Clinical Chemistry 59:5
846–849 (2013)

Brief Communication

Feasibility Study of Semiconductor Sequencing for Noninvasive Prenatal Detection of Fetal Aneuploidy

Yuan Yuan,[1†] Fuman Jiang,[1†] Sang Hua,[1†] Bole Du,[1] Yibin Hao,[2] Lili Ye,[1] Jiucheng Liu,[1] Kaiyan Feng,[1] Xinjie Huang,[1] Xin Yi,[1,3] Wei Wang,[1] Ling Yang,[3] Feng Mu,[4] Caixia Liu,[5*] and Yu Liang[1*]

[1] Shenzhen Birth Defect screening project laboratory, BGI-Shenzhen, Shenzhen, China; [2] Zhengzhou Central Hospital Affiliated with Zhengzhou University, Zhengzhou, China; [3] BGI-Tianjin, Tianjin, China; [4] BGI-Beijing, Beijing, China; [5] Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, China; * address correspondence to: C.L. at Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Liaoning, Shenyang, China. Fax +86-24-96615-1#-43221; e-mail LiuCX@sj-hospital.org. Y.L. at BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen, China. Fax +86-755-25016585; e-mail liangyu@genomics.cn. † Yuan Yuan, Fuman Jiang, and Sang Hua contributed equally to this work and should all be considered first authors.

BACKGROUND: Noninvasive prenatal detection of common fetal aneuploidies with cell-free DNA from maternal plasma has been achieved with high-throughput next-generation sequencing platforms. Turnaround times for previously tested platforms are still unsatisfactory for clinical applications, however, because of the time spent on sequencing. The development of semiconductor sequencing technology has provided a way to shorten overall run times. We studied the feasibility of using semiconductor sequencing technology for the noninvasive detection of fetal aneuploidy.

METHODS: Maternal plasma DNA from 13 pregnant women, corresponding to 4 euploid, 6 trisomy 21 (T21), 2 trisomy 18 (T18), and 1 trisomy 13 (T13) pregnancies, were sequenced on the Ion Torrent Personal Genome Machine sequencer platform with 318 chips. The data were analyzed with the *T* statistic method after correcting for GC bias, and the *T* value was calculated as an indicator of fetal aneuploidy.

RESULTS: We obtained a mean of 3 524 401 high-quality reads per sample, with an efficiency rate of 77.9%. All of the T21, T13, and T18 fetuses could be clearly distinguished from euploid fetuses, and the time spent on library preparation and sequencing was 24 h.

CONCLUSIONS: Semiconductor sequencing represents a suitable technology for the noninvasive prenatal detection of fetal aneuploidy. With this platform, sequencing times can be substantially reduced; however, a further larger-scale study is needed to determine the imprecision of noninvasive fetal aneuploidy detection with this system.

Chromosomal aneuploidy is a common cause of birth defects. Unfortunately, the diagnostic sensitivity and specificity of current screening programs for fetal chromosomal aneuploidy have been unsatisfactory *(1)*. The discovery of cell-free fetal DNA in maternal plasma *(2)* has led to the introduction of a new method of noninvasive fetal aneuploidy detection that uses next-generation sequencing *(3–5)*. This method has been tested widely in clinical applications in recent years *(6–8)*.

The current next-generation sequencing platforms have several weaknesses, however, and these limitations need to be addressed before these platforms can be used in routine clinical applications. These weaknesses include poor sample scalability, high cost, and a long sequencing time (2 to 3 days). A new benchtop sequencing instrument developed by Ion Torrent (owned by Life Technologies), based on semiconductor sequencing technology, can solve many of these problems *(9, 10)*. The Ion Torrent Personal Genome Machine (PGM),[6] using this new sequencing technology, can generate approximately 5×10^6 reads with a read length up to 200 bp on 318 chips within several hours, and the data output is expected to increase further with the Ion Proton instrument, which is expected in the near future. Because of its advantages with respect to flexibility, speed of sequencing, and cost of the instrument, compared with other next-generation sequencing platforms, this type of platform would be more suitable for noninvasive fetal aneuploidy detection in clinical settings. To test the feasibility of using semiconductor sequencing technology for noninvasive detection of fetal aneuploidy, we used the PGM to sequence 13 previously characterized plasma DNA samples from pregnant women.

Plasma samples from 13 individuals—6 with trisomy 21 (T21) pregnancies, 2 with T18 pregnancies, 1 with a T13 pregnancy, and 4 with euploid pregnancies (all with karyotype-confirmed results)—were recruited from BGI-Health. All plasma samples were obtained during the second trimester (range, 13+6 weeks to 29 weeks; for details, see Table 1 in the Data Supplement that accompanies the online version of this Brief Communication at http://www.clinchem.org/content/

[6] Nonstandard abbreviations: PGM, Ion Torrent Personal Genome Machine; T21, trisomy 21.

vol59/issue5). Cell-free DNA was isolated from 600 μL plasma with the TIANamp Micro DNA Purification Kit (Tiangen Biotech). The sequencing library was prepared according to the manufacturer's instructions. The library was loaded onto an Ion 318 chip, and single-end sequencing was performed on the PGM with a standard 110-cycle run.

The data were initially processed with the Ion Torrent platform–specific pipeline software (Torrent Suite, version 2.0.1) to generate sequence reads, trim adapter sequences, and filter out low-quality reads. We obtained a mean read length of approximately 169 nucleotides (see Fig. 1 in the online Data Supplement), which is consistent with the length distribution of plasma cell-free DNA reported by Lo et al. *(11)*.

Ion Torrent's mapping program (TMAP, version 0.2.3; https://github.com/iontorrent/TMAP) was used to align the generated sequence data to the hg19 human reference genome. The data were filtered further by removing duplicate reads (SAMtools, version 0.1.18), unmapped reads, reads with a mapping quality <10, and reads with lengths <35 bp (because information provided by reads >35 bp is more reliable). We obtained a mean of 4 506 952 (SD, 695 493) raw reads for each sample. A mean of 3 524 401 (SD, 728 849) reads per sample, representing 77.9% (SD, 9.71%) of the raw data, remained after filtering.

As with the Illumina platform *(3)*, a GC bias was observed in the sequence data. The tag density was overrepresented in regions with a moderate GC content, and it was underrepresented in regions with a higher or lower GC content (see Fig. 2 in the online Data Supplement; for details of the calculation method see the Supplemental Methods file). The correction method reported by Fan and Quake *(12)* was introduced to eliminate the effects of GC bias. In brief, for each sample we divided all chromosomes into 60-kb segments, called "bins." The read number and GC content were calculated for each bin. A smooth spline was applied to fit the read number of all the bins to the GC content; bins with no reads or overabundant reads were ignored. The assigned weights (w) were calculated as:

$$w = \frac{\bar{M}}{M_{ip}},$$

where $\bar{M}$ is the global mean read number per 60-kb bin and M_{ip} is the predicted read number of the ith bin for the GC content fitted by the smooth spline. The read number of each bin was corrected by multiplying by the w factor; all read numbers presented in this report are corrected read numbers. The GC bias was eliminated almost completely after these processes (see Fig. 3 in the online Data Supplement).

Samples from T13, T18, or T21 fetuses are expected to have higher read numbers for chromosomes 13, 18, or 21 than for euploid autosomes, separately. The T statistic was used to quantify the overrepresentation of clinically interesting chromosomes (e.g., chromosome 21). After the GC correction, the baseline was constructed by selecting 2 normal samples as a reference. We then normalized the corrected number of sequencing reads for each bin of a test sample to this baseline (see Supplemental Methods file in the online Data Supplement). The T value for each sample was calculated with the formula:

$$T = \frac{\mu_i - \mu_j}{\sigma \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}},$$

where μ_i and μ_j are the mean corrected read numbers per bin for chromosome i and chromosome j, and:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{22}(n_i - 1)\sigma_i^2}{\sum_{i=1}^{22}(n_i - 1)}},$$

where σ_i is the SD of the read number per bin of chromosome i and n is the number of bins. For each sample, the T statistic was used to quantify the overrepresentation of each clinically interesting chromosome (e.g., chromosome 21) relative to the other 21 autosomes, and then the mean T values for chromosomes 13, 18, and 21 were calculated. To minimize the bias, we removed the 2 maximum values and the 2 minimum values before calculating the mean T values, because some extreme conditions, such as aneuploidy at 2 or 3 autosomes, can occur. The T value calculated by comparing 2 aneuploidy autosomes will be abnormal and is not suitable for the calculation in the next step. The absolute value of a mean T value ≥3.3 was used as the threshold for the detection of abnormal chromosomes, with values above this threshold indicating that the mean number of reads per bin for that chromosome was significantly different from that of the other chromosomes at an α level <0.001 (2-tailed test).

Our data show that women carrying T21 fetuses can be correctly distinguished from women carrying disomy 21 fetuses (Fig. 1). The chromosome 21 T values for the 6 T21 cases (cases 8–13) were >3.3, whereas these values were <3.3 for the 4 euploid cases (cases 1–4) (see Table 2 in the online Data Supplement). The T18 pregnancies (cases 6 and 7) and the T13 pregnancy (case 5) were also identified correctly. The calculation of the proportion of fetal DNA for the aneuploid samples was based on the overrepresentation of the abnor-

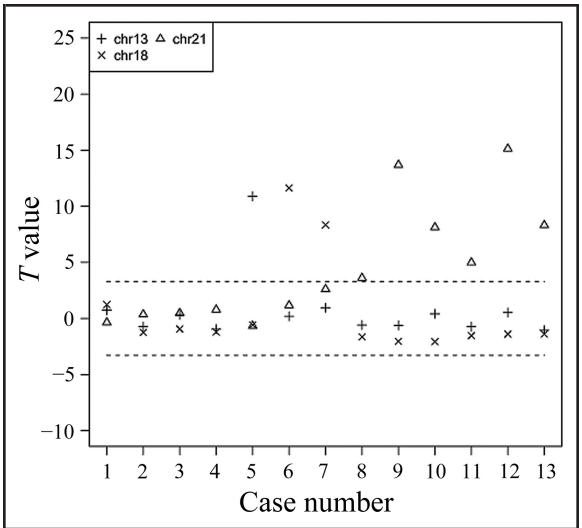

Fig. 1. *T* values of chromosomes (chr) 13, 18, and 21 in normal (cases 1–4), T13 (case 5), T18 (cases 6 and 7), and T21 (cases 8–13) pregnancies.

Cases with *T* values >3.3 for a specific chromosome (13, 18, or 21) were classified as T18, T13, or T21.

mal chromosome, and the values ranged from 7% to 20% (see Table 2 in the online Data Supplement).

The percentage of reads for the Y chromosome for male fetuses was >0.034%, whereas this percentage was <0.034% for female fetuses (see Fig. 4 in the online Data Supplement). The Y chromosome signal observed for female fetus samples could be due to false alignment of a small number of sequences with the Y chromosome, which may be caused by the presence of homologous sequences in the X and Y chromosomes. In this study, samples from male fetuses and those from female fetuses could be distinguished when we introduced a cutoff value of approximately 0.034%; however, whether this method can be used for sex determination needs to be evaluated in further studies.

The processing time was 24 h for the PGM, compared with 61 h for the HiSeq 2000 system (Illumina); thus, the time required for the PGM analysis was substantially shorter (see Table 3 in the online Data Supplement).

In this study, we confirmed that semiconductor sequencing technology can be used for noninvasively detecting fetal T21, T18, and T13. Because of the advantages with respect to the sequencing speed, the lower cost of the semiconductor sequencing instrument, and the flexibility in the throughput, this technology should be suitable for clinical applications. Although semiconductor sequencing has difficulty detecting homopolymer stretches, this drawback did not have an obvious effect on aneuploidy detection in our study, because the low-quality reads were filtered out before the analysis. GC bias was also observed with the semiconductor sequencing platform. After the correction, the GC bias was removed effectively for reads with GC contents ranging from 0.3 to 0.55; however, noise can still be observed in the regions with a GC content >0.55 (see Fig. 3 in the online Data Supplement). This noise may be caused by the insufficient number of reads in each bin in that region, which is only approximately 1.7% of the total reads on average. Because only a few reads had a GC content >0.55, the results were not affected. Although only 5×10^6 raw reads can be produced with the PGM and 318 chips (which means 1 sample per chip), the positive and negative samples could still be distinguished in this study. The forthcoming Ion Torrent Proton platform, which is based on the same sequencing technology, can produce more data. Sixteen samples can be analyzed per run with the Ion Torrent Proton I, which may improve the statistical power further by allowing multiple samples to be run on 1 chip.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *No authors declared any potential conflicts of interest.*

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Acknowledgments: We thank our colleagues at BGI for library preparation and sequencing, our colleagues at BGI-Health for sample acquisition, and personnel at Life Technologies for technical assistance in library preparation and sequencing.

References

1. Wapner R, Thom E, Simpson JL, Pergament E, Silver R, Filkins K, et al. First-trimester screening for trisomies 21 and 18. N Engl J Med 2003;349:1405–13.
2. Wright CF, Burton H. The use of cell-free fetal nucleic acids in maternal blood for non-invasive prenatal diagnosis. Hum Reprod Update 2009;15:139–51.
3. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci U S A 2008;105:16266–71.
4. Chiu RWK, Sun H, Akolekar R, Clouser C, Lee C, McKernan K, et al. Maternal plasma DNA analysis with massively parallel sequencing by ligation for noninvasive prenatal diagnosis of trisomy 21. Clin Chem 2010;56:459–63.
5. Oever JME, Balkassmi S, Verweij EJ, Iterson M, Scheltema PNA, Oepkes D, et al. Single molecule sequencing of free DNA from maternal plasma for

noninvasive trisomy 21 detection. Clin Chem 2012;58:699–706.
6. Ehrich M, Deciu C, Zwiefelhofer T, Tynan JA, Cagasan L, Tim R, et al. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. Am J Obstet Gynecol 2011;204:205e1–11.
7. Lau TK, Chan MK, Lo PS, Chan HY, Chan WS, Koo TY, et al. Clinical utility of noninvasive fetal trisomy (NIFTY) test—early experience. J Matern Fetal Neonatal Med 2012;25:1856–9.
8. Chen EZ, Chiu RW, Sun H, Akolekar R, Chan KC, Leung TY, et al. Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. PLoS One 2011;6:e21791.
9. Elliott AM, Radecki J, Moghis B, Li X, Kammesheidt A. Rapid detection of the ACMG/ACOG-recommended 23 CFTR disease-causing mutations using Ion Torrent semiconductor sequencing. J Biomol Tech 2012;23:24–30.
10. Vogel U, Szczepanowski R, Claus H, Jünemann S, Prior K, Harmsen D. Ion Torrent Personal Genome Machine sequencing for genomic typing of *Neisseria meningitidis* for rapid determination of multiple layers of typing information. J Clin Microbiol 2012;50:1889–94.
11. Lo YMD, Chan KCA, Sun H. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med 2010;2:61ra91.
12. Fan HC, Quake SR. Sensitivity of noninvasive prenatal detection of fetal aneuploidy from maternal plasma using shotgun sequencing is limited only by counting statistics. PLoS One 2010;5:e10439.

Previously published online at
DOI: 10.1373/clinchem.2012.196725