

# Noninvasive prenatal exome sequencing inefficient for detecting single-gene disorders – problems and possible solutions

Dayne L Filer<sup>1,2,\*</sup>, Piotr A Mieczkowski<sup>1</sup>, Alicia Brandt<sup>1</sup>, Kelly L Gilmore<sup>3</sup>, Bradford C Powell<sup>1,2</sup>, Jonathan S Berg<sup>1</sup>, Kirk C Wilhelmsen<sup>1,2,4</sup>, and Neeta L Vora<sup>1,3</sup>

<sup>1</sup>Department of Genetics, School of Medicine, UNC Chapel Hill, Chapel Hill, NC

<sup>2</sup>Renaissance Computing Institute, Chapel Hill, NC

<sup>3</sup>Department of Obstetrics & Gynecology, School of Medicine, UNC Chapel Hill, Chapel Hill, NC

<sup>4</sup>Department of Neurology, School of Medicine, UNC Chapel Hill, Chapel Hill, NC

\*Corresponding author: 100 Europa Drive Suite 540, Chapel Hill, NC 27517,  
dayne\_filer@med.unc.edu

## 1 Abstract

2 What's already known about this topic?

- 3 • Sequencing-based noninvasive testing can detect large copy number abnormalities and some auto-  
4 somal dominant single-gene disorders
- 5 • Exome sequencing (ES) on fetal samples provides 20% diagnostic yield for structural abnormalities  
6 after normal karyotype & microarray

7 What does this study add?

- 8 • ES on cell-free DNA in three gravid patients with suspected genetic disease in the fetus
- 9 • We demonstrate broad sequencing approaches are limited by sampling and technical difficulties,  
10 concluding broad sequencing is currently inappropriate for noninvasive testing

## 11 Letter

12 The beneficial health outcomes from newborn screening programs (NBS) are indisputable. We envision  
13 future NBS will begin with prenatal genetic testing to enable care in the immediate newborn period,  
14 and open up new possibilities for *in utero* and genetic therapies. During pregnancy placental DNA is  
15 released into maternal circulation, enabling noninvasive interrogation of fetal genetics (noninvasive pre-  
16 natal testing, NIPT). NIPT has a well-established clinical utility in screening for common chromosomal  
17 abnormalities such as Down syndrome with high sensitivity and specificity.<sup>1</sup> More recently, efforts have  
18 demonstrated sequencing-based testing for *de novo* pathogenic variants in a list of 30 genes associated  
19 with dominant Mendelian disorders<sup>2</sup> and PCR-based testing for a small number of recessive Mendelian  
20 disorders.<sup>3</sup> To date, no one has reported reliable fetal genotyping purely from maternal cell-free DNA  
21 using a sequencing-based approach.

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

	GA	Clinical findings	Genetic diagnosis	FF	Depth	%Dup	%Filt
1	32w2d	5 prior pregnancies affected with X-linked recessive Menke's syndrome	Menke's syndrome; del. ATP7A exon 1	0.117	241	42.8	21.96
2	24w5d	Fetal sonogram at 21w5d showed femoral bowing with shortened length (<3% for GA) bilaterally	Osteogenesis imperfecta type VIII; P3H1 c.1120G>T (rs140468248)	0.122	152	33.32	22.09
3	34w0d	Fetal sonogram at 19w0d showed bilateral club foot with bilateral upper limb arthrogryposis	None, to date, despite exome and genome sequencing of newborn	0.169	330	53.67	32.65

Table 1: Case summaries. GA: gestational age at the time of blood draw for cfES. FF: estimated fetal fraction. Depth: median depth used to estimate genotypes (does not include duplicated/filtered reads). %Dup: percentage of total mapped read pairs discarded as PCR and/or optical duplicates. %Filt: percentage of total mapped read pairs discarded for improper pairing and/or mapping quality.

22 To begin NBS with prenatal genetic testing, we believe we first need a reliable noninvasive test only  
23 requiring a maternal sample. Others could reasonably argue the availability of carrier screening, and  
24 the immeasurably small risk of invasive testing,<sup>4</sup> removes the need for the noninvasive test. Such an  
25 argument, however, dismisses (1) the ethical and practical issues surrounding the necessity of involving  
26 the biological father, (2) the fact that many genetic disorders arise due to *de novo* mutations, and  
27 (3) the understandable fear and apprehension around invasive testing (especially for rare conditions).  
28 Additionally, we believe the prenatal diagnosis community should focus work on sequencing-based (as  
29 opposed to PCR-based) approaches. Sequencing-based approaches generalize across disorders more easily  
30 than PCR-based approaches, multiplex to a degree not feasible using PCR, and will only continue to  
31 decrease in cost.

32 Previously, Kitzman et al. performed whole-genome fetal sequencing from maternal plasma by combining  
33 whole genome data from the father, mother, and maternal plasma,<sup>5</sup> but illustrate the cost-infeasibility in  
34 their subsequent review article and suggest more targeted approaches such as exome sequencing (ES).<sup>6</sup>  
35 As an exploratory exercise, we performed ES on cell-free DNA (cfES) from three pregnant women with  
36 singleton fetuses.

37 Briefly, we collected cell-free DNA from maternal plasma, prepared sequencing libraries for the Illumina  
38 platform, and performed exome capture using the IDT xGen Exome Research Panel v1.0 (Cases 1 & 2)  
39 or Agilent SureSelect Human All Exon v7 (Case 3). All participants were consented and enrolled at UNC  
40 Hospitals by certified genetic counselors with approval from the UNC Institutional Review Board (IRB  
41 Number: 18-2618); we do not include any identifying information in this manuscript. We processed  
42 the data using a novel analytic pipeline developed in Snakemake using Anaconda environments for  
43 reproducibility. Sequencing reads were aligned to hg38 (excluding alternate contigs) using BWA-MEM,  
44 then base quality scores were re-calibrated using GATK4. We only retained non-duplicate, properly-  
45 paired reads with unambiguous mapping and mapping quality >30 for each read. We called variants  
46 using the bcftools software requiring basepair quality scores >20, 5 alternate allele-supporting fragments,  
47 and 80 total fragments. Analyses were restricted to the regions overlapping between the IDT and Agilent  
48 capture platforms. Using the identified single-nucleotide variants, we applied a novel empirical Bayesian  
49 procedure to estimate the fetal fraction (FF; the proportion of placental/fetal to maternal sequencing  
50 reads). We then estimated fetal and maternal genotypes using a maximal likelihood model incorporating  
51 the FF estimate and observed proportion of minor allele (alternate) reads (PMAR). Full analytic pipeline  
52 available upon request.

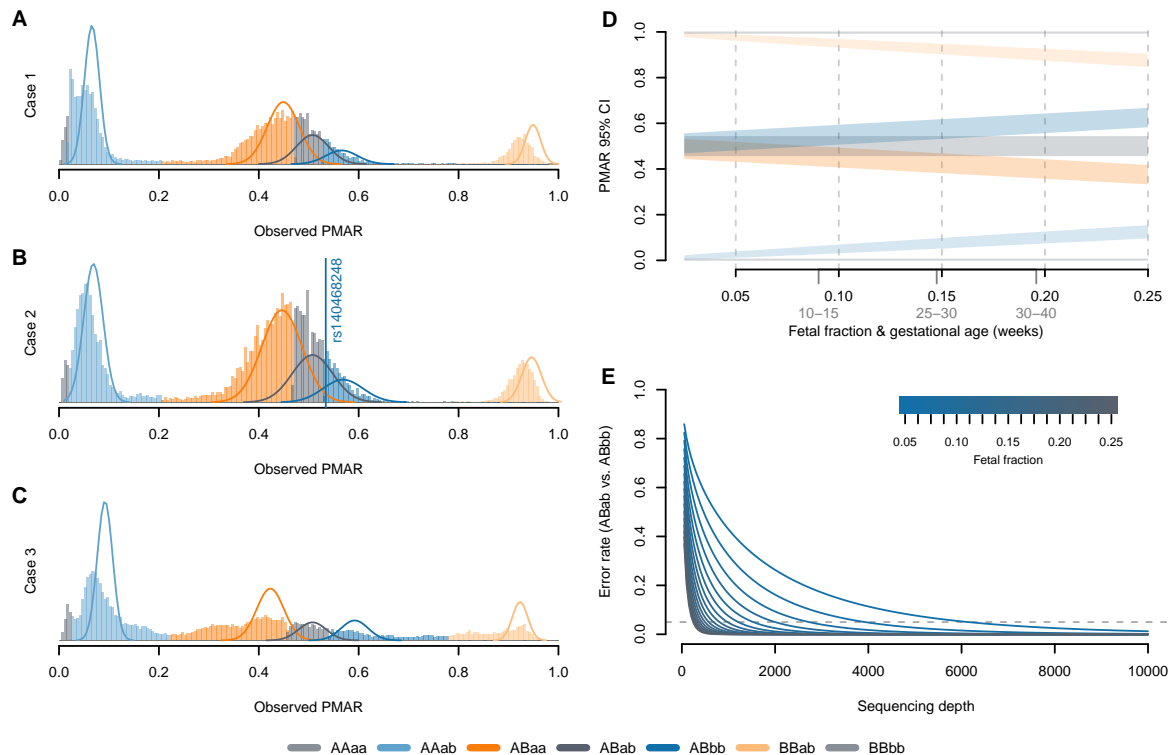


Figure 1: [A-C] Distribution of observed PMAR values for the three cases across the possible maternal-fetal genotype pairs. Uppercase letters give the estimated maternal genotype, lowercase letters give the estimated fetal genotype; ‘A/a’ indicates the reference allele, ‘B/b’ indicates the alternate allele. Solid lines show the normal approximation for the theoretical distribution of binomial probabilities, given the frequency of the estimated genotypes. The vertical line in [B] shows the observed PMAR for the known pathogenic variant, rs140468248. [D] 95% confidence intervals on the binomial proportions for possible maternal-fetal genotype pairs across increasing fetal fractions; represents a sequencing depth of 500x. Average fetal fractions by gestational age (in weeks) given in light gray.<sup>7</sup> [E] Expected misclassification rate (Weitzman overlapping coefficient; i.e. the area of overlapping distributions in [D]) considering ABab versus ABbb as a function of sequencing depth and fetal fraction. The dashed horizontal line shows 5% error. The theoretical error rates for ABab vs ABaa are symmetric and equal; however, the frequency of errors will depend on the population frequency of the reference versus alternate allele.

53 Table 1 lists the known genetic diagnoses for the three cases presented. Genetic counselors recruited  
 54 the three participants with investigators and cfES analysis blinded to the eventual genetic diagnoses.  
 55 In Cases 1 & 2, specific gene sequencing based on family history and sonographic findings, respectively,  
 56 provided genetic diagnoses. To date, Case 3 does not have a specific genetic diagnosis despite whole-  
 57 genome sequencing of the newborn and ES on the trio. Afterwards, we learned the mother in Case 1  
 58 carries a deletion of exon 1 in the gene most-often responsible for Menke’s syndrome (ATP7A). Neither  
 59 exome capture platform targets ATP7A exon 1; therefore, cfES could not have identified the diagnosis  
 60 for Case 1 with the platform used. In Case 2, we identified the causal variant using cfES. In this case,  
 61 we correctly genotyped the fetus, but lacked the power to make the genotyping call with any level of  
 62 confidence acceptable for clinical use (fig. 1B, note the widely-overlapping distributions at the causal  
 63 variant). We did not identify any known pathogenic variants in the sequencing of Case 3, and despite  
 64 performing whole-genome sequencing on the newborn, we still do not have a genetic diagnosis for the  
 65 family.

66 Without the ability to reliably exclude maternal DNA fragments, noninvasive sequencing-based methods  
 67 to genotype the fetus either require additional sequencing of parental samples or distinguishing genotypes  
 68 by the proportion of minor allele reads (PMAR). Here, we make no attempt to utilize parental genetic

69 information and demonstrate the difficulty of inferring the genotypes directly from the PMAR. We model  
70 the PMAR as a binomial proportion; given the fetal fraction, one can prove the true PMAR defines the  
71 maternal-fetal genotype combination.

72 For illustration, consider watching two people randomly place balls into an urn. We know each person  
73 either has all white balls, all black balls, or equal numbers of white and black balls; we also know the  
74 number, but not the color of balls each person places. We count 60 black balls and 40 white balls in  
75 the urn. Given Person A placed 80 balls, the maximum likelihood estimate suggests Person A had equal  
76 white and black balls ( $0.5 \times 80 = 40$ ) and Person B had all black balls ( $1.0 \times 20 = 20$ ).

77 The theoretical bounds of the binomial distribution, therefore, confine our ability to discriminate maternal-  
78 fetal genotypes. Using the normal approximation for the binomial variance (valid when the number of  
79 observations (sequencing depth),  $N$ , times the binomial proportion (PMAR),  $p$  is greater than 10), we  
80 can clearly explain the poor results we observed (fig. 1D-E). At sequencing depths up to 500x, the 95%  
81 confidence intervals on PMAR distributions still overlap for fetal fractions up to roughly 0.17 (fig. 1D).  
82 When we calculate the degree of distribution overlap (a proxy for classification error rate), we see required  
83 sequencing depths in excess of 8,000x for low fetal fraction samples.

84 The sequencing herein likely suffers from three problems: (1) inadequate sequencing depth; (2) biased  
85 PMAR values from the removal of duplicate reads; (3) errors in sequencing and/or PCR. We have already  
86 illustrated the inadequate depth, but emphasize that the theoretical results we present speak to the final  
87 depths (not the raw sequencing depth). In our three cases, we excluded over half the reads taken off the  
88 sequencer due to sequencing quality thresholds (table 1). We observe the evidence of problems (2) and  
89 (3) by observing the high proportion of both duplicate reads and PMAR values outside the theoretic  
90 distributions. Additionally, for Case 3 only, we can assess the accuracy of the genotype estimates. In  
91 Case 3, we have ES from newborn cord-blood; if we examine variants from both the cfES and ES of  
92 newborn cord-blood, we observe a 50.9% genotyping accuracy (data not shown).

93 Typical sequencing workflows start with randomly fragmenting DNA molecules to build sequencing li-  
94 braries. Standard bioinformatic practices suggest we remove read-pairs with identical endpoints, because  
95 the duplicate read-pairs more likely represent PCR amplification of a single molecule than two molecules  
96 with the same fragmentation. Cell-free DNA molecules are shorter than nuclear DNA, not requiring  
97 manual fragmentation, and have a non-random distribution of endpoints.<sup>8</sup> Therefore, compared to stan-  
98 dard sequencing libraries, the likelihood of observing true duplicates in cell-free libraries increases and  
99 we cannot necessarily assume duplicates represent PCR amplification. However, for this work we have  
100 no way of differentiating reads representing true duplicate molecules versus PCR duplicates and thus  
101 excluded duplicate reads from our analysis.

102 To solve the above issues, we are currently developing and testing a more targeted approach with se-  
103 quencing depths in excess of 10,000x and unique molecular identifiers to estimate accurately sequencing  
104 errors and differentiate true versus artifactual duplicate reads. Given the depth requirements for esti-  
105 mating fetal genotypes by the PMAR, and the challenge of variants of uncertain clinical significance, we  
106 advocate against broad sequencing modalities on noninvasive samples. Despite the challenges ranged by  
107 this letter, we have good reason to believe we can assess hundreds to thousands of basepairs, rather than  
108 the tens of millions targeted in ES, economically and reliably. In doing so, we hope to foster population-  
109 level screening for Mendelian disorders during the prenatal period and, ultimately, unlock new avenues  
110 in the treatment of these disorders.

## Acknowledgements

We thank Dr. James Evans for providing review and feedback of this manuscript. We especially thank the authors of the software packages that we did not have room in this brief letter to cite directly. Neeta Vora and this work was supported by NICHD (K23HD088742). Dayne Filer was supported by NICHD (F30HD101228) and by NIGMS (5T32GM067553).

## References

- [1] F L Mackie, K Hemming, S Allen, R K Morris, and M D Kilby. The accuracy of cell-free fetal dna-based non-invasive prenatal testing in singleton pregnancies: a systematic review and bivariate meta-analysis. *BJOG*, 124(1):32–46, Jan 2017.
- [2] Jinglan Zhang, Jianli Li, Jennifer B Saucier, Yanming Feng, Yanjun Jiang, Jefferson Sinson, Anne K McCombs, Eric S Schmitt, Sandra Peacock, Stella Chen, Hongzheng Dai, Xiaoyan Ge, Guoli Wang, Chad A Shaw, Hui Mei, Amy Breman, Fan Xia, Yaping Yang, Anne Purgason, Alan Pourpak, Zhao Chen, Xia Wang, Yue Wang, Shashikant Kulkarni, Kwong Wai Choy, Ronald J Wapner, Ignatia B Van den Veyver, Arthur Beaudet, Sheetal Parmar, Lee-Jun Wong, and Christine M Eng. Non-invasive prenatal sequencing for multiple mendelian monogenic disorders using circulating cell-free fetal dna. *Nat Med*, Jan 2019.
- [3] David S Tsao, Sukrit Silas, Brian P Landry, Nelda P Itzep, Amy B Nguyen, Samuel Greenberg, Celeste K Kanne, Vivien A Sheehan, Rani Sharma, Rahul Shukla, Prem N Arora, and Oguzhan Atay. A novel high-throughput molecular counting method with single base-pair resolution enables accurate single-gene nipt. *Sci Rep*, 9(1):14382, Oct 2019.
- [4] L J Salomon, A Sotiriadis, C B Wulff, A Odibo, and R Akolekar. Risk of miscarriage following amniocentesis or chorionic villus sampling: systematic review of literature and updated meta-analysis. *Ultrasound Obstet Gynecol*, 54(4):442–451, Oct 2019.
- [5] Jacob O Kitzman, Matthew W Snyder, Mario Ventura, Alexandra P Lewis, Ruolan Qiu, Lavone E Simmons, Hilary S Gammill, Craig E Rubens, Donna A Santillan, Jeffrey C Murray, Holly K Tabor, Michael J Bamshad, Evan E Eichler, and Jay Shendure. Noninvasive whole-genome sequencing of a human fetus. *Sci Transl Med*, 4(137):137ra76, Jun 2012.
- [6] Matthew W Snyder, LaVone E Simmons, Jacob O Kitzman, Donna A Santillan, Mark K Santillan, Hilary S Gammill, and Jay Shendure. Noninvasive fetal genome sequencing: a primer. *Prenat Diagn*, 33(6):547–554, Jun 2013.
- [7] Sarah L Kinnings, Jennifer A Geis, Eyad Almasri, Huiquan Wang, Xiaojun Guan, Ron M McCullough, Allan T Bombard, Juan-Sebastian Saldivar, Paul Oeth, and Cosmin Deciu. Factors affecting levels of circulating cell-free fetal dna in maternal plasma and their implications for noninvasive prenatal testing. *Prenat Diagn*, 35(8):816–822, Aug 2015.
- [8] K C Allen Chan, Peiyong Jiang, Kun Sun, Yvonne K Y Cheng, Yu K Tong, Suk Hang Cheng, Ada I C Wong, Irena Hudcová, Tak Y Leung, Rossa W K Chiu, and Yuk Ming Dennis Lo. Second generation noninvasive fetal genome analysis reveals de novo mutations, single-base parental inheritance, and preferred dna ends. *Proc Natl Acad Sci U S A*, 113(50):E8159–E8168, Dec 2016.

## Word count

File: ResearchLetter.tex

Encoding: utf8

Words in text: 1577

Words in headers: 18

Words outside text (captions, etc.): 235

Number of headers: 5

Number of floats/tables/figures: 2

Number of math inlines: 5

Number of math displayed: 0

Subcounts:

text+headers+captions (#headers/#floats/#inlines/#displayed)

57+13+0 (1/0/0/0) \_top\_

79+1+0 (1/0/0/0) Section: Abstract

1385+1+235 (1/2/5/0) Section: Letter

56+1+0 (1/0/0/0) Section: Acknowledgements

0+2+0 (1/0/0/0) Section: Word count