



**HAL**  
open science

# Bayesian inference of parental allele inheritance in fetus for noninvasive prenatal diagnosis

Ghislain Durif, Cathy Liautard-Haag, Marie-Claire Vincent

► **To cite this version:**

Ghislain Durif, Cathy Liautard-Haag, Marie-Claire Vincent. Bayesian inference of parental allele inheritance in fetus for noninvasive prenatal diagnosis. StatLearn 2023, Apr 2023, Montpellier, France. , <https://statlearn.sciencesconf.org/459803>. hal-04059861

**HAL Id: hal-04059861**

**<https://hal.science/hal-04059861>**

Submitted on 5 Apr 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Bayesian inference of parental allele inheritance in fetus for noninvasive prenatal diagnosis

Ghislain Durif<sup>1,2</sup> Cathy Liautard-Haag<sup>3</sup> Marie-Claire Vincent<sup>3,4</sup>

<sup>1</sup>IMAG, Université de Montpellier, CNRS UMR 5149, Montpellier, France

<sup>2</sup>LBMC, ENS de Lyon, CNRS UMR 5239, Inserm U1293, Université Claude Bernard Lyon 1, Lyon, France

<sup>3</sup>Laboratoire de Génétique Moléculaire, Institut Universitaire de Recherche Clinique, Université de Montpellier, CHU Montpellier, Montpellier, France

<sup>4</sup>PhyMedExp, Université de Montpellier, CNRS UMR 9214, INSERM U1046, Montpellier, France

## More info

Publication [1]



<https://hal.science/hal-03716132>

Source code



<https://github.com/gdurif/nipd>

Contact



<https://gdurif.perso.math.cnrs.fr>

## The problem (c.f. figure 1)

Noninvasive prenatal diagnosis (NIPD) [2] used to detect pathogenic genetic variation potentially transmitted to the fetus by a parent at risk without a hazardous invasive prenatal diagnosis (PND), i.e. risk of miscarriage

NIPD successfully applied for single-gene disorders [3] but still challenging for triplet-repeat expansion diseases, a.k.a. trinucleotide repeat disorder (e.g., myotonic dystrophy type 1, Huntington's disease, fragile X syndrome)

Fetal genotyping using circulating cell-free fetal DNA (cff-DNA) from maternal blood may not be able to detect complex genetic patterns in the fetal DNA that could have been inherited from a parent at risk

## The proposed method (c.f. figure 1)

Directly infer which haplotypes among the pair of parental homologous chromosomes have been inherited by the fetus

→ to determine if the haplotype region carrying the pathogenic variation was transmitted to the fetus or not

## The data (c.f. figure 2)

- parental phased haplotypes for the targeted chromosomal region
- maternal plasma cell-free DNA (cf-DNA) genotype for the targeted chromosomal region

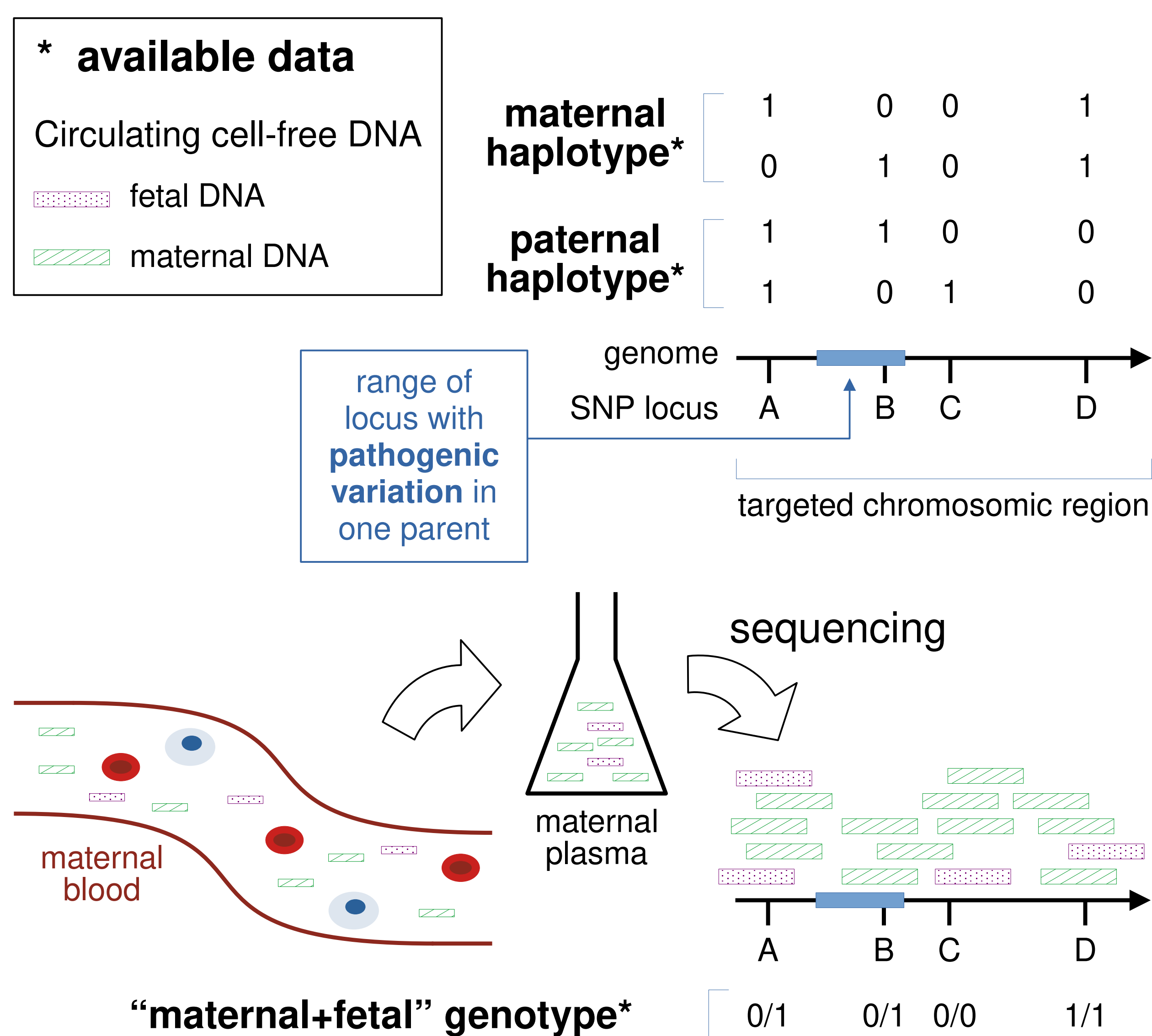


Figure 2. Available data to infer the parental allele origin in the fetus

## The perspectives

- Integrate the fetal fraction (=proportion of fetal DNA in sampled maternal plasma DNA) estimation [4] in the Bayesian model
- Run a larger-scale study with a different phasing protocol

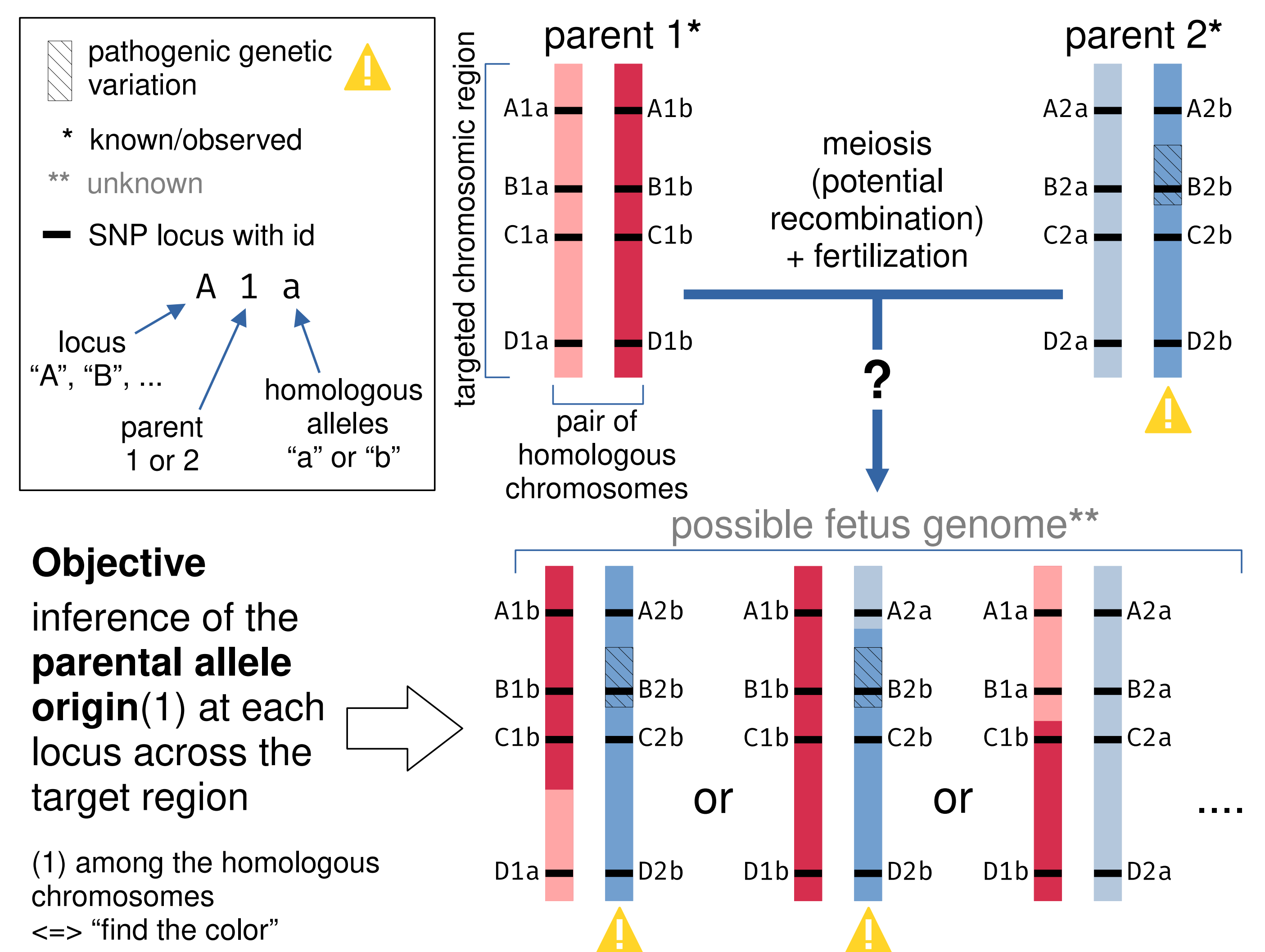


Figure 1. Inheritance pattern of the pathogenic genetic variation

## The Bayesian model

Output:  $O_\ell = (\text{mat}X, \text{pat}Y)$  fetal allele origin among the pair of homologous haplotypes from each parent ( $\text{mat}$  = maternal,  $\text{pat}$  = paternal and  $X, Y \in \{1, 2\}$ ) at locus  $\ell$

Objective: infer  $(O_1, \dots, O_\ell, \dots, O_L)$  over all locus inside the targeted region

Tool: estimate the full posterior jointly over all locus

$$P(O_1, \dots, O_\ell, \dots, O_L | \text{data})$$

Issue: impossible because of combinatorial cost (no independence between locus)

Workaround: use a Markov chain Monte Carlo (MCMC) algorithm to estimate the full posterior, specifically a Gibbs sampler based on iterative sampling from the conditional marginal posterior

$$P(O_\ell | \text{data}, O_{\ell-1}) \sim P(\text{data at locus } \ell | O_\ell) \times P(O_\ell | O_{\ell-1})$$

- $P(O_\ell | O_{\ell-1})$  = transition probability (depending on recombination probability and potential phasing error in both parents between locus  $\ell$  and  $\ell - 1$ )
- $P(\text{data at locus } \ell | O_\ell)$  = conditional likelihood that can be explicitly derived from fetal genotype model using read counts and fetal fraction estimation (as in the locus-independent fetal genotyping Hoobari model [3])

## The results

Analysis: a cohort of 14 families, each including one parent with a pathogenic genetic variation	Non conclusive <sup>(2)</sup> 9/14
Comparison between Bayesian NIPD and invasive PND results (=ground truth)	Accordance 4/14
	Error 1/14

(2) because of either poor polymorphism or phasing quality

## References

- C. Liautard-Haag et al. en. In: *Scientific Reports* 12.1 (July 2022), p. 11423. ISSN: 2045-2322. DOI: 10.1038/s41598-022-15307-2.
- Lisa Hui. eng. In: *Methods in Molecular Biology (Clifton, N.J.)* 1885 (2019), pp. 45–58. ISSN: 1940-6029. DOI: 10.1007/978-1-4939-8889-1\_3.
- Tom Rabinowitz et al. In: *Genome Research* 29.3 (Mar. 2019), pp. 428–438. ISSN: 1088-9051. DOI: 10.1101/gr.235796.118.
- Landon L. Chan and Peiyong Jiang. eng. In: *Clinical Biochemistry* 48.15 (Oct. 2015), pp. 962–975. ISSN: 1873-2933. DOI: 10.1016/j.clinbiochem.2015.04.022.