Novel sequencing-based framework for non-invasive fetal genotyping

Dayne L Filer[1,2], Kirk C Wilhelmsen[1,2,3], Bradford C Powell[1,2], Kelly L Gilmore[4], Cynthia M Powell[1], Jonathan S Berg[1], Neeta L Vora[4]

1) Department of Genetics, UNC School of Medicine; 2) Renaissance Computing Institute;
3) Department of Neurology, UNC School of Medicine; 4) Department of Obstetrics and Gynecology, UNC School of Medicine

Approach

Figure 1: Using maternal cell-free DNA, make probabilistic genotype calls based on fetal fraction estimates and observed base-pair proportions. In theory, given the proportion of fetal to maternal DNA (fetal fraction), a true estimate of the allele frequency defines the maternal and fetal genotypes at that locus. We are employing two genotyping panels to interrogate selectively specific genetic variants: (1) an off-the-shelf panel of probes with even density throughout the genome to define the fetal fraction and detect chromosomal abnormalities; (2) a custom panel of probes covering the most common variants in diverse populations.

Notation & expected values

Represent maternal and fetal genotype pairs, given by the random variable G, with capital and lowercase letters, where 'A' and 'B' represent the major and minor alleles (e.g. 'AAab' represents the fetus uniquely heterozygous for the minor allele). Let X, Y be random variables for major and minor allele read counts. Define the fraction of fetal DNA and proportion of minor allele reads (PMAR) as the random variables F and M. Then, by definition, $\mathrm{E}[M] = \mathrm{E}[Y/(X+Y)]$. It's easily proven:

$$\mathrm{E}[M|G = \mathrm{AAab}, F = f] = \frac{f}{2} \quad (1)$$

$$\mathrm{E}[M|G = \mathrm{ABaa}, F = f] = \frac{1-f}{2} \quad (2)$$

$$\mathrm{E}[M|G = \mathrm{ABab}, F = f] = \frac{1}{2} \quad (3)$$

$$\mathrm{E}[M|G = \mathrm{ABbb}, F = f] = \frac{1+f}{2} \quad (4)$$

$$\mathrm{E}[M|G = \mathrm{BBab}, F = f] = \frac{2-f}{2} \quad (5)$$

Binomial distribution bounds

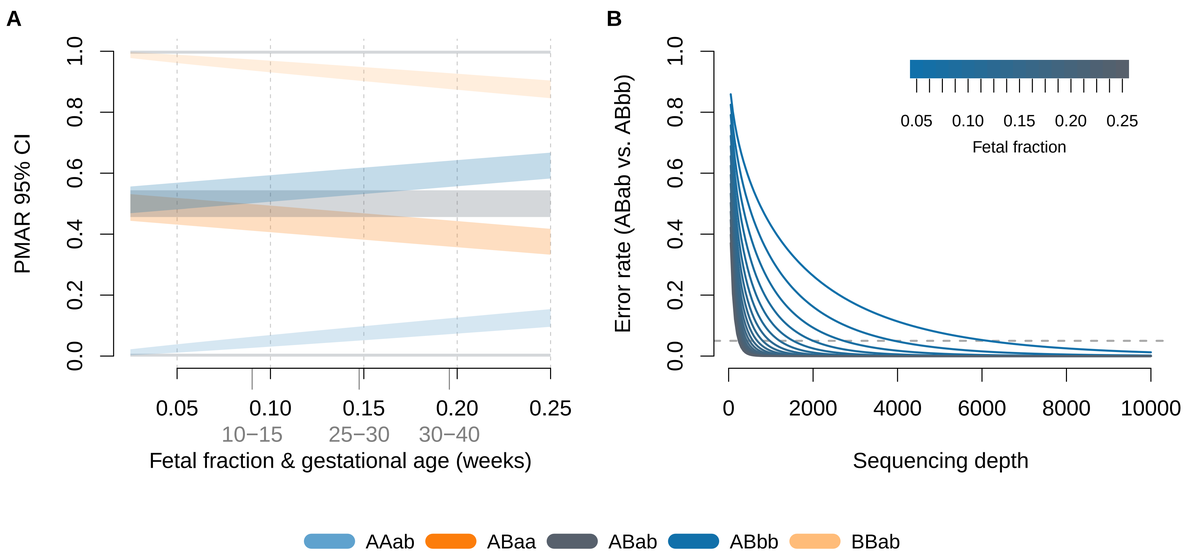

Figure 2: Binomial distribution bounds. (A) 95% confidence intervals for expected PMAR of maternal-fetal genotypes under the binomial distribution for a sequencing depth of 500x. (B) Expected missclassification rate (Weitzman overlapping coefficient) considering ABab versus ABbb; error rates identical for ABab versus ABaa.

Methodology

- Simulations performed based on user-defined values for the fetal fraction mean (μ_f), fetal fraction variance (σ^2_f), sequencing depth mean (μ_d), sequencing depth dispersion (ϕ_d), and the population minor allele frequency (q). Sequencing counts at each locus simulated as follows:
  1. Sample maternal-fetal genotype using Hardy-Weinberg assumptions.
  2. Draw the *true* fetal fraction, f, from $\mathrm{N}(\mu_\mathrm{f}, \sigma^2_\mathrm{f})$; $\sigma^2_\mathrm{f} := \mu_\mathrm{f}/10$.
  3. Draw the sequencing depth, d, from $\mathrm{NegativeBinomial}(\mu_\mathrm{d}, \phi_\mathrm{d})$; $\phi_\mathrm{d} := \mu_\mathrm{d}/200$.
  4. Draw y from $\mathrm{Binomial}(d, f)$; note $x = d - y$.
- $\hat{F}$ estimated using an empirical Bayesian approach to identify unique fetal alleles.
- $\hat{G}$ estimated as:

$$\hat{g}_i = \underset{g \in G}{\mathrm{argmax}} \left\{ \mathcal{L}(g|m_g, x_i, y_i) \right\}, Y_i \sim \mathrm{Bin}(x_i + y_i, m_g) \quad (6)$$

Results & conclusions

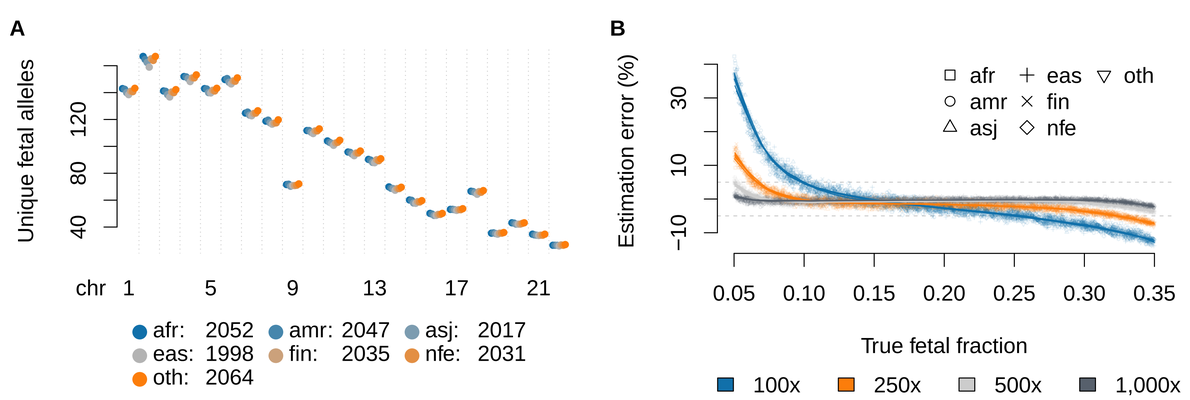

Figure 3: Fetal fraction estimation across diverse gnomAD populations. (A) Expected unique fetal alleles per chromosome in panel (1); expected genome totals given in legend. (B) Simulation results for fetal fraction estimation. Points represent individual simulations; lines show Friedman's super smoother for each population & depth.

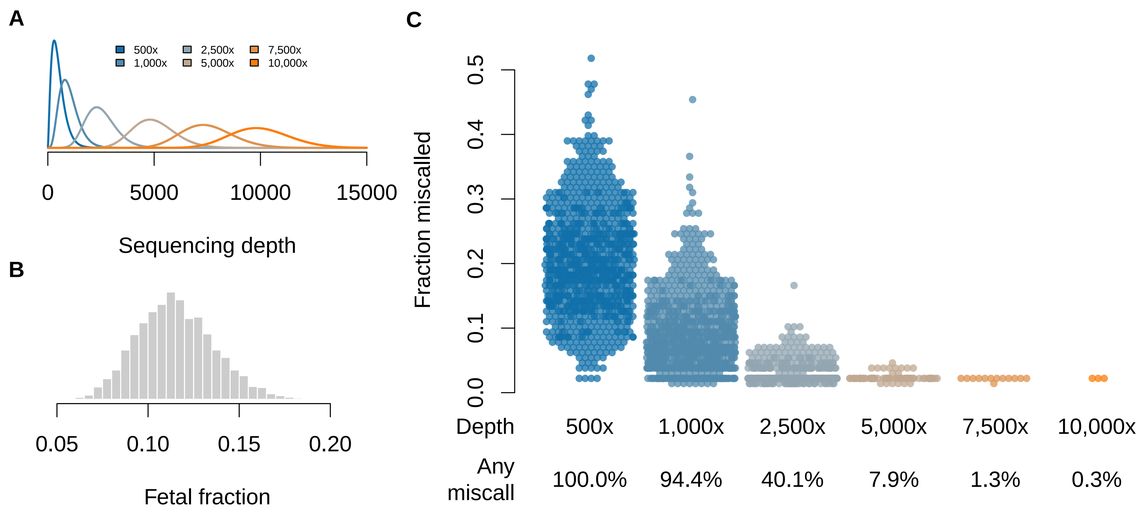

Figure 4: Genotyping accuracy by sequencing depth. (A) Read depth distributions used for simulations. (B) Fetal fraction distribution used for simulations. (C) Number of miscalled fetal genotypes when the mother is heterozygous. Each depth represents 1,0000 independent simulations.

- Using an existing bait panel, simulations suggest we can reliably estimate the fetal fraction within 5% error across diverse genetic populations.
- Deep sequencing (>7,500x) coverage is necessary to accurately genotype the fetus at sites with maternal heterozygosity. Limiting the scope of variants interrogated, we expect to affordably achieve >99% fetal genotyping accuracy.
- Extending these simulation results to rare conditions (average homozygote incidence of 1:2500), we approximate the positive predictive value for fetal homozygosity to be >55%.

Funding Sources:
F30HD10122801 (Filer)
UNC-Illumina Pilot Award (Filer)
K23HD088742 (Vora)

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

renci

dayne_filer@med.unc.edu
daynefiler.com
produced with LaTeX