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

Ghislain Durif*†[1,2], Cathy Liautard Haag[3], and Marie-Claire Vincent[3,4]

[1]Institut Montpelliérain Alexander Grothendieck – Centre National de la Recherche Scientifique : UMR5149, Université de Montpellier, LBMC – France

[2]Laboratoire de biologie et modélisation de la cellule – École Normale Supérieure - Lyon, Université Claude Bernard Lyon 1, Institut National de la Santé et de la Recherche Médicale : U1210, Centre National de la Recherche Scientifique : UMR5239, Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique – France

[3]Laboratoire de Génétique Moléculaire, Institut Universitaire de Recherche Clinique, Université de Montpellier, CHU Montpellier – CHRU Montpellier – France

[4]Physiologie médecine expérimentale du Cœur et des Muscles [U 1046] – Institut National de la Santé et de la Recherche Médicale : U1046, Centre National de la Recherche Scientifique : UMR9214, Université de Montpellier : UMR9214, Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Université de Montpellier – France

Abstract

The field of noninvasive prenatal diagnosis (NIPD) has undergone significant progress over the last decade. Direct haplotyping has been successfully applied for NIPD of a few single-gene disorders. However, technical issues remain for triplet-repeat expansion diseases (a.k.a. trinucleotide repeat disorder). Developping an NIPD approach for couples at risk of transmitting dynamic mutations is thus challenging but crucial. For instance, fetal genotyping using circulating cell-free fetal DNA (cff-DNA) from maternal blood might not be able to detect complex genetic patterns in the fetal DNA that could have been inherited from a parent at risk. In such family, a workaround would be to directly detect which haplotypes among the pair of parental homologous chromosomes have been inherited by the fetus, for an entire targeted region of the genome. In combination with haplotype phasing of the parent(s) at risk, it would allow to determine if the haplotype region carrying the pathogenic variation was transmitted to the fetus or not.

We present a Bayesian approach that is able, not only to infer the fetal genotype, but more importantly to directly infer the fetal allele origin from the parental phased haplotypes at each locus in a target chromosome region. In particular, our model aims to identify the parental haplotype of origin for the genetic material inherited by the fetus. To do so, only haplotype data from both parents and genotype data from circulating cell-free DNA (cf-DNA) in maternal plasma (i.e. a mix of maternal and fetal DNA) are used. On contrary to existing fetal genotyping models which consider all loci independently, we infer the allele origin jointly on all loci in the targeted region. Because of combinatorial issues, we cannot directly derive the joint posterior but we rather use a Markov chain Monte Carlo (MCMC)

*Speaker

†Corresponding author: ghislain.durif@ens-lyon.fr

procedure (specifically a Gibbs sampler) to estimate the full posterior over the entire region, and a maximum a posteriori (MAP) estimation to determine the allele inheritance patterns over the entire region.

We performed analyses using blood samples from families with Huntington's disease or myotonic dystrophy type 1. We were able to perform the Bayesian inference of parental haplotype transmissions for five fetuses. The predicted variant status of four of these fetuses was in agreement with the invasive prenatal diagnosis findings. Conversely, no conclusive result was obtained for the NIPD of fragile X syndrome. Although improvements should be made to achieve clinically acceptable accuracy, our study shows that linked-read sequencing and parental haplotype phasing can be successfully used for NIPD of triplet-repeat expansion diseases.

Our approach is implemented as a Python package with a command line interface (CLI). The source code can be found in a dedicated repository (https://github.com/gdurif/nipd), and the corresponding work has been published (https://hal.science/hal-03716132).

Keywords: noninvasive prenatal diagnosis, bayesian inference, MCMC, Gibbs sampling, haplotyping, genotyping