Supplemental figures for

Analysis of a cell-free DNA-based cancer screening cohort links fragmentomic profiles, nuclease levels and plasma DNA concentrations

Yasine Malki[1,2,3,5], Guannan Kang[1,2,3,5], W. K. Jacky Lam[1,2,3,4,5], Qing Zhou[1,2,3], Suk Hang Cheng[1,2,3], Peter P.H. Cheung[2,3], Jinyue Bai[1,2,3], Ming Lok Chan[2,3], Chui Ting Lee[2,3], Wenlei Peng[1,2,3], Yiqiong Zhang[2,3], Wanxia Gai[1,2,3], Winsome W.S. Wong[1,2,3], Mary-Jane L. Ma[1,2,3], Wenshuo Li[1,2,3], Xinzhou Xu[2,3], Zhuoran Gao[2,3], Irene O. L Tse[2,3], Huimin Shang[1,2,3], L.Y. Lois Choy[1,2,3,4], Peiyong Jiang[1,2,3,4], K. C. Allen Chan[1,2,3,4], Y.M. Dennis Lo[1,2,3,4*]

[1]Centre for Novostics, Hong Kong Science Park, Pak Shek Kok, Hong Kong SAR, China; [2]Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China; [3]Department of Chemical Pathology, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China; [4]State Key Laboratory of Translational Oncology, Sir Y.K. Pao Centre for Cancer, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China.

[5]These authors contributed equally to this work.

*Corresponding author: loym@cuhk.edu.hk.

Supplemental Figure 1

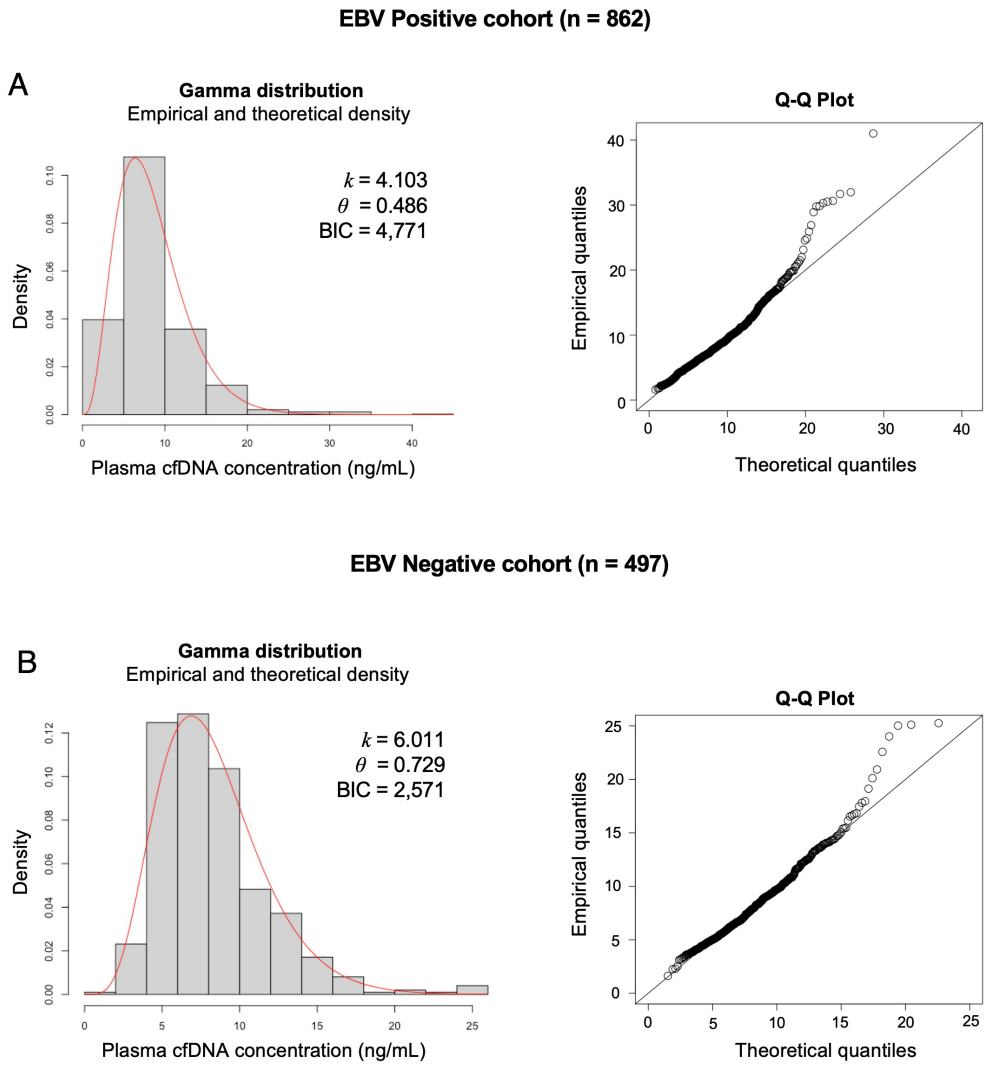

Figure S1. Plasma cfDNA concentration of individuals from the study cohorts represented as a gamma function distribution, with respective Q-Q plots, as shown for the (A) EBV-positive cohort and (B) EBV-negative cohort.

Supplemental Figure 2

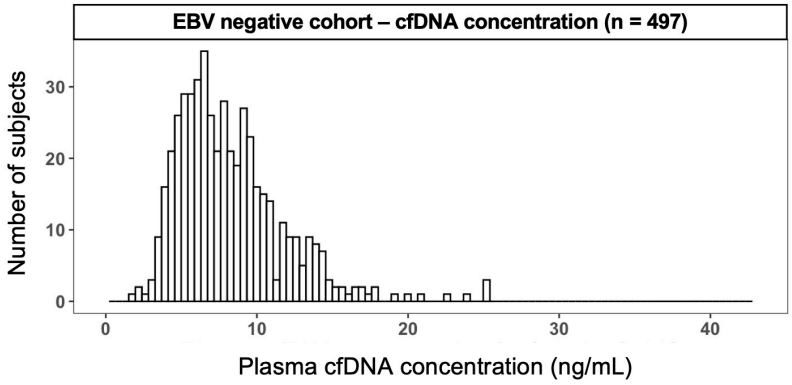

Figure S2. Frequency distribution histogram plot showing the distribution of plasma cfDNA concentration from 497 individuals of the EBV-negative cohort.

Supplemental Figure 3

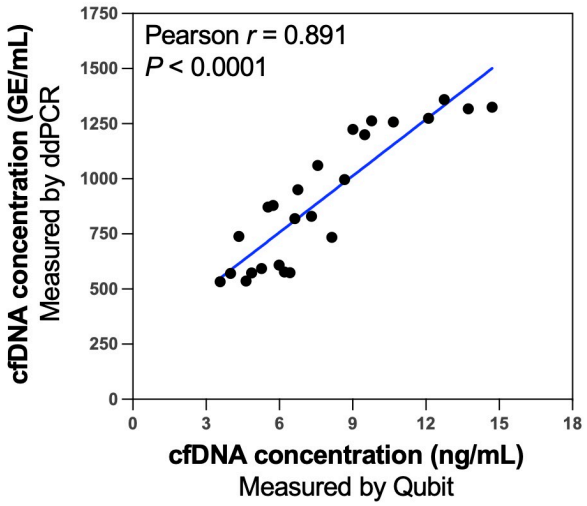

Figure S3. Correlation of the cfDNA concentration from 25 individuals of the EBV-negative cohort, from paired measurements using Qubit and digital droplet PCR (ddPCR). ddPCR was performed with a primer pair within the VCP gene, producing a short amplicon of 68-bp. The y-axis represents cfDNA concentration measured by ddPCR (Genomic equivalents, GE, per mL of plasma), while the x-axis is cfDNA concentration measured by Qubit (ng of DNA per mL of plasma).

Supplemental Figure 4

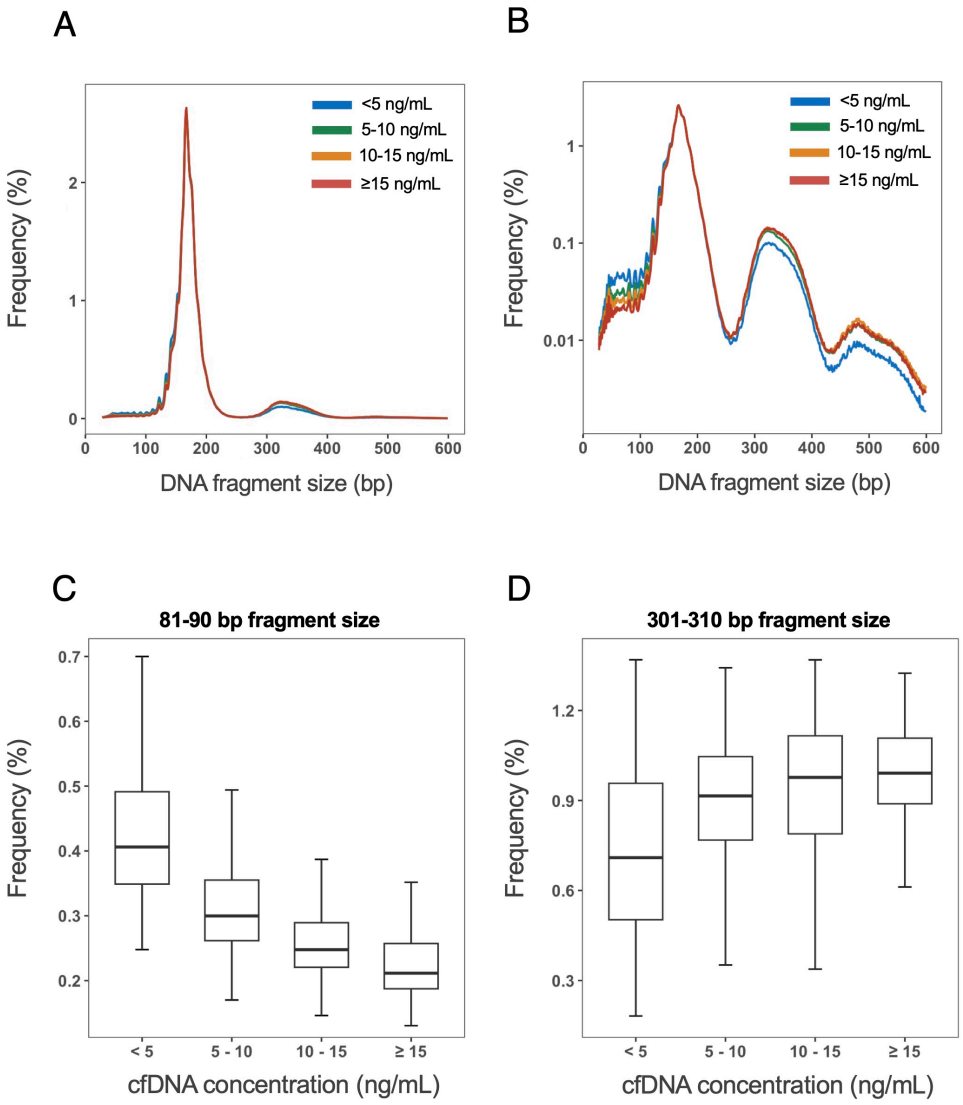

Figure S4. Analysis of the size profile of cfDNA with different plasma cfDNA concentration ranges. The mean size profile of plasma DNA fragments are plotted in (A) linear scale and (B) logarithmic scale. Boxplots showing the frequency of cfDNA fragments for the (C) 81-90 bp fragment size range, and (D) 301-310 bp fragment size range.

Supplemental Figure 5

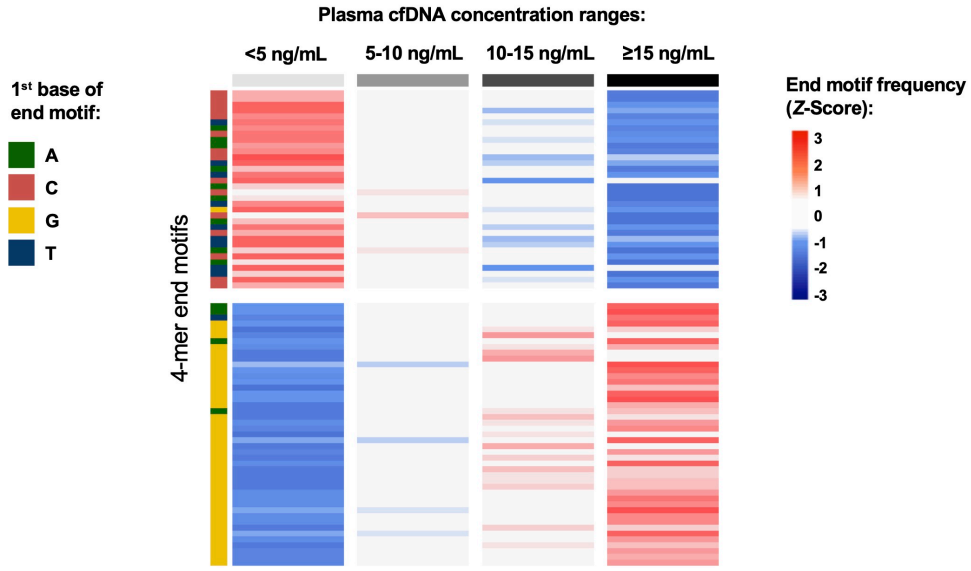

Figure S5. Heatmap analysis showing frequency *z*-scores of the selected 79 end motifs with different plasma cfDNA concentration ranges. The mean motif ratio of all individuals from each plasma cfDNA concentration range was used, and the frequency *z*-score was calculated for each end motif, as shown by the color scale.

Supplemental Figure 6

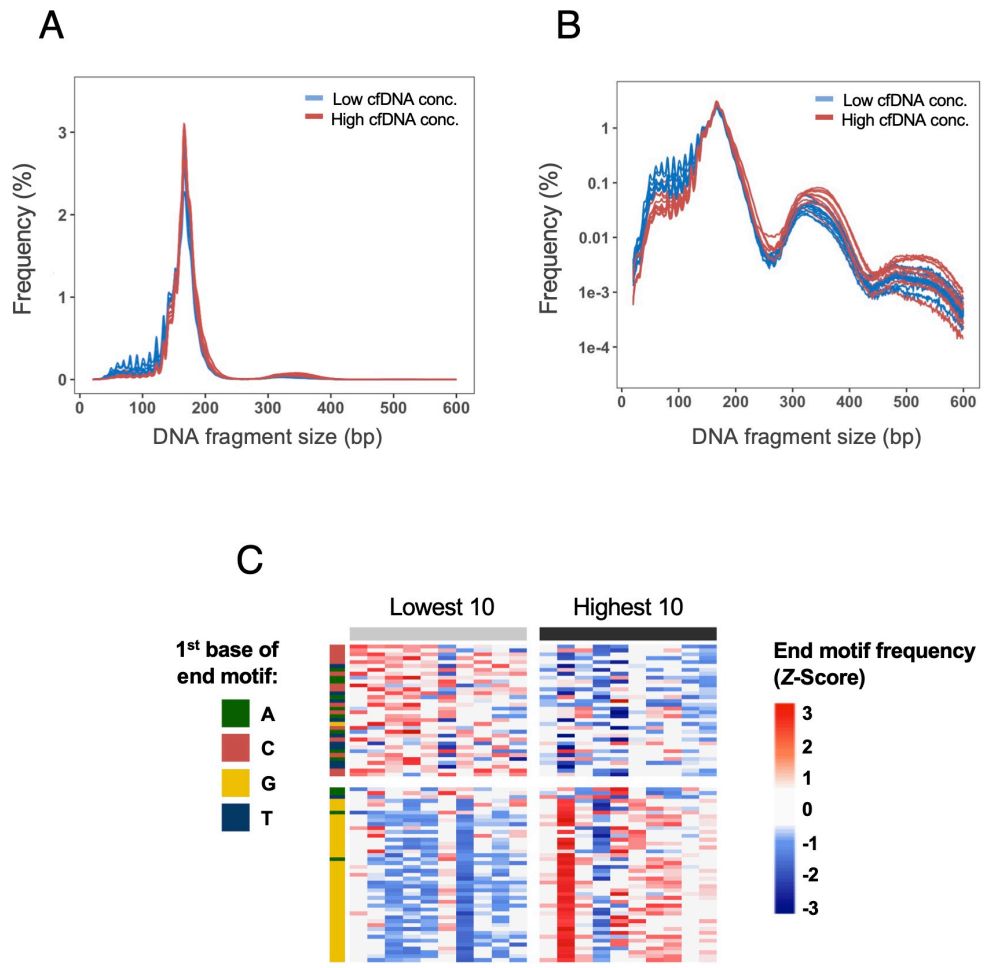

Figure S6. Fragmentomic characteristics of the highest and lowest 10 subjects of the EBV-negative cohort. (A) Size profiles of plasma cfDNA on a linear scale. (B) Size profiles of cfDNA on a logarithmic scale. (C) A heatmap analysis showing a frequency *z*-score calculated for 79 significantly correlated end motifs.

Supplemental Figure 7

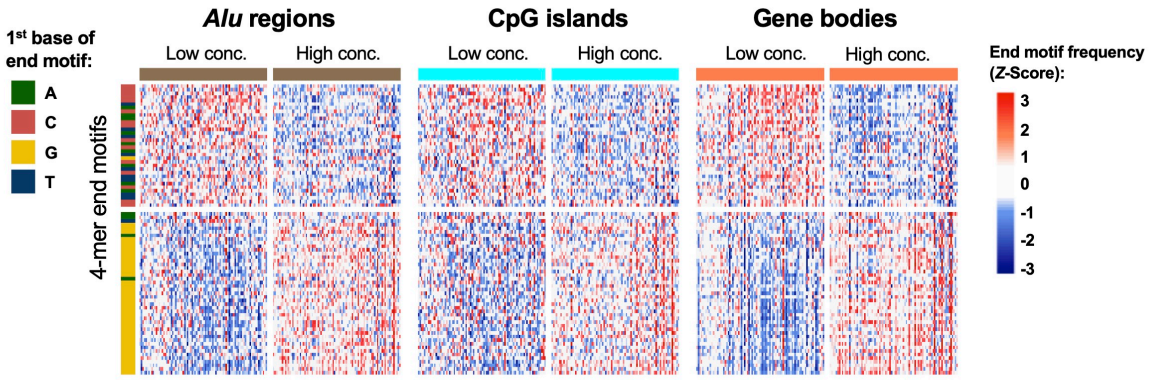

Figure S7. Heatmap analysis showing *z*-scores of informative end motif frequencies between the lowest and highest 10% of subjects across different genomic regions (*Alu* regions, CpG islands and gene bodies). A sequence context-based normalization method (O/E ratio – observed to expected end motif frequency) was used to minimize the potential biases in end motif analysis across different genomic regions.

Supplemental Figure 8

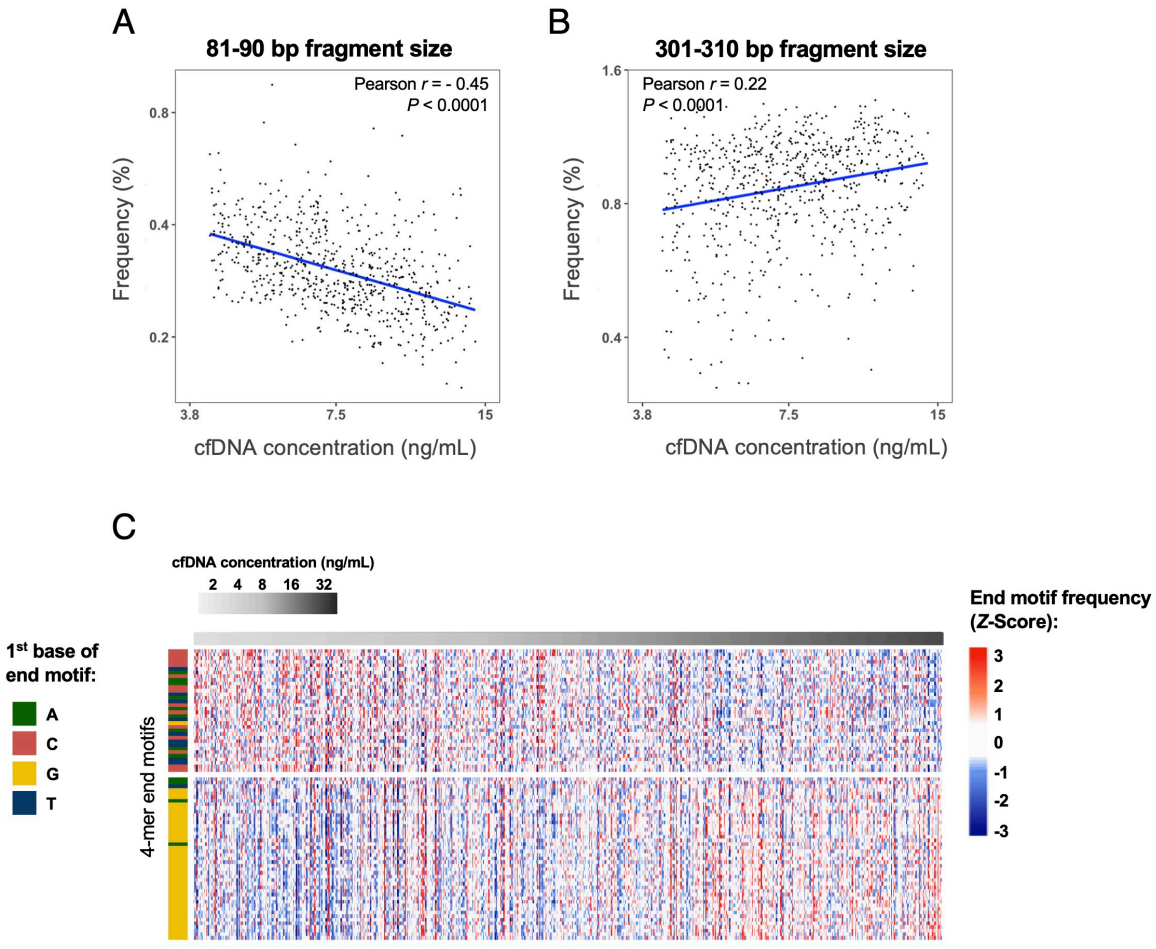

Figure S8. Analysis of fragmentomic features with the removal of subjects within the highest and lowest 10% of cfDNA concentration. Correlation between the frequency of DNA fragments between 10-bp windows and cfDNA concentration, for (A) 81-90 bp DNA fragment size, and (B) 301-310 bp DNA fragment size. (C) Heatmap analysis showing frequency *z*-scores of the selected 79 end motifs.

Supplemental Figure 9

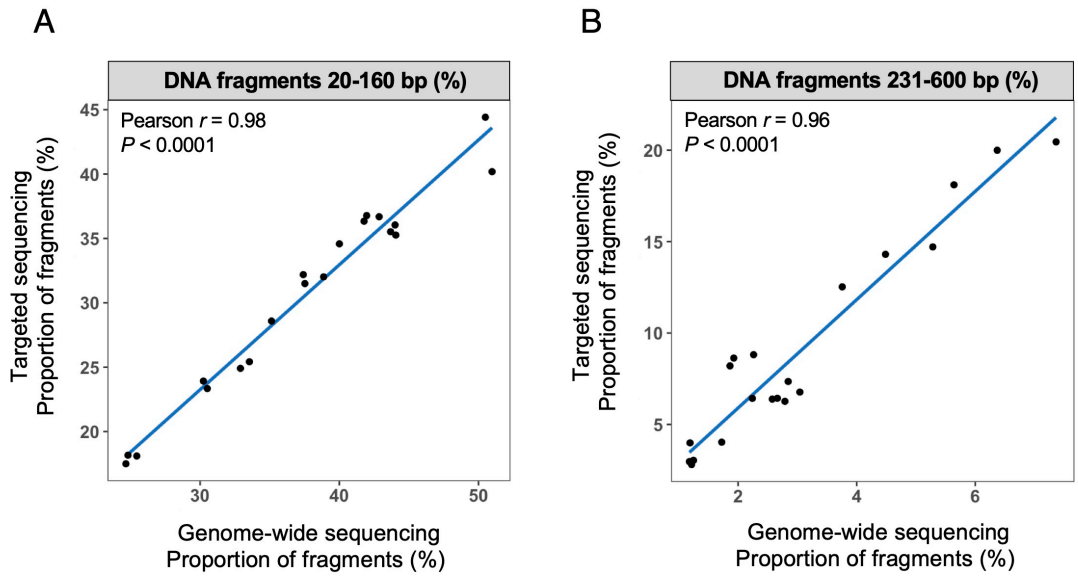

Figure S9. Correlation of size profile analysis between targeted sequencing and genome-wide sequencing dataset from paired-matched samples. Correlation of the proportion of DNA fragments within the size range of (A) 20-160 bp and (B) 231-600 bp, between the two sequencing methods.

Supplemental Figure 10

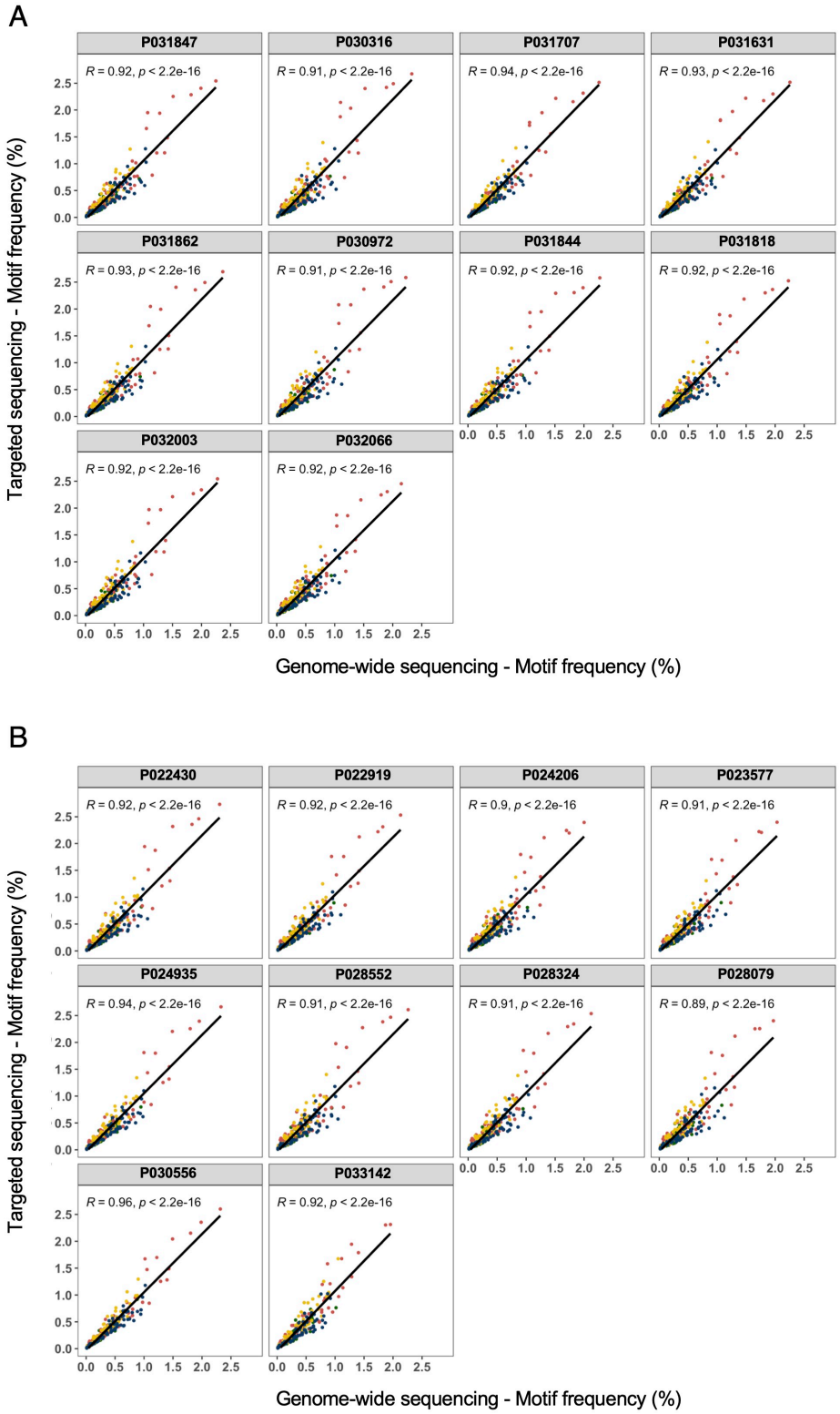

Figure S10. Correlation of end motif frequencies between targeted sequencing and genome-wide sequencing dataset from paired-matched samples. Correlation of end motif frequencies of all 256 end motifs between the two sequencing methods, for (A) the 10 subjects with the lowest cfDNA concentrations and (B) the 10 subjects with the highest cfDNA concentrations.

Supplemental Figure 11

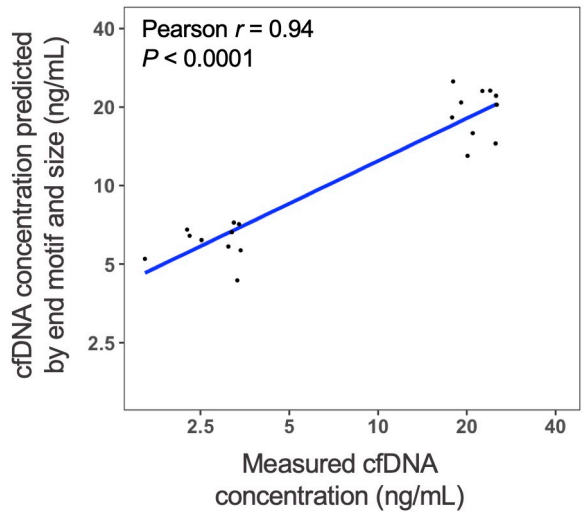

Figure S11. Correlation between measured and predicted cfDNA concentration of the sequenced samples from the EBV-negative individuals (20 subjects) using the SVR model trained using the size distribution and end motif profiles.

Supplemental Figure 12

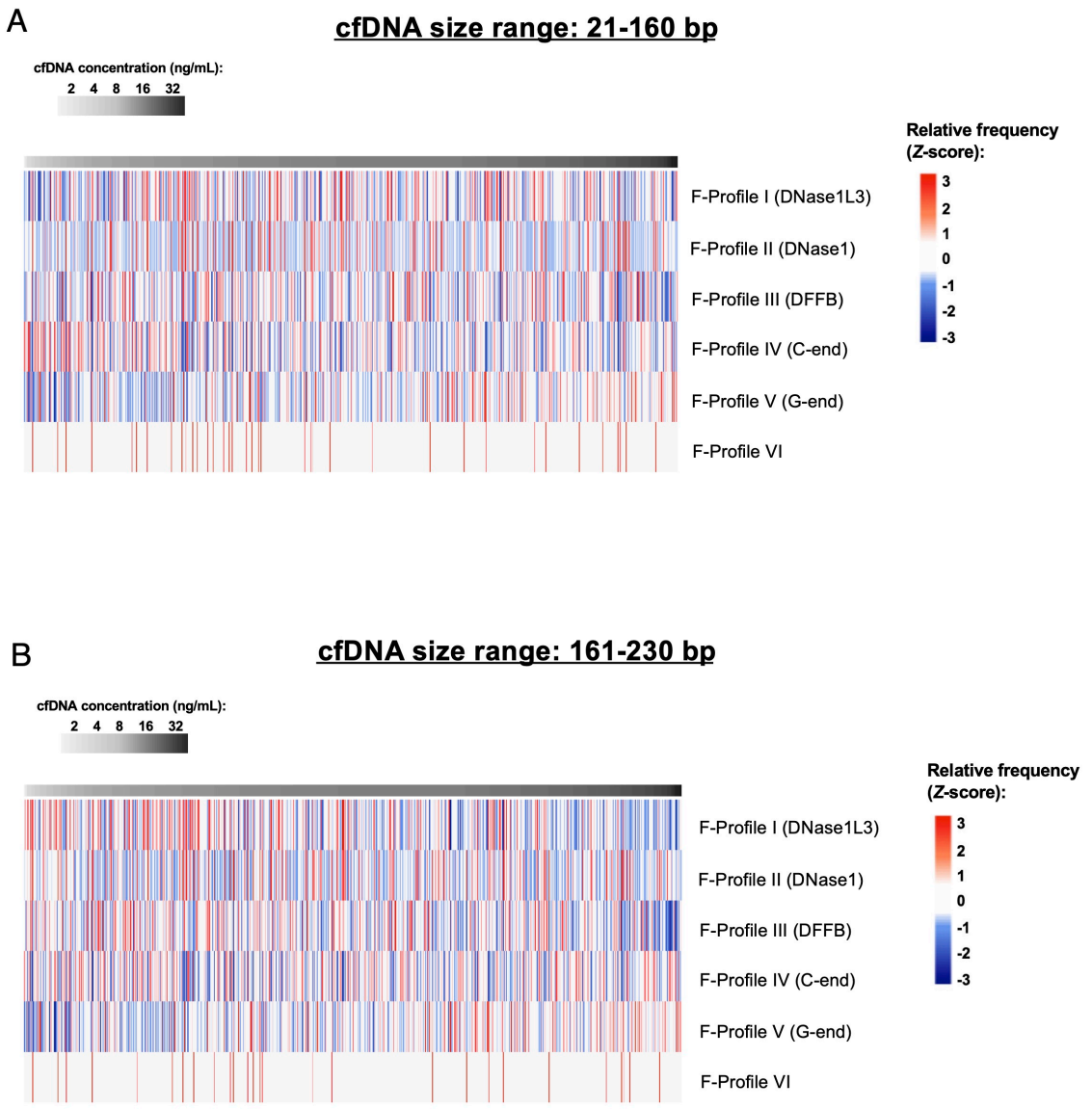

Figure S12. "Founder" end motif profile (F-profile) contributions of plasma DNA between subjects of different cfDNA concentrations. The F-profile contribution was normalized by *z*-score (calculated for each F-profile) across all 862 subjects, and was performed for the following size ranges of plasma DNA: (A) 21 – 160 bp and (B) 161 – 230 bp.

Supplemental Figure 13

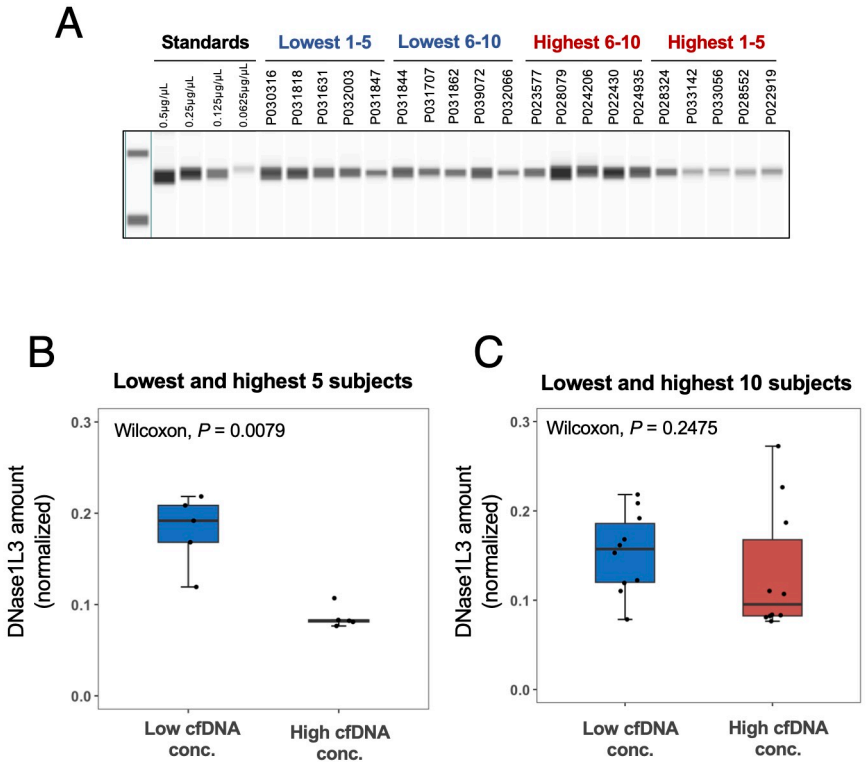

Figure S13. Quantification of the concentration of DNase1L3 in plasma of subjects with the lowest and highest cfDNA concentration. Automated Western blotting was performed for each individual's plasma sample using anti-DNase1L3 antibodies. The DNase1L3 band intensity area was quantified for each subject. (A) Representative immunoblot of the highest and lowest 10 subjects. Comparisons of the DNase1L3 concentration in plasma between (B) the highest and lowest five subjects and (C) the highest and lowest 10 subjects. Quantification of band intensity area was normalized to standards with different amounts of plasma proteins.

Supplemental Figure 14

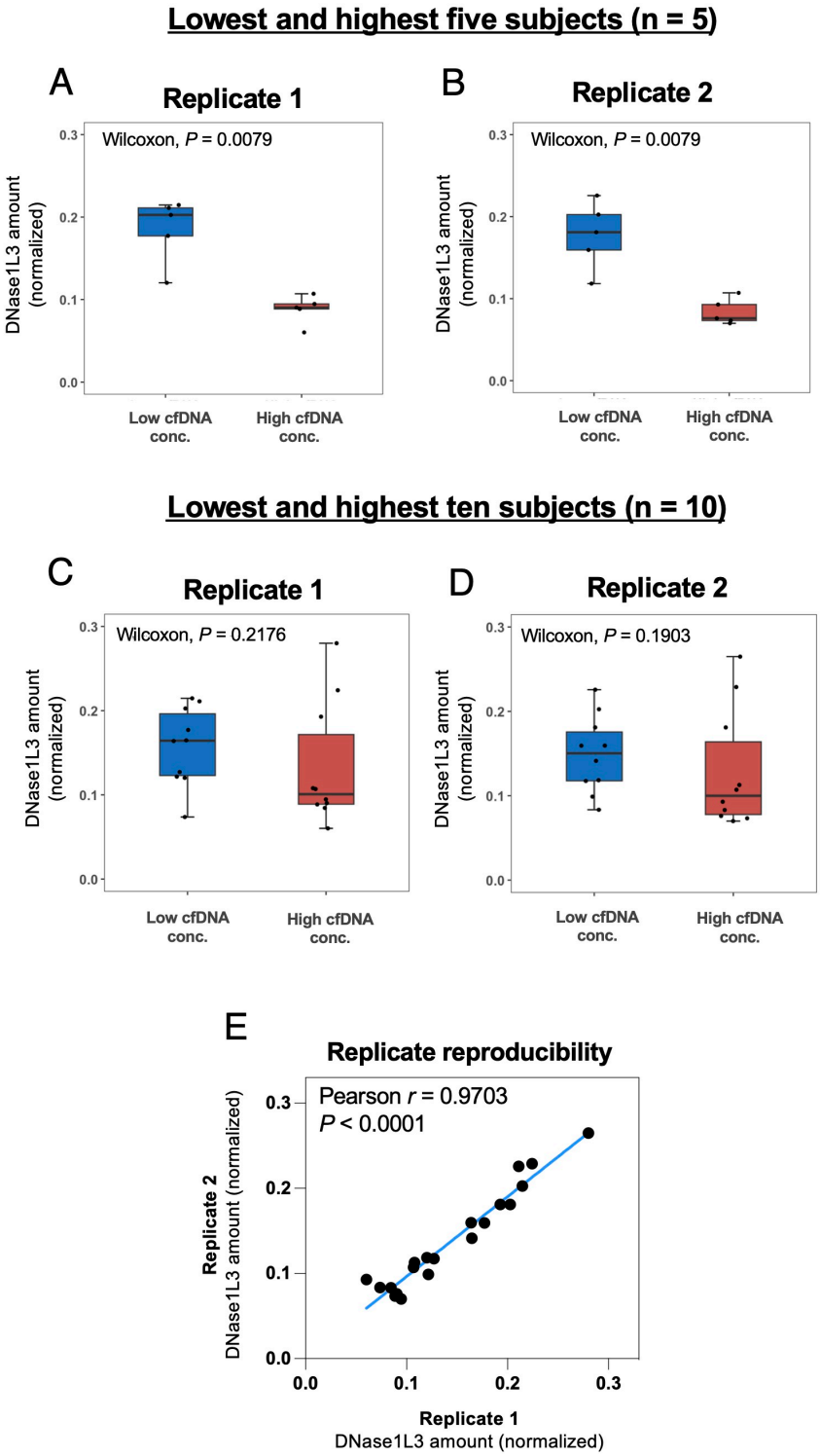

Figure S14. Immunoblotting for DNase1L3 plasma protein levels in subjects from both cohorts. Immunoblotting was performed in replicates. The normalized DNase1L3 concentrations for each replicate are shown for (A-B) the highest and lowest five subjects and (C-D) the highest and lowest 10 subjects. (E) Correlation between DNase1L3 concentration measured in two replicates.

Supplemental Figure 15

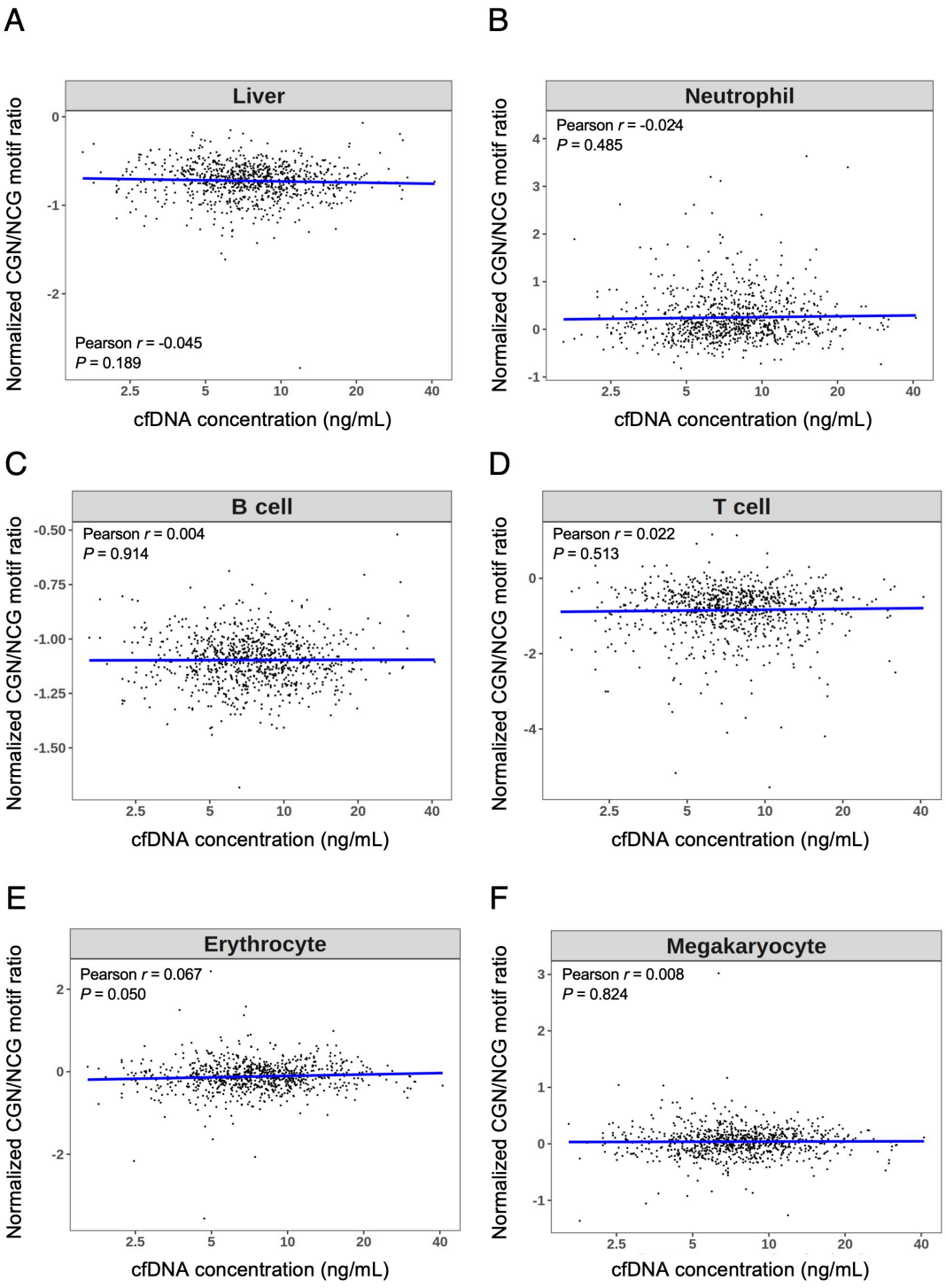

Figure S15. Tissue origins of cfDNA of subjects with different cfDNA concentrations deduced by Fragmentomics-Based Methylation Analysis (FRAGMA). The deduced tissue contribution, expressed as the normalized CGN/NCG motif ratio from selected tissue/cell-type specific CpG sites, was correlated with cfDNA concentration (n = 862 subjects) for (A) Liver, (B) Neutrophil, (C) B cell, (D) T cell, (E) Erythroblast and (F) Megakaryocyte.

Supplemental Figure 16

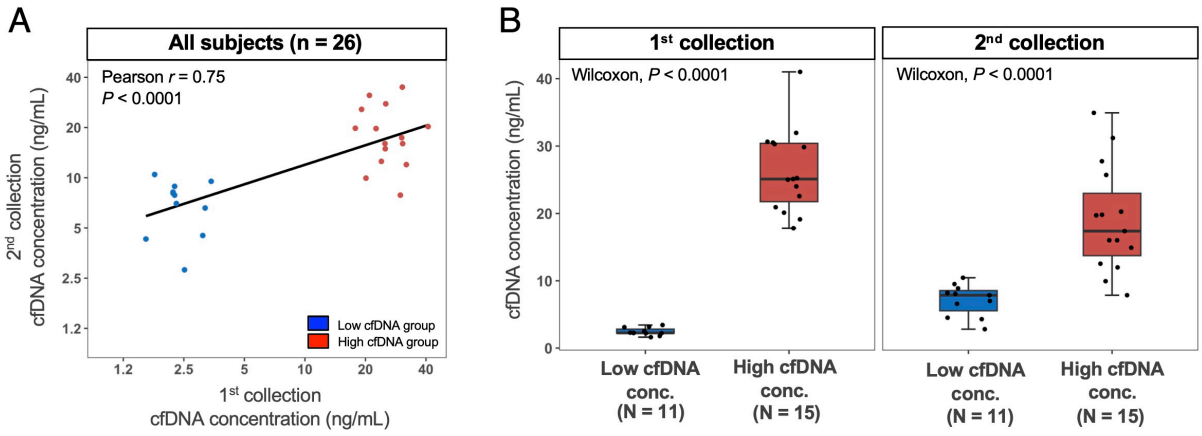

Figure S16. Comparison of cfDNA concentrations from the 1st collection (2017-2020 NPC cohort) and 2nd collection (Follow-up blood collection), with a median time interval of 76 months. (A) Correlation of cfDNA concentration from paired matched samples from the same individuals from both collection time points. (B) Box plots of cfDNA concentration from the low and high cfDNA groups between paired matched samples from both cohorts.

Supplemental Figure 17

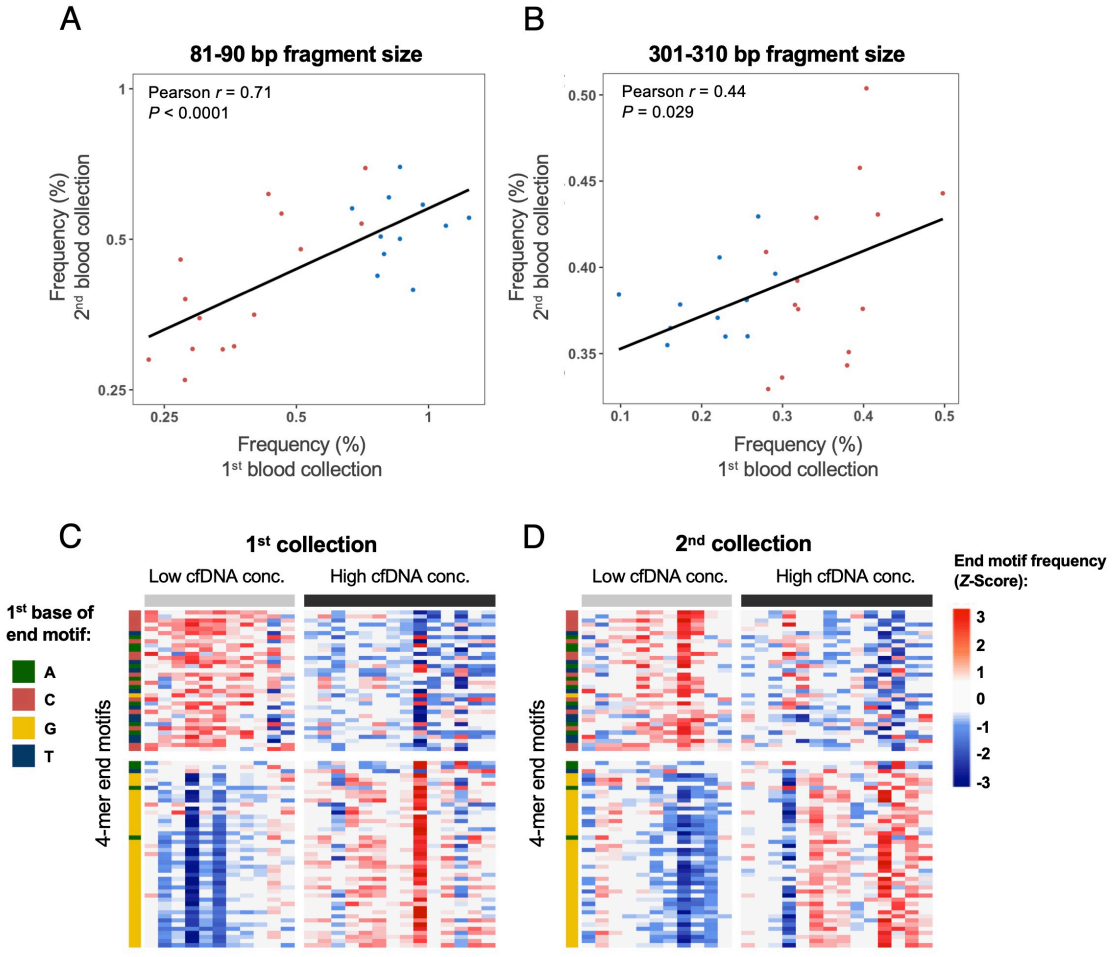

Figure S17. Fragmentomic features of plasma cfDNA from paired matched individuals from the 1st blood collection (NPC screening study) and 2nd blood collection (subsequent collection after a median of 76-month interval). Correlation between the proportion of DNA fragments between the two blood collections within the (A) 81-90 bp DNA fragment size and (B) 301-310 bp DNA fragment size. Heatmap frequency *z*-score of the selected 79 end motifs in (C) the 1st blood collection and (D) 2nd blood collection.

Supplemental Figure 18

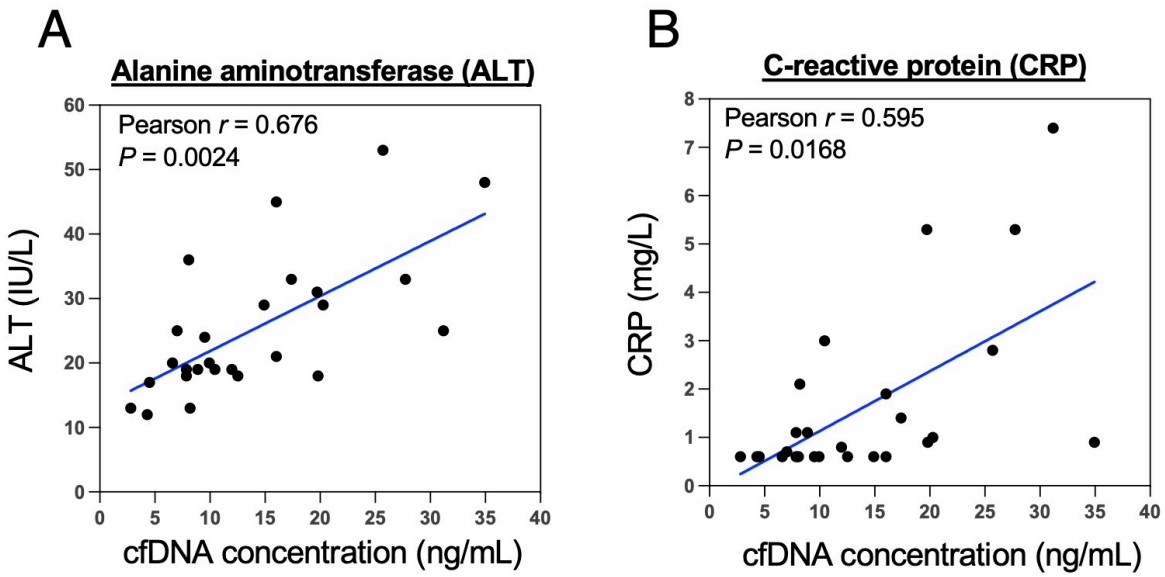

Figure S18. Correlation of the levels of (A) alanine aminotransferase (ALT) and (B) C-reactive proteins (CRP) to plasma cfDNA concentration in subjects of the follow-up collection (n = 26)

Supplemental Figure 19

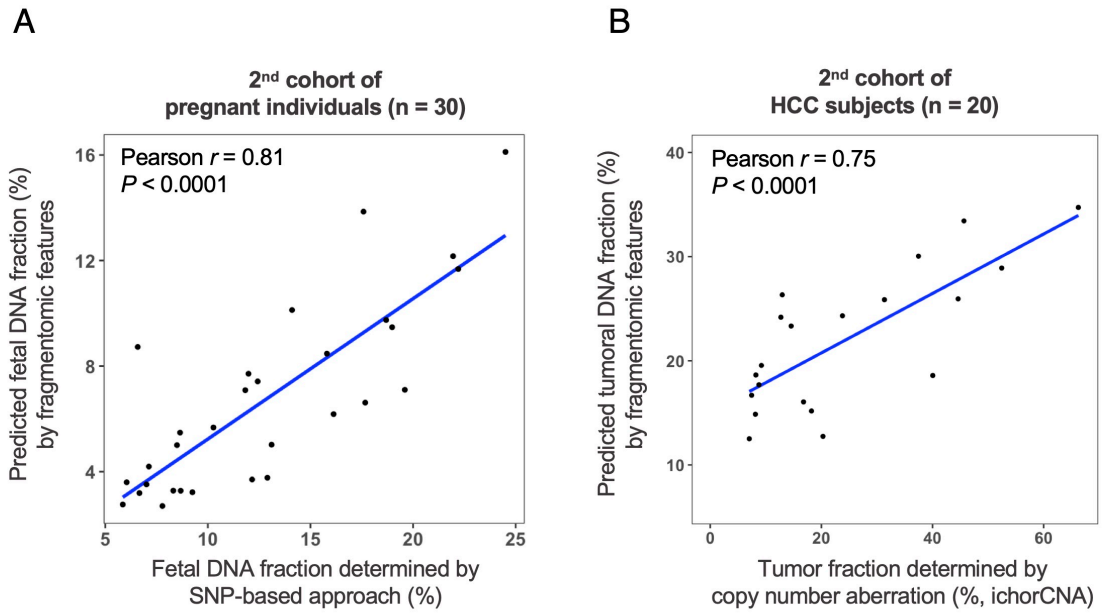

Figure S19. Validation of fragmentomic based deduction of fractional DNA concentration from specific tissue types. (A) Correlation between the fetal fraction predicted by fragmentomic features and SNP-based approach, using a test cohort of 30 pregnant subjects. (B) Correlation between tumor DNA fraction predicted by fragmentomic features and copy number aberration (ichorCNA), using a test cohort of 20 HCC patients.