Dnase1l3 deletion causes aberrations in length and end-motif frequencies in plasma DNA

Lee Serpas,1, 2, Rebecca W. Y. Chan,1 Peiyong Jiang,1, 2, Meng Ni,1, 2, Kun Sun,1, 2, Ali Rashidfarrokhii,3 Chetna Soni,3 Vanja Sisirak,4, 5, Wing-Shan Lee,6, 7 Suk Hang Cheng,1, 2, Wenlei Peng,1, 2, K. C. Allen Chan,1, 2, Rossa W. K. Chiu,1, 2, Boris Reizis,1, 2, and Y. M. Dennis Lo1, 2

1Department of Pathology, New York University School of Medicine, New York, NY 10016; 2Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, NT, Hong Kong SAR, China; 3Department of Chemical Pathology, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, NT, Hong Kong SAR, China; and 4UMR 5164 Immunologie Conceptuelle, Expérimentale et Translationnelle, ImmunoConcEpT, Université de Bordeaux, 33076 Bordeaux, France

Circulating DNA in plasma consists of short DNA fragments. The biological processes generating such fragments are not well understood. DNASE1L3 is a secreted DNASE1-like nuclease capable of digesting DNA in chromatin, and its absence causes anti-DNA responses and autoimmunity in humans and mice. We found that the deletion of Dnase1l3 in mice resulted in aberrations in the fragmentation of plasma DNA. Such aberrations included an increase in short DNA molecules below 120 bp, which was positively correlated with anti-DNA antibody levels. We also observed an increase in long, multinucleosomal DNA molecules and decreased frequencies of the most common end motifs found in plasma DNA. These aberrations were independent of anti-DNA response, suggesting that they represented a primary effect of DNASE1L3 loss. Pregnant Dnase1l3−/− mice carrying Dnase1l3−/− fetuses showed a partial restoration of normal frequencies of plasma DNA end motifs, suggesting that DNASE1L3 from Dnase1l3-proficient fetuses could enter maternal systemic circulation and affect both fetal and maternal DNA fragmentation in a systemic as well as local manner. However, the observed shortening of circulating fetal DNA relative to maternal DNA was not affected by the deletion of Dnase1l3. Collectively, our findings demonstrate that DNASE1L3 plays a role in circulating plasma DNA homeostasis by enhancing fragmentation and influencing end-motif frequencies. These results support a distinct role of DNASE1L3 as a regulator of the physical form and availability of cell-free DNA and may have important implications for the mechanism whereby this enzyme prevents autoimmunity.

Significance

Circulating DNA in plasma has many diagnostic applications, including noninvasive prenatal testing and cancer liquid biopsy. Plasma DNA consists of short fragments of DNA. However, there is little information about mechanisms that are involved in the fragmentation of plasma DNA. We showed that mice in which Dnase1l3 had been deleted showed aberrations in the fragmentation of plasma DNA. We also observed a change in the ranked frequencies of end motifs of plasma DNA caused by the Dnase1l3 deletion. In Dnase1l3−/− mice pregnant with Dnase1l3−/− fetuses, we observed a partial reversal of the plasma DNA aberrations. This study has thus linked the fields of nucleae biology and circulating nucleic acids and has opened up avenues for future research.

Author contributions: Y.M.D.L. initiated the study; Y.M.D.L., R.W.Y.C., L.S., and B.R. designed research; L.S., R.W.Y.C., A.R., and S.H.C. performed research; P.J., M.N., K.S., W.L., and W.P. performed bioinformatics analysis; R.W.Y.C., P.J., K.S., K.C.A., R.W.C., B.R., and Y.M.D.L. analyzed data; L.S., A.R., C.S., V.S., and B.R. generated the gene-targeted mice; and R.W.Y.C., P.J., and Y.M.D.L. wrote the paper.

Reviewers: M.R.S., Medical University of Graz; and A.R.T., U896 INSEMER, Institut Recherche en Cancérologie de Montpellier.

Conflict of interest statement: Y.M.D.L. is a scientific cofounder and a member of the scientific advisory board for Grail. Y.M.D.L., R.W.K.C., and K.C.A.C hold equity in Grail and receive research funding from Grail/Cirina for other projects. Y.M.D.L., R.W.K.C., and K.C.A.C. are co-founders and board members of DNA Company Limited. Y.M.D.L., R.W.Y.C., R.W.K.C., K.C.A.C., P.J., L.S., and B.R. plan to file a patent application based on the data generated from this work. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: Sequence data for this work have been deposited in the European Genome-Phenome Archive (EGA), https://www.ebi.ac.uk/ega/, hosted by the European Bioinformatics Institute (accession no. EGA500001003174).

1L.S. and R.W.Y.C. contributed equally to this work.
2To whom correspondence may be addressed. Email: Boris.Reizis@nyulangone.org or loym@cuhk.edu.hk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1815031116/-/DCSupplemental.

Published online December 28, 2018.

www.pnas.org/cgi/doi/10.1073/pnas.1815031116

PNAS | January 8, 2019 | vol. 116 | no. 2 | 641–649
Results

Size Profiling of Plasma DNA Using Electrophoresis. Mice with targeted Dnase1l3 deletion were the primary focus of the present work. To test for potential genetic redundancy with Dnase1, we included mice with targeted Dnase1 deletion as well as mice doubly deficient in Dnase1 and Dnase1l3. Dnase1l3-deficient mice develop anti-DNA IgG antibodies and progressive systemic lupus erythematosus (SLE)-like disease (20), and we had previously demonstrated that human subjects with SLE have a variety of plasma DNA aberrations that were related to binding of anti-DNA antibodies to plasma DNA (21). As CD40 ligand (CD40LG) is important for mounting a humoral immune response (22, 23), we examined mice doubly deficient in DNASE1L3 and CD40LG to distinguish between the primary effects of DNASE1L3 deficiency and secondary effects of anti-DNA response in such mice.

SI Appendix. Fig. S1 shows the electrophoretic patterns of plasma DNA sequencing libraries prepared from WT mice and gene-targeted mice with different genotypes. The key observation was that Dnase1l3 deletion was associated with increased prominence of the bands with mean peak values at 497 bp, 683 bp, and 906 bp (marked e, f, and g, respectively, in SI Appendix, Fig. S1). These signals, after deducting the length of the adaptor primers ligated to the DNA molecules (125 bp), correspond to di-, tri-, and tetranucleosomal sizes. A very weak dinucleosomal signal was observed for the WT mice and those with only the Dnase1 gene deleted. No difference was observed between mice with just the Dnase1l3 gene deleted and those in which Dnase1l3 was deleted together with Dnase1 or Cd40lg.

Sequencing of Plasma DNA. Paired-end sequencing was performed for the plasma DNA libraries from the WT mice and the four types of gene-targeted mice. Following DNA sequence alignment, the deletion of Dnase1, Dnase1l3, and Cd40lg genes for the respectively targeted mice could be observed (SI Appendix, Fig. S2).

Paired-end sequencing allowed the plasma DNA sizes to be determined at single-nucleotide resolution. Fig. 1A shows the overall size distribution of plasma DNA from the WT and four types of targeted mice. A peak frequency at ~165 bp was reminiscent of the nucleosomal nature of plasma DNA. Fig. 1B is a plot of the size distribution in which the y axis is plotted on a logarithmic scale so that the frequencies of the longer DNA fragments, which are a minority population, can be seen more clearly. Fig. 1B shows that Dnase1l3 deletion is associated with an increase in the frequencies of plasma DNA molecules above 250 bp. Mice with Dnase1 deletion did not show any notable difference in the frequencies of plasma DNA molecules above 250 bp compared with WT mice, as reported previously (17). Mice with only Dnase1l3 deletion, those with Dnase1/Dnase1l3 double deletion, and those with Dnase1l3/Cd40lg double deletion had similarly increased frequencies of plasma DNA molecules above 250 bp. The differences in the plasma DNA size profiles between the WT mice and mice with Dnase1l3−/− are illustrated in SI Appendix, Fig. S3. Fig. 2 indicates that Dnase1l3 deletion, either alone or in combination with Dnase1 or Cd40lg deletion, was associated with an increased percentage of plasma DNA molecules >250 bp. We observed a larger than twofold increase in the percentage of plasma DNA molecules of over 250 bp in mice with Dnase1l3 deletion (median: 4.9%; range: 2.9–10.4%) compared with WT mice (median: 2.3%; range: 1.37–6.20%).

Fig. 1C shows the frequencies of relatively short plasma DNA molecules in the size range of 20–120 bp in the plasma of the WT and targeted mice. Notably, there was an increase in the frequencies of short plasma DNA molecules in the plasma of mice with Dnase1l3 deletion, and those with double deletion of both Dnase1l3 and Dnase1, compared with WT mice and those with Dnase1 deletion or Dnase1l3/Cd40lg double deletion. In addition, we observed a 40% increase in the median percentage of plasma DNA molecules of less than 120 bp in mice with Dnase1l3 deletion (median: 18.6%; range: 11.6–28.9%) compared with WT mice (median: 13.3%; range: 7.0–35.6%; SI Appendix, Fig. S4).

In our previous work with human subjects with SLE, the presence of anti-DNA antibodies was associated with an increase...
in the frequencies of short plasma DNA molecules (21). We thus explored the correlation between the levels of anti-dsDNA antibody and the frequencies of short DNA molecules in the plasma of the studied mice. SI Appendix, Fig. S5A shows a positive correlation between the percentage of short plasma DNA molecules (≤120 bp) and the anti-dsDNA antibody levels (r = 0.5577, P = 0.0009, Spearman correlation). In contrast, we observed no correlation between the percentage of long plasma DNA molecules (>250 bp) and the anti-dsDNA antibody levels (SI Appendix, Fig. S5B; r = 0.0079, P = 0.9657, Spearman correlation). These results suggest that the increase in short DNA fragments may be caused by the anti-DNA response in Dnase1l3-deficient animals, whereas the increase in long DNA fragments may be a primary effect of DNASE1L3 deficiency.

The data in Figs. 1 and 2 were generated using the Illumina platform, which was designed for sequencing short reads and has a practical limit of readout at ~600 bp. In view of this limitation, we also generated data using the single-molecule, real-time technology from Pacific Biosciences, which could generate sequencing data for molecules kilobases in length (24, 25). Using the platform from Pacific Biosciences, we further sequenced two plasma DNA libraries from each of the groups. We observed that plasma DNA molecules of 600 bp to 2,000 bp in size accounted for less than 4.8% of the circulating DNA molecules. SI Appendix, Fig. S6 shows data from the Pacific Biosciences platform and indicates that the proportion of long plasma DNA fragments in mice with Dnase1l3 deletion is higher than that of the WT mice or mice with Dnase1 deletion.

Frequency Distribution of Plasma DNA End Motifs and Dnase1l3 Deletion. We next investigated if there might be DNA sequence motifs that would be preferentially cleaved by DNASE1L3. We reasoned that if DNASE1L3 contributed to the generation of plasma DNA fragments, then the preferred DNA sequence motifs would be present at high frequencies in the plasma of WT mice and reduced in the plasma of Dnase1l3−/− mice. Hence, we ranked the end motifs of plasma DNA in WT mice in descending order of frequencies (Dataset S1, tab “Wildtype”). Then, we removed from this ranked order any end motif that did not show significant reduction in the Dnase1l3−/− mice (Dataset S1, tab “Dnase1l3 del”). The top 25 end motifs that were found by such a procedure were all within the top 42 motifs defined in WT mice (Dataset S1), and are listed in Table 1. The top six motifs, CCCA, CCTG, CCAG, CCAA, CCAT, and CCTC, all started with CC and thus were considered together in subsequent analyses. These six motifs were all within the top eight motifs defined in WT mice, and together account for 7.43% (range: 6.05–8.10%) of plasma DNA ends in WT mice but only for 4.22% (range: 3.46–4.34%) in mice with Dnase1l3 deletion. Fig. 3 shows the frequencies of the combined percentages of the top six motifs in the plasma of WT mice and gene-targeted mice of different genotypes. We observed that Dnase1l3 deletion was associated with a 45.3% reduction in the median frequency of these motifs compared with WT mice. Frequencies for each of the six motifs are shown in SI Appendix, Fig. S7 A–F and are consistent with the combined results shown in Fig. 3. Thus, the loss of DNASE1L3 reduced the frequencies of the most common sequence motifs of plasma DNA ends, suggesting that DNASE1L3 is primarily responsible for generating them in the steady state.

Size Profiling of Plasma DNA in Pregnant Mice. Plasma of a pregnant subject contains a mixture of DNA from the pregnant mother and her fetus(es) (1). By crossing female B6 mice with male BALB/c mice, the fetuses would have one half of their genomes with a B6 genetic background and the other half with a BALB/c background. Any DNA molecules in the plasma of the pregnant mothers that had a BALB/c genetic signature could be classified as fetal-derived.

Table 2 shows the characteristics of these pregnant mice. Fig. 4 shows the mean size profile of plasma DNA from two pregnant WT mice carrying WT fetuses, four Dnase1l3−/− pregnant mice carrying Dnase1l3−/− fetuses, and three Dnase1l3−/− pregnant mice carrying Dnase1l3−/− fetuses. An increase in the frequencies of relatively long plasma DNA molecules above 250 bp as well as an increase in the frequencies of relatively short plasma DNA molecules below 120 bp were observed in all Dnase1l3−/− pregnant mice irrespective of the fetal genotype. These characteristics are similar to those of plasma samples from nonpregnant mice (Fig. 1B).

We also sought to explore if differential size distributions between circulating fetal and maternal DNA molecules were still observable when the pregnant mother was deficient in DNASE1L3. Such analyses would also allow one to measure the percentage of fetal DNA in a particular maternal plasma sample (Table 2). However, circulating DNA molecules with genetic signatures of the B6 strain were predominantly, but not completely, of maternal origin. Fetal-derived DNA with genetic signatures of the B6 strain would be expected to comprise a small proportion of circulating DNA molecules as well.

Fig. 5A shows the size profile of BALB/c (i.e., fetal) and B6 (i.e., predominantly maternal) DNA in the plasma of a pregnant Dnase1l3−/− mouse carrying WT Dnase1l3−/+ fetuses. As expected, the fetal DNA exhibited a relatively shorter size profile compared with that of the maternal DNA. SI Appendix, Fig. S8 shows that the percentage of fetal DNA below 150 bp (median: 32.8%; range: 20.5–40.6%) was lower than that of maternal DNA (median: 54.3%; range: 36.3–65.6%) in the pregnant murine plasma. Fig. 5B shows cumulative frequency plots of BALB/c and B6 DNA in maternal plasma. These cumulative frequency plots show the progressive accumulation of DNA molecules from the short to long ones. We further define a parameter, called ΔS, which represents the difference in the cumulative frequencies between the fetal and maternal DNA at a particular size (Fig. 5B). A positive value of ΔS for a particular size suggests that a larger amount of the fetal DNA molecules is present below that particular size compared with the maternal DNA molecules. Fig. 5C shows ΔS across different sizes between the BALB/c and B6 DNA in maternal plasma relative to DNA fragment size. ΔS peaks were observed at the nucleosomal (163 bp) and dinucleosomal (334 bp) sizes.
These results suggested that fetal DNA was present with a relatively higher proportion of fragments below 163 bp compared with the situation for maternal DNA. Furthermore, this difference between the size profiles of circulating fetal and maternal DNA was observed with or without Dnase1l3 deletion.

Fig. 5 D–F show the size distribution, cumulative frequency, and ΔS plots for a pregnant Dnase1l3−/− mouse carrying Dnase1l3−/− fetuses. These plots were similar to those for pregnant WT (Dnase1l3+/+) mice carrying WT (Dnase1l3+/+) fetuses shown above. Hence, the shortening of circulating fetal DNA in maternal plasma was observed even in a pregnant mother that was Dnase1l3−/−. The size difference between the maternal and fetal DNA of all other pregnant WT or Dnase1l3−/+ mice is shown in SI Appendix, Fig. S9.

In this pregnancy model, when both the mothers and fetuses were of the B6 genetic background, one could identify the fetal DNA in the maternal plasma by analyzing DNA originating from the Y chromosome. As multiple fetuses were typically present in a murine pregnancy, there was a high probability that at least one of the fetuses would be male. SI Appendix, Fig. S10 shows the DNA size profiles of the Y chromosome (i.e., fetal) and autosomes (i.e., predominantly maternal) in the plasma of the pregnant WT (Dnase1l3+/+) mice carrying WT (Dnase1l3+/+) fetuses (SI Appendix, Fig. S10A) and the pregnant Dnase1l3−/+ mice carrying Dnase1l3−/+ fetuses (SI Appendix, Fig. S10B). The fetal DNA molecules were consistently found to be shorter than maternal DNA molecules. These findings were consistent with those of the nonpregnant model in which an increase in the frequencies of both short (<120 bp) and long (>250 bp) DNA molecules was observed in Dnase1l3−/+ mothers compared with WT (Dnase1l3+/+) mice (SI Appendix, Fig. S11). Thus, the aberrant size of plasma DNA in Dnase1l3-deficient dams cannot be rescued by the presence of Dnase1l3 in the fetus.

### Table 1. Top 25 motifs with the highest frequency in WT mice and statistically significant reduction in Dnase1l3−/− mice

| Motif | Motif frequency of WT, % (a) | Motif frequency of Dnase1l3−/−, % (b) | Fold change (a/b) | P value |
|-------|-----------------------------|-------------------------------------|-------------------|---------|
| CCCA  | 1.51                        | 0.76                                | 1.98              | <0.0001 |
| CCTG  | 1.45                        | 0.80                                | 1.81              | <0.0001 |
| CCAG  | 1.37                        | 0.56                                | 2.47              | <0.0001 |
| CCAA  | 1.12                        | 0.51                                | 2.21              | <0.0001 |
| CCAT  | 1.11                        | 0.67                                | 1.66              | <0.0001 |
| CCTC  | 1.10                        | 0.93                                | 1.18              | <0.0001 |
| CAAA  | 1.02                        | 0.78                                | 1.32              | <0.0001 |
| TGGT  | 0.98                        | 0.51                                | 1.92              | <0.0001 |
| TGGT  | 0.96                        | 0.58                                | 1.65              | <0.0001 |
| CCTA  | 0.87                        | 0.52                                | 1.68              | <0.0001 |
| TATT  | 0.84                        | 0.61                                | 1.39              | <0.0001 |
| CCAC  | 0.81                        | 0.45                                | 1.81              | <0.0001 |
| TCTT  | 0.78                        | 0.56                                | 1.39              | <0.0001 |
| CCCC  | 0.77                        | 0.54                                | 1.44              | <0.0001 |
| TGAG  | 0.76                        | 0.44                                | 1.73              | 0.0001  |
| CAAG  | 0.73                        | 0.50                                | 1.45              | <0.0001 |
| CATG  | 0.73                        | 0.60                                | 1.20              | <0.0001 |
| TATA  | 0.72                        | 0.45                                | 1.61              | <0.0001 |
| GGAA  | 0.68                        | 0.55                                | 1.25              | <0.0001 |
| TGTA  | 0.68                        | 0.37                                | 1.83              | <0.0001 |
| CATA  | 0.66                        | 0.59                                | 1.13              | <0.0001 |
| TACA  | 0.65                        | 0.42                                | 1.57              | <0.0001 |
| TCTG  | 0.65                        | 0.45                                | 1.45              | <0.0001 |
| CAAT  | 0.65                        | 0.58                                | 1.12              | <0.0001 |
| TGCT  | 0.64                        | 0.41                                | 1.57              | 0.0001  |

Plasma DNA End Motifs in Pregnant Mice. The murine pregnancy model provided an opportunity to investigate if the contribution of Dnase1l3 to the fragmentation of plasma DNA might occur at a systemic level (e.g., through DNASE1L3 activity in plasma) or at a tissue level (e.g., within the placenta). Hence, for a pregnancy involving a Dnase1l3−/− female pregnant mouse carrying Dnase1l3−/− fetuses, the pregnant mother did not express Dnase1l3 whereas the fetuses would be expected to express Dnase1l3, albeit at a lower level than Dnase1l3+/+ fetuses. If the Dnase1l3−/− fetuses could release DNASE1L3 into the circulation of their Dnase1l3−/+ pregnant mother, then one would expect that there might be a partial reversal of the phenotype associated with Dnase1l3 deletion in the pregnant mother.

We pursued this question of phenotype reversal based on the six highest-ranked plasma DNA end motifs associated with Dnase1l3 deletion described above. Fig. 6 illustrates the reduction in plasma DNA fragments with these six highest-ranked end motifs in nonpregnant Dnase1l3−/− mice compared with WT (Dnase1l3+/+) mice. Pregnancies involving pregnant mothers and fetuses that were both Dnase1l3−/− had plasma DNA end-motif frequencies that were similar to those of the nonpregnant WT mice. Using the strain difference between the pregnant mothers (i.e., B6) and the impregnating males (i.e., BALB/c), we were able to separately analyze the end-motif frequencies of plasma DNA fragments derived from the fetuses and the pregnant mothers. Hence, plasma DNA fragments with BALB/c genetic signatures were derived from the fetuses while plasma DNA fragments with B6 genetic signatures were derived predominantly from the mother, with a small proportion contributed by the fetuses. Using fetal DNA fragments, we could also calculate the percentage of fetal DNA that was present in the maternal plasma. Using the percentage of fetal DNA in maternal plasma, we mathematically adjusted the plasma DNA motif frequencies for the circulating DNA of B6 background such that they represented just those contributed by the pregnant mother, but excluding those contributed by the fetuses [under the heading “(adjusted)” in Fig. 6]. As fetal DNA represented a minority population in maternal plasma, the effect of such an adjustment was relatively minor.

For pregnancies involving Dnase1l3−/− pregnant mice and Dnase1l3−/+ fetuses, we observed that the frequencies of the six highest-ranked plasma DNA end motifs were in between those of Dnase1l3−/− and Dnase1l3+/+.
Mean size distributions of plasma DNA fragments in pregnant mice. Dnase1l3 mice January 8, 2019. Dnase1l3 no. 2 fetuses. Solid red line represents data from three PNAS and nonpregnant Dnase1l3+ fetuses were higher than those derived − fetuses. − > 0.0286, Wilcoxon rank-sum test). > 0.0286, Wilcoxon rank-sum test). fetuses. This observation suggests that there might be a local DNASE1L3 effect within placental cells. Discussion In this work, we demonstrated that Dnase1l3+/− mice exhibited a number of specific aberrations in the fragmentation of plasma DNA. First, we observed an increase in the amount of longer DNA molecules, especially in the di-, tri-, and tetranucleosomal sizes (SI Appendix, Fig. S1). DNA sequencing analyses also confirmed these results for the di- and trinucleosomal-sized fragments (Fig. 1). Due to the inefficiency of the Illumina platform in sequencing DNA molecules >600 bp, we did not plot sequencing data above that size in Fig. 1B. Second, there was a 40% increase in the amount of short plasma DNA molecules below 120 bp (Fig. 1C and SI Appendix, Fig. S3). There was a positive correlation between the amount of short DNA and anti-dsDNA antibody levels (SI Appendix, Fig. S5A). Dnase1l3+/− mice are prone to develop antibodies to chromatin and DNA and a SLE-like syndrome (20, 26). Furthermore, an increase in the amounts of short plasma DNA molecules was also observed in human subjects with SLE (21). We proposed that this increase in the amounts of short plasma DNA was related to such autoantibodies. Our hypothesis is supported by a reduction in the amounts of short plasma DNA in Dnase1l3+/−/C(4)O4− mice that failed to develop autoantibody response (SI Appendix, Fig. S4).

We hypothesized that there might be DNA motifs that DNASE1L3 might preferentially cleave. If such DNA motifs existed and if DNASE1L3 was one of the enzymes involved in the fragmentation of DNA that eventually was found in plasma, we conjectured that such DNA motifs should be present at relatively high levels in the plasma of WT mice and would be reduced in Dnase1l3+/− mice. Indeed, we found a number of such motifs (Table 1), the top six of which all started with CC. These results were consistent with the observations of Chandrananda et al. (27), who observed an overrepresentation of CC at the ends of plasma DNA sequenced from human pregnant women. Our data suggest that DNASE1L3 may contribute to this phenomenon.

One actively pursued area in the field of circulating fetal DNA involves the use of circulating fetal DNA in maternal plasma for noninvasive prenatal testing (1, 2). It has been known for a number of years that circulating fetal DNA molecules have a size distribution that is shorter than their maternal counterparts (5, 8, 9). Paired-end massively parallel sequencing enables high-resolution measurement of DNA size at single-base resolution by measuring the number of nucleotides between the ends of the sequenced fragment. Using paired-end sequencing, circulating fetal and maternal DNA molecules have been shown to have peak frequencies at 143 bp and 166 bp, respectively (5). Indeed, such a subtle difference in the size between maternal and fetal DNA has already been used in noninvasive prenatal diagnostics (9, 14).

Considering that the Illumina platform has not been designed to sequence molecules >600 bp, we used the single-molecule, real-time sequencing platform from Pacific Biosciences to further sequence two cases from each of the groups. This platform was designed to generate DNA sequences in the kilobases range (24, 25). Using the Pacific Biosciences platform, we were able to analyze plasma DNA reads well into the 1,500-bp range, and

Table 2. Characteristics of the pregnant mice

| Case no. | Maternal genome background | Maternal Dnase1l3 genotype | Age, wk | Approximate days of pregnancy | Fetal genome background | Fetal Dnase1l3 genotype | No. of fetuses | Fetal percentage*, % |
|----------|---------------------------|---------------------------|--------|-----------------------------|------------------------|------------------------|---------------|---------------------|
| Mu31     | B6                        | +/-                       | 12.0   | 15                           | B6, BALB/c             | +/-                    | 7             | 6.4                 |
| Mu45     | B6                        | +/-                       | 10.9   | 14                           | B6, BALB/c             | +/-                    | 8             | 8.4                 |
| Mu46     | B6                        | +/-                       | 11.4   | 18                           | B6, BALB/c             | +/-                    | 18            | 20.7                |
| Mu47     | B6                        | +/-                       | 31.2   | 18                           | B6, BALB/c             | +/-                    | 12            | 15.1                |
| Mu48     | B6                        | +/-                       | 34.4   | 20                           | B6, BALB/c             | +/-                    | 10            | 12.3                |
| Mu49     | B6                        | +/-                       | 38.6   | 22                           | B6, BALB/c             | +/-                    | 8             | 12.8                |
| Mu104    | B6                        | +/-                       | 13.6   | 16                           | B6                      | +/-                    | 8             | 14.9                |
| Mu105    | B6                        | +/-                       | 17.6   | 16                           | B6                      | +/-                    | 12            | 18.7                |
| Mu106    | B6                        | +/-                       | 31.2   | 18                           | B6                      | +/-                    | 8             | 12.8                |

*The fetal percentages in a pregnant mouse carrying fetuses with the B6 and BALB/c genomic background were deduced using a SNP-based approach. The fetal percentages in a pregnant mouse carrying fetuses with the B6 genomic background (i.e., Mu104, Mu105, and Mu106) were estimated using reads aligned to the Y chromosome.

One of the enzymes involved in the fragmentation of DNA is DNASE1L3.
DNASE1L3 gene in mouse did not appear to alter the length of DNA molecules from 600 bp to 2,000 bp in size for WT and Dnase1l3−/− mice carrying Dnase1l3 mice (long, Dnase1l3−/−) and the fetuses were present at very low frequencies with the average percentage of DNA molecules from 600 bp to 2,000 bp in size for WT and Dnase1l3−/− mice being present at 0.5% and 2.1%, respectively.

In particular, such pregnancies would thus be of value to investigate the role of other nucleases in this process. In this regard, we had previously demonstrated deleting both the Dnase1 and Dnase1l3 genes together appears to increase the amount of short DNA below 120 bp (Fig. 1C). This occasional difference in both short (<120 bp) and long (>250 bp) DNA between pregnant Dnase1l3−/− mice carrying Dnase1l3+/− fetuses and nonpregnant Dnase1l3−/− mice (long, P = 0.33; short, P = 0.85; Wilcoxon rank-sum test; SI Appendix, Fig. S11). We concluded that pregnant Dnase1l3−/− mice carrying Dnase1l3+/− fetuses had an overall size distribution of plasma DNA that was similar to nonpregnant Dnase1l3−/− mice. In particular, such pregnancies maintained the relative size shortening of fetal DNA molecules in maternal plasma, compared with their maternal counterparts.

However, using the plasma DNA end motif preference data (Table 1), one could further obtain insights into the site of action of DNASE1L3. Hence, in pregnancies in which the pregnant mothers were Dnase1l3−/− and the fetuses were Dnase1l3+/− one would expect that the mothers would not be able to produce DNASE1L3 while the fetuses would still be able to produce DNASE1L3. If the fetally produced DNASE1L3 enzyme was able to gain access into the maternal circulation and acted systemically, then one would observe a partial normalization of the six highest-ranked motifs in plasma DNA molecules. This was indeed observed from the data shown in Fig. 6 in which plasma DNA molecules of B6 origin (mainly of maternal origin) exhibited a partial normalization of the DNASE1L3-preferred end motifs toward the levels observed in WT mice. Because the majority of plasma DNA was derived from hematopoietic cells and the DNASE1L3 in this setting was derived from the fetus, these data suggested that DNASE1L3 might mediate plasma DNA fragmentation in a cell-extrinsic manner.

It is also notable that, for each of the six DNASE1L3-preferred end motifs, the degree of normalization of the circulating BALB/c DNA in maternal plasma, which was all fetally derived, was higher than that of circulating B6 DNA (which was predominantly of maternal origin) (SI Appendix, Fig. S12 A-F). This observation suggests that there might be a higher activity of the DNASE1L3 enzyme at the site of its production from the fetuses. As fetal DNA in maternal plasma in humans has been shown to be derived from the placenta (28, 29), one site of action of the fetally derived DNASE1L3 might be in the placenta. This possibility is supported by recent data from a mouse cell atlas project in which Dnase1l3 gene expression has indeed been found in the murine placenta (30).

Our work reported here has demonstrated that DNASE1L3 plays a role in the fragmentation of plasma DNA. It is likely that there might be other players in this fragmentation process. It would thus be of value to investigate the role of other nucleases in this process. In this regard, we had previously demonstrated deleting the Dnase1 gene in mouse did not appear to alter the length of circulating DNA (17). In the present work, our data suggest that deleting both the Dnase1 and Dnase1l3 genes together appears to increase the amount of short DNA below 120 bp (Fig. 1C). This...
observation would require further confirmation and exploration in future work. It is also possible that the loss of DNASE1 enzymatic function, either on its own or in conjunction with DNASE1L3, might result in aberrations of circulating DNA that are not seen with the techniques used in the present work. We would continue to explore if DNASE1L3 and DNASE1 may work synergistically on DNA fragmentation (31). Further work using other sequencing platforms, such as nanopore sequencing (32), might be used to explore this issue. It would also be interesting to further explore the Dnase1 deletion mouse model in the pregnancy context, as has been done for Dnase1l3 in the present work and in other experimental systems, for example sepsis (33).

Homozygous mutations of the DNASE1L3 gene in humans have been described to be associated with SLE, often preceded by hypocomplementemic urticarial vasculitis syndrome (34–36). It would be interesting to see if the plasma DNA aberrations observed in the mouse models described in this paper can also be seen in these human subjects. It also remains to be examined how the observed lengthening of circulating DNA is related to the increased DNA load of circulating microparticles in DNASE1L3-deficient patients and experimental animals (20), and how it may contribute to the breach of tolerance and anti-DNA responses in these conditions.

In the present study, it is intriguing that the plasma DNA aberrations observed in the mice with the deletion of Dnase1l3 only accounted for a small proportion of circulating DNA molecules (<10%). We postulated that DNASE1L3 may only be one of a number of nucleases that are involved in the homeostasis of circulating DNA. To identify these additional players in fragmentation mechanisms, further studies using the experimental approach described in this work might be a fruitful avenue of research.

In summary, we have demonstrated that deletion of the Dnase1l3 gene in mice was associated with aberrations in plasma DNA fragmentation. These results suggested that DNASE1L3 was one of the components of the mechanism for generating circulating DNA fragments. We believe that it would be a productive direction of research to explore the effects of other nucleases in the characteristics of circulating nucleic acids. An enhanced understanding of the biology of circulating DNA might help us to design optimal strategies for their use in molecular diagnostics.

---

Fig. 6. Percentages of fragments with the six selected motifs, including CCCA, CCTG, CCAG, CCAA, CCAT, and CCTC, in plasma DNA of nonpregnant mice; maternal and fetal DNA fractions in plasma of pregnant mice with different Dnase1l3 genotypes. Each circle represents the data from one mouse. The genotypes of the nonpregnant mice as well as the fetal and maternal DNA subset of pregnant mice are annotated under the x axis. The first two groups are nonpregnant WT and Dnase1l3−/− mice, respectively. For pregnant mice, the strains of the maternal and fetal DNA subset in plasma are also annotated. B6 and BALB/c are two different mouse strains with different genetic backgrounds. As the motifs of fetal DNA sequences would theoretically affect the measurement of motifs in the shared sequences, the abundance of a motif from maternally derived sequences was adjusted by the fetal DNA fraction as described in Materials and Methods. The adjusted percentages of the selected motifs of the maternal DNA subset are labeled with “(adjusted)” on the x axis. Statistical difference in the percentage between maternal and fetal DNA subsets of pregnant mice with Dnase1l3 deletion was calculated by using the Wilcoxon rank-sum test.
Materials and Methods

Animals. All animal studies were performed according to the investigator’s protocol approved by the Institutional Animal Care and Use Committees of New York University School of Medicine. The Dnase13−/− mouse model on C57BL/6 (B6) background was described previously (20). Mice carrying a targeted allele of Dnase1 (Dnase1tm1.1(EMK)Fli) on B6 background were obtained from the Knockout Mouse Project and crossed with WT B6 and Dnase13−/− mice to obtain Dnase1−/− and Dnase1+/−/Dnase1−/− mice, respectively. Mice carrying a targeted allele of Cd40 (B6.129S2-Cd40tm2miv/J) on B6 background were obtained from the The Jackson Laboratory and crossed to obtain Dnase13−/−/Cd40−/− mice. WT control mice of B6 background were bred in the same animal facility or purchased from Taconic, Inc. and maintained in the same facility. WT mice of BALB/c background for the pregnancy studies were obtained from Taconic, Inc. Pregnant dams were killed and exsanguinated for plasma collection, and pregnancy terms were estimated by embryo morphology.

Sample Processing and DNA Extraction. Animals were killed and exsanguinated by cardiac puncture. Blood was transferred into EDTA-containing collection tubes. The blood samples were first centrifuged at 1,600 × g for 10 min at 4 °C. The plasma portion was further subjected to centrifugation at 16,000 × g for 10 min at 4 °C to pellet the residual cells and platelets. The resulting plasma was harvested. Circulating cell-free DNA was extracted from 0.2 to 0.5 mL of plasma using the DSP Blood Mini Kit (Qiagen) as previously described (37).

DNA Library Preparation and Electrophoresis. Circulating DNA libraries were constructed by using the KAPA HTP Library Preparation Kit (Roche) and purified using a MinElute Reaction Cleanup Kit (Qiagen) according to the manufacturer’s instructions. Adaptor-ligated libraries were analyzed on an Agilent 4200 TapeStation (Agilent Technologies) using the High Sensitivity D1000 ScreenTape System (Agilent Technologies) for quality control and gel-based size determination. Before sequencing, the libraries were quantified by qPCR using a KAPA Library Quantification Kit (Roche) on a LightCycler 96 instrument (Roche).

DNA Sequencing Using the Illumina Platform. The DNA libraries were sequenced for 75 bp for each end in a paired-end format on a NextSeq 500 instrument (Illumina). Real-time image analysis and base calling were performed using the NextSeq Control Software v2.1.0 and Real Time Analysis Software v2.4.11 (Illumina). After base calling, adapter sequences and low quality bases (i.e., quality score < 5) on the 3’ ends of the reads were removed.

For the analysis of sequencing data, the sequenced reads were aligned to the non-repetitive masked mouse reference genome (NCBI build 37/UCSC mm9) using the Short Oligonucleotide Alignment Program 2 as previously described (37, 38). Only paired-end reads which were aligned to the same chromosome in a correct orientation with an insert size less than 5,000 bp were used for downstream analysis. For paired-end reads sharing the same start and end aligned genomic coordinates only one would be kept for further analysis, while the rest were deemed to be PCR duplicates and were discarded. SI Appendix, Table S1 summarizes the number of sequenced fragments of each sample detected by using the Illumina platform.

DNA Sequencing Using the Pacific Biosciences Sequencing Platform. Sequencing templates were constructed using SMRTbell Template Prep Kit 1.0 - SPV3 (Pacific Biosciences) according to the manufacturer’s instructions, except that the ampiclon templates were purified with AMPure PB beads (Pacific Biosciences). Sequencing primer annealing and polymerase binding conditions were calculated with SMRT Link v5.1.0 software (Pacific Biosciences). Briefly, sequencing primer v3 was annealed to sequencing template, then polymerase was bound to templates using Sequel Binding and Internal Control Kit 2.1 (Pacific Biosciences). Sequencing was performed on a Sequel SMRT Cell 1M v2 for each template. Each sequencing movie was collected on the Sequel system for 10 h with Sequel sequencing kit 2.1 (Pacific Biosciences).

Molecular Size Determination of Circulating DNA. Following paired-end sequencing, both ends sequences of each DNA molecule were aligned to the mouse reference genome for the B6 strain (NCBI build 37/UCSC mm9). The genome coordinates of the aligned ends were then used to deduce the sizes of the sequenced DNA molecules (8). SI Appendix, Table S2 summarizes the percentage and number of long DNA molecules detected using the Illumina and the Pacific Biosciences sequencing platforms.

Motif Analysis of Circulating DNA. To study whether deletion of the Dnase13 gene would alter the cleavage pattern of circulating DNA, the first 4-bp sequence (i.e., a 4-mer motif) on each 5’ fragment end of plasma DNA molecules was determined. The frequency of each of the 256 possible motifs (i.e., 4 × 4 × 4 × 4) was calculated and normalized by the total number of ends. For each motif, the difference in the frequency between Dnase13−/− mice and WT mice was tested by the Wilcoxon rank-sum test and its P value was adjusted by the Holm-Bonferroni procedure because of multiple comparisons (39).

Size and Motif Analyses for Pregnancy Model Involving both the B6 and BALB/c Genetic Backgrounds. The female Dnase13−/− and WT (Dnase13+/−) mice of the B6 genetic background were impregnated by male WT (Dnase13+/−) mice of the BALB/c genetic background. Hence, their fetuses would inherit the SNP signatures of the mother (B6) and BALB/c (father) strains. The maternal plasma DNA of a pregnant B6 mouse was composed of molecules from both the maternal B6 genome and the paternal BALB/c genome. In the bioinformatics data analysis, the paired-end reads from plasma DNA of pregnant mice were initially aligned to the B6 reference genome.

Based on the alignment results, the sequenced fragments bearing the paternal-specific variants (i.e., fetal DNA molecules) were identified using the 4,991,500 SNPs that were different between B6 and BALB/c strains, which were annotated in the Mouse Genomes Project (https://www.sanger.ac.uk/science/data/mouse-genomes-project). Since there were a number of non-single nucleotide genetic differences between the two strains, for example small insertions/deletions and copy number variations, the genomic coordinates of the two strains were not consistent. Hence, the accuracy in determining fragment sizes and motifs would be affected when analyzing BALB/c sequences using the B6 genome as a reference. To overcome this issue, DNA fragments carrying the paternal-specific variants were realigned to the BALB/c reference genome and the actual fragment size of fetal DNA molecules was determined based on the realigned genomic coordinates. The motifs of fetal DNA molecules were also deduced from realigned results in the BALB/c reference genome. However, the plasma DNA molecules bearing the shared variants would generally reflect the maternal characteristics because the majority of those molecules were likely derived from B6 maternal hematopoietic cells. Thus, for those DNA fragments carrying shared alleles, the fragment size and motif analyses were performed directly using paired-end reads aligned to the B6 reference genome.

The fetal DNA fraction (f) was calculated by using the following formula:

\[ F = \frac{2p}{q + 2p} \times 100\% \]

where p is the number of sequenced reads carrying fetal-specific variants and q is the number of sequenced reads carrying variants shared by the mother and the fetuses.

Motif Correction for Shared Sequences in Pregnant Mice Involving both the B6 and BALB/c Genetic Backgrounds. We analyzed the motifs derived from maternal and fetal sequences, the plasma DNA fragments carrying the informative SNPs, in which the maternal and paternal genotypes were both homozygous but for a different variant each. In this scenario, the maternal and paternal genotypes could be denoted by AA and BB, respectively, making the fetal genotype AB. There was a small proportion of fetal DNA fragments contributing to those shared sequences carrying variants of A.

The motifs of fetal DNA sequences would theoretically affect the measurement of motifs in the shared sequences. Thus, to infer the abundance of a motif present in maternally derived sequences, we needed to adjust the abundance of motifs of maternal sequences according to the fetal DNA fraction (f). Assuming that the observed motifs abundance (M) in shared sequences was a linear combination of motifs contributed by the maternally derived sequences (M1) and the fetally derived sequences (M2), respectively, we derived the following formula:

\[ M = M_1 \times (1 - P) + M_2 \times P = M_1 - P + M_2 \]

where P was the percentage of shared sequences released by the fetuses, which was

\[ P = \frac{f}{2 - f} \]

Thus, M2 could be deduced according to Eq. 1:

\[ M_2 = \frac{2M - (M_1 + M_2) \times f}{2 - 2f} \]

M1 was the abundance of a motif in shared sequences adjusted by the fetal DNA fraction and would thus correctly reflect the motifs derived from maternal-derived sequences.

Serpas et al.
Size and Motif Analyses of Pregnancy Model Involving only the B6 Genetic Background. To test whether physiological changes in pregnancy would confound the motif analysis, we allowed three Dnase1(-/-) B6 females to mate with the Dnase1(+/+) B6 males, resulting in pregnancies carrying Dnase1(+/+) fetuses of the B6 background. Fetal DNA reads were taken as those aligned to the Y chromosome, which were originated from the male fetuses. The size analysis of the fetal DNA was performed using such Y chromosome-aligned reads. As the size analysis of these three pregnant mice was the same as that for the nonpregnant mice.

Quantification of Anti-dsDNA IgG Antibodies. Plasma was obtained from mice using the same protocol as described above. Anti-dsDNA IgG titers were determined by ELISA using plates precoated with poly-γ-lysine (0.01% wt/vol in PBS) for 1 h at room temperature and then coated with 0.01 mg/mL calf thymus DNA as an antigen. After incubation with plasma, the amount of bound IgG was measured with an alkaline phosphatase-conjugated goat anti-mouse IgG antibody (1:1,000; Jackson Immunoresearch). Antigen-specific IgG levels were determined using serial dilutions of the plasma from a positive animal as a standard and expressed as units per milliliter.

Statistical Analysis. Analysis was performed using custom-built programs written in Perl and R languages. Statistical difference was calculated using the Wilcoxon rank-sum test unless otherwise specified. A P value of less than 0.05 was considered as statistically significant and all probabilities were two-tailed.

ACKNOWLEDGMENTS. This work was supported by the Research Grants Council of the Hong Kong SAR Government under Theme-Based Research Scheme T12-403/15-N; the Colton Center for Autoimmunity; the Lupus Research Alliance; and NIH Grants AR071703, AR070591, GM007308, and AI100853. Y.M.D.L. is supported by an endowed chair from the Li Ka Shing Foundation.