

Tissue of Origin Characterization of Cell Free DNA in Seminal Plasma: Implications for New Liquid Biopsies.

Stephanie Huang^{a1}, James C. Hart^{a1}, James F Smith^{a,b}, Shellie Bench^a, Laura Rivas Yepes^a, Bailey Griscom^a and Kim M. Clark-Langone^{a*}.

^aFellow Health Inc. San Leandro, CA, 94577

^bUniversity of California, San Francisco, Dept. of Urology, San Francisco, CA 94143.

¹S.H. and J.C.H contributed equally to this work.

*Corresponding author: Kim Clark-Langone, Fellow Health, 1933 Davis Street, Suite 249, San Leandro, CA, 94577. Email: kclangone@meetfellow.com

Author Contributions: S.H. designed and executed wet lab activities, J.C.H developed the tissue deconvolution algorithm and performed bioinformatics analysis of the data. J.F.S. was the Principal Investigator on the Research Study that enabled the collection of semen samples. S.B. and L.R.Y. performed experiments and generated data. B.G. led the Research Study that enabled collection of the semen samples. K.M.C.L supervised the work and wrote the paper.

Competing Interest Statement: S.H., J.C.H., J.F.S., S.B., L.R.Y., B.G. and K.M.C.L. are employees of Fellow Health Inc. and have stock ownership. S.H., J.C.H. and K.M.C.L. are inventors on patent applications related to cfDNA isolation from seminal plasma.

Abstract

Liquid biopsies are becoming increasingly used for the detection and monitoring of disease states. While cell free DNA (cfDNA) in blood and urine have been well studied, much less is known about the composition of cfDNA in seminal fluid. We sought to characterize cfDNA in seminal fluid through tissue of origin studies using methylation analysis in men aged 21-60 yrs. We confirmed the observations of others that seminal fluid contains an abundance of cfDNA that is both nucleosomal and >1 kb. However, here we demonstrate for the first time that the high molecular weight (HMW) DNA harbors a lower sperm signal and higher somatic cell signal compared to the nucleosomal fraction. Prostate, granulocytes and kidney showed a mean predicted increased contribution of 6.2%, 4.9% and 2.9%, respectively in the HMW fraction. While sperm was the predominant signal in most men without vasectomies, the proportion of predicted prostate contribution reached as high as 26.5% in the HMW fraction. In other subjects without vasectomies, granulocyte cfDNA made up most of the signal. We also observed subject-specific cfDNA size distribution patterns that were reproducible over time, irrespective of abstinence times. These results suggest that seminal fluid is a rich source of cfDNA from various somatic cell types, and enriching for the HMW fraction would yield even higher sensitivity for somatic cfDNA detection. Considering these novel findings, it appears that seminal fluid may be able to serve as liquid biopsy for the detection and monitoring of prostate cancer, benign prostate hyperplasia, prostatitis and infertility.

Introduction

Seminal fluid is a complex mixture of secretions from the seminal gland, prostate, testes, epididymis, and bulbourethral glands. It not only serves as a medium for sperm transport but also provides nutrients and a protective environment for spermatozoa during their journey through the female reproductive tract. Seminal fluid contains a complex range of organic and inorganic

constituents, including proteins, enzymes, lipids, carbohydrates, and various ions, which play crucial roles in sperm motility and viability. Despite the emerging interest in liquid biopsies in cancer diagnostics and beyond, seminal fluid has not been extensively studied as a potential specimen type. Here we characterize the composition of cell-free DNA (cfDNA) in seminal fluid, highlighting its potential use in the liquid biopsy space.

Extracellular DNA (exDNA) or cfDNA is thought to be released into the extracellular environment through apoptosis, necrosis, NETosis, or active secretion by cells (1, 2). Given the abundance of sperm in semen, one could expect the large majority of cfDNA in seminal fluid to arise from apoptosis of sperm cells either undergoing spermatogenesis in close proximity to the seminiferous tubule lumen, or during storage of mature sperm in the epididymis. Such cfDNA could serve as biomarkers for fertility as reported by Chou et al. (3), whereby they demonstrated correlations between cfDNA size and important sperm parameters, and Di Pizio et al. (4) who found significantly higher cfDNA levels in patients with sperm abnormalities compared to controls. Other sources of cfDNA in semen could arise from the prostate, bladder, kidney, reproductive tract cells, or resident immune cells. One group proposed seminal plasma cfDNA as a potential biomarker for prostate cancer given their observation of significantly higher concentrations of cfDNA in prostate cancer compared to controls, and larger cfDNA fragment sizes in prostate cancer patients compared to those with benign prostate hyperplasia (BPH) and healthy controls (5–7). In their study on cfDNA levels and its association with sperm abnormalities, Di Pizio et al. (4) concluded that it may be of interest to study the cfDNA's origin and clearance and its methylation profile. One group showed LGALS3 cfDNA methylation status in seminal fluid to be able to discriminate between prostate cancer and benign prostate hyperplasia (8), but to date, no-one to our knowledge has fully characterized the cfDNA methylation signatures present in seminal fluid to understand the relative abundance of cfDNA from the various cell/tissue types that make up the male genitourinary system. Moreover, to our knowledge, no-one has characterized the cell/tissue of origin of the various cfDNA fragment lengths that have previously been correlated with prostate cancer (7). We sought to further investigate these findings by characterizing cfDNA yield, fragment size, and cell/tissue of origin from healthy volunteers. Once the composition of cfDNA in seminal fluid is better understood, we can begin to think about its potential use as a liquid biopsy for fertility, prostate cancer, testicular cancer, benign prostate hyperplasia and prostatitis.

Finally, liquid biopsy test manufacturers working with blood and urine have adopted the use of various preservatives to stabilize cells and nucleic acids so samples can be shipped without fear of losing sensitivity (9, 10). If blood is collected without the use of a suitable preservative, cell free nucleic acids will degrade, and white blood cells will die and lyse releasing high molecular weight DNA. This will result in a massive amount of background signal that would likely mask the signal of the already low analyte of interest. To date however, there are no described preservatives that have been shown to stabilize sperm and nucleic acids in seminal fluid. In this study we also sought to identify such a preservative.

Results

Stabilization of Sperm/cfDNA

To enable a semen sample to be produced at home as a liquid biopsy, it would be necessary to use a preservative that stabilizes sperm in addition to maintaining the integrity of the cfDNA during shipment to the laboratory. In this study we identified a preservative that could successfully achieve this. Figure 1 illustrates a comparison of the same sample with or without preservative. Figure 1A shows that on Day 0 the profiles look largely the same (both were processed within 2 hours of collection). At Day 3 in the oven (with oscillating temperatures from 10°C- 35°C), the sample with preservative maintains a very similar profile (Figure 1B), while the one without preservative shows

evidence of cell lysis and a massive increase in the amount of DNA recovered (Figure 1C). Note that when cfDNA was extracted from seminal fluid without preservative but within 2 hours of collection, we usually observed nucleosomal peaks (~150-200 bp) in addition to a peak >1 kb (Figure 1A). However, in some individuals, only the peak >1 kb was observed (see cfDNA size distribution profile section below).

cfDNA Yield:

Similar to other studies, we found the cfDNA yield in seminal plasma to be very high. The average total yield was 1.2 ug and there was a weak correlation between total yield and the volume of seminal plasma used in DNA extraction (Figure 2A). When normalized to the volume of seminal plasma, the average was 1003 ng per ml seminal plasma, which is at least an order of magnitude higher than that seen in blood plasma or urine. Values ranged from 260 ng - 3229 ng per ml seminal plasma (Figure 2B). Not surprising, the vasectomy subjects tended to have lower yields of cfDNA, however they were not clear outliers (Figures 2A, 2B), again demonstrating that not all cfDNA in seminal fluid is sperm derived.

cfDNA Size Distribution Profile by Electrophoresis

In this study we observed several different characteristic cfDNA size-distribution profiles that appear to be subject specific and may be associated with interesting underlying biology. In most cases the predominant nucleosomal peak appeared to be closer to 200 bp by TapeStation analysis (see Figure 3A) rather than the typical 166 bp characteristic of nucleosomal cfDNA, but upon sequencing, the mean peak insert size was 146 bp. In some cases, there was a single clean nucleosomal peak (Figure 3A), while others had clear multi-nucleosomal peaks (Figure 3B) along with the >1 kb peak. In the case of vasectomy participants, we observed no prominent distinct nucleosomal cfDNA peak, but only the peak >1 kb (Figure 3C). Given the absence of an abundant distinct peak of small (nucleosomal) cfDNA in men having undergone vasectomy, this suggests that most of the small cfDNA is coming from sperm. It is worth noting that in every participant, including those with vasectomies, there was a peak >1 kb. This was of great interest as it indicated that the cfDNA >1 kb was derived, at least in part, from somatic cells. To investigate whether the variability between subjects could also be associated with abstinence times, two subjects provided samples after various periods of abstinence. As can be seen in Figure 3D, abstinence time did not appear to be a factor in the overall profile, but did seem to impact the total amount of cfDNA. After 6 days of abstinence the yield was 2297 ng/ml plasma as compared to only 989 ng/ml plasma after 1 day of abstinence. However, the percent nucleosomal DNA, (i.e. between 50 and 700 bp as measured by TapeStation), remained at approximately 50%. Figure 3E represents the cfDNA profile from a different subject abstaining for either 1 day or 2 days. In this case, there was no difference in either the cfDNA profile or the yield. The images in Figures 3D and 3E also illustrate the often unique, but reproducible, cfDNA size-distribution profiles from the same subject.

Methylation Profile

The main objectives of this study were to determine the tissue of origin of the cfDNA present in seminal fluid and to understand whether the source of high molecular weight cfDNA was the same, or different, from that of the small (presumably nucleosomal) cfDNA. To do this, we performed size selection of the cfDNA, followed by enzymatic methylation analysis and tissue deconvolution. Figure 4A illustrates the tissue deconvolution results of the size selected small DNA (SSD) for 15 men without vasectomies (see Supplemental Figure 1 for results for men with vasectomies). In 13 of the 15 subjects, sperm was by far the most predominant signal contributing to over 70% of the total cfDNA signal. The other 2 subjects had a very high granulocyte signal and less than 20% of the signal was from sperm. The overall predicted prostate fraction for SSD samples was high compared to blood (11), with a median of 2.7% and a maximum of 20%. Interestingly, the tissue deconvolution results from the size selected high molecular weight (HMW) cfDNA showed an increased proportion of somatic cell signal and decreased proportion of sperm signal (Figure 4B). The proportion of predicted prostate contribution reached as high as 26.5%, with a median of 10.9%, which is of notable significance for the field of liquid biopsy. The HMW DNA showed a

marked reduction in sperm signal, with a median reduction of 28.7% compared to matched SSD samples (Figure 4C). Somatic tissues increased in relative predicted fraction in HMW vs SSD samples, with prostate (median +6.2%), granulocyte (median +4.9%), and kidney (median +2.9%) showing predicted increased contribution. These findings demonstrate that the HMW fraction of cfDNA from seminal fluid harbors a higher abundance of somatic cell cfDNA than the nucleosomal fraction.

Discussion

During apoptosis, DNA is fragmented at inter-nucleosomal linker sites and thus the resulting cfDNA is typically around 166 bp, which represents the size of the nucleosome DNA bound to the histone core (146 bp) and a (20 bp) linker connecting the nucleosomes (12). However, cfDNA fragments from tumors and within the fetal fraction of pregnant women tend to be shorter (13). In this study, the electropherograms indicated nucleosomal peaks closer to 200 bp rather than 166 bp, but upon sequencing SSD the mean peak insert size was 146 bp. We believe the insert size to be a more reliable marker of fragment size and believe the larger size seen on electropherogram to be some sort of artifact. Given that 85% of mature sperm DNA is associated with protamines as opposed to histones (14), it also begs the question of whether the sperm signal within the SSD fraction is coming from mature sperm or cells prior to spermiogenesis. Protamine associated DNA forms large toroidal structures of approximately 50,000 kb (15), therefore the majority of cfDNA from apoptotic mature sperm would not be expected to be in the SSD fraction. We also do not expect the HMW fraction to contain an abundance of protamine associated DNA since we deliberately performed a high speed centrifugation step to pellet the protamine toroidal structures. It was interesting to see various unique, but reproducible, cfDNA profiles between subjects. For example, some subjects had a single clean peak of mono-nucleosomal cfDNA, while others had very clear multi-nucleosomal peaks. The underlying biology of this warrants further investigation, but it could be related to the different amounts of DNases present in semen or the efficiency of cfDNA clearing, which could be of urological clinical significance. Our data also suggests that while increased abstinence times do not change the overall profile of cfDNA (or proportion of small to HWM cfDNA), the total amount of cfDNA increases as a function of abstinence time.

Previous studies have demonstrated that the yield of cfDNA obtained from seminal fluid is higher than that from other bodily fluids (16), and that fragment size and yield could potentially serve as a prostate cancer biomarker (5–7). Ponti, et al (5–7) postulated that the high molecular weight DNA seen in prostate cancer patients was derived from necrotic prostate cancer cells (6). However, the exact tissue source of seminal fluid cfDNA has not, until now, been determined. Here we demonstrate for the first time that in men aged 21-60 without vasectomy, most cfDNA usually arises from sperm (either mature or immature), but that there is an abundance of cfDNA that comes from various other cell types, including prostate, granulocytes and monocytes. In previous studies, researchers were careful to ensure no sperm lysis during sample collection and processing, and thus claimed that the DNA >1 kb was indeed “cell free” DNA as opposed to an artifact created during collection (16). Another way to rule this out is to look at cfDNA profiles from men who have undergone vasectomy, thus eliminating the chance that sperm is lysed during collection and processing. In this study we observed the absence of a discrete peak of small, nucleosomal cfDNA in participants that had undergone vasectomy, but still a large amount of cfDNA >1 kb. This essentially proves that the larger cfDNA (>1 kb) is coming from cells, at least in part, other than sperm. It is worth noting however, that the majority of this cohort were under 40 yrs of age, and as such, unlikely to have prostate cancer. It is therefore unlikely that in this cohort, the >1 kb cfDNA is coming from necrotic prostate cancer cells.

Methylation analysis of cfDNA has emerged as a powerful tool for determining the tissue of origin (TOO) of cfDNA fragments, which has significant implications for diagnostics and monitoring of

various diseases. To fully understand the TOO of cfDNA in seminal plasma, we performed a series of size selections to purify fragments <500 bp (small cfDNA, SSD) and >800 bp (HMW) and then performed methylation analysis on these two fractions. Our results demonstrate that in men having undergone a vasectomy there was no (significant) sperm signal in either fraction, and that in men not having undergone a vasectomy, the main source of cfDNA in both fractions was usually from sperm. In blood from healthy individuals, more than 90% of cfDNA is derived from white blood cells, with vascular endothelial cells and hepatocytes being the only detectable solid tissue source (11). In this study we demonstrate that prostate derived cfDNA can be highly abundant in healthy controls demonstrating seminal fluid to be a potentially more useful sample type for studying disorders of the prostate. The most striking observation was that the somatic cell signal was usually higher in the HMW fraction than in the SSD fraction. This result is counterintuitive to the world of blood liquid biopsies, whereby the presence of HMW DNA is seen negatively in that it indicates a higher amount of background noise. From a diagnostics perspective, our data suggests that one could improve the detection sensitivity for various pathological conditions (for example, prostate cancer, prostatitis or benign prostate hyperplasia) by enriching the HMW portion of cfDNA in seminal fluid.

One limitation of our study is that we did not have access to reference datasets for all tissue/cell types present in the male reproductive tract (for example, epididymis or Sertoli cells) in building the deconvolution algorithm. Consequently, the proportion estimates of the various tissue/cell types are likely overestimated (given that the algorithm sums the signal to 1). This likely explains why in vasectomy subjects, the level of sperm was not zero. Also, given almost 78% of the testes is thought to consist of spermatogonial stem cells (17), which have almost identical methylation patterns as mature sperm (18), we are likely capturing cfDNA from these cells within the sperm signal. Another limitation is that the cohort of subjects included in this study was limited to males 21-60 yrs of age and limited medical history was obtained. To understand the potential utility of this approach for detection of prostate associated diseases such as cancer, prostatitis or benign prostate hyperplasia, further studies will be required using the relevant patient populations.

Materials and Methods

Seminal Fluid Collection: Fresh semen samples were obtained from men aged 21-60 yrs as part of a study approved by an Institutional Review Board. Informed patient consent was obtained for the use of seminal fluid samples and vasectomy status was captured as part of the enrollment. Participants were instructed to collect seminal fluid through masturbation after a period of 2-6 days abstinence. In some cases, participants dropped the sample off at the Fellow laboratory within 2 hours of collection and seminal plasma was then prepared immediately, or preservative added to the seminal fluid sample and left for up to 3 days prior to seminal plasma preparation. In other cases, preservative was added immediately after collection (within 30 minutes) and the sample left for up to 3 days prior to seminal plasma preparation. As part of a stability study, samples were placed in a temperature-controlled chamber and the temperature cycled from 10°C to 35°C to mimic shipping conditions.

Seminal Plasma Preparation:

Seminal plasma was prepared by centrifuging the semen at 400 x g for 15 mins at room temperature to pellet sperm cells and somatic cells. The supernatant was transferred to a new tube and a second spin performed at 16,000 x g for 10 mins to pellet cellular debris and the protamine associated cfDNA from mature sperm. Seminal plasma was stored at -80°C prior to cfDNA extraction.

cfDNA Extraction and Size Selection

cfDNA was extracted from seminal plasma following Qiagen's Circulating Nucleic Acids Extraction kit instructions using 1-3 ml seminal plasma. cfDNA was then quantitated using Qubit

(Thermo Fisher) and the cfDNA profile obtained from the TapeStation 2200 (Agilent). High molecular weight cfDNA (>800 bp) and small molecular weight cfDNA (<500 bp) were separated as follows: a double size selection was performed using a 0.6x/1.8x ratio of SPRI reagent to sample volume. The high molecular weight DNA was eluted from the beads and underwent a further size selection using 0.5x ratio of SPRI reagent to sample volume. The high molecular weight DNA was subject to 250 seconds of sonication using a Covaris ultrasonicator, followed by a SPRI clean up using 1.5x ratio of SPRI reagent to sample volume. After size selection, there was sufficient small sized selected cfDNA (SSD) from 18 subjects (14 non vasectomy and 4 vasectomy) to take into library prep. After size selection, sonication and clean up for the high molecular weight (HMW) DNA, there was sufficient DNA from 15 subjects (11 non vasectomy and 4 vasectomy). Fourteen subjects (10 non vasectomy and 4 vasectomy) had a paired SSD and HMW sample.

Methylation Library Preparation and Sequencing

Methylation libraries were prepared using 7-40 ng of DNA for both SSD and HMW using New England Biolab's NEBNext® Enzymatic Methyl-Seq kit. Target enrichment was then performed using Twist's Human Methylome Panel. Next Generation Sequencing was performed on a NovaSeqX (either on a 10B or 25B flowcell).

Methylation Computational Processing

Following demultiplexing, reads were trimmed to remove adapters and low quality sequences using fastp (v0.23.4) (19) (extra options: '--trim_poly_g -f 1'). Reads were aligned to GRCh38 using bwameth (v0.2.7) (20) and bwa mem2 (v2.2.1) (21). Following alignment, reads were sorted and indexed using samtools (v1.3) (22), and read duplicates were marked using picardtools v3.1.0 ([Picard](#)). CpG and CHH methylation content was tabulated using MethylDackel (v0.6.1) ([MethylDackel](#)), only within regions covered by the Twist Human Methylome Panel, using the options '--CHH --nOT 3,0,0,3 --nOB 0,3,3,0' to exclude read ends with observed decreased methyl conversion in control materials. For each CpG, the *Beta* value (methylated reads)/(methylated reads + unmethylated reads) was tabulated.

Methylation Tissue Deconvolution

Reference tissue datasets were processed from previously published studies (Table S1) (11, 23–25). When raw sequencing reads were available, the same methylation processing pipeline described above was used. Otherwise, the processed CpG methylated/unmethylated counts supplied by the study were used to compute *Beta* values. CpGs were grouped into regions based on the Twist Human Methylome Panel, with the overall methylation *Beta* value calculated using the median of the CpG level *Beta* values. For each tissue, 25 marker regions were selected using a one-vs-all approach detailed in Loyfer et al. (11). In short, tissue specific hypomethylated markers were selected based on the difference between the 75th percentile within the given tissue vs the 2.5th percentile for the remaining samples in other tissues. The top 25 markers based on this score were selected, and the tissue signature methylation profile was calculated as the median region level *Beta* value for all reference samples from the given tissue, restricted to all markers selected across all reference tissues (Table S2).

A methyl region *Beta* value was tabulated for each seminal plasma sample, again using the median CpG level *Beta* value for the given sample. Using the above tissue signature methylation profiles, SciPy non-negative least squares implementation was used to find the optimal coefficients of each tissue contributing to the given sample's profile. All coefficients were normalized to sum to 1, ensuring estimations could be interpreted as proportions.

Acknowledgments

Funding was provided by Fellow Health Inc. We thank Mark Ferguson for his support in generating graphics, Daniel Colin and Awesta Froz for their contributions to cfDNA extraction,

and William Mathews, Meredith Halks Miller, Tara Maddala, James Wingrove and Alice Chen for their thoughtful discussions. We also thank the study participants without whom, this research would not have been possible.

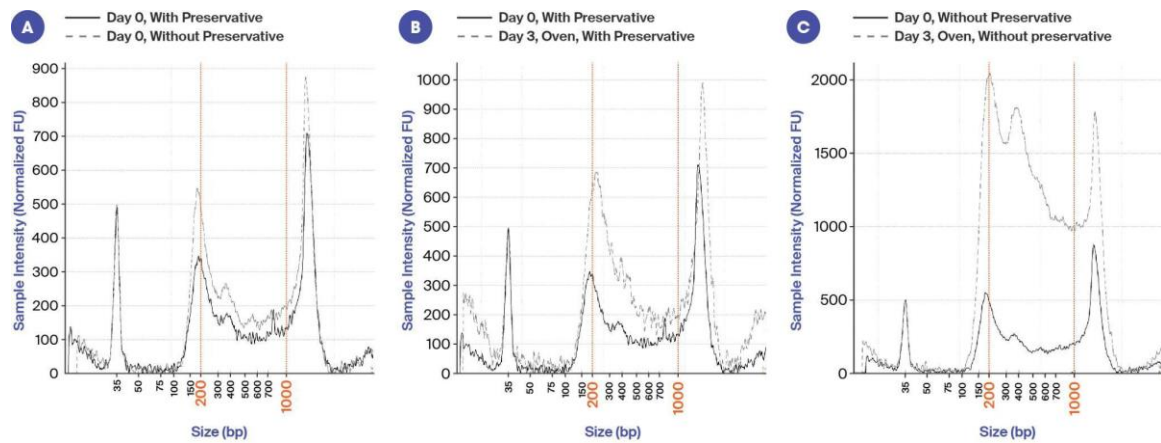
References

1. Z. Hu, H. Chen, Y. Long, P. Li, Y. Gu, The main sources of circulating cell-free DNA: Apoptosis, necrosis and active secretion. *Crit. Rev. Oncol. Hematol.* **157**, 103166 (2021).
2. E. Pisareva, *et al.*, Neutrophil extracellular traps have auto-catabolic activity and produce mononucleosome-associated circulating DNA. *Genome Med.* **14**, 135 (2022).
3. J. S. Chou, J. D. Jacobson, W. C. Patton, A. King, P. J. Chan, Modified isocratic capillary electrophoresis detection of cell-free DNA in semen. *J. Assist. Reprod. Genet.* **21**, 397–400 (2004).
4. P. Di Pizio, *et al.*, Seminal cell-free DNA and sperm characteristic's: An added biomarker for male infertility investigation. *Andrologia* **53**, e13822 (2021).
5. G. Ponti, *et al.*, Seminal cell free DNA concentration levels discriminate between prostate cancer and benign prostatic hyperplasia. *Anticancer Res.* **38**, 5121–5125 (2018).
6. G. Ponti, *et al.*, Seminal Cell-Free DNA Assessment as a Novel Prostate Cancer Biomarker. *Pathol. Oncol. Res.* **24**, 941–945 (2018).
7. G. Ponti, *et al.*, Quick assessment of cell-free DNA in seminal fluid and fragment size for early non-invasive prostate cancer diagnosis. *Clin. Chim. Acta* **497**, 76–80 (2019).
8. I. Abramovic, *et al.*, LGALS3 cfDNA methylation in seminal fluid as a novel prostate cancer biomarker outperforming PSA. *Prostate* **84**, 1128–1137 (2024).
9. I. Medina Diaz, *et al.*, Performance of Streck cfDNA Blood Collection Tubes for Liquid Biopsy Testing. *PLoS ONE* **11**, e0166354 (2016).
10. I. M. Diaz, *et al.*, Pre-Analytical Evaluation of Streck Cell-Free DNA Blood Collection Tubes for Liquid Profiling in Oncology. *Diagnostics (Basel)* **13** (2023).
11. N. Loyfer, *et al.*, A DNA methylation atlas of normal human cell types. *Nature* **613**, 355–364 (2023).
12. A. R. Thierry, S. El Messaoudi, P. B. Gahan, P. Anker, M. Stroun, Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev.* **35**, 347–376 (2016).
13. J. Shi, R. Zhang, J. Li, R. Zhang, Size profile of cell-free DNA: A beacon guiding the practice and innovation of clinical testing. *Theranostics* **10**, 4737–4748 (2020).
14. J. M. Gatewood, G. R. Cook, R. Balhorn, E. M. Bradbury, C. W. Schmid, Sequence-specific packaging of DNA in human sperm chromatin. *Science* **236**, 962–964 (1987).
15. R. Balhorn, The protamine family of sperm nuclear proteins. *Genome Biol.* **8**, 227 (2007).

16. H.-G. Li, S.-Y. Huang, H. Zhou, A.-H. Liao, C.-L. Xiong, Quick recovery and characterization of cell-free DNA in seminal plasma of normozoospermia and azoospermia: implications for non-invasive genetic utilities. *Asian J. Androl.* **11**, 703–709 (2009).
17. R. Islam, *et al.*, T cells in testicular germ cell tumors: new evidence of fundamental contributions by rare subsets. *Br. J. Cancer* **130**, 1893–1903 (2024).
18. J. Guo, *et al.*, Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development. *Cell Stem Cell* **21**, 533-546.e6 (2017).
19. S. Chen, Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**, e107 (2023).
20. B. S. Pedersen, K. Eyring, S. De, I. V. Yang, D. A. Schwartz, Fast and accurate alignment of long bisulfite-seq reads. *arXiv* (2014) <https://doi.org/10.48550/arxiv.1401.1129>.
21. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
22. P. Danecek, *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).
23. R. Schrott, *et al.*, Refraining from use diminishes cannabis-associated epigenetic changes in human sperm. *Environ. Epigenet.* **7**, dvab009 (2021).
24. X. Chen, *et al.*, Whole genome bisulfite sequencing of human spermatozoa reveals differentially methylated patterns from type 2 diabetic patients. *J. Diabetes Investig.* **11**, 856–864 (2020).
25. K. Skvortsova, *et al.*, DNA hypermethylation encroachment at cpg island borders in cancer is predisposed by H3K4 monomethylation patterns. *Cancer Cell* **35**, 297-314.e8 (2019).

Figures and Tables

Figure 1.



Electropherogram images illustrating the stabilization of sperm and cfDNA with preservation solution.

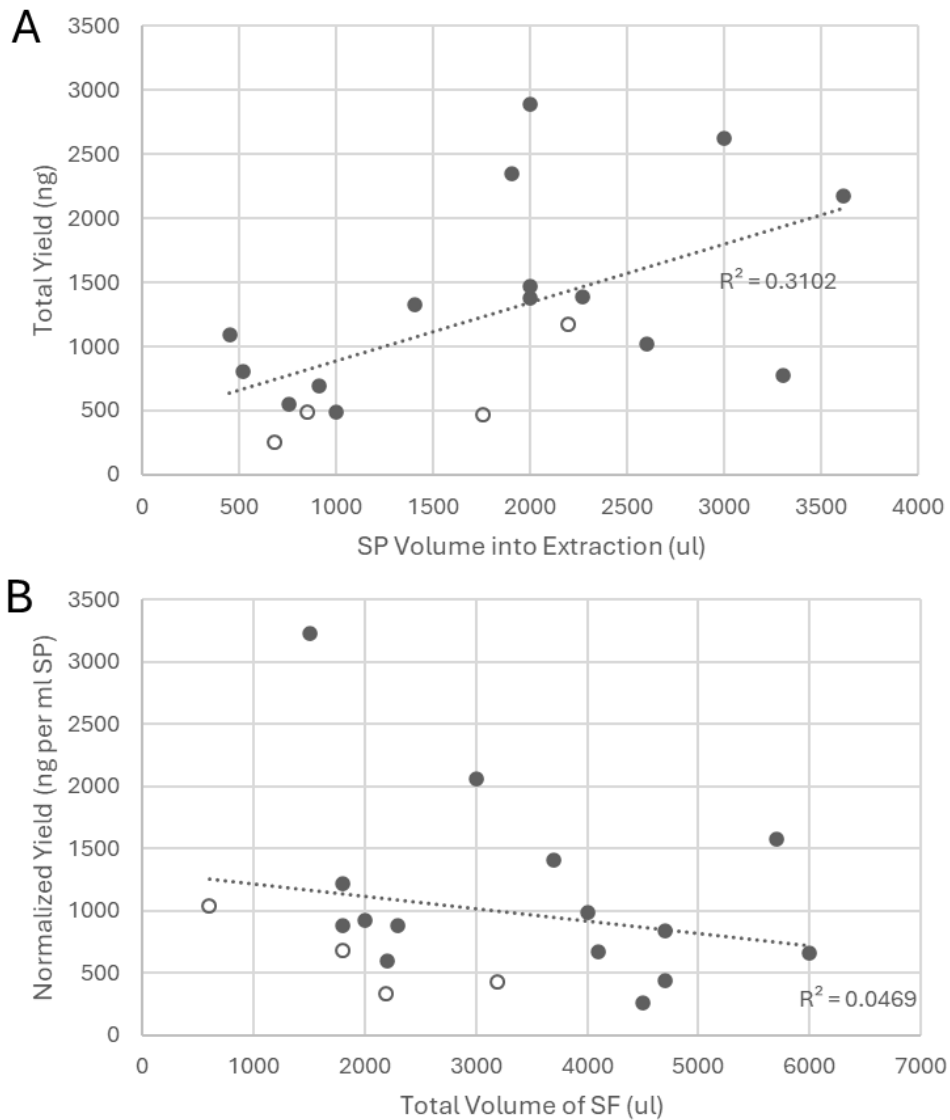
x-axis is size (bp) and y-axis is signal intensity.

A: Comparison of the same sample with and without preservative on Day 0 (seminal fluid processed within 2 hrs of collection).

B: Sample with preservative on Day 0 and Day 3 in the oven.

C: Sample without preservative on Day 0 and Day 3 in the oven.

Figure 2.

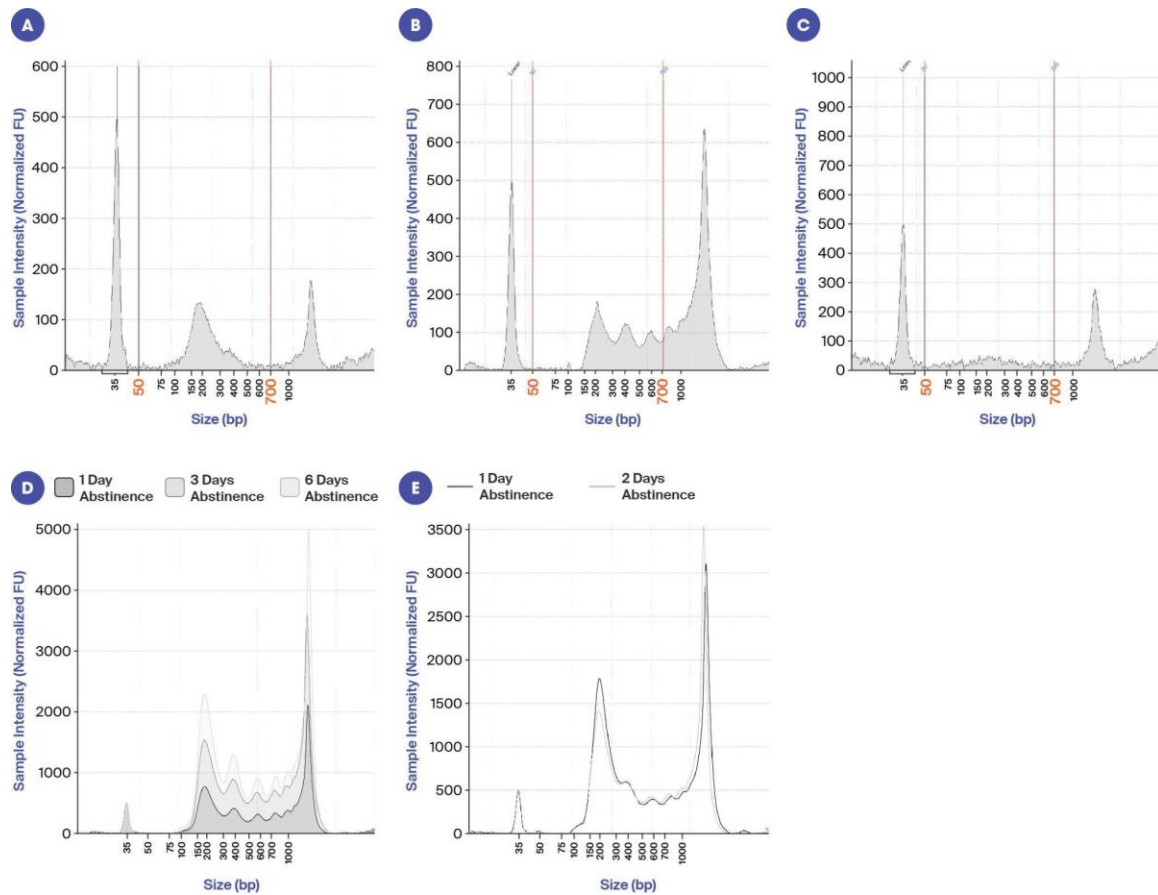


cfDNA yields correlated weakly with volume

A: Total yield (ng) by volume of seminal plasma (SP) into extraction. Solid circles are non-vasectomy subjects, open circles are vasectomy subjects.

B: Normalized yield (ng per ml SP) by volume of seminal fluid (SF). Solid circles are non-vasectomy subjects, open circles are vasectomy subjects.

Figure 3



Electropherogram images from various subjects illustrating the variability in size profiles and impact of abstinence time.

A: Example of a participant with two dominant cfDNA peaks, one nucleosomal cfDNA peak and one >1 kb.

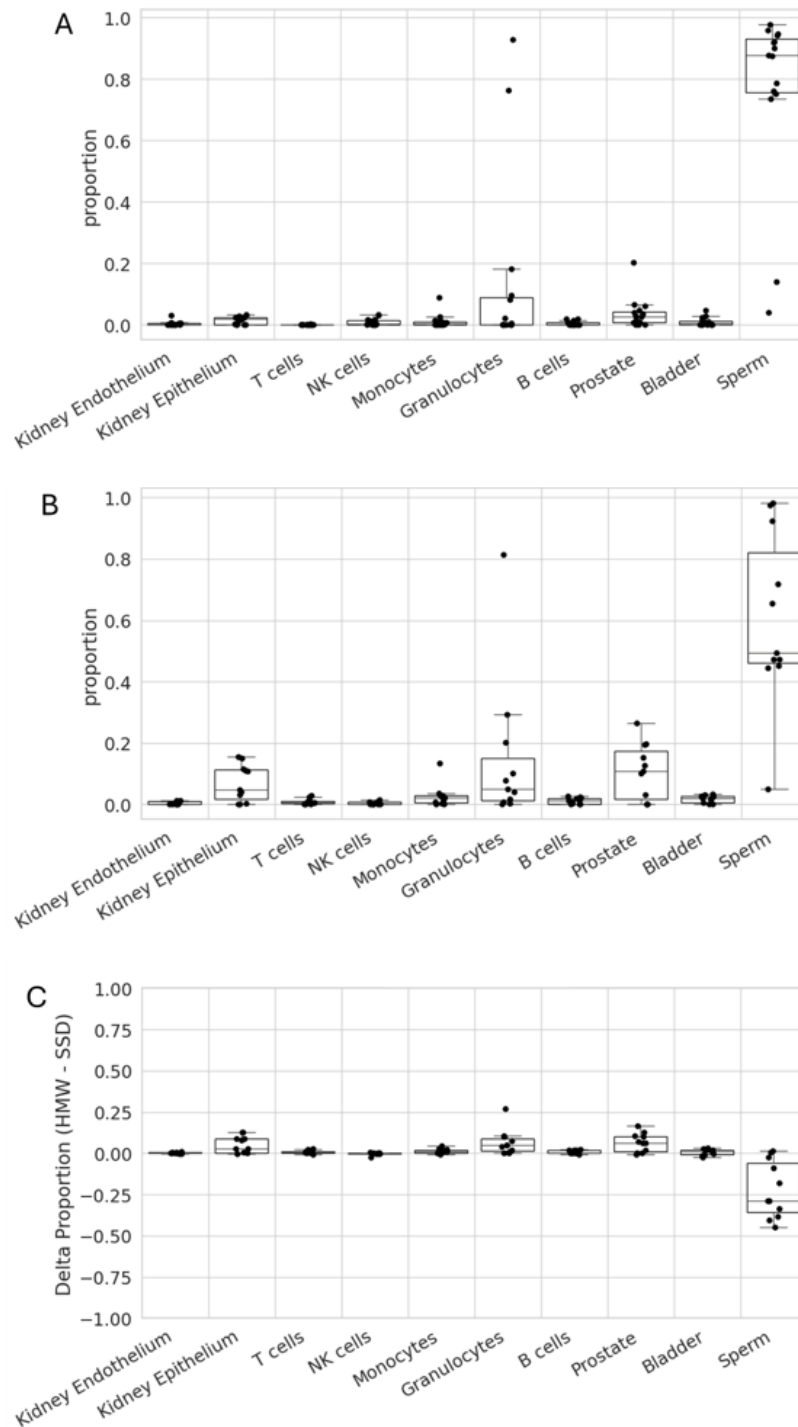
B: Example of a participant with what appears to be multi-nucleosomal cfDNA peaks and one peak >1 kb.

C: Example of a vasectomy participant just one peak >1 kb. While these profiles were typically seen in subjects having undergone vasectomy, they were also seen occasionally in subjects not having undergone vasectomy.

D: Three different seminal fluid cfDNA samples from the same participant after either 1, 3 or 6 days of abstinence.

E: Two different seminal fluid cfDNA samples from the same patient after either 1 or 2 days of abstinence.

Figure 4.



Tissue deconvolution results illustrate HMW has lower proportion of sperm signal and higher proportion of somatic cell signal

A: Tissue deconvolution results for SSD (n=15)

B: Tissue deconvolution results for HMW (n=11)

C: Difference in proportion of signal between HMW and SDD (n=11)