Clinical Chemistry 56:3
459–463 (2010)

Brief Communications

Maternal Plasma DNA Analysis with Massively Parallel Sequencing by Ligation for Noninvasive Prenatal Diagnosis of Trisomy 21

Rossa W.K. Chiu,[1,2] Hao Sun,[1,2] Ranjit Akolekar,[3] Christopher Clouser,[4] Clarence Lee,[4] Kevin McKernan,[4] Daixing Zhou,[4] Kypros H. Nicolaides,[3] and Y.M. Dennis Lo[1,2*]

[1] Centre for Research into Circulating Fetal Nucleic Acids, Li Ka Shing Institute of Health Sciences, and [2] Department of Chemical Pathology, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China; [3] Harris Birthright Research Centre for Fetal Medicine, King's College Hospital, London, UK; [4] Life Technologies, Beverly, MA; * address correspondence to this author at: Department of Chemical Pathology, The Chinese University of Hong Kong, Rm. 38061, 1/F, Clinical Sciences Bldg., Prince of Wales Hospital, 30-32 Ngan Shing St., Shatin, Hong Kong SAR, China. Fax 852-2636-5090; e-mail loym@cuhk.edu.hk.

BACKGROUND: Noninvasive prenatal diagnosis of trisomy 21 (T21) has recently been shown to be achievable by massively parallel sequencing of maternal plasma on a sequencing-by-synthesis platform. The quantification of several other human chromosomes, including chromosomes 18 and 13, has been shown to be less precise, however, with quantitative biases related to the chromosomal GC content.

METHODS: Maternal plasma DNA from 10 euploid and 5 T21 pregnancies was sequenced with a sequencing-by-ligation approach. We calculated the genomic representations (GRs) of sequenced reads from each chromosome and their associated measurement CVs and compared the GRs of chromosome 21 (chr21) for the euploid and T21 pregnancies.

RESULTS: We obtained a median of 12×10^6 unique reads (21% of the total reads) per sample. The GRs deviated from those expected for some chromosomes but in a manner different from that previously reported for the sequencing-by-synthesis approach. Measurements of the GRs for chromosomes 18 and 13 were less precise than for chr21. *z* Scores of the GR of chr21 were increased in the T21 pregnancies, compared with the euploid pregnancies.

CONCLUSIONS: Massively parallel sequencing-by-ligation of maternal plasma DNA was effective in identifying T21 fetuses noninvasively. The quantitative biases observed among the GRs of certain chromosomes were more likely based on analytical factors than biological factors. Further research is needed to enhance the precision for measuring for the representations of chromosomes 18 and 13.

The possibility of fetal chromosomal aneuploidy, particularly trisomy 21 (T21)[5] (Down syndrome), is a major reason why couples consider prenatal diagnostic studies. Current definitive practices of prenatal diagnosis rely on obtaining fetal genetic material via chorionic villus sampling or amniocentesis, both of which have associated risks of fetal miscarriage. In 1997, we reported the presence of cell-free fetal DNA in the circulation of pregnant women *(1, 2)*. Analysis of fetal DNA in maternal plasma is useful for the prenatal assessment of sex-linked diseases, Rhesus D genotyping, and certain monogenic diseases *(3, 4)*.

The direct diagnosis of fetal chromosomal aneuploidy via nucleic acid analysis of maternal plasma has been more challenging *(5)*. Chromosomal aneuploidy refers to a quantitative imbalance in the dosage of particular chromosomes in a genome. Thus, approaches need to be developed to detect the quantitative perturbations associated with the aneuploid chromosome among the fetal DNA molecules within maternal plasma. Precise quantification of circulating fetal DNA in maternal plasma has proved difficult, however, owing to its low fractional and absolute concentrations against the high background concentrations of maternal DNA *(6, 7)*. Recently, our group *(8)* and Fan et al. *(9)* demonstrated the feasibility of noninvasive prenatal diagnosis of T21 by massively parallel sequencing of maternal plasma DNA.

It is possible to identify the chromosomal origin of each sequenced plasma DNA molecule by comparing its nucleotide sequence with the reference human genome. Because a T21 fetus carries an additional copy of chromosome 21 (chr21) in its genome, this copy would contribute additional amounts of chr21 DNA fragments into the maternal plasma. The small increments in the proportional amounts, termed the "genomic representation" (GR), of chr21 sequences in the plasma of women carrying T21 fetuses compared with euploid fetuses could be detected with high precision by massively parallel sequencing *(5, 8)*.

We reported, however, that the precision for measuring the GR varied among human chromosomes and tended to be worse for chromosomes with GC contents at either end of the GC-abundance spectrum *(8)*. Fan

[5] Nonstandard abbreviations: T21, trisomy 21; chr21, chromosome 21; GR, genomic representation; %GR, percentage for a given chromosome of the total number of unique reads obtained for the sample.

et al. *(9)* observed a quantitative bias in the relative amounts of plasma DNA molecules sequenced from each human chromosome that bore a relationship to the chromosome's GC content. Chromosomes with a low GC content were underrepresented, whereas chromosomes with a high GC content were overrepresented. Both previous studies *(8, 9)* were performed with the Genome Analyzer from Illumina, which uses a sequencing-by-synthesis approach *(10)*. In particular, the measurements of the proportions of plasma DNA molecules originating from chromosomes 13 and 18 were less precise and had greater negative biases than those for chr21. Because chromosomes 13 and 18 are involved in trisomy 13 and trisomy 18, respectively, it would be of diagnostic interest to develop protocols that have more uniform performance across chromosomes.

We collected 5 mL of blood from 15 women presenting for first trimester aneuploidy screening: 5 women with T21 pregnancies (2 female fetuses) and 10 women with euploid pregnancies (3 female fetuses). Maternal plasma DNA libraries were prepared using the low-input fragment DNA protocol, without shearing the DNA *(11)*. Massively parallel sequencing-by-ligation was performed on a SOLiD™ 3 System (Applied Biosystems/Life Technologies) according to the manufacturer's protocol *(11)*. Details of the methods are given in the Supplemental Data file in the Data Supplement that accompanies the online version of this Brief Communication at http://www.clinchem.org/content/vol56/issue3.

We obtained a median of 59×10^6 raw reads (range, $31–78 \times 10^6$) from each sample. A median of 12×10^6 reads/sample (range, $7–16 \times 10^6$), representing 21% (range, 17%–23%) of the raw reads, could be aligned uniquely to a single location on the reference genome with up to 1 color space mismatch. We termed these reads "perfectly matched unique reads" because despite the 1 color space mismatch, the correct sequence of the read could be deduced because each sequenced base was interrogated by 2 adjacent colored dinucleotides. With a read length of 50 bp, a median of 12×10^6 unique reads is equivalent to 20% coverage of the haploid human genome. The number of unique reads aligned to each human chromosome was expressed as a percentage of the total number of unique reads obtained for the sample, which we termed the "%GR" of that chromosome. We determined the median, mean, and SD of the %GR for each chromosome of the sequenced maternal plasma samples. The corresponding values for chr21 were based on the 10 euploid samples only, whereas those for the other chromosomes were calculated from all 15 samples.

We compared the median %GR of each chromosome to that expected for a repeat-masked haploid female genome, because the majority of DNA molecules in maternal plasma originated from a female, i.e., the mother. The degree of deviation from the expected %GR is estimated by: {[(median of the experimentally derived %GR for chromosome N) − (expected %GR for chromosome N)]/(expected %GR for chromosome N)} × 100%. The data are presented in Fig. 1 in the online Data Supplement. The CVs for %GR measurements were calculated for each chromosome; the results are presented in Fig. 2 in the online Data Supplement.

For T21 diagnosis, 4 euploid pregnancies of male fetuses were selected as the control group. The mean and SD values of the chr21 %GR for these 4 controls were used to determine the chr21 *z* scores of the remaining 11 samples, as previously described *(8)*. The chr21 *z* score for a test case is the difference between the %GR of chr21 for the test case and the mean %GR for the same chromosome of the control group, divided by the SD. Because +3 SDs is the 99.9th percentile from the mean of the reference group for a 1-tailed distribution, a *z* score of +3 was used as the cutoff value for determining overrepresentation of chr21 sequences in a maternal plasma sample. The chr21 *z* scores for the 5 T21 cases were 8.6, 15.3, 15.6, 18.6, and 19.9. The chr21 *z* scores for the 6 euploid cases were −2.1, −1.5, −1.0, −0.7, −0.7, and 0.8. *z* Scores for the other autosomes were also calculated; these results are presented in Fig. 1.

To further confirm the robustness of this approach for T21 diagnosis, we used a bootstrapping procedure on 9 of the 10 euploid cases at a time as the control group to reanalyze the chr21 *z* scores for the 5 T21 cases and the remaining euploid case. Because there were 10 euploid cases, we had 10 different possible combinations of 9 control cases. Fig. 3 in the online Data Supplement shows that for all 10 analyses the euploid case had a *z* score <3, whereas, all of the T21 cases had *z* scores >3.

We previously showed that a small number of sequences would be falsely aligned to chromosome Y, even for female fetuses. In the present study, the %GR values of chromosome Y were 0.009% for 4 female fetuses and 0.023% for the remaining female fetus. The median %GR of chromosome Y for the male fetuses was 0.034% (range, 0.019%–0.045%). Thus, the %GR of chromosome Y for this 1 female fetus was as high as that for the male fetuses. The chorionic villus of this fetus was negative when tested by a quantitative PCR assay targeting the *SRY* gene (sex determining region Y) *(6)*. We cannot exclude the possibility of a mix-up of the maternal plasma sample with that from a pregnancy with a male fetus. The accuracy of the use of the chromosome Y %GR for fetal sex determination would need to be further confirmed in future studies.

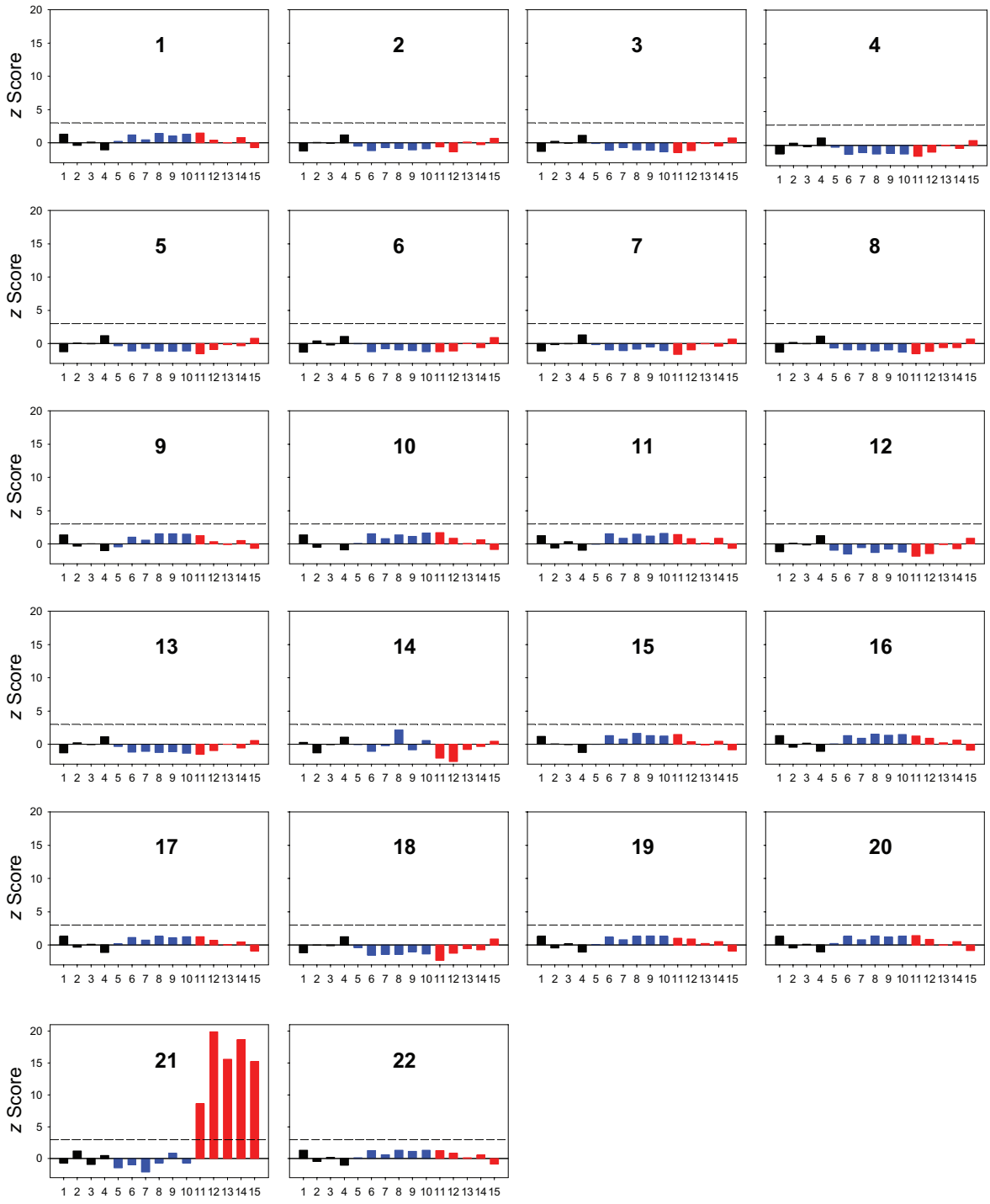

Fig. 1. *z* **Scores for the autosomes (chromosome number indicated by the number within each panel) in the control (cases 1–4, in black), euploid (cases 5–10, in blue), and T21 (cases 11–15, in red) pregnancies.**

A *z* score of 3 (dashed line) was used as a cutoff to determine the presence of overrepresentation of sequences from the corresponding chromosome.

Similarly with the sequencing-by-synthesis approach *(9)*, the quantitative representation of each chromosome was not uniform. It is interesting that unlike Fan et al. *(9)*, who reported a positive bias for chromosomes with a high GC content, such as chromosomes 19 and 22, we observed a negative bias (see Fig. 1 in the online Data Supplement). These data suggest that the nonuniform representation is more likely explained by analytical factors than biological factors *(12, 13)*. For example, the amplification efficiency of the DNA libraries may be nonuniform across sequences with different GC contents, although the form and extent of biases are different among different massively parallel sequencing platforms. In fact, Chu et al. *(14)* suggested that such PCR-related biases should be included as a parameter in the algorithms for calculating the chromosomal GR. To compute such sophisticated algorithms, however, would first require elucidation of the exact mechanism and extent of analytical bias introduced by the sequencing protocols. There are key differences between the protocols for the Genome Analyzer and the SOLiD 3 System. For clonal amplification of DNA libraries, the former uses solid-phase bridge amplification, whereas the latter uses emulsion PCR. Interestingly, the sequencing-by-ligation approach demonstrated less bias for sequences from chromosomes with a low GC content, as exemplified by the data for chromosome 13.

The precision profiles for chromosomes 18 and 13 are still worse than for chr21. In our previous study *(8)*, the CVs for measuring the %GR of chromosomes 21, 18, and 13 ranked (from most to least precise) 3rd, 10th, and 17th, respectively. In the present study, the corresponding rankings are 4th, 8th, and 16th. Thus, despite the change in the sequencing platform, the %GR values of chromosomes 18 and 13 were still more difficult to quantify as precisely as the %GR of chr21. Fan et al. *(9)* performed massively parallel sequencing of maternal plasma from 1 pregnancy each for trisomy 18 and trisomy 13 to study the feasibility of their noninvasive prenatal diagnosis. Yet, our current and previous data *(8)* suggest that noninvasive prenatal diagnosis of trisomy 18 and 13 with this approach would likely be less precise than for T21. Conversely, the sequencing-by-ligation approach appears to be a robust method for the direct detection of T21 fetuses. Large-scale clinical trials would be needed to confirm these findings. Currently, the complete protocol on the SOLiD 3 System, from DNA extraction to data interpretation, takes 7.5 days; however, we obtained 5 times more unique reads per sample than in our earlier study, when the equivalent processing time was 5 days *(8)*. Higher sequencing counts would allow the analysis of >1 sample per reaction chamber if bar codes were used. Hence, there is an opportunity for cost reduction and increased sample throughput, which would bring this new technology closer to clinical application.

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 of Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: Y.M.D. Lo, Sequenom.
Stock Ownership: C. Lee, Life Technologies; K. McKernan, Life Technologies; D. Zhou, Life Technologies; Y.M.D. Lo, Sequenom.
Honoraria: None declared.
Research Funding: R.W.K. Chiu and Y.M.D. Lo, Life Technologies, sponsor of the sequencing runs as part of a grant from the Innovation and Technology Fund (ITS/054/09); Y.M.D. Lo, Sequenom.
Expert Testimony: None declared.
Other Remuneration: R.W.K. Chiu and Y.M.D. Lo hold patents and have filed patent applications on the detection of fetal nucleic acids in maternal plasma for noninvasive prenatal diagnosis. Y.M.D. Lo was supported by an endowed professorship from the Li Ka Shing Foundation.

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 Elizabeth Levandowsky and Tristen Weaver for preparing the libraries and performing the sequencing.

References

1. Lo YMD, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, Wainscoat JS. Presence of fetal DNA in maternal plasma and serum. Lancet 1997;350:485–7.
2. Lo YMD, Chiu RWK. Prenatal diagnosis: progress through plasma nucleic acids. Nat Rev Genet 2007;8:71–7.
3. Lun FMF, Tsui NBY, Chan KCA, Leung TY, Lau TK, Charoenkwan P, et al. Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. Proc Natl Acad Sci U S A 2008; 105:19920–5.
4. 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.
5. Chiu RWK, Cantor CR, Lo YMD. Non-invasive prenatal diagnosis by single molecule counting technologies. Trends Genet 2009;25:324–31.
6. Lo YMD, Tein MS, Lau TK, Haines CJ, Leung TN, Poon PM, et al. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. Am J Hum Genet 1998;62:768–75.
7. Lo YMD, Lun FMF, Chan KCA, Tsui NBY, Chong KC, Lau TK, et al. Digital PCR for the molecular detection of fetal chromosomal aneuploidy. Proc Natl Acad Sci U S A 2007;104:13116–21.
8. Chiu RWK, Chan KCA, Gao Y, Lau VYM, Zheng W, Leung TY, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc Natl Acad Sci U S A 2008;105: 20458–63.
9. 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.
10. Schuster SC. Next-generation sequencing transforms today's biology. Nat Methods 2008;5:16–8.
11. McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, et al. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. Genome Res 2009;19:1527–41.
12. Hillier LW, Marth GT, Quinlan AR, Dooling D, Fewell G, Barnett D, et al. Whole-genome sequencing and variant discovery in *C. elegans*. Nat Methods 2008;5:183–8.
13. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res 2008;36:e105.
14. Chu T, Bunce K, Hogge WA, Peters DG. Statistical model for whole genome sequencing and its application to minimally invasive diagnosis of fetal genetic disease. Bioinformatics 2009;25:1244–50.

Previously published online at
DOI: 10.1373/clinchem.2009.136507