning a small number of events counted at  
355 each locus), or low fetal fraction (families 2\_33 and 2\_40) (meaning a small allelic imbalance to  
356 be detected). For these families, the block score  $S_b$  returned between 0.55 and 0.7 (Figure 3B).  
357 Finally, 3 samples returned inconclusive (5.1%). For family 2\_39, a recombination event occurred  
358 in the block adjacent to the position of the maternal variant, not allowing a confident haplotype  
359 inference at this position (Figure 3C). Family 2\_12 had a low fetal fraction and a low overall cfDNA  
360 sequencing depth, so that not enough positions could be studied. Similarly, not enough SNP4  
361 were included in family 2\_14's RHDO analysis as few SNP4 were identified in the *CFTR* locus  
362 and the sample returned with a relatively low fetal fraction (5.9%).

363 **Figure 3. Examples of results with varying quality scores.**

364 **(A) Family 2\_42 at risk of transmitting Duchenne/Becker muscular dystrophy. A high number of tested SNP4**  
365 **(n=1270) associated with a deep sequencing depth (mean=211X, data not shown) allowed determination of a**  
366 **great number of blocks (59 haplotype blocks in forward direction and 59 haplotype blocks in reverse direction)**  
367 **with high concordance in favor of the transmission of Hapl (Sc=1.00), despite a relatively low fetal fraction**  
368 **(f=4.8%).  $S_b$  was evaluated at 0.89.**

369 **(B) Family 2\_5 at risk of transmitting Cystic Fibrosis. This family presented with a reasonable number of SNP4**  
370 **(n=396) and fetal fraction (f=7.5%) but only a few haplotype blocks could be reconstructed (n=39 blocks in**  
371 **forward + reverse directions) with a relatively high number of SNP4 per block (na=17, nb=28.1), resulting in a**  
372 **lower block score ( $S_b$ =0.62). On note that the high concordance between forward and reverse analysis**  
373 **(Sc=0.93), as well as the position of the maternal variant along the CFTR locus (black dashed line) still allows**  
374 **to conclude in favor or the transmission of HapII with high confidence.**

375 **(C) Family 2\_24 at risk of maternal transmission of neurofibromatosis type I. Only 26 haplotype blocks could**  
376 **be defined with a high mean number of SNPs per block (na=40.5, nb=29,6), resulting in longer blocks and**  
377 **therefore a lower block score ( $S_b$ =0.54). Because 2/26 blocks were uncorrectly classified HapII instead of Hapl,**



378 suggesting a SPRT error, the concordance score estimated at 0.84. Although the maternal variant is located in  
379 a concordant, Hapl region, the low number of haplotype blocks, the haplotypic change at the 3' end of the  
380 locus and the unclassified blocks at both extremities prevent us from concluding on the fetal status for NF1. In  
381 a diagnosis setting, analysis could be repeated on a subsequent sample.

382 **S<sub>b</sub> : Bloc Score; S<sub>c</sub> : Concordance score; n<sub>a</sub> : mean number of SNP4a per haplotype bloc; n<sub>b</sub> : mean number**  
383 **of SNP4b per haplotype bloc; f : fetal fraction**

384  
385 We suggest that the use of thresholds determined by the user could help them when interpreting  
386 their result. In our hands, we observed that scores >0.7 were associated with high quality results,  
387 while scores <0.55 were associated with poor quality results.

## 388 389 **Computer simulation for quantification of the respective effect of** 390 **quality-influencing factors**

391 Thanks to both training and test cohorts' study, we picked out biological and analytical critical  
392 factors that appear crucial for the quality of the analysis, namely fetal fraction, number of SNP4  
393 available for eRHDO analysis, and cfDNA sequencing depth. To evaluate their respective  
394 relevance, we simulated sequencing data on which we tested the quality scores. In each  
395 simulation, one of the parameters was altered, the statistical risks being fixed to 1/1000.

396 As shown in Figure S9, these three parameters have a confirmed major influence on the quality  
397 of the results: the higher, the easier the conclusions of the SPRT tests, as well as the smaller and  
398 the more conclusive the blocks, the higher  $\mathcal{S}_c$  and  $\mathcal{S}_b$ .

399 To quantify the effect of all these parameters on the result quality, we also performed sensitivity  
400 analysis through the computation of Sobol sensitivity indices. In brief, these indices measure how  
401 much the variance in the model output is due to each parameter to be fixed. A low sensitivity index  
402 (close to 0) reflects that variations of the associated parameter will lead to low variations in the

403 results whereas a high sensitivity index (close to 100) implies huge changes. Sensitivity analysis  
404 also gives the opportunity to explore the interaction effects of parameters on the model. In addition  
405 to the computation of Sobol indices, we finally performed ANalysis Of VAriance (ANOVA) tests to  
406 identify the statistically significant effects of the parameters on both scores. Sobol indices of fetal  
407 fraction, number of SNPs, sequencing depth, combinations of them and the associated ANOVA  
408  $p$ -values are provided in Table S7.

409 These further analyses confirm the major impact of the fetal fraction on both  $\mathcal{S}_c$  and  $\mathcal{S}_b$ . The effect  
410 of the number of SNPs and the sequencing depth, while still being highly significant, is however  
411 less strong. Only interactions between parameters including the fetal fraction leads to significant  
412 variations of the  $\mathcal{S}_c$ , which emphasizes again its importance. On the contrary, the  $\mathcal{S}_b$  is sensitive to  
413 simultaneous changes of all parameters.

## 414 Discussion

415 Since the first description of RHDO, no practical recommendations or guidelines concerning data  
416 generation and analysis, fetal fraction evaluation, quality thresholds and interpretation have been  
417 proposed.<sup>27</sup> In this report, we sought to develop a standardized workflow for clinical  
418 implementation of a reliable SGD-NIPD. To that effect, we worked on each step of the previously  
419 described RHDO and optimized this workflow on a training cohort, composed of families where  
420 the fetal gDNA was used to infer parental haplotypes. Then, we tried it on a test cohort, as close  
421 as possible to a clinical setting with various transmission modes and different types of molecular  
422 abnormalities as well as multiple proband and couple families' schemes.

423 As we aim to provide the most direct computational analysis, we tried to review every situation  
424 that could be encountered during genetic counselling in our workflow construction. We detailed  
425 each SNP category by taking into account the proband's relationship to the couple (previous child

426 or close relative) as well as his status towards the parental variants (affected/carrier or unaffected).

427 That way, haplotype phasing and decisional threshold calculations can be run automatically  
428 without the need for a bespoke adaptation of the informatics analysis pipeline for specific  
429 situations.

430 A key point of our approach is that our workflow has been designed in a way that users are able  
431 to adjust several input and output data variables at diverse steps of the process, according to their  
432 needs. The investigation of the training cohort as well as simulated data analysis enabled to test  
433 different values for these variables, to examine objectively their impact on the quality scores.  
434 Thereby, the removal or not of PCR duplicates, the extensive or restrictive thresholds used to  
435 specify a heterozygous position in gDNA, the minimum sequencing depth of cfDNA, the minimum  
436 interSNP distance, and the statistical risks associated with SPRT have been investigated and  
437 adjusted.

438 Quality control points have been implemented at all stages of the workflow, to minimize the risk of  
439 errors, but also to allow the users to assess the robustness and confidence they can have in their  
440 result. During SNP categorization, we added two "WARNING" categories for uninformative and  
441 impossible genotype combinations. The former may help expand the range of family members  
442 that may be tested, while the latter will be of use in identifying the potential causes of an unusual  
443 result. As an example, too many SNPs excluded from the analysis because of uninformative  
444 combinations may explain an inconclusive result on qualitatively correct experimental data that  
445 might reflect analysis from a consanguineous parental couple. An over-representation of  
446 impossible combinations may indicate an error in the input data for sequencing analysis, a sample  
447 swap or contamination during the technical steps.

448 Because ADO may hinder analysis, or even cause a false result, we added an evaluation of the  
449 mean ADO rate based on the SNP1 analysis in the output data. Finally, sequencing error rate  
450 density for each family was estimated using SNP2, corresponding to positions where both parents

451 are homozygous for the same allele. Significant ADO or sequencing error rate should result in a  
452 cautious interpretation of the result.

453 The main quality-influencing factors are the fetal fraction, the number of SNP4 tested and the  
454 cfDNA sequencing depth, which impact the number of events tested. Computer simulations  
455 allowed identifying the fetal fraction as the most preponderant of them. As the number of SNP4  
456 depends on the parental genotypes combination and fetal fraction is biologically determined in a  
457 given cfDNA sample, we recommend that sequencing conditions should be optimized in order to  
458 obtain sufficient cfDNA sequencing depth, higher than 100x in this study.

459 As fetal fraction needs to be evaluated precisely, we chose to use the minor allele frequency  
460 distribution of all the SNPs targeted in the capture panel rather than relying on the previous  
461 knowledge of their category, as it is done for SNP1-based evaluation. By increasing the number  
462 of events, we hope to obtain a more accurate estimate of the fetal fraction than the SNP1-based  
463 evaluation.

464 Paternal inheritance determination relies on the qualitative detection of fetal-specific SNPs that  
465 are present at a very low allelic fraction in maternal plasma. We introduced a minimal allelic  
466 fraction threshold for considering a SNP3 as significantly detected, depending on the sequencing  
467 error rate and the fetal fraction of the sample, to discriminate between background noise and  
468 SNPs that were indeed transmitted to the fetus. We also divided the analysis between "non-at-risk  
469 haplotype detected" and "at-risk haplotype detected" and require concordance between these two  
470 analyses.

471 Finally, we designed our output to be as simple and unambiguous as possible, while retaining as  
472 much information as possible. To that effect, we included a graphical visualization that  
473 summarizes the input data, reports the result with a two-color code, and includes several quality  
474 parameters to facilitate biological interpretation. In order to address the risk of misclassification

475 due to an undetected recombination event, we represented the distance between two adjacent  
476 SNPs included in the analysis as a black curve in this final visualization, as the risk of  
477 recombination between two loci increases with genomic distance.<sup>28,29</sup>

478 Our workflow proved very robust, as we obtained 94.9% conclusive and correctly inferred fetal  
479 genotypes in our real-life-like cohort, and no false negative or false positive result was returned  
480 (56/59 cases conclusive and correctly inferred + 3/59 inconclusive cases). Rather, our quality  
481 controls proved stringent enough for low quality analyses to return inconclusive, mostly in the  
482 situation of an insufficient number of tested events, either because of a low fetal fraction, low  
483 sequencing depth or because of a small number of informative SNPs in the families. Our objective  
484 quality scores corroborated consistently our biological interpretation, being higher than 0.7 for  
485 high-quality analyses and lower than 0.55 for low-quality analyses. Nevertheless, we chose not to  
486 set a threshold for analysis, as we developed these quality scores as a support for interpretation  
487 rather than strict quality criteria. Indeed, they only refer to the quality of the haplotype identification  
488 in maternal plasma and do not consider the variant's position along the target locus. They might  
489 therefore be close to 1 in a family for whom no SNPs 5' or 3' of the variant can be identified, a  
490 situation in which the maternal haplotype at the variant's locus cannot be determined with  
491 confidence as a recombination event cannot be excluded. Likewise, if a recombination event is  
492 identified close to the parental variant position, the analysis may be of overall high quality but the  
493 interpretation of the SGD-NIPD should return inconclusive regardless of the quality scores.

494 This recombination risk is a well-described limitation of the MPS-based NIPD. Because of its  
495 reliance on an indirect approach without the possibility of the fetal sequences in the cfDNA, the  
496 occurrence of a recombination event close to the parental variant position may not allow to  
497 conclude.<sup>23</sup> While this limitation cannot be entirely addressed, we hoped to assist interpretation  
498 by adding information on the parental variant position along the haplotype. In our test cohort, we  
499 observed a total of 7 recombination events, as expected from the known recombination

500 frequencies in the genomic regions investigated. Five of them occurred at sufficient genomic  
501 distance of the parental variant and did not affect the SGD-NIPD result, but the other two hindered  
502 SGD-NIPD interpretation. On note, while a recombination event on the maternal allele close to the  
503 familial variant in the context of maternal autosomal dominant or X-linked transmission would not  
504 allow NIPD, if both parents carry different pathogenic variants in an autosomal recessive setting,  
505 paternal allele may be sought for directly in maternal cfDNA sequences in a qualitative manner.

506 Another major limitation of haplotype-based SGD-NIPD is the need for a proband's genotype  
507 information for haplotype phasing. Caution must be made when choosing this proband. In some  
508 cases, one of the parents and their unaffected sibling may not share any haplotype. Similarly,  
509 when parents carry the same variant in an autosomal recessive disorder, the parental origin of the  
510 variant of an heterozygous first child will most likely be unknown, unless an indirect technique has  
511 been previously performed in these families. Parental haplotypes will be correctly phased but  
512 identification of the at-risk and non-at-risk haplotypes will not be possible. Although we anticipated  
513 most situations, this workflow cannot be offered to a family where a variant appeared *de novo*, or  
514 if no other family member is available for testing. Whereas *de novo* haplotyping methods have  
515 been reported in the context of NIPD,<sup>30-35</sup> experimental approaches often require specific  
516 equipment and reagents, making them costly and inappropriate for implementation in a clinical  
517 setting. Computational approaches, while accurate and less expensive, may require specific  
518 bioinformatics skills and rely on population database, making them unreliable for rare or private  
519 variants phasing.<sup>36,37</sup> An exception would be the Targeted Locus Amplification approach,  
520 described by Vermeulen et al, which does not depend on any specific equipment but may still  
521 require adaptations for implementation in laboratories unused to chromosome conformation  
522 capture techniques.<sup>38</sup> Yet, experiences in preimplantation genetic testing showed that declining a  
523 referral because of the impossibility of analyzing a proband seldom happens.<sup>39</sup>

524 MPS-based techniques also require a careful study of the genomic environment of the tested gene  
525 before designing the enrichment panel. In our study, we tested hemophilia A-carriers in our training  
526 cohort, but had to exclude them from our test cohort, as our capture probes did not target enough  
527 SNPs at the 3' extremity of the *F8* gene, which is localized at the telomeric extremity of the X  
528 chromosome. Specific adaptations may be needed for particular disorders, for example increasing  
529 the number of targeted SNPs in the locus of interest by lowering the minor allele frequency  
530 threshold used in the capture probes design, or increasing the size of the target region around the  
531 gene.

532 With the wide implementation of non-invasive prenatal testing for fetal aneuploidy, maternal  
533 incidental findings have been reported, especially maternal malignancy.<sup>40,41</sup> To our knowledge,  
534 this situation has not been reported yet in the context of NIPD for SGD. The occurrence of an  
535 abnormal profile of our density plot of MAF for fetal fraction estimation may be suggestive of a  
536 maternal underlying disorder. To date, disclosure of results suggesting maternal malignancy  
537 remains controversial,<sup>41,42</sup> but the possibility of maternal incidental findings during cfDNA analysis  
538 should be mentioned during pretest counseling.

539 As a conclusion, we report an easy-to-use workflow for SGD-NIPD. We adapted the previously  
540 published relative haplotype dosage analysis approach for a straightforward, easy to perform  
541 workflow. Although thresholds for statistical risks and quality parameters can be easily adjusted  
542 depending on the biological expertise, we propose an optimized value for each parameter. The  
543 accuracy and reliability of the whole process was then validated on a large real-life-like cohort. By  
544 testing multiple transmission modes, variant types and proband's relationship, we hope to review  
545 a broad spectrum of situations frequently encountered in the PND context. This safe, non-invasive  
546 approach can be readily implemented as an accredited diagnosis service in a public health  
547 laboratory.

## 548 Declaration of interests

549 The authors declare no competing interests

## 550 Acknowledgments

551 We are grateful to the patients, patients' associations and the professional, geneticists,  
552 obstetricians, mid wives, nurses and laboratory technicians for supporting us throughout our  
553 developments. Thank you very much to Pr. Jean-Claude Kaplan and Dr. Christine Bellanne  
554 Chantelot for their wise and beneficial proofreading. The DANNI study received financial  
555 support from Agence de la Biomédecine, the association Vaincre la Mucoviscidose and the  
556 Association Française contre les Myopathies. The NID study received financial support from  
557 Agence de la Biomédecine and the association Gaétan Saleün.

## 558 Author contributions

559 MP: conceptualization, methodology, validation, investigation, data curation, writing, visualization,  
560 funding acquisition; CV: conceptualization, methodology, validation, investigation, data  
561 curation, writing, visualization. MC: methodology , software, validation, resources, formal  
562 analysis, data curation, writing, visualization; LO: investigation; AP, EG, FL, DV, CF, TB:  
563 writing - review and editing; RD: methodology, software, validation, resources, data curation,  
564 validation, writing, visualization; JN: conceptualization, methodology, validation, resources,  
565 writing, visualization, supervision, project administration, funding acquisition.

## 566 Data and code availability

567 Our NIPD pipeline is available on Github at [https://github.com/rdaveau/nipt\\_sprt\\_erhdo](https://github.com/rdaveau/nipt_sprt_erhdo).



## 568 References

- 569 1. Lo, Y.M.D., Corbetta, N., Chamberlain, P.F., Rai, V., Sargent, I.L., Redman, C.W., and  
570 Wainscoat, J.S. (1997). Presence of fetal DNA in maternal plasma and serum. *The Lancet*  
571 *350*, 485–487. [10.1016/S0140-6736\(97\)02174-0](https://doi.org/10.1016/S0140-6736(97)02174-0).
- 572 2. Taylor-Phillips, S., Freeman, K., Geppert, J., Agbebiyi, A., Uthman, O.A., Madan, J., Clarke,  
573 A., Quenby, S., and Clarke, A. (2016). Accuracy of non-invasive prenatal testing using cell-  
574 free DNA for detection of Down, Edwards and Patau syndromes: a systematic review and  
575 meta-analysis. *BMJ Open* *6*, e010002. [10.1136/bmjopen-2015-010002](https://doi.org/10.1136/bmjopen-2015-010002).
- 576 3. Bianchi, D.W., and Chiu, R.W.K. (2018). Sequencing of Circulating Cell-free DNA during  
577 Pregnancy. *N Engl J Med* *379*, 464–473. [10.1056/NEJMra1705345](https://doi.org/10.1056/NEJMra1705345).
- 578 4. Familiari, A., Boito, S., Rembouskos, G., Ischia, B., Accurti, V., Fabietti, I., Volpe, P., and  
579 Persico, N. (2021). Cell-free DNA analysis of maternal blood in prenatal screening for  
580 chromosomal microdeletions and microduplications: a systematic review. *Prenatal Diagnosis*  
581 *41*, 1324–1331. [10.1002/pd.5928](https://doi.org/10.1002/pd.5928).
- 582 5. Christiaens, L., Chitty, L.S., and Langlois, S. (2021). Current controversies in prenatal  
583 diagnosis: Expanded NIPT that includes conditions other than trisomies 13, 18, and 21 should  
584 be offered. *Prenatal Diagnosis* *41*, 1316–1323. [10.1002/pd.5943](https://doi.org/10.1002/pd.5943).
- 585 6. Kater-Kuipers, A., Bakkeren, I.M., Riedijk, S.R., Go, A.T.J.I., Polak, M.G., Galjaard, R.-J.H.,  
586 de Beaufort, I.D., and Bunnik, E.M. (2021). Non-invasive prenatal testing (NIPT): societal  
587 pressure or freedom of choice? A vignette study of Dutch citizens' attitudes. *Eur J Hum Genet*  
588 *29*, 2–10. [10.1038/s41431-020-0686-9](https://doi.org/10.1038/s41431-020-0686-9).
- 589 7. Bjerregaard, L., Stenbakken, A.B., Andersen, C.S., Kristensen, L., Jensen, C.V., Skovbo, P.,  
590 and Sørensen, A.N. (2017). The rate of invasive testing for trisomy 21 is reduced after  
591 implementation of NIPT. *Dan Med J* *64*, A5359.
- 592 8. Palomaki, G.E., Kloza, E.M., O'Brien, B.M., Eklund, E.E., and Lambert-Messerlian, G.M.  
593 (2017). The clinical utility of DNA-based screening for fetal aneuploidy by primary obstetrical  
594 care providers in the general pregnancy population. *Genetics in Medicine* *19*, 778–786.  
595 [10.1038/gim.2016.194](https://doi.org/10.1038/gim.2016.194).
- 596 9. Devaney, S.A., Palomaki, G.E., Scott, J.A., and Bianchi, D.W. (2011). Noninvasive fetal sex  
597 determination using cell-free fetal DNA: a systematic review and meta-analysis. *JAMA* *306*,  
598 627–636. [10.1001/jama.2011.1114](https://doi.org/10.1001/jama.2011.1114).
- 599 10. Hill, M., Lewis, C., Jenkins, L., Allen, S., Elles, R.G., and Chitty, L.S. (2012). Implementing  
600 noninvasive prenatal fetal sex determination using cell-free fetal DNA in the United Kingdom.  
601 *Expert Opinion on Biological Therapy* *12*, S119–S126. [10.1517/14712598.2012.666522](https://doi.org/10.1517/14712598.2012.666522).
- 602 11. Orhant, L., Rondeau, S., Vasson, A., Anselem, O., Goffinet, F., Allach El Khattabi, L., Leturcq,  
603 F., Vidaud, D., Bienvenu, T., Tsatsaris, V., et al. (2016). Droplet digital PCR, a new approach  
604 to analyze fetal DNA from maternal blood: application to the determination of fetal RHD  
605 genotype. *Annales de biologie clinique* *74*, 269–277. [10.1684/abc.2016.1139](https://doi.org/10.1684/abc.2016.1139).

- 606 12. Zhu, Y., Zheng, Y., Li, L., Zhou, H., Liao, X., Guo, J., and Yi, P. (2014). Diagnostic accuracy  
607 of non-invasive fetal RhD genotyping using cell-free fetal DNA: a meta analysis. *The Journal*  
608 *of Maternal-Fetal & Neonatal Medicine* 27, 1839–1844. 10.3109/14767058.2014.882306.
- 609 13. Gruber, A., Pacault, M., El Khattabi, L.A., Vaucouleur, N., Orhant, L., Bienvenu, T., Girodon,  
610 E., Vidaud, D., Leturcq, F., Costa, C., et al. (2018). Non-invasive prenatal diagnosis of  
611 paternally inherited disorders from maternal plasma: detection of NF1 and CFTR mutations  
612 using droplet digital PCR. *Clin Chem Lab Med* 56, 728–738. 10.1515/cclm-2017-0689.
- 613 14. Chitty, L.S., Mason, S., Barrett, A.N., McKay, F., Lench, N., Daley, R., and Jenkins, L.A.  
614 (2015). Non-invasive prenatal diagnosis of achondroplasia and thanatophoric dysplasia: next-  
615 generation sequencing allows for a safer, more accurate, and comprehensive approach.  
616 *Prenat Diagn* 35, 656–662. 10.1002/pd.4583.
- 617 15. Pacault, M., Verebi, C., Lopez, M., Vaucouleur, N., Orhant, L., Deburgrave, N., Leturcq, F.,  
618 Vidaud, D., Girodon, E., Bienvenu, T., et al. (2022). Non-invasive prenatal diagnosis of single  
619 gene disorders by paternal mutation exclusion: 3 years of clinical experience. *BJOG*, 1471-  
620 0528.17201. 10.1111/1471-0528.17201.
- 621 16. Lo, Y.M.D., Chan, K.C.A., Sun, H., Chen, E.Z., Jiang, P., Lun, F.M.F., Zheng, Y.W., Leung,  
622 T.Y., Lau, T.K., Cantor, C.R., et al. (2010). Maternal Plasma DNA Sequencing Reveals the  
623 Genome-Wide Genetic and Mutational Profile of the Fetus. *Science Translational Medicine* 2,  
624 61ra91-61ra91. 10.1126/scitranslmed.3001720.
- 625 17. Lam, K.-W.G., Jiang, P., Liao, G.J.W., Chan, K.C.A., Leung, T.Y., Chiu, R.W.K., and Lo,  
626 Y.M.D. (2012). Noninvasive Prenatal Diagnosis of Monogenic Diseases by Targeted  
627 Massively Parallel Sequencing of Maternal Plasma: Application to  $\beta$ -Thalassemia. *Clinical*  
628 *Chemistry* 58, 1467–1475. 10.1373/clinchem.2012.189589.
- 629 18. New, M.I., Tong, Y.K., Yuen, T., Jiang, P., Pina, C., Chan, K.C.A., Khattab, A., Liao, G.J.W.,  
630 Yau, M., Kim, S.-M., et al. (2014). Noninvasive prenatal diagnosis of congenital adrenal  
631 hyperplasia using cell-free fetal DNA in maternal plasma. *J Clin Endocrinol Metab* 99, E1022-  
632 1030. 10.1210/jc.2014-1118.
- 633 19. Parks, M., Court, S., Cleary, S., Clokie, S., Hewitt, J., Williams, D., Cole, T., MacDonald, F.,  
634 Griffiths, M., and Allen, S. (2016). Non-invasive prenatal diagnosis of Duchenne and Becker  
635 muscular dystrophies by relative haplotype dosage. *Prenat Diagn* 36, 312–320.  
636 10.1002/pd.4781.
- 637 20. Parks, M., Court, S., Bowns, B., Cleary, S., Clokie, S., Hewitt, J., Williams, D., Cole, T.,  
638 MacDonald, F., Griffiths, M., et al. (2017). Non-invasive prenatal diagnosis of spinal muscular  
639 atrophy by relative haplotype dosage. *Eur J Hum Genet* 25, 416–422. 10.1038/ejhg.2016.195.
- 640 21. Hudecova, I., Jiang, P., Davies, J., Lo, Y.M.D., Kadir, R.A., and Chiu, R.W.K. (2017).  
641 Noninvasive detection of F8 int22h-related inversions and sequence variants in maternal  
642 plasma of hemophilia carriers. *Blood* 130, 340–347. 10.1182/blood-2016-12-755017.
- 643 22. Chandler, N.J., Ahlfors, H., Drury, S., Mellis, R., Hill, M., McKay, F.J., Collinson, C., Hayward,  
644 J., Jenkins, L., and Chitty, L.S. (2020). Noninvasive Prenatal Diagnosis for Cystic Fibrosis:  
645 Implementation, Uptake, Outcome, and Implications. *Clin Chem* 66, 207–216.  
646 10.1373/clinchem.2019.305011.

- 647 23. Young, E., Bowns, B., Gerrish, A., Parks, M., Court, S., Clokie, S., Mashayamombe-  
648 Wolfgarten, C., Hewitt, J., Williams, D., Cole, T., et al. (2020). Clinical Service Delivery of  
649 Noninvasive Prenatal Diagnosis by Relative Haplotype Dosage for Single-Gene Disorders.  
650 *The Journal of Molecular Diagnostics* 22, 1151–1161. 10.1016/j.jmoldx.2020.06.001.
- 651 24. Chan, K.C.A., Jiang, P., Sun, K., Cheng, Y.K.Y., Tong, Y.K., Cheng, S.H., Wong, A.I.C.,  
652 Hudcovova, I., Leung, T.Y., Chiu, R.W.K., et al. (2016). Second generation noninvasive fetal  
653 genome analysis reveals de novo mutations, single-base parental inheritance, and preferred  
654 DNA ends. *Proc Natl Acad Sci U S A* 113, E8159–E8168. 10.1073/pnas.1615800113.
- 655 25. Sun, K., Jiang, P., Wong, A.I.C., Cheng, Y.K.Y., Cheng, S.H., Zhang, H., Chan, K.C.A., Leung,  
656 T.Y., Chiu, R.W.K., and Lo, Y.M.D. (2018). Size-tagged preferred ends in maternal plasma  
657 DNA shed light on the production mechanism and show utility in noninvasive prenatal testing.  
658 *Proc Natl Acad Sci U S A* 115, E5106–E5114. 10.1073/pnas.1804134115.
- 659 26. Stoler, N., and Nekrutenko, A. (2021). Sequencing error profiles of Illumina sequencing  
660 instruments. *NAR Genomics and Bioinformatics* 3, lqab019. 10.1093/nargab/lqab019.
- 661 27. Scotchman, E., Chandler, N.J., Mellis, R., and Chitty, L.S. (2020). Noninvasive Prenatal  
662 Diagnosis of Single-Gene Diseases: The Next Frontier. *Clin Chem* 66, 53–60.  
663 10.1373/clinchem.2019.304238.
- 664 28. Kong, A., Gudbjartsson, D.F., Sainz, J., Jonasdottir, G.M., Gudjonsson, S.A., Richardsson, B.,  
665 Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G., et al. (2002). A high-resolution  
666 recombination map of the human genome. *Nat Genet* 31, 241–247. 10.1038/ng917.
- 667 29. Kong, A., Thorleifsson, G., Gudbjartsson, D.F., Masson, G., Sigurdsson, A., Jonasdottir, A.,  
668 Walters, G.B., Jonasdottir, A., Gylfason, A., Kristinsson, K.Th., et al. (2010). Fine-scale  
669 recombination rate differences between sexes, populations and individuals. *Nature* 467,  
670 1099–1103. 10.1038/nature09525.
- 671 30. Hui, W.W.I., Jiang, P., Tong, Y.K., Lee, W.-S., Cheng, Y.K.Y., New, M.I., Kadir, R.A., Chan,  
672 K.C.A., Leung, T.Y., Lo, Y.M.D., et al. (2017). Universal Haplotype-Based Noninvasive  
673 Prenatal Testing for Single Gene Diseases. *Clin Chem* 63, 513–524.  
674 10.1373/clinchem.2016.268375.
- 675 31. Jang, S.S., Lim, B.C., Yoo, S.-K., Shin, J.-Y., Kim, K.-J., Seo, J.-S., Kim, J.-I., and Chae, J.H.  
676 (2018). Targeted linked-read sequencing for direct haplotype phasing of maternal DMD  
677 alleles: a practical and reliable method for noninvasive prenatal diagnosis. *Sci Rep* 8, 8678.  
678 10.1038/s41598-018-26941-0.
- 679 32. Lee, J.-S., Lee, K.B., Song, H., Sun, C., Kim, M.J., Cho, S.I., Lee, Y.K., Park, S.S., and Seong,  
680 M.-W. (2021). Noninvasive prenatal test of single-gene disorders by linked-read direct  
681 haplotyping: application in various diseases. *Eur J Hum Genet* 29, 463–470. 10.1038/s41431-  
682 020-00759-9.
- 683 33. Jiang, F., Liu, W., Zhang, L., Guo, Y., Chen, M., Zeng, X., Wang, Y., Li, Y., Xian, J., Du, B., et  
684 al. (2021). Noninvasive prenatal testing for  $\beta$ -thalassemia by targeted nanopore sequencing  
685 combined with relative haplotype dosage (RHDO): a feasibility study. *Sci Rep* 11, 5714.  
686 10.1038/s41598-021-85128-2.

- 687 34. Rabinowitz, T., Polsky, A., Golan, D., Danilevsky, A., Shapira, G., Raff, C., Basel-Salmon, L.,  
688 Matar, R.T., and Shomron, N. (2019). Bayesian-based noninvasive prenatal diagnosis of  
689 single-gene disorders. *Genome Res.* 29, 428–438. 10.1101/gr.235796.118.
- 690 35. Li, H., Du, B., Jiang, F., Guo, Y., Wang, Y., Zhang, C., Zeng, X., Xie, Y., Ouyang, S., Xian, Y.,  
691 et al. (2019). Noninvasive prenatal diagnosis of  $\beta$ -thalassemia by relative haplotype dosage  
692 without analyzing proband. *Mol Genet Genomic Med* 7. 10.1002/mgg3.963.
- 693 36. Browning, S.R., and Browning, B.L. (2011). Haplotype phasing: existing methods and new  
694 developments. *Nat Rev Genet* 12, 703–714. 10.1038/nrg3054.
- 695 37. Snyder, M.W., Adey, A., Kitzman, J.O., and Shendure, J. (2015). Haplotype-resolved genome  
696 sequencing: experimental methods and applications. *Nat Rev Genet* 16, 344–358.  
697 10.1038/nrg3903.
- 698 38. Vermeulen, C., Geeven, G., de Wit, E., Verstegen, M.J.A.M., Jansen, R.P.M., van  
699 Kranenburg, M., de Bruijn, E., Pulit, S.L., Kruisselbrink, E., Shahsavari, Z., et al. (2017).  
700 Sensitive Monogenic Noninvasive Prenatal Diagnosis by Targeted Haplotyping. *The American*  
701 *Journal of Human Genetics* 101, 326–339. 10.1016/j.ajhg.2017.07.012.
- 702 39. ESHRE PGT-M Working Group, Carvalho, F., Moutou, C., Dimitriadou, E., Dreesen, J.,  
703 Giménez, C., Goossens, V., Kakourou, G., Vermeulen, N., Zuccarello, D., et al. (2020).  
704 ESHRE PGT Consortium good practice recommendations for the detection of monogenic  
705 disorders†. *Human Reproduction Open* 2020, hoaa018. 10.1093/hropen/hoaa018.
- 706 40. Bianchi, D.W. (2018). Cherchez la femme: maternal incidental findings can explain discordant  
707 prenatal cell-free DNA sequencing results. *Genetics in Medicine* 20, 910–917.  
708 10.1038/gim.2017.219.
- 709 41. Lannoo, L., Lenaerts, L., Van Den Bogaert, K., Che, H., Brison, N., Devriendt, K., Amant, F.,  
710 Vermeesch, J.R., and Van Calsteren, K. (2021). Non-invasive prenatal testing suggesting a  
711 maternal malignancy: What do we tell the prospective parents in Belgium? *Prenatal Diagnosis*  
712 41, 1264–1272. 10.1002/pd.6031.
- 713 42. Benn, P., Plon, S.E., and Bianchi, D.W. (2019). Current Controversies in Prenatal Diagnosis  
714 2: NIPT results suggesting maternal cancer should always be disclosed. *Prenatal Diagnosis*  
715 39, 339–343. 10.1002/pd.5379.
- 716
- 717



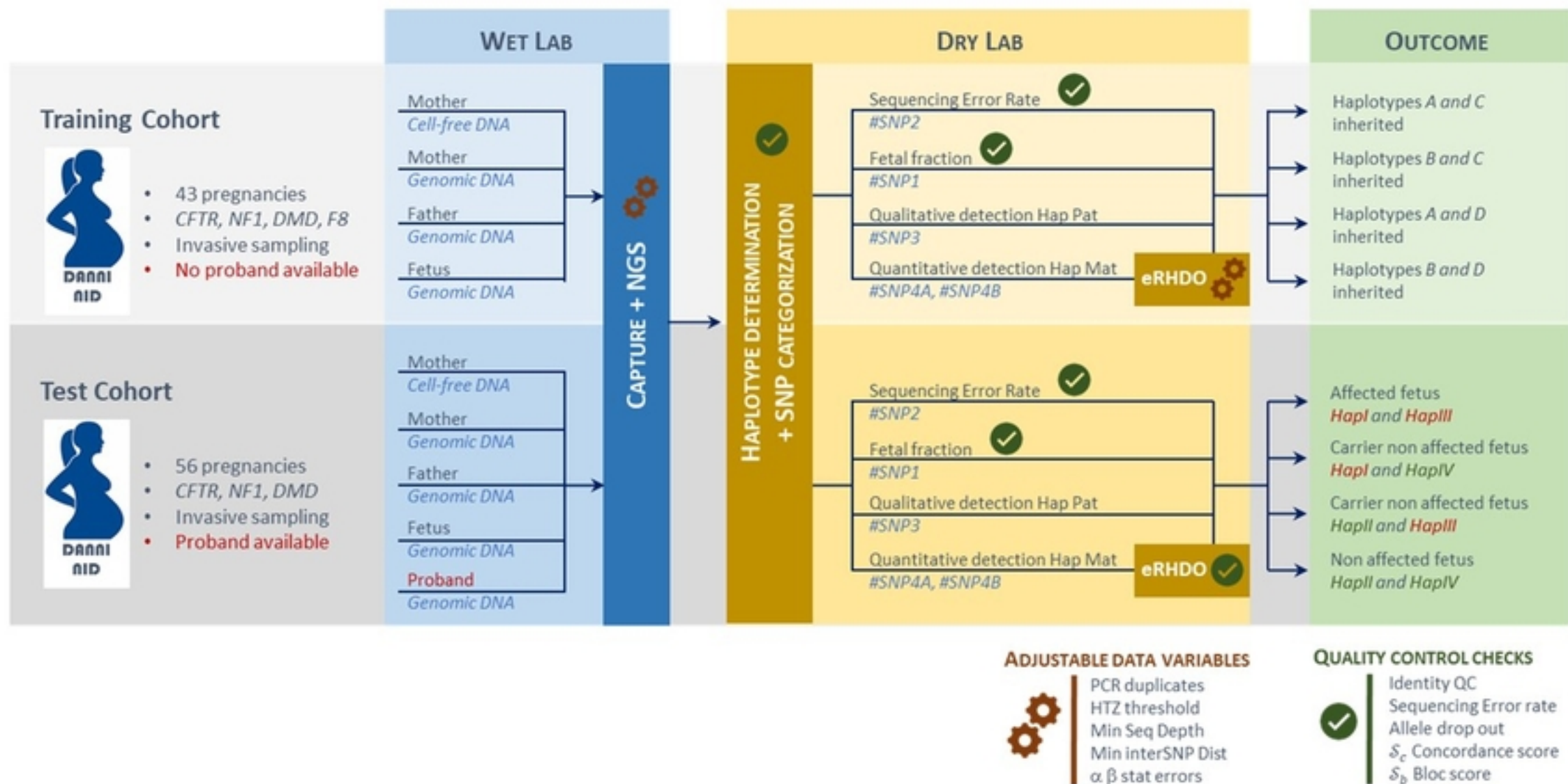


Figure 1

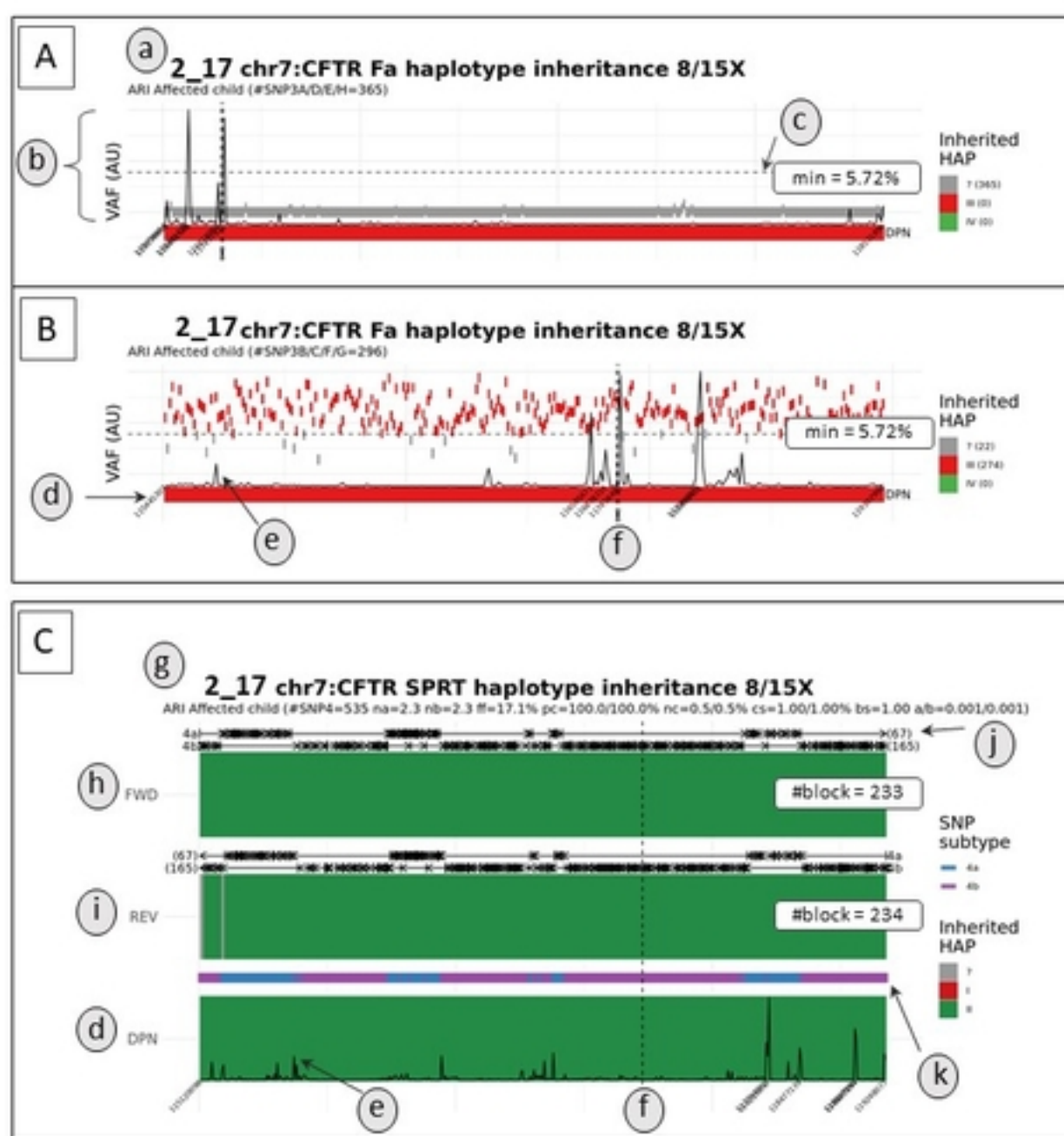


Figure 2

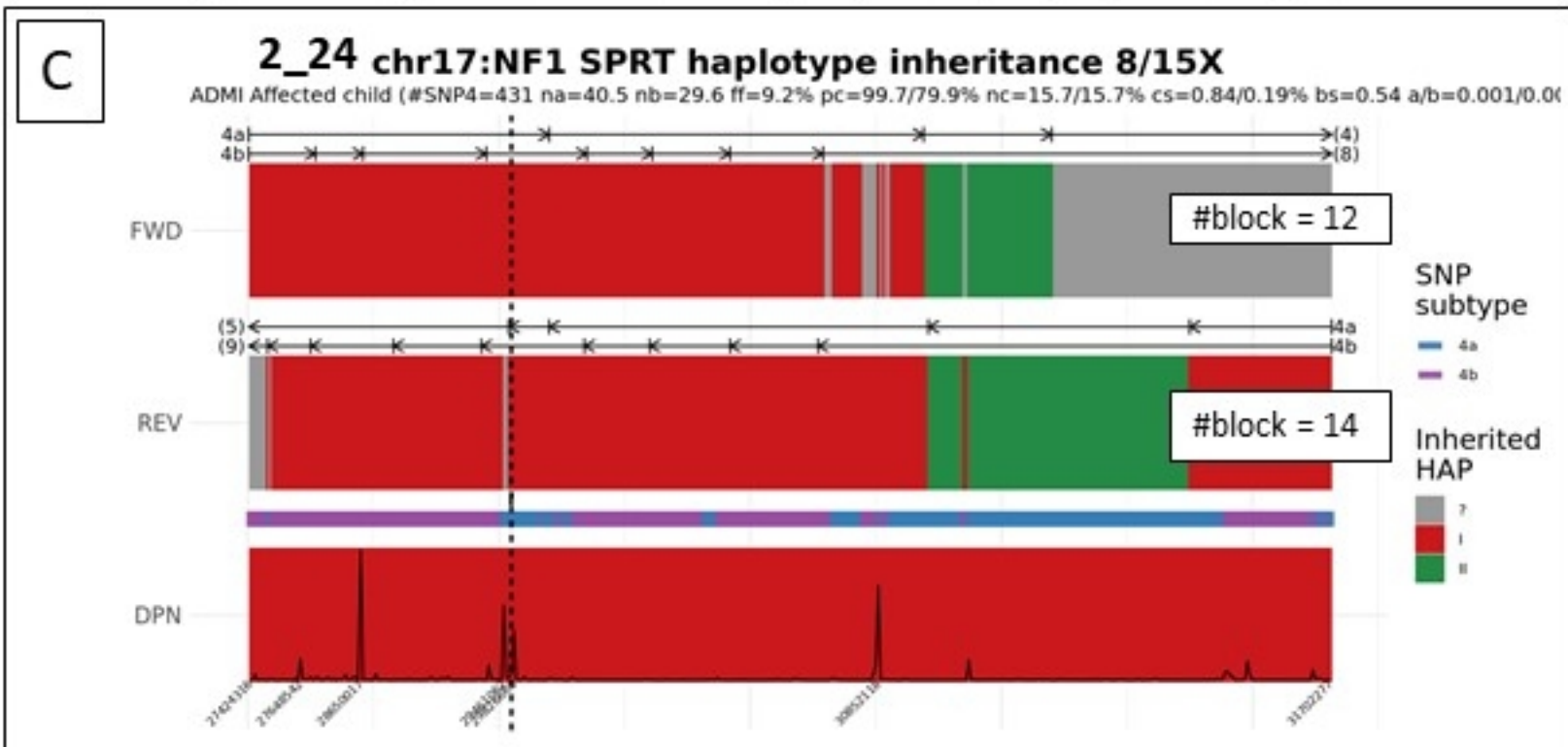
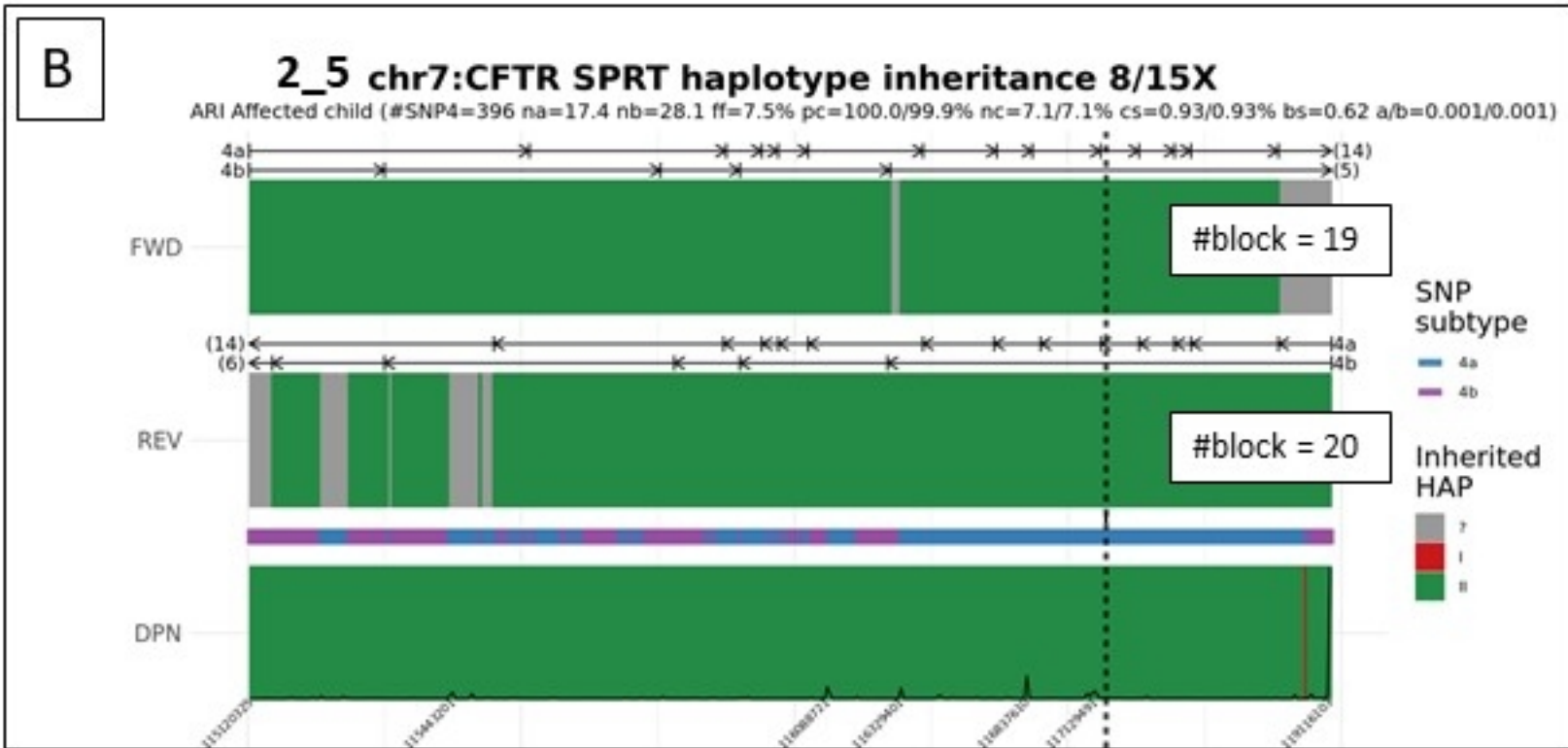
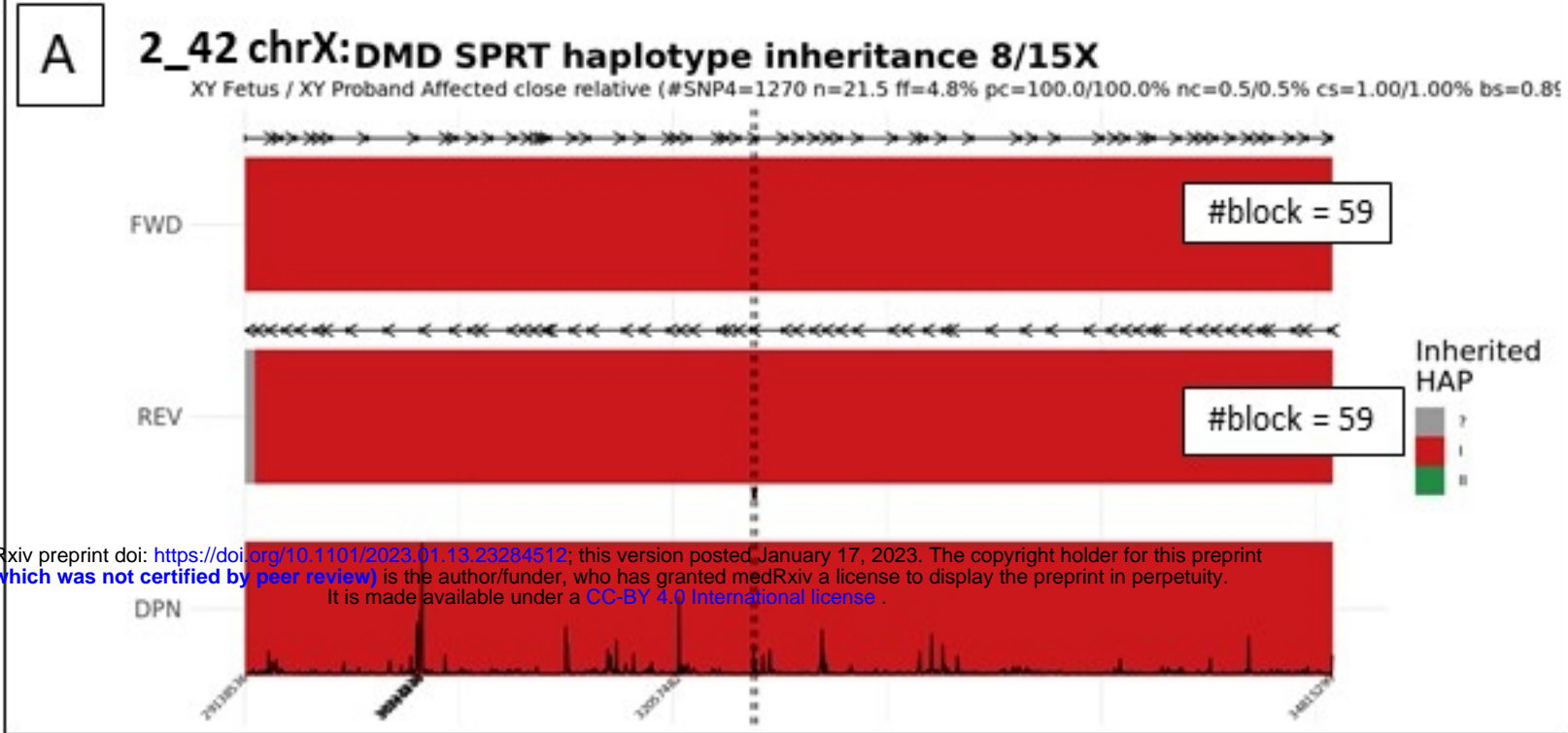


Figure 3