Non-invasive fetal genotyping from deep sequencing of maternal plasma: Proof of concept study

Dayne L Filer[1,2], Piotr A Mieczkowski[1], Amber R Ivins[3], Kirk C Wilhelmsen[1,2,4], Karen F Dorman[3], Bradford C Powell[1,2], Kelly L Gilmore[3], Cynthia M Powell[1], Neeta L Vora[3]

1) Department of Genetics, UNC School of Medicine; 2) Renaissance Computing Institute;

3) Department of Obstetrics and Gynecology, UNC School of Medicine; 4) WVU Rockefeller Neuroscience Institute

Approach

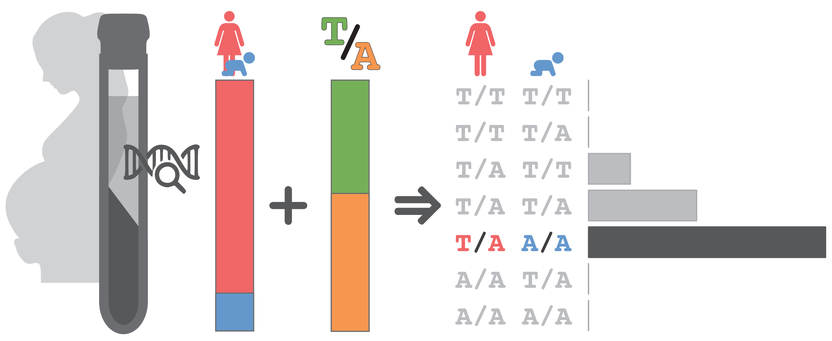

Figure 1: Using maternal cell-free DNA, make probabilistic genotype calls based on fetal fraction estimates and observed base-pair proportions. In theory, given the proportion of fetal to maternal DNA (fetal fraction), a true estimate of the allele frequency defines the maternal and fetal genotypes at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

- We performed targeted sequencing on cell-free DNA (cfDNA) from 60 pregnancies with paired germline sequencing of the mother and newborn.
- Sequencing libraries created using the IDT xGen Prism platform with dual-index UMI-containing adapters. We captured cell-free samples with two panels: (1) IDT xGen Human Copy Number Variant Backbone Hybridization Panel (CNV panel) to estimate fetal fraction; (2) IDT xGen Human Identification Hybridization Panel (ID panel) covering 72 common SNVs to test genotyping accuracy. Germline samples captured only with the ID panel.
- We combined the four libraries from each duo (maternal, newborn, cfDNA ID panel, cfDNA CNV panel) in a 3:3:39:80 ratio for sequencing. Paired-end, 150 basepair sequencing performed on Illumina NovaSeq with one lane of an S4 chip for the first 20 duos and 3 lanes of an S4 chip for the remaining 40.
- Sequencing reads grouped by UMI to accurately count molecules without discarding true duplicates.
- We employed our novel algorithm for estimating the maternal and fetal genotypes from cell-free DNA, described previously (`https://doi.org/10.1002/pd.6009`; `PMID: 34265090`). Briefly, the algorithm uses Bayesian expectation-maximization to identify loci with unique fetal heterozygosity and estimate the fetal fraction. Then, given the fetal fraction estimate and the observed proportion of minor allele reads, the algorithm uses a maximum-likelihood estimator based on expected binomial proportions to estimate maternal-fetal genotype pairs.

Genotyping accuracy

Table 1: Fetal genotyping accuracy by duo. 'DEP' indicates the median depths of the genotyped loci for the plasma and germline samples on the left and right, respectively; 'ALL' indicates the overall fetal genotyping accuracy; 'HET' indicates the fetal genotyping accuracy at loci with matenral heterozygosity. Numbers in parentheses indicate the number of loci evaluated.

	DEP		ALL	HET		DEP		ALL	HET
D009	487	225	0.90 (68)	0.78 (28)	D086	2874	2	0.48 (44)	0.50 (4)
D012	1613	297	0.77 (68)	0.64 (43)	D088	1430	2	0.56 (28)	0.44 (17)
D014	4216	380	0.76 (69)	0.60 (37)	D089	2454	5	0.82 (50)	0.86 (7)
D016	1193	267	0.87 (70)	0.79 (35)	D091	1743	4	0.60 (61)	0.61 (23)
D017	891	268	0.88 (67)	0.81 (39)	D093	2371	6	0.86 (63)	0.79 (28)
D018	3020	300	0.81 (68)	0.64 (34)	D099	3126	4	0.44 (60)	0.29 (17)
D020	3042	353	0.89 (68)	0.81 (37)	D101	3538	3	0.56 (46)	0.63 (19)
D021	2063	252	0.89 (69)	0.73 (27)	D102	2170	3	0.67 (49)	0.63 (27)
D024	3443	277	0.79 (69)	0.64 (37)	D104	1403	4	0.88 (33)	1.00 (3)
D025	6539	322	0.79 (69)	0.57 (34)	D105	1537	4	0.63 (42)	1.00 (2)
D026	2327	207	0.94 (69)	0.84 (27)	D107	1400	8	0.84 (64)	0.75 (24)
D027	2157	190	0.88 (68)	0.82 (35)	D110	2432	6	0.82 (62)	0.64 (22)
D029	2321	140	0.79 (68)	0.56 (29)	D112	1486	6	0.69 (63)	0.70 (20)
D030	1633	8	0.73 (66)	0.57 (28)	D113	1977	8	0.63 (66)	0.52 (25)
D032	2499	5	0.64 (61)	0.46 (28)	D115	1700	5	0.78 (60)	0.64 (25)
D035	2015	98	0.90 (67)	0.79 (30)	D116	1175	4	0.74 (57)	0.79 (20)
D039	2400	234	0.91 (68)	0.84 (33)	D117	1454	3	0.66 (52)	0.70 (20)
D040	1657	166	0.80 (70)	0.66 (37)	D122	1948	5	0.72 (55)	0.61 (18)
D041	2881	4	0.77 (51)	0.74 (23)	D123	1612	2	0.50 (44)	0.17 (6)
D044	2884	186	0.75 (71)	0.53 (35)	D124	946	4	0.79 (60)	0.79 (28)
D046	2734	9	0.92 (66)	0.86 (29)	D128	3336	5	0.78 (50)	0.71 (24)
D047	1565	72	0.86 (69)	0.76 (30)	D131	1134	3	0.78 (52)	0.72 (19)
D049	2076	7	0.79 (64)	0.75 (32)	D132	1186	6	0.85 (65)	0.76 (27)
D051	2735	5	0.74 (61)	0.64 (25)	D134	1284	4	0.82 (52)	0.90 (10)
D054	3196	3	0.50 (36)	0.41 (17)	D135	2112	8	0.75 (60)	0.62 (21)
D063	1971	187	0.77 (71)	0.56 (33)	D136	1861	4	0.71 (57)	0.75 (18)
D065	814	3	0.53 (51)	0.29 (7)	D138	1790	5	0.82 (61)	0.75 (20)
D070	3293	8	0.87 (63)	0.92 (24)	D141	2282	14	0.86 (70)	0.64 (23)
D071	682	6	0.53 (66)	0.52 (23)	D143	2168	8	0.59 (65)	0.61 (28)
D080	1756	5	0.49 (64)	0.43 (30)					

Expanded results

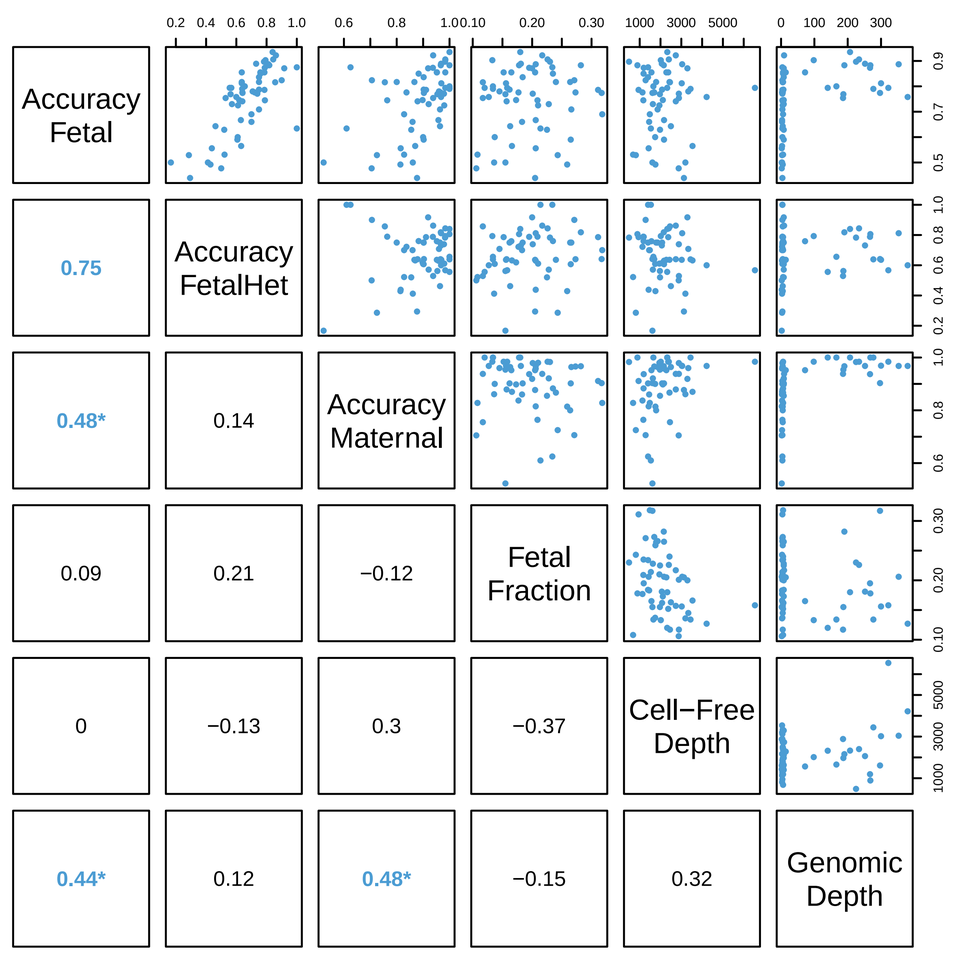

Figure 2: Genotyping accuracy does not correlate to fetal fraction or median sequencing depths. Each point represents a single duo. Numbers in the lower triangle represent the correlation coefficient; highlighted numbers indicate Bonferroni-adjusted significance at $\alpha < 0.05$, astricies indicate non-significant when controlling for batch (first 20 duos versus second 40).

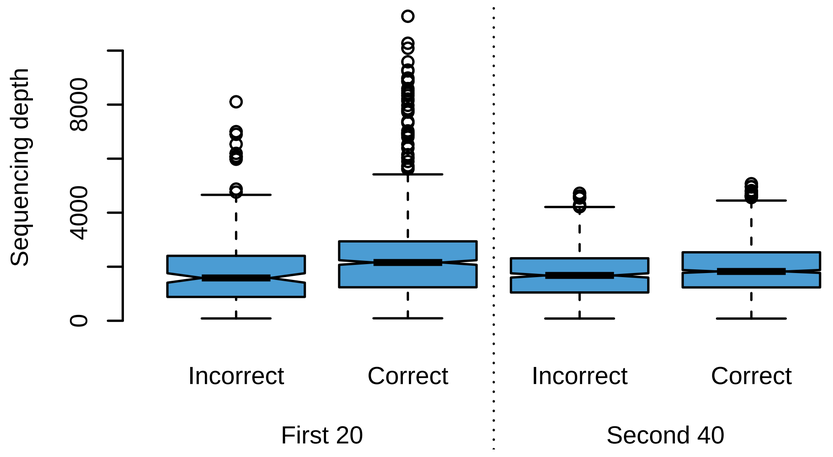

Figure 3: Difference in loci-specific sequencing depth between correct and incorrect fetal genotype estimates seen in the first batch, not in the second. Non-overlapping notches indicate significantly different distributions.

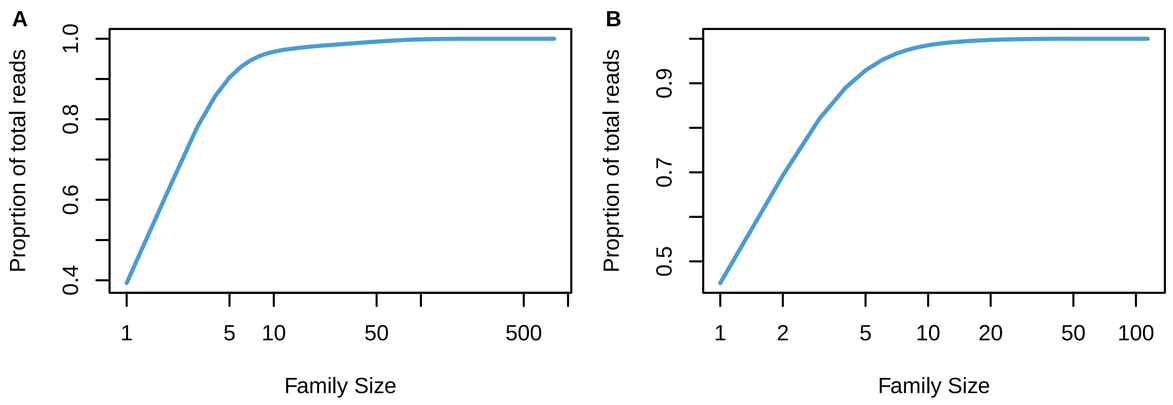

Figure 4: Roughly 40% of molecules sequenced only once. Lines represent empiric cumulative distribution function of duplicates reads (family size) by UMI. (A) cell-free DNA samples; (B) genomic samples.

Summary & conclusions

- Overall, we achieved maternal and fetal genotyping accuracies of 90.2% and 75.5%, respectively. At loci with maternal heterozygosity, we achieved an accuracy of 67.4%.
- We did not obtain sufficient depth on germline samples for the second batch of 40 duos. Compared to the first 20 duos, with a median sequencing depth of 221x, we only obtained a median depth of 5x for germline samples.
- At loci with germline sequencing depths greater than 50x, we achieved overall maternal and fetal genotyping accuracies of 97.8% and 85%, with 71.6% accuracy at loci with maternal heterozygosity.
- We failed to reach our goal sequencing depths of >7,500x required for reliable genotyping accuracy; this approach will require more efficient library construction and capture methodologies.

Funding Sources:
F30HD10122801, NICHD; UNC-Illumina Pilot Award
K23HD088742, NICHD; 550KR231934, NC TraCS

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

renci

dfiler@renci.org
daynefiler.com

produced with LaTeX