

<https://doi.org/10.1038/s42003-025-07567-1>

EV DNA from pancreatic cancer patient-derived cells harbors molecular, coding, non-coding signatures and mutational hotspots



Appolinaire A. Olou^{1,4} , Wesley A. Tom^{1,4}, Gary Krzyzanowski¹, Chao Jiang¹, Dinesh S. Chandel¹, Nirmalee Fernando¹, Adrian W. Draney², Joel Destino², Danny R. Welch³ & M. Rohan Fernando¹ 

DNA packaged into cancer cell-derived EV is not well appreciated. Here, we uncovered signatures of EV DNA secreted by pancreatic cancer cells. The cancer cells and non-cancer counterparts exhibit distinct low vs. high molecular weight (LMW vs. HMW) EV DNA fragments distribution, respectively. Genome sequencing and Single Nucleotide Variants analysis revealed that 95% of reads and 94% of SNVs map to noncoding regions of the genome. Given that ~1% of the human genome represents coding regions, the 5% mapping rate to coding regions suggests a non-random enrichment of certain coding regions and mutations. The LMW DNA fragments not only set cancer cells apart, but also harbor cancer specific enrichment of unique coding regions, the top nine being *FAM135B*, *COL22A1*, *TSNARE1*, *KCNK9*, *ZFAT*, *JRK*, *MROH5*, *GSDMD*, and *MIR3667HG*. Additionally, the cancer cells' LMW DNA fragments exhibit dense centromeric mapping more strikingly on chromosomes 3, 7, 9, 10, 11, 13, 17, and 20. Mutational profiling turned up close to 200 mutations specific for the cancer cells. Altogether, our analyses suggest that centromeric regions might hold clues to EV DNA content from pancreatic cancer, the molecular, mutational signatures thereof, and rationalizes the need for a new approach to DNA biomarker research.

Pancreatic cancer is a deadly disease of still unknown etiology. The mechanisms of pancreatic cancer progression, including metastasis, are not fully known. Emerging evidence hints at exosome cargo as a driving factor of pancreatic cancer metastasis and a prognostic indicator for metastasis^{1,2}. Exosomes are a class of extracellular vesicles (EVs) that were initially dismissed as just a way of cellular waste disposal³. However, that idea was quickly sidelined by studies showing that EVs are reservoirs of biomolecules with pleiotropic effects on cellular functions in both normal physiology and disease state. EVs such as exosomes have even been touted to harbor potential diagnostic markers for cancer⁴ but to reach such a diagnostic milestone, an appreciation of EV molecular content is imperative. One component of EV cargo is double-stranded (dsDNA), an intriguing molecule fit for a non-invasive disease diagnosis. DNA is a relatively stable biomolecule, and it has been known for some time now that most cell-free

DNA is enclosed in exosomes⁵. The stability of DNA is desirable not just for ensuring the preservation and accurate transmission of genetic information, but also for a potential diagnostic marker for diseases. EVs from tumor cells carry genomic dsDNA, with segments spanning the entire genome⁶. Moreover, DNA molecules have been detected in exosomes from a spectrum of cell/system types^{6,7}. These and similar several other findings indicated the potential for leveraging exosome-enriched EV and DNA composition to differentiate, at the basic level, cancer vs. non-cancer cells.

Previous studies on EV DNA were not all carried out in the same cell/system types. Likewise, the studies' approaches were often different, for example, targeted mutations detection versus a whole genome sequencing approach, or exosomes vs. other classes of EV. Some of the data seem to be conflicting as well. For example, while it has been known for some time now that exosomes carry DNA^{5,6,8,9} a recent study suggests that exosomes might

¹Molecular Diagnostic Research Laboratory, Center for Sensory Neuroscience, Boys Town National Research Hospital, Omaha, NE, USA. ²Department of Chemistry, Creighton University, Omaha, NE, USA. ³Department of Cancer Biology, Kansas University Medical Center, and the University of Kansas Comprehensive Cancer Center, Kansas City, KS, USA. ⁴These authors contributed equally: Appolinaire A. Olou, Wesley A. Tom. ✉e-mail: appolinaire.olou@boystown.org; m.rohan.fernando@boystown.org

not contain DNA¹⁰. Though, that study did not provide evidence for where DNA is in EVs if it were not in exosomes. Moreover, a recent review article¹¹ pointed to some of the limitations of that study. The variation in published data above could also be a reflection of the complexity of different cell/system models, whereby cell type-to-cell type distinction might result in significant differences in the pool of molecules packaged into the respective exosomes. For cells that do package DNA in exosomes, the genomic origination of these DNA molecules has not been well established. Likewise, cell-specific exosomal DNA signatures have yet to be characterized but are critical to differentiate different cell types. Here, we profile EVs DNA using high depth next-generation sequencing (NGS) of dsDNA from pancreatic cancer patient-derived cells. Our analyses suggest that DNA packaged in EVs is not random.

Results

Characterization of the vesicles by Nanoparticles Tracking Analysis, transmission electron microscopy, and antibody array

We isolated EVs from pancreatic cancer patient-derived cells and non-cancer counterparts, all grown in cell culture media containing exosome-depleted FBS. We first assessed the size distribution of the EVs using NanoSight (A.K.A Nanoparticles Tracking Analysis) (Fig. 1A and Supplementary Fig. 1). Next, we used transmission electron microscopy (TEM) to

visualize the morphology of the nanostructure which revealed a cup-shaped morphology (Fig. 1B). Lastly, we performed a molecular characterization of the vesicles by carrying out an antibody array that simultaneously evaluates the expression of 8 EV protein markers and contaminant. These protein markers are cluster of differentiation 63 (CD63)^{5,12}, cluster of differentiation 81 (CD81)¹², tumor susceptibility gene 101 (TSG101)^{12,13}, ALG-2-interacting Protein X (ALIX)¹², intracellular adhesion molecule (ICAM)¹², Annexin5 (ANXA5)¹², epithelial cell adhesion molecule (EpCAM)¹², and Flotilin1 (FLOT1)¹². The antibody array included two positive controls for HRP detection, and a negative control for cellular contamination of the EVs. Importantly, the array confirms the expression of EV protein markers (CD63, CD81, ALIX, FLOT1, ICAM1, EpCam, ANXA5, and TSG101) to varying degrees (Fig. 1C).

Distinct DNA fragments packaged in EVs released from pancreatic cancer and non-cancer cells

Next, we extracted double-stranded (ds) DNA molecules from the vesicles and first validated the molecules using the Qubit Fluorometer which employs a dsDNA sensitive dye for dsDNA detection and measurement (Fig. 2A; Supplementary Fig. 2A). DNA fragments size was assessed using the Agilent 4200 TapeStation. A prominent DNA fragment distribution in the pancreatic ductal epithelial non-cancer cells

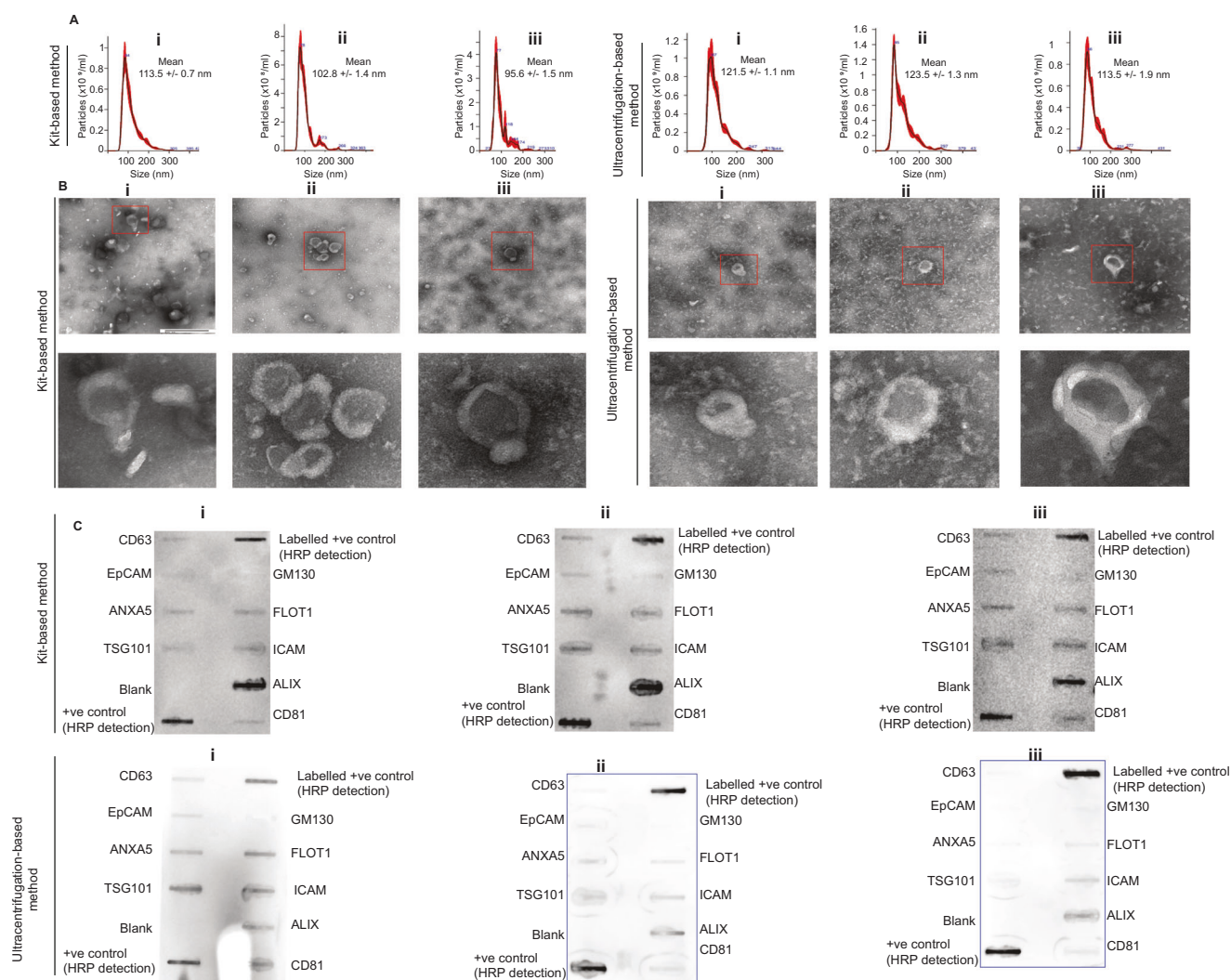


Fig. 1 | Characterization of the vesicles by Nanoparticles Tracking Analysis, transmission electron microscopy, and antibody array. A EVs from H6C7 (i), Panc1 (ii), and HPAF-II (iii) cells isolated by the indicated methods were analyzed for size distribution by NanoSight (NTA) or B visualized with Transmission electron

microscopy (TEM); scale bar is 500 nm. The bottom panels represent the zoom-in on the boxed area in the image. C Exosomes antibody array was carried out on the EVs from the cells above isolated by the indicated methods, showing expression of various EV protein markers.

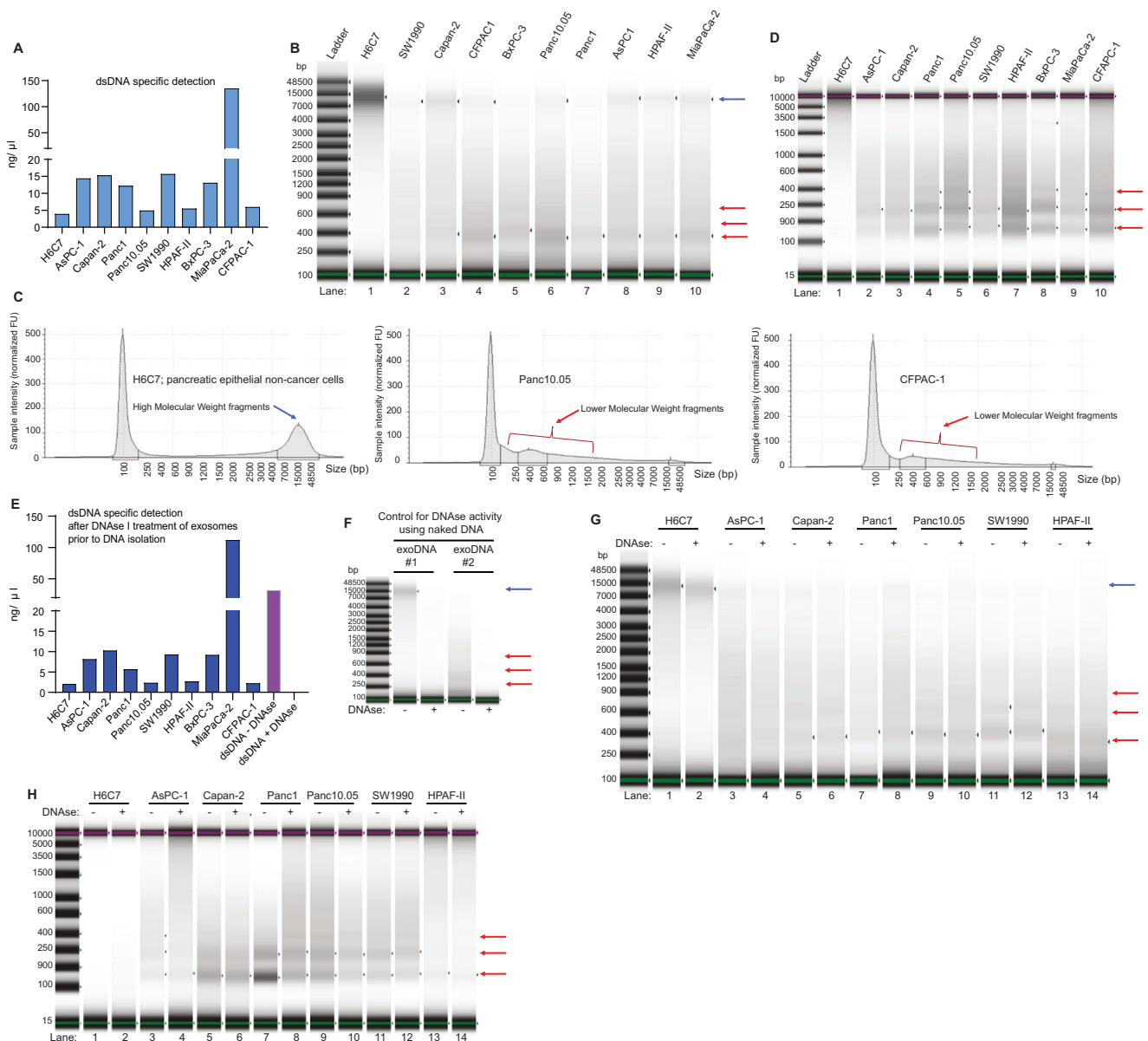


Fig. 2 | Distinct DNA fragments packaged in EVs released from pancreatic cancer and non-cancer cells. **A** Following DNA extraction from the EVs prepared by the kit-based method, a dsDNA-specific detection method was used to validate the presence of DNA molecules in the samples; representative of 3 experiments. Equal amount of dsDNA, 2 ng, was loaded on the Agilent 4200 TapeStation system using the genomic (**B**) or the D5000 (**D**) screen tape protocol; experiments were done at least three times. **C** The electropherograms showing distinct peaks for the control

and cancer cells; 3 representatives from the 10 cell lines. **E** EV preparations were treated with DNase I before DNA was extracted. A dsDNA-specific detection method was used to confirm the presence of DNA molecules as in (**A**). **F–H** The DNA fragments in samples from (**E**) were analyzed as described in (**B**, **D**). The activity of the DNase I was confirmed by analyzing EV DNA from H6C7 (#1) or MiaPaCa2 (#2) digested with the DNase I.

(H6C7) appeared at about 15,000 base-pairs (bp) in length, suggesting that most of the EV DNA in those cells is high molecular weight (HMW) (Fig. 2B; Supplementary Fig. 2B). Importantly, such patterns of DNA fragments were mostly absent in DNA samples from pancreatic cancer cell-derived vesicles (Fig. 2B; Supplementary Fig. 2B). Instead, these cells produced consistent DNA banding patterns at 200, 350, and 600 bp with varied degrees of intensity, and low levels of HMW DNA representation compared to the non-cancer cell counterpart (Fig. 2B; Supplementary Fig. 2B). The electropherograms (Fig. 2C; Supplementary Fig. 2C) summarize these observations and show two salient patterns: HMW DNA fragment mostly from the non-cancer cells (H6C7) but low molecular weight (LMW) DNA fragment from the cancer cells. To better visualize and evince the LMW DNA fragments, we used a different bioanalytical protocol that focuses on smaller fragment sizes and

provides better band resolution. This approach showed more evident bands that represent smaller DNA fragments from the cancer cell-derived vesicles with varying degrees, but not from the non-cancer cell-derived vesicles (Fig. 2D; Supplementary Fig. 2D). This DNA banding pattern was prominent in the EVs but absent in DNA isolated directly from parental cells secreting the EVs (Supplementary Fig. 2E, F). Lastly, we re-evaluated the results above after first treating the vesicular preparation with DNase I before proceeding to DNA extraction. Again, as before, here too dsDNA molecules were detected by the Qubit Fluorometer (Fig. 2E). Control naked DNA molecules treated with the DNase were degraded (Fig. 2E, F) suggesting that the other DNA molecules being analyzed (Fig. 2E) are protected from DNase I treatment. Importantly, the fragment analysis showed a pattern (Fig. 2G, H) like that described previously (Fig. 2B, D). To further assess that result,

we treated the EVs with a membrane-solubilizing detergent to lyse the EVs (Supplementary Fig. 2G). DNase treatment of the pre-lysed EV preparations before DNA extraction led to a loss of DNA recovery (Supplementary Fig. 2H).

Whole genome sequencing analysis demonstrated a large proportion of EV DNA molecules from non-coding regions of the genome

To further investigate the DNA banding features, the EV DNA was fractionated using magnetic bead size selection into HMW and LMW DNA sub-populations (Fig. 3A). The HMW DNA was fractionated using sonication to ~500 bp insert sizes, while shearing was not necessary for the LMW DNA. Each cell line had two libraries, an HMW and an LMW, each receiving its own barcode for sequencing. Samples were sequenced on the Illumina NextSeq (150 bp paired-end sequencing) to an average depth of 48,152,253 reads per sample and 98.1% reads mapping to the GRCh38 human genome build (Supplementary Table 1); only ~0.01% of the reads map to the mitochondrial genome (Supplementary Table 2). The consensus from the human genome project is that only about 1–2% of the human genome is coding genes, and ~99% of the genome is non-coding nucleic acids. We reasoned that if the packaging of DNA in EVs is random and uniformly distributed across the genome as previous studies report^{6,14}, then it would be expected that similar ratios of coding to non-coding read mapping would be observed in EV DNA. Thus, the proportion of reads mapping to coding and non-coding regions were determined from the DNA fragments. Irrespective of the molecular weight fraction of the DNA sample (HMW or LMW), roughly 5% of reads mapped to coding regions (Fig. 3B–D) whereas about 95% of the DNA fragments fall into non-coding regions of the genome (Fig. 3B–D). This suggests that a large proportion of the DNA molecules are composed of mostly non-coding regions of the genome with a small fraction containing coding regions, but the frequency of reads mapping to coding regions (~5%) is slightly elevated as compared to the anticipated 1–2% coding region mapping that would be expected if the DNA was packaged uniformly across the entire genome.

The low molecular weight DNA fragments distinguish cancer cell from non-cancer cell counterpart

To address how similar the read profiles of each fraction were to one another, Pearson correlation analysis of the mapped read features was performed for all pairwise comparisons in LMW DNA fractions as well as HMW DNA from all cell lines. When the LMW fragments from the non-cancer cells (H6C7) are compared to the LMW fragments from the cancer cells, the correlation coefficient is farther from 1 (Fig. 4A) than when the HMW fragments from the non-cancer cells are compared to the HMW fragments from the cancer cells in which case the coefficient is closer to 1 (Fig. 4B). That observation indicates that the LMW DNA fragments from the cancer cells' EV DNA samples are distinct from the non-cancer counterparts' whereas the HMW DNA fragments are more similar among the non-cancer and cancer cells. To further assess that inference, we evaluated the principal components of the mapped read profiles for each DNA fraction (LMW and HMW). Again, most of the cancer cells' LMW EV DNA samples cluster together and further away from the non-cancer cells' samples (Fig. 4C) whereas that clustering pattern is much less evident for the HMW fragments (Fig. 4D). The analysis indicates that the LMW DNA fragments might have more clues as to the distinction between cancer cell and the non-cancer cell counterpart.

Unique coding DNA signatures for the LMW fragments

DNA mapping signatures among the LMW DNA fragments could further set the cancer cells apart from their non-cancer cell counterpart. To investigate this, the LMW reads which mapped to coding regions were extracted, the read counts per gene normalized to approximate abundance, and differential abundance analysis was performed to determine which genes had differential read mapping. In total, 44 coding regions had

significantly upregulated read abundances in the cancer cells as compared to the non-cancer cells (Table 1 and Fig. 5).

The top 10 genes which had elevated read mapping were Family with sequence similarity 135 member B (*FAM135B*), Collagen type XXII alpha 1 chain (*COL22A1*), t-SNARE domain containing 1 (*TSNARE1*), Potassium two pore domain channel subfamily K member 9 (*KCNK9*), Zinc finger and AT-hook domain containing (*ZFAT*), Jrk helix-turn-helix protein (*JRK*), Maestro heat like repeat family member 5 (gene/pseudogene) (*MROH5*), Gasdermin D (*GSDMD*), Trafficking protein particle complex subunit 9 (*TRAPPC9*), and MIR3667 host gene (*MIR3667HG*) (Table 1 and Fig. 5). The evaluation turned ATPase phospholipid transporting 11C (*ATP11C*) gene as a top gene with elevated read mapping in the non-cancerous cells (Table 1 and Fig. 5). Further studies are needed to evaluate any correlation of those EV DNA fragments with patient outcomes.

The low molecular weight DNA fragments demonstrate dense centromeric mapping for cancer cell

Even though a majority of the human genome is non-coding, we now know that those regions are full of components that play a crucial role in regulating gene expression¹⁵. It is often the case in genome-wide association studies, that genetic markers associated with disease risk or phenotype fall within uncharacterized regions of the genome¹⁶. The non-coding portions of the genome are important and merit further investigation but remain difficult to interpret. To this end, a focus was narrowed on the patterns of elevated read coverages from the EV DNA in sections of the genome with elevated read coverage (>50× coverage) which might distinguish the cancer cells from the non-cancer cells. We previously observed that the HMW DNA from the cancer cells and non-cancer counterparts were similar (Fig. 4B), with reads spanning the entire genome at very low coverage. However, centromeric regions (Fig. 6; Supplementary Figs. 4–25) and mitochondrial DNA (Supplementary Fig. 3) exhibited high levels of read coverage exceeding 50X coverage. In LMW fragments, the read distribution across the entire genome was higher in the pancreatic non-cancer cells, while the cancer cells' LMW DNA shows sparse whole genome coverage with peaks of high coverage (>50×) in distinct areas near the centromeres (Fig. 6), mitochondrial DNA (Supplementary Fig. 3), and certain chromosome-specific regions (Supplementary Figs. 4–25).

Additionally, our analysis reveals that the LMW fragments of the non-cancer cells' EV DNA uniformly mapped throughout the genome whereas, remarkably, the cancer cells densely map to very similar regions, especially near the centromeres of most chromosomes (Fig. 6A; Supplementary Figs. 5–25). The HMW fractions (Supplementary Fig. 4) all exhibited similar mapping across the genome, again with distinct peaks of high read coverages, but this time the non-cancerous cells HMW fraction was more similar to the cancer HMW fractions (Supplementary Fig. 4) when compared to the LMW profiles (Fig. 6A). The observations above were consistent across most of the chromosomes (Supplementary Figs. 5–25) but were more striking on chromosomes 7 (Fig. 6A), 3, 9, 10, 11, 13, 17, and 20 (Supplementary Figs. 7, 12, 13, 14, 16, 20, and 23, respectively). Chromosome 17 even shows a larger accumulation of the LMW reads from the cancer cells to the centromere (Supplementary Fig. 20), whereas chromosomes 9 and 11 reveal centromeric mapping uniquely for the cancer cells' LMW only (Supplementary Figs. 12 and 14). A zoom-in on the centromeric region revealed evident mapping signatures that distinguish all the cancer cells from their non-cancer counterparts (Fig. 6B).

Mutational analysis reveals cancer cell-specific mutations in low-molecular-weight DNA fragments

Mutations in cancer remain a topic for discussion. Our observations are that (i) a large proportion (95%) of cell-derived EV DNA molecules are non-coding (Fig. 3) and that (ii) the LMW DNA fragments distinguish the cancer cells from their non-cancer counterparts (Fig. 4A, C) and showed a distinct centromeric mapping signature (Fig. 6; Supplementary Figs. 7, 12, 13, 14, 16, 20, and 23). Accordingly, we tested the hypothesis that the LMW DNA

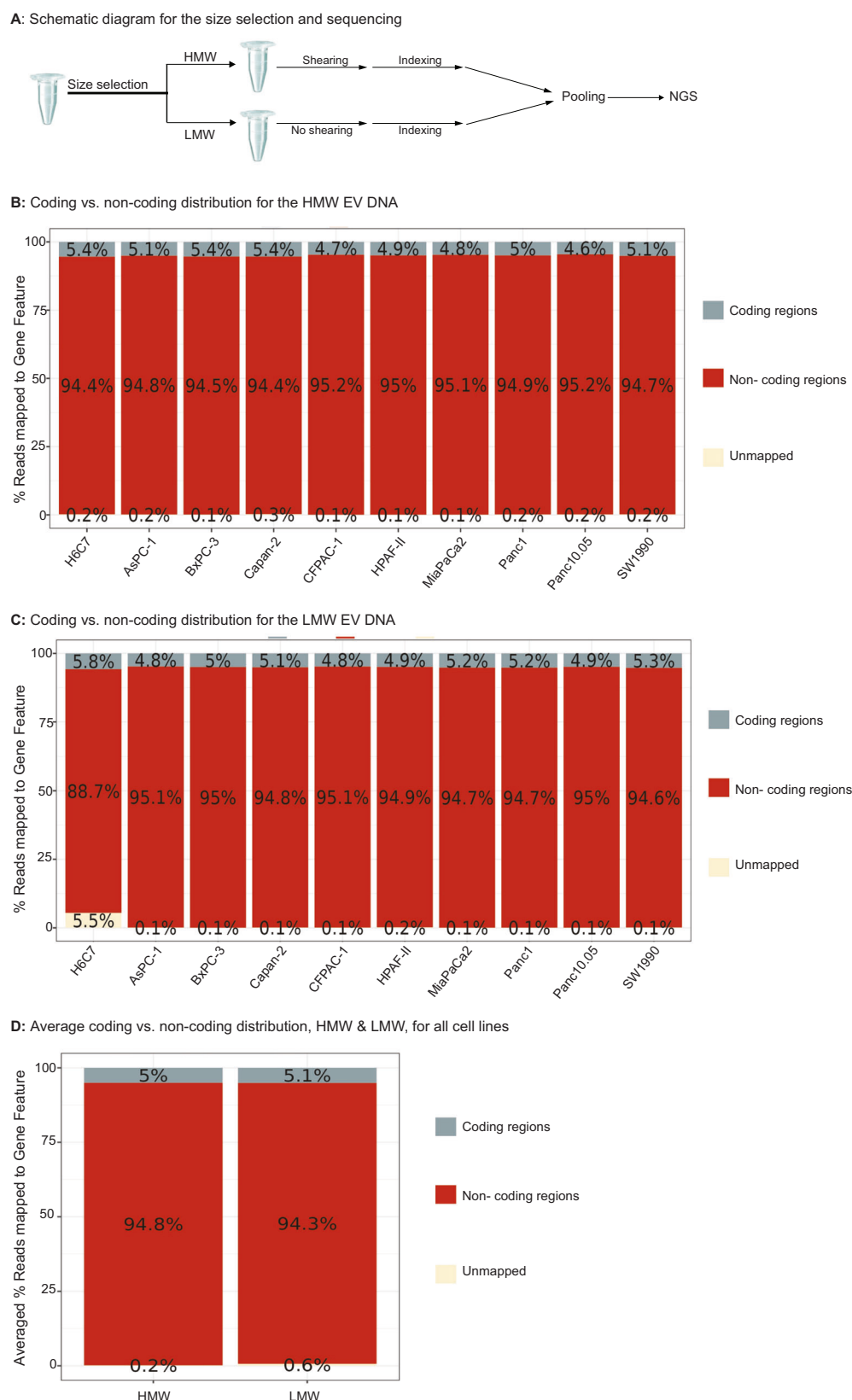


Fig. 3 | Whole genome sequencing analysis demonstrated a large proportion of EV DNA molecules from non-coding regions of the genome. **A** Schematic diagram of the size selection. Following the sequencing of the DNA from the EVs (isolated by the kit-based method), the reads were re-aligned against the human

genome, the percentage of reads from the high (**B**) or low (**C**) molecular weight fraction mapping to the coding or non-coding regions was determined bioinformatically, and graphed. **D** Average of the reads mapping to coding vs. non-coding regions from all ten cell lines.

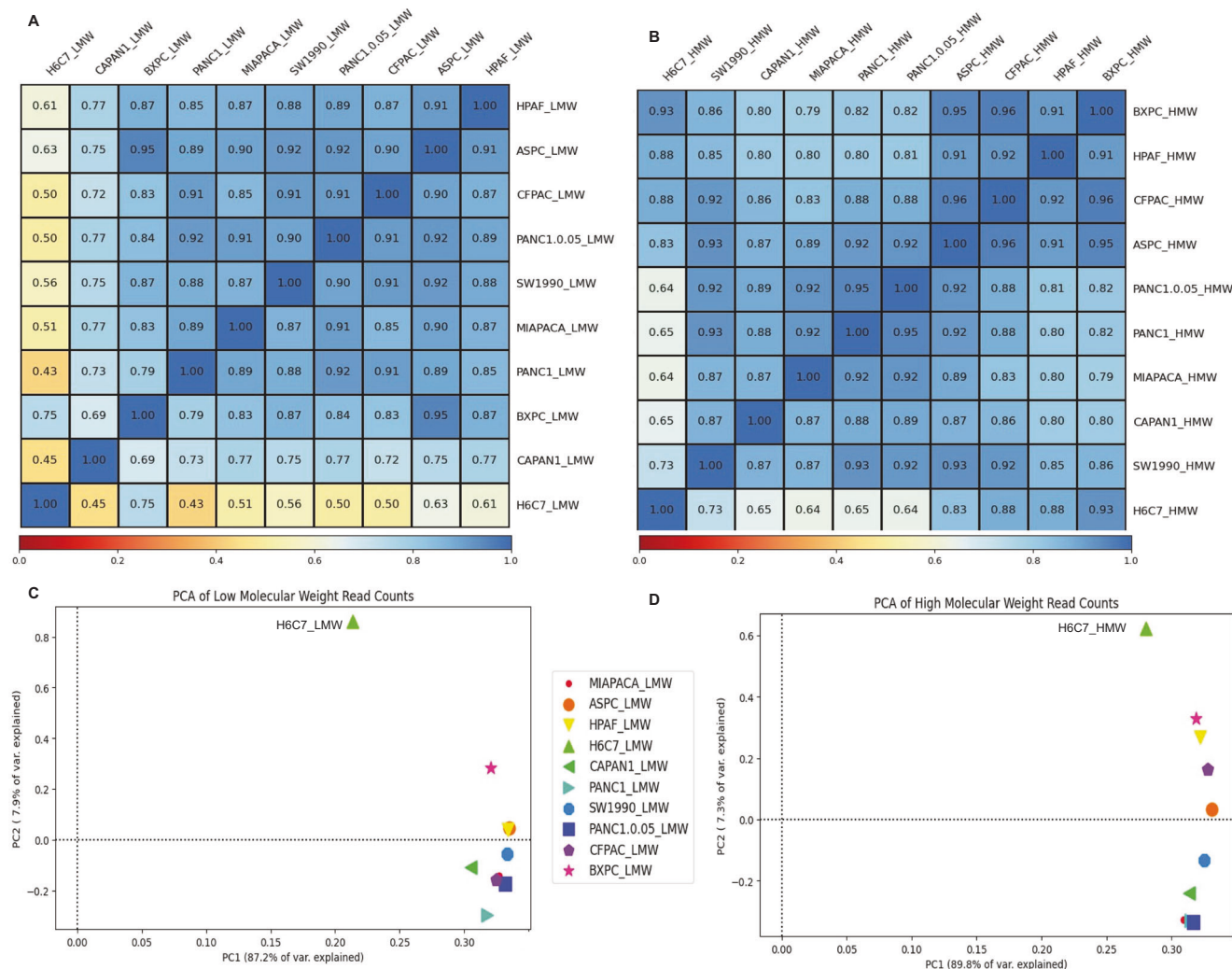


Fig. 4 | The low molecular weight DNA fragments distinguish the cancer cells from their non-cancer cells counterparts. A, B Pearson correlation coefficient showing that the LMW fragments (A) are more distinctive between the cancer cells

and the non-cancer cells. **C, D** Principal components analysis showing that the LMW fragments (C) from the cancer cells cluster away from the non-cancer cells.

fragments from cancer cells' EVs might harbor unique distinguishing mutational marks. Our mutational analysis demonstrates that a large proportion (>94%) of the mutations harbored in the DNA are in intergenic regions (Fig. 7A, B), a trend that aligns with our previous observations that a similar proportion of EV DNA fragments (~95%) originates from non-coding regions of the genome (Fig. 3B–D). Further analyses showed that A>G/T>C and C>T/G>A are the most frequent mutation types (Fig. 7C). Likewise, 1 bp and 2–4 bp insertion or deletion were in higher proportion (Fig. 7D). Lastly, we identified close to 200 mutations unique to the LMW DNA samples from the cancer cells only. These mutations, again, were largely intergenic with precise genomic locations (Table 2).

Discussion

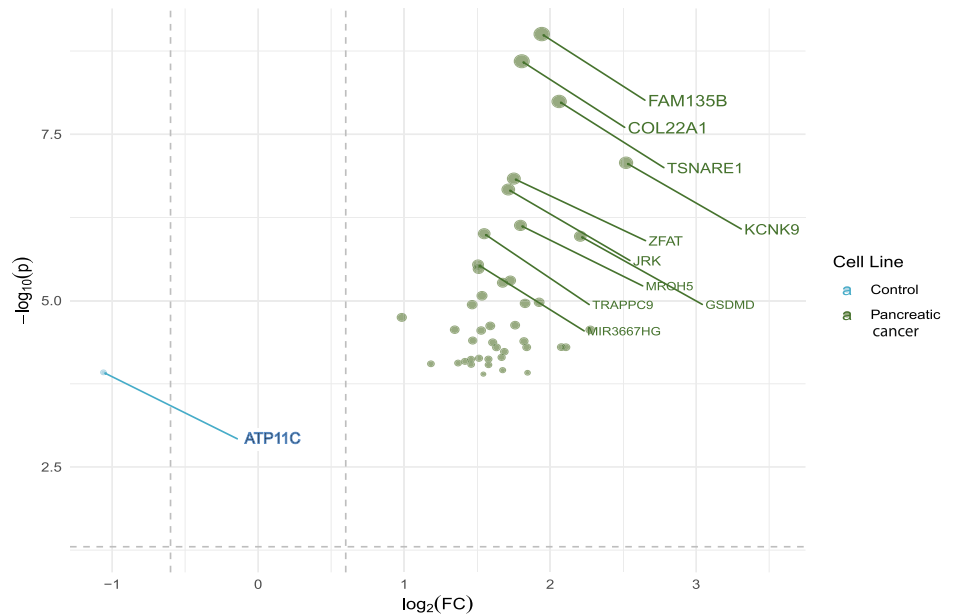
The molecular, coding, non-coding, and mutational profiling of EV DNA from different cell types has not yet been previously appreciated. In this study, we demonstrated that pancreatic cancer patient-derived cells vs. pancreatic non-cancer cells can be differentiated based on their EV DNA signatures. The state-of-the-art Agilent 4200 TapeStation-based DNA electrophoretic analysis reveals that pancreatic cancer patient-derived cell and their pancreatic non-cancer cell counterpart secrete EV DNA fragments with distinct signatures whereby HMW DNA fragments distinguish non-cancer cell from cancer cell which feature mostly LMW DNA fragments (Fig. 2B–D). The facts that the patterns above were absent in genomic DNA directly from the parental cells

secreting the EVs (Supplementary Fig. 2E, F) indicate that the fragments observed might be selectively packaged into EVs from the respective cell lines. NGS analysis revealed that dsDNA from the EVs from pancreatic cancer cell lines cluster together and are distinguishable from non-cancer cells based on their LMW read mapping profiles as well as dsDNA fragment size distribution (Fig. 4A, C). We further uncovered specific DNA fragments unique to the cancer cells (Table 1; Fig. 5). The biology of these DNA molecules, in the context of pancreatic cancer, remains to be elucidated especially given emerging functions of DNA aptamers in cell biology, such as their growth factor-like activities^{17–19}. Though, an overview of the literature suggests potential biomarker-like applicability for these molecules. For instance, the family with sequence similarity 135 member B (*FAM135B*) (Table 1; Fig. 5) has been identified as mutated and promotes esophageal squamous cell carcinoma²⁰. Likewise, the Collagen type XXII alpha 1 chain (*COL22A1*) is commonly mutated in lung adenocarcinoma and is associated with increased tumor mutations and poorer prognosis in the cancer²¹. The Potassium two pore domain channel subfamily K member 9 (*KCNK9*) is notoriously amplified in triple-negative breast and is associated with poor patient outcome²². Conversely, ATPase phospholipid transporting family proteins have tumor-suppressor-like biology²³. In our study, a DNA fragment corresponding to *ATP11C* was differentially highly abundant in the non-cancer cells only (Table 1; Fig. 5). Further investigations are necessary to evaluate a correlation of these DNA fragments with pancreatic cancer outcome.

Table 1 | Differentially abundant DNA fragments mapping to their corresponding genes

Entrez gene ID	Base mean	Fold change	log2 fold change	lfcSE	p value	p adj	Gene name
51305	52.5293	5.356389	2.519976	0.470462	8.49E-08	0.000298	potassium two pore domain channel subfamily K member 9(KCNK9)
340385	40.83501	4.19331	2.274789	0.542481	2.75E-05	0.017206	zinc finger protein 517(ZNF517)
79792	44.5353	4.876449	2.206439	0.452468	1.08E-06	0.001686	gasdermin D(GSDMD)
203054	43.64905	4.054393	2.10785	0.519893	5.03E-05	0.023628	aarF domain containing kinase 5(ADCK5)
286102	35.9425	4.055266	2.076666	0.512091	5.01E-05	0.023628	transmembrane p24 trafficking protein 10 pseudogene 1(TMED10P1)
203062	93.95212	5.727662	2.061427	0.359907	1.02E-08	4.77E-05	t-SNARE domain containing 1(TSNARE1)
51059	123.5015	6.111224	1.942543	0.317865	9.89E-10	1.39E-05	family with sequence similarity 135 member B(FAM135B)
57713	94.51126	4.404351	1.924556	0.436967	1.06E-05	0.009523	Scm like with four mbt domains 2(SFMBT2)
2765	32.24195	3.843083	1.84521	0.480138	0.000121	0.03969	glycosylphosphatidylinositol anchored molecule like(GML)
90990	51.15059	4.054229	1.839221	0.453655	5.03E-05	0.023628	kinesin family member C2(KIFC2)
340390	73.73653	4.396036	1.828317	0.415901	1.10E-05	0.009523	WD repeat domain 97(WDR97)
7587	54.95983	4.103096	1.820575	0.443708	4.08E-05	0.022905	zinc finger protein 37A(ZNF37A)
169044	217.9104	5.959427	1.80555	0.302974	2.53E-09	1.78E-05	collagen type XXII alpha 1 chain(COL22A1)
389690	96.78981	4.950125	1.796476	0.362915	7.42E-07	0.001488	maestro heat like repeat family member 5 (gene/pseudogene)(MROH5)
575	98.13776	4.229418	1.760145	0.416167	2.34E-05	0.016868	adhesion G protein-coupled receptor B1(ADGRB1)
57623	92.20973	5.256394	1.751295	0.333174	1.47E-07	0.000413	zinc finger and AT-hook domain containing(ZFAT)
23513	94.37445	4.565079	1.727285	0.378369	4.99E-06	0.005828	scribble planar cell polarity protein(SCRIB)
8629	97.41843	5.18646	1.713219	0.330325	2.14E-07	0.000502	Jrk helix-turn-helix protein(JRK)
10656	50.00904	4.018281	1.68615	0.41962	5.86E-05	0.026564	KH RNA binding domain containing, signal transduction associated 3(KHDRBS3)
11156	50.81729	3.864983	1.674666	0.433292	0.000111	0.038063	protein tyrosine phosphatase 4A3(PTP4A3)
727957	157.3975	4.548848	1.674413	0.368096	5.39E-06	0.005828	maestro heat like repeat family member 1(MROH1)
66004	58.00909	3.971569	1.667592	0.419882	7.14E-05	0.030669	Ly6/neurotoxin 1(LYNX1)
286077	81.26694	4.053481	1.632136	0.402651	5.05E-05	0.023628	family with sequence similarity 83 member H(FAM83H)
9831	48.19577	4.093406	1.6065	0.392461	4.25E-05	0.022966	zinc finger protein 623(ZNF623)
114822	71.65165	4.223848	1.589839	0.376396	2.40E-05	0.016868	rhophilin Rho GTPase binding protein 1(RHPN1)
340371	62.80066	3.909885	1.577613	0.403493	9.23E-05	0.032428	nuclear receptor binding protein 2(NRBP2)
90987	57.30882	3.956916	1.577231	0.398601	7.59E-05	0.030669	zinc finger protein 251(ZNF251)
83696	125.1061	4.89394	1.548243	0.316359	9.88E-07	0.001686	trafficking protein particle complex subunit 9(TRAPPC9)
9401	49.10256	3.831467	1.542665	0.402631	0.000127	0.040667	RecQ like helicase 4(RECQL4)
4796	76.4573	4.453242	1.533215	0.344292	8.46E-06	0.008487	tonsoku like, DNA repair protein(TONSL)
7553	62.97017	4.187755	1.527551	0.364766	2.82E-05	0.017206	zinc finger protein 7(ZNF7)
83481	154.6704	3.964448	1.512488	0.381513	7.36E-05	0.030669	epiplakin 1(EPPK1)
27161	189.1098	4.648852	1.509881	0.324786	3.34E-06	0.004262	argonaute RISC catalytic component 2(AGO2)
348645	106.8914	4.678868	1.506052	0.321884	2.88E-06	0.004052	MIR3667 host gene(MIR3667HG)
286101	58.56147	4.108216	1.469352	0.357662	3.99E-05	0.022905	zinc finger protein 252, pseudogene(ZNF252P)
5339	206.222	4.386394	1.465785	0.334166	1.15E-05	0.009523	plectin(PLC)
80728	82.81046	3.912744	1.45889	0.372856	9.13E-05	0.032428	Rho GTPase activating protein 39(ARHGAP39)
286075	56.63698	3.955367	1.458306	0.36869	7.64E-05	0.030669	zinc finger protein 707(ZNF707)
50626	69.84278	3.938497	1.416377	0.359624	8.20E-05	0.031993	zinc finger TRAF-type containing 1(ZFTRAF1)
5747	149.4065	3.924691	1.369411	0.348922	8.68E-05	0.032428	protein tyrosine kinase 2(PTK2)
22898	130.743	4.194187	1.346491	0.321037	2.74E-05	0.017206	DENN domain containing 3(DENND3)
58500	83.49485	3.918417	1.183341	0.301995	8.91E-05	0.032428	zinc finger protein 250(ZNF250)
10071	320.1541	4.289752	0.983784	0.229333	1.79E-05	0.013959	mucin 12, cell surface associated(MUC12)
286410	97.42421	-3.84638	-1.05959	0.275479	0.00012	0.03969	ATPase phospholipid transporting 11 C(ATP11C)

Fig. 5 | Unique coding DNA signatures for the LMW fragments. Volcano plot showing the differential abundance of the DNA fragments mapping to their corresponding coding genes.



Our analysis further revealed that EV DNA might originate from non-coding regions of the genome. A large proportion of the DNA fragments (~95%) across all cell lines used originate from non-coding regions of the genome (Fig. 3), somewhat following a similar general trend in the human genome project reports. It was assessed that almost all human genome, or 99% to be accurate, is non-coding, while only a very small fraction (1%) is coding genes. Nonetheless, our result that 5% of EV DNA reads mapped to coding regions of the genome suggests that EV DNA distribution (coding vs. non-coding) follows a different pattern than the expected 1% coding vs. 99% non-coding distribution reported in the human genome project. This indicates that perhaps the coding DNA packaged in EVs might be amplified from 1% to 5%. Despite the similar distribution of the EV DNA (coding vs. non-coding) across all cell lines assayed (Fig. 3), we observed a distinctive mapping pattern of the LMW DNA fragments: that from cancer cells densely mapped to the centromeric region of the genome while non-cancer cell-derived EV DNA sparsely mapped throughout the genome (Fig. 6; Supplementary Figs. 5–25).

Some chromosomes showed the trend above more clearly than others, such as chromosomes 7 (Fig. 6A), 3, 9, 10, 11, 13, 17, and 20 (Supplementary Figs. 7, 12, 13, 14, 16, 20, and 23, respectively). These chromosomes revealed distinct mapping signatures, setting aside the cancer cells in this study from their control non-cancer cells or vis-versa. Why this centromere-dense mapping pattern for the cancer cell-derived EV DNA is an interesting question. We speculate that the chromosome-dependent centromeric mapping signature of EV DNA from cancer cells represents unique genomic events at those chromosomes. The fact that these DNA signatures are mostly from non-coding regions, not coding regions, further hints at higher genomic activities at these non-coding regions. The 0.01% of the sequencing reads mapping to the mitochondrial genome (Supplementary Table 2) suggests very negligible mitochondrial DNA (mtDNA) proportional to the rest of the reads in the dataset; however, this was enough to achieve high coverage of the mitochondrial genome (Supplementary Fig. 3) potentially due to the relatively small size of the mitochondrial (~16569 bp) genome compared to the whole human genome (~3,200,000,000 bp). A relatively small size of the mitochondria genome might make high coverage of the mitochondrial genome more achievable compared to the rest of the human genome because there are many fewer bases in the mitochondrial chromosome. The significance of this small fraction of mtDNA fragments remains to be elucidated as well. Previous reports indicate that an association might exist between mtDNA and chemoresistance³⁴ or tumor growth²⁵. This merits further investigation.

Ultimately, we demonstrated that EV dsDNA fragments and the mutations therein originate largely from non-coding centromeric regions of the genome. The 5% fractions of the DNA molecules mapping to coding regions hints at a potentially elevated packaging of coding DNA fragments in EV DNA relative to what would have been expected based on the 1% coding region of the genome. The LMW DNA fragments (i) hold clues as to the distinction between pancreatic cancer patient-derived cells and their non-cancer counterparts, (ii) harbor unique coding DNA fragments and mutation hotspots, and (iii) demonstrate dense centromeric mapping for the cancer cells. Overall, the study rationalizes the need for a new approach to clinically relevant DNA biomarker discoveries and applications in cancer.

As prefaced in the introduction, conflicting data currently exists in the literature pertaining to DNA in EVs such as exosomes. On one hand, exosomes harbor DNA^{5,6,8,9}. On the other hand, exosomes might not contain DNA¹⁰. As we highlighted in the introduction, the differences in the various cell/system types used in those studies might explain the seemingly contradicting findings in those studies. Our study aligns with those studies that found that exosomes contain DNA^{5,6,8,9}. Like our study, pancreatic cancer cells were used^{16,9}, unlike the other study that focused on other cancer types¹⁰. In that study, DNA might be contained in a fraction of particles that the authors stated “in some ways...resemble...exosomes”¹⁰. Though, a recent review article pointed to some of the limitations of that study, such as the limited number of cell lines used, as well as the lack of details about how many exosomes were used in the study¹¹. Importantly, the previous study that first discovered and characterized exomeres⁹ also reported DNA in not just the exomeres but also in exosomes⁹, indicating that DNA might be differentially distributed in EVs such as exosomes and exomeres. It is even likely that exosomes have more DNA than exomeres depending on the cell line⁹. Of note, the methods used to purify the exomeres in those two studies above are also different, one conventional method¹⁰ while the other⁹ employed an advanced extraction technique. Regardless, exomeres are said to be non-membranous particles⁹. If so, in one’s mind, they lack defined structural boundaries within which hypothetical molecules such as DNA are to be found in the first place, let alone “exclusively”. On the other hand, exosomes are membranous particles, making exosomes more likely to harbor molecules such as DNA.

On the question of DNA in exosomes, it is most likely that the DNA is mostly membrane-enclosed. If it was not, a remaining possibility would be that it is tethered/attached to the exosomes. However, the biophysical properties of exosomes make DNA attaching to them very unlikely in the first place. Namely, exosomes are said to have a negatively charged surface⁹.

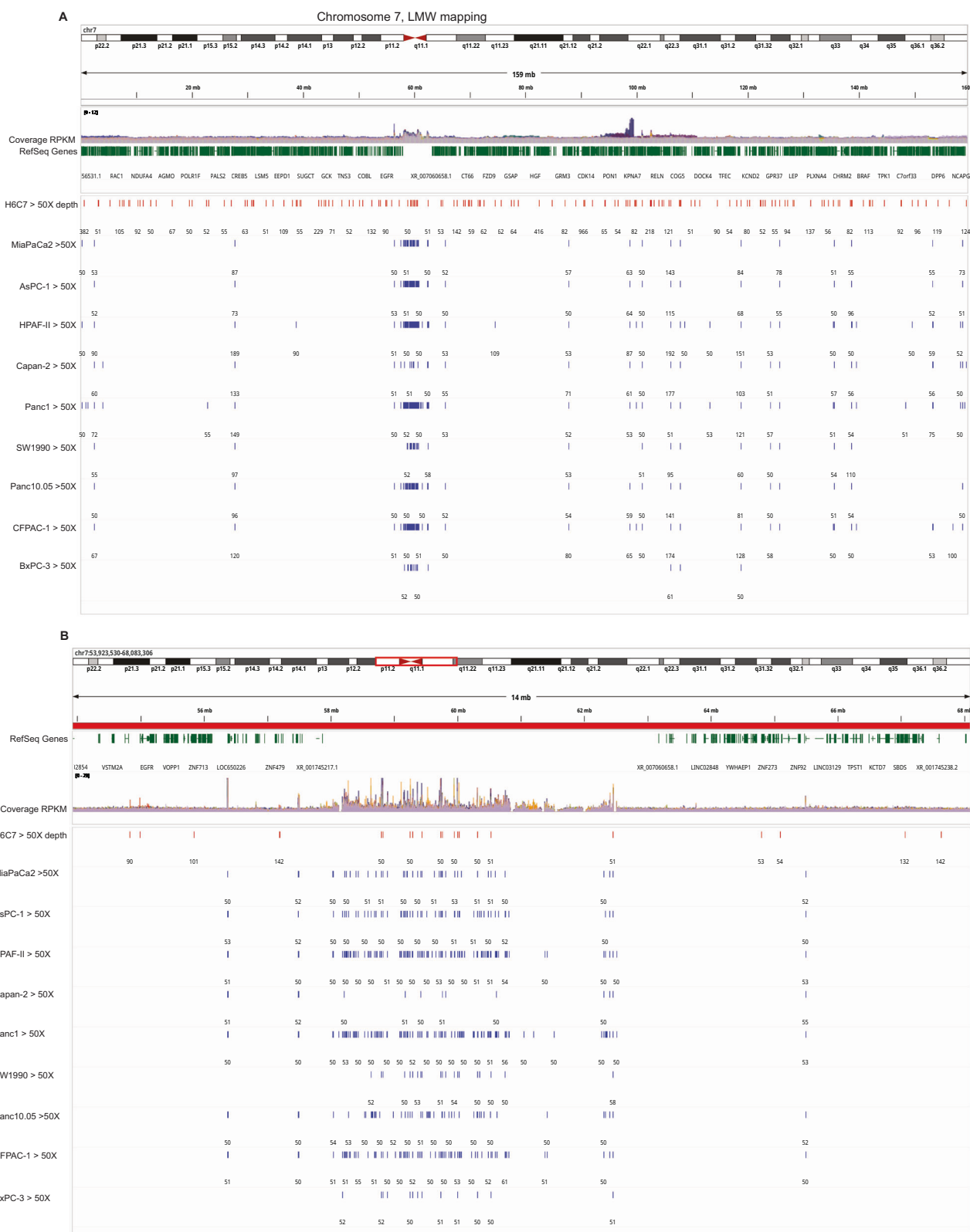


Fig. 6 | The low molecular weight DNA fragments demonstrate dense centromeric mapping for the cancer cells. A A mapping of the reads showing a dense centromeric mapping of all cancer cell lines but a sparse mapping for the control

non-cancer cells. **B** A zoom-in of the centromeric region demonstrating clear distinct mapping signatures for the cancer vs. control cells.

DNA molecules have long been widely thought to be negatively charged as well, making direct exosome-DNA association unlikely since particles/molecules with similar charges would be less likely to interact. Still, we acknowledge that biology is not fully predictable. In a more complex

environment such as in vivo, the dynamics might be different where interaction between countless molecules/particles might alter charges in a way that enables attraction/association between molecules/particles that would otherwise not associate with each other in the first place. Yet, so far,

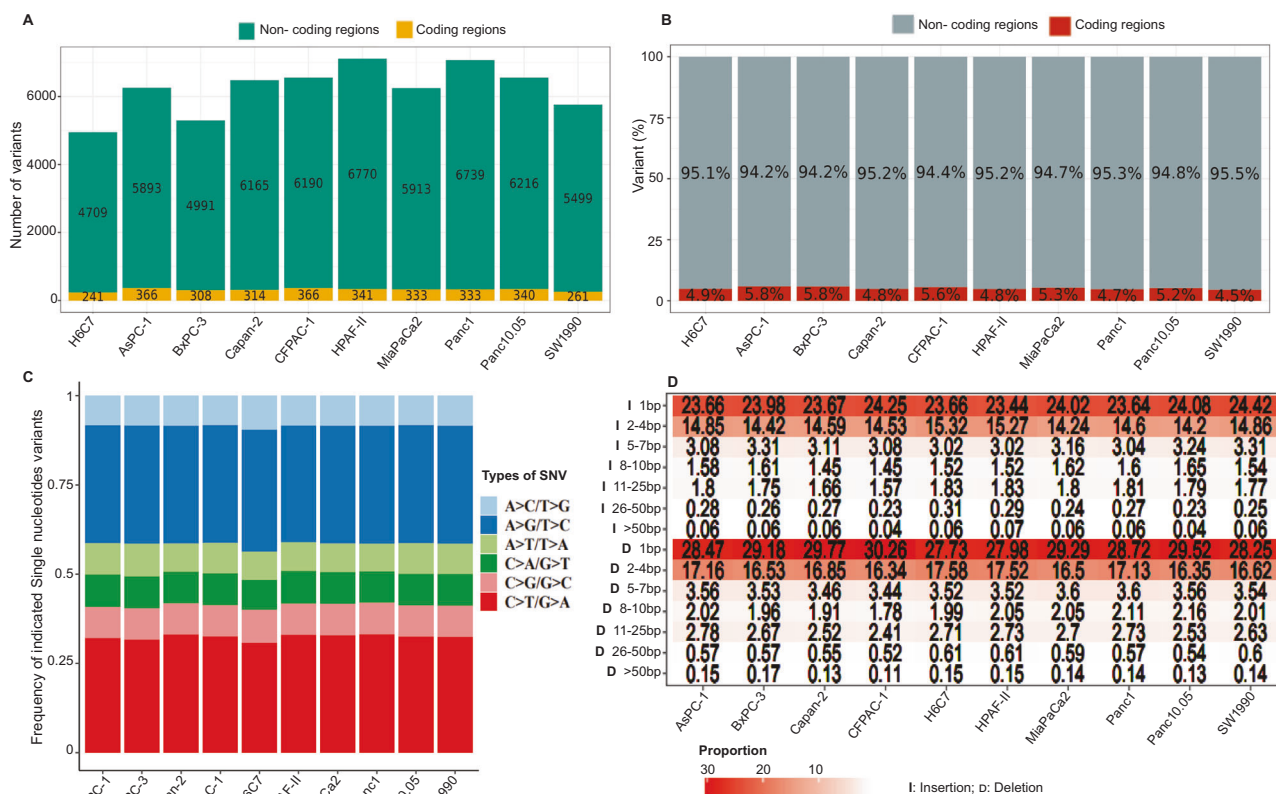


Fig. 7 | Mutational analysis reveals cancer cell-specific mutations in low-molecular-weight DNA fragments. Number (A) or percentage (B) of variants that fall into coding (intragenic) or non-coding (intergenic) regions of the genome. Types

of mutations, single nucleotide variant and frequency (C) or insertion, deletion, and proportion (D).

even in that in vivo environment, Fernando et al. found out that most serum/plasma-derived DNA is inside, not outside, exosomes⁵.

Just like any study, ours has its limitations. While the use of nine pancreatic cancer-derived cell lines demonstrated more than just the rigor in the study, there was only one control cell line derived from the pancreatic duct of a non-cancer individual, which offers the best comparison to pancreatic ductal cancer cells. This limitation is inherent to the lack of a large pool of pancreatic non-cancer cell lines for use in pancreatic cancer research. Currently, there are only two such cell lines widely used, the Human Pancreatic Ductal Epithelial (HPDE) and the Human Pancreatic Nestin-expressing (HPNE) cell lines. The H6C7 control cell line is derived from the HPDE cell line²⁶. The highly priced HPNE cell line (over \$6K/vial at ATCC and over \$7K/vial at Fisher) prevented the use of that cell line as another control cell line. Second, the study lacked pancreatic cancer patients' data. We actively sought to produce such data but were limited by access to the needed patients' plasma/serum samples.

Methods

Cell lines and culture

H6C7, a normal HPDE cell line^{27–29} was purchased from Kerafast (Shirley, MA, USA; cat#: ECA001-FP) and maintained in Keratinocyte serum-free media [dermal cell basal medium (ATCC PCS-200-030) supplemented with keratinocyte growth kit (PCS-200-040)] from American Type Culture Collection (ATCC; Manassas, VA, USA), streptomycin and gentamycin. Of note, the H6C7 cell line is derived from the normal primary HPDE cell line by immortalization with human papillomavirus-16 E6E7 gene²⁶, is morphologically similar to and demonstrated near normal genotype and phenotype of the parental HPDE cell line, per the vendor (Kerafast). Pancreatic cancer cells (Panc1, Panc10.05, CFPAC-1, Capan-2, MiaPaCA2, BxPC-3, HPAF-II, AsPC1, and SW1990) were purchased from ATCC (cat#: TCP-1026; CRL-1469) and grown in DMEM (cat#: ATCC 30-2002; ATCC) supplemented

with 10% FBS (cat#: 16000044; Thermo Scientific) and streptomycin and gentamycin. All cell lines were used immediately after purchase.

EV collection

Cells, at about 60% confluency, were switched from their growth media (with 10% FBS) to a media supplemented with 10% exosome-free FBS (Thermo Scientific; cat#: A2720803). The culture media was collected 48 h later for EV isolation and the cells were harvested for genomic DNA extraction (below). The media was centrifuged at 2 K × g for 30 min to remove cells and debris and then at 10 K × g for 30 min to remove macromolecules/microparticles. EVs were collected from the supernatant made of the cell/debris-free media by two methods:

- Ultracentrifugation was described³⁰ and used in other studies^{1,31–33}. Briefly, the cell/debris-free media above was centrifuged at 120 K × g for 4 h. The EV pellet was washed in 1X PBS followed by a re-centrifugation as above. The resulting pellet was dissolved in PBS and used immediately or stored at –80 °C.
- The PEG (Polyethylene Glycol)-based precipitation using the Thermo Fisher reagent for collecting total exosomes from culture media (cat#: 4478359) and previously reported in other studies^{34–37}. The cell/debris-free media above was mixed with the reagent at 1:0.5 ratio and incubated overnight at 4 °C. The next day, the mixture was centrifuged at 10,000 × g for 1 h at 4 °C, the supernatant removed and the exosomal pellet re-suspended in 1x PBS.

Extraction of EV DNA, cellular total DNA, DNA detection and quantification by Qubit Fluorometer

EV DNA extraction was carried out using the Qiagen circulating nucleic acid kit (cat# 55114) used in ref. 5. In brief, 1 ml of EV preparation + 100 µl of Proteinase K and 800 µl of lysis buffer (ACL) was vortexed (30 s) and digested for 30 min at 60 °C, followed by

Table 2 | Mutational analysis reveals cancer-specific mutations in low-molecular-weight DNA fragments

1:125179330 — A/T	4:49154498 — T/C
1:143217266 — C/T	4:49154558 — C/T
1:143217802 — A/C	4:49635250 — C/G
1:143219615 — A/T	4:49635266 — T/A
1:143220903 — C/G	4:49650841 — A/G
1:143251869 — G/C	5:49602423 — C/G
1:143267064 — G/C	5:49602513 — T/G
1:143267252 — A/T	5:49602777 — C/G
2:90385331 — A/G	5:49656779 — T/G
2:90385654 — C/A	5:49657626 — C/A
2:90396631 — C/A	5:49657660 — C/T
2:90401008 — C/T	5:49657719 — T/C
2:90401025 — C/T	5:49658067 — A/G
2:90401090 — C/T	5:49658498 — T/G
3:75669148 — G/A	5:49659796 — A/T
3:75669168 — C/A	5:49660093 — G/C
3:75669202 — T/C	5:49660289 — T/C
3:75669230 — C/T	5:49661373 — A/C
3:75669364 — G/A	5:49661840 — C/T
3:75669388 — G/C	5:49661860 — C/T
4:49092994 — A/C	8:85811224 — T/G
4:49093016 — G/A	10:41858331 — A/G
4:49097326 — A/G	10:41858334 — A/C
4:49097372 — T/A	10:41858335 — C/G
4:49097387 — T/A	10:41858521 — A/G
4:49097392 — A/G	10:41904378 — G/T
4:49097736 — G/T	10:41914771 — T/G
4:49118986 — C/G	10:42067184 — G/A
4:49119646 — T/C	10:42067722 — A/G
4:49139651 — A/G	16:34572792 — T/A
4:49139654 — G/C	16:34582986 — G/A
4:49139675 — C/T	16:34582987 — G/C
4:49139685 — G/A	16:34583017 — C/G
4:49139734 — C/G	16:34583045 — G/A
4:49139752 — T/A	16:34589463 — G/C
4:49139797 — C/T	16:34592621 — A/T
4:49147308 — ATT/-	16:34594656 — /CAC
4:49147313 — /TGT	16:34595771 — C/G
4:49147350 — T/G	16:46383623 — G/A
4:49147362 — CAT/-	16:46393017 — T/A
4:49147368 — G/T	16:46395221 — A/T
4:49149434 — C/G	16:46395228 — T/A
4:49149456 — A/C	16:46395271 — C/A
4:49149457 — T/C	16:46401112 — C/T
4:49149459 — C/G	17:21969336 — G/A
4:49149816 — G/A	17:21969340 — /GCAGG
17:21969376 — A/-	1_KI270709v1_random:5741 — T/C
17:21969378 — A/T	1_KI270709v1_random:6655 — G/A
17:21969382 — AGTGGAATGGAATA/-	1_KI270709v1_random:7687 — A/T
17:21969411 — A/G	4_GL000008v2_random:2140 — G/A
17:21969456 — A/C	14_GL000225v1_random:85723 — A/T
17:21973067 — T/G	14_GL000225v1_random:85724 — G/T

Table 2 (continued) | Mutational analysis reveals cancer-specific mutations in low-molecular-weight DNA fragments

17:25389843—G/T	14_GL000225v1_random:85734—G/T
18:108344—G/C	14_GL000225v1_random:85741—C/A
18:108364—T/C	14_GL000225v1_random:120676—T/G
20:31053843—G/A	14_GL000225v1_random:120680—C/T
20:31053891—A/G	17_KI270729v1_random:1062—C/T
20:31053894—C/G	17_KI270729v1_random:3631—A/G
20:31054913—G/T	17_KI270729v1_random:3633—T/A
20:31054969—AATGGAATGGAATGGAATGGTATGG/-	17_KI270729v1_random:3640—A/C
20:31055110—A/C	17_KI270729v1_random:4271—T/G
20:31055127—GGAA/-	17_KI270729v1_random:4297—TT/-
20:31055133—G/C	17_KI270729v1_random:5955—A/G
20:31055151—T/C	17_KI270729v1_random:5965—T/G
20:31055599—-/AACGGAATGC	17_KI270729v1_random:5968—C/T
20:31058131—T/G	17_KI270729v1_random:25118—A/T
20:31059029—A/C	17_KI270729v1_random:27853—G/A
20:31059040—A/G	17_KI270729v1_random:27864—T/A
20:31059050—A/G	22_KI270733v1_random:144773—A/G
20:31061329—T/C	22_KI270734v1_random:121558—C/A
20:31064466—A/C	Un_GL000216v2:720—C/-
20:31064542—T/C	Un_GL000216v2:723—TTCCATTCCATTCCATTCC/-
20:31067331—-/GGAAC	Un_GL000216v2:166796—C/A
20:31068343—C/G	Un_GL000224v1:8200—A/G
20:31072340—G/C	Un_JTFH01001405v1_decoy:1532—T/G
20:31073450—A/G	Un_JTFH01001405v1_decoy:1589—C/G
20:31075803—G/C	Un_KI270337v1:838—G/C
20:31075831—G/A	Un_KI270438v1:101410—T/G
20:31185113—C/A	Un_KI270438v1:101699—C/G
21:10692816—C/G	Un_KI270438v1:101775—G/T
21:10692818—-/ACTCGGGTTG	Un_KI270438v1:102025—G/C
21:10692881—G/C	Un_KI270438v1:104641—A/G
21:10692923—A/C	Un_KI270438v1:105240—C/G
21:10692929—T/C	Un_KI270438v1:105248—C/G
21:10692936—C/A	Un_KI270438v1:105387—C/T
22:18730652—C/G	Un_KI270438v1:107131—A/G
Un_KI270744v1:108459—T/C	Un_KI270438v1:108797—-/AT
Un_KI270744v1:110645—G/A	Un_KI270744v1:108413—A/T

incubation with binding buffer (ACB). The nucleic acids were inserted into a column, washed, and eluted. *Total cellular DNA* (including genomic and mitochondrial) was extracted from cell pellet using the Zymo Research genomic DNA isolation kit (cat#: D4068) following vendor's instructions. Concisely, cell pellets re-suspended in 200 µl of PBS were digested with 200 µl of BioFluid & Cell Buffer + 20 µl Proteinase K at 55 °C for 10 min followed by addition of genomic binding buffer, insertion into column, wash, and elution.

DNA detection and quantification were performed by adding 2 µl of sample to 198 µl of dsDNA detection buffer containing 1 µl of reagent (dye) followed by reading in the Qubit Fluorometer.

Transmission electron microscopy (TEM) and nanoparticle tracking analysis (NTA)

Were performed at the University of Nebraska Medical Center. During the last revision, additional TEM images were taken at the University of Nebraska, Lincoln, and appear in Supplementary Fig. 2.

Exosome markers array

Protein quantification was run on EV samples using the Pierce BCA protein assay (cat#: 23227; Thermo Fisher) and exosome marker proteins were assayed using the Exo-Check Exosome Array antibodies (cat#: EXOR-AY200B-4; System Biosciences, CA, USA) according to vendor's manual. Briefly, 50 µg of EV were lysed in lysis buffer, incubated with labeling reagent for 30 min at room temperature (RT), followed by blocking and incubation with the membrane overnight at 4 °C. Next, the membrane was washed, incubated with detection buffer for 30 min at RT and developed on the Azure 600 imager using Clarity Western ECL detection substrate (cat #: 1705060, Bio-Rad, USA).

Exosomal DNA analysis by Agilent 4200 TapeStation system

The D5000 and Genomic TapeStation Screen Tapes were run according to the manufacturer's protocol. 2 ng of DNA was mixed with 10 µl of sample buffer (containing SYBR green) and run on the TapeStation instrument per manufacturer's instructions.

DNase I treatment

Exosomal preparation in PBS or dsDNA in TE buffer was treated with DNase I (Qiagen) 0.2 unit per μL of EVs preparation plus reaction buffer with MgCl_2 , and incubated at 37 °C for 30 min, then at 65 °C for 10 min to inactivate the DNase.

Size selection

DNA fragments in the EVs exhibited unique banding patterns based on tapestation screening. This observation led us to split NGS samples into two fractions using ~1 kb as the dividing threshold, HMW > 1 kb, and LMW < 1 kb. Samples were split using a modified AMPure XP magnetic bead size selection (Beckman Coulter, PN A63881). Initially, to each cell line exosomal DNA extraction, a 0.5X volume AMPure XP beads was added and thoroughly mixed and incubated at RT for 5 min. Samples were then placed on a magnetic stand and allowed to clear (takes ~2 min). The supernatant containing LMW DNA was transferred to a new tube, while the HMW DNA remained bound to the beads. HMW samples were washed with two rounds 200 μL of 80% ethanol. Following the second wash, all remaining ethanol was removed, and the pellet was allowed to dry (takes ~2 min) and HWM samples were eluted in 15 μL 0.1X TE buffer. HMW DNA was then sheared to ~500–700 bp using a Covaris M220 (PN 500295) focused-Ultrasonicator in microTUBE AFA Fiber Snap-Cap (PN 520045), peak incident power = 50, Duty factor = 10, cycles per burst = 200, and a treatment time of 50 s. LMW DNA fractions did not need to be sheared, already falling within the desired library insert size range. LMW and sheared HMW DNA fractions were quantified on a Thermo Fisher Qubit 4.0 fluorometer (cat# Q33238) using Invitrogen's Broad Range dsDNA assay (cat# Q33265). DNA insert size was verified using an Agilent's Tapestation 4150 (PN G2992AA) using a D1000 assay (PN 5067-5582, 5067-5583).

DNA library preparation and NGS

HMW and LMW NGS libraries were prepared using New England Biolab's NEBNext Ultra II Library Prep Kit for Illumina (PN E7645S) according to the manufacturer's instructions. A 1:10 dilution of the adaptor at step 2.1 was used for all samples, and cleanup of adaptor-ligated DNA was performed without size selection (protocol section 3B). PCR amplification of barcoded libraries at step 4.1.3 was performed using 7 cycles. Resulting libraries cleanup was performed using a 0.65X AMPure XP bead cleanup. LMW and HMW libraries were then sequenced on an Illumina NextSeq 550DX High output flowcell (PN20028871), with an average read depth of 48 million and average of 98% reads mapping to the human genome reference (HG38).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Bioinformatics, statistical analysis and reproducibility Read mapping, variant calling, genome coverage, and visualization

Read processing, mapping and variant calling were performed using Sention's DNAScope pipeline for variant calling (Sention Inc., San Jose, California USA)³⁸. Reads from all samples were mapped to the GRCh38 build of the human reference genome using BWA-MEM algorithm, creating BAM files for high and LMW samples in each cell line. Duplicate reads were removed with Sention's Dedup algorithm, and variant calling was performed with the DNAScope algorithm. Genome coverage for each sample was calculated using the genomeCoverageBed function in the BEDTools software suite³⁹. Regions of the genome with an average of 50X coverage or greater were documented in bedfiles using covtobed command in the BamToCov toolkit⁴⁰. The Integrative Genomics Viewer (IGV) genome browser was used to visualize genomic regions with high read coverage (> 50X) from the sequenced EV DNA⁴¹.

Variant analysis

Variant analysis was performed using VarSeq™ v2.5.0 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com) (ref: VarSeq™ (Version 8.x)

[Software]. Bozeman, MT: Golden Helix, Inc. Available from <http://www.goldenhelix.com>). Overall, 10,066,263 variants were detected in the dataset. Variants were quality-filtered to a minimum read depth of at least 100, without a LowQual flag, as well as have a variant allele fraction ≥ 0.1 . Additionally, variants with allele frequencies ≤ 0.01 and ≥ 0.99 in gnomAD and 1 kG Phase3 databases were kept^{42,43}. Variants remaining after filtration are summarized and visualized using ggplot2 in R.

Investigation of reads mapped to exon regions

Reads mapped to coding regions of the genome were calculated using the featureCount() function in the subreadR package in R. The percentage of reads which mapped to coding regions of the genome versus non-coding are summarized and graphed. Gene feature counts that were differentially abundant in cancer cell lines and control cell lines were performed with DESeq2 package in R and visualized using ggplot2 in R.

BAM comparisons

Comparison of BAM alignment profiles was performed using the command line interface of deepTools⁴⁴. The multiBamSummary command was used to summarize all reads with mapping quality scores over 30. DeepTool's plotPCA function was used to perform a Principal Component Analysis (PCA) of read count profiles generated in the multi-Bam summaries. Similarly, Pearson correlation coefficient matrices and heatmaps were generated from the multi-Bam summaries with deepTools2 plotCorrelation function.

Data availability

All data/images supporting the figures presented are in this manuscript. The sequencing data is deposited into NCBI (BioProject ID: PRJNA1075045).

Received: 13 March 2024; Accepted: 17 January 2025;

Published online: 05 March 2025

References

- Hoshino, A. et al. Tumour exosome integrins determine organotropic metastasis. *Nature* **527**, 329–335 (2015).
- Costa-Silva, B. et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat. Cell Biol.* **17**, 816–826 (2015).
- Johnstone, R. M. The Jeanne Manery-Fisher Memorial Lecture 1991. Maturation of reticulocytes: formation of exosomes as a mechanism for shedding membrane proteins. *Biochem. Cell Biol.* **70**, 179–190 (1992).
- Makler, A. & Asghar, W. Exosomal biomarkers for cancer diagnosis and patient monitoring. *Expert Rev. Mol. Diagn.* **20**, 387–400 (2020).
- Fernando, M. R., Jiang, C., Krzyzanowski, G. D. & Ryan, W. L. New evidence that a large proportion of human blood plasma cell-free DNA is localized in exosomes. *PLoS ONE* **12**, e0183915 (2017).
- Thakur, B. K. et al. Double-stranded DNA in exosomes: a novel biomarker in cancer detection. *Cell Res* **24**, 766–769 (2014).
- Kalluri, R. & LeBleu, V. S. Discovery of Double-Stranded Genomic DNA in Circulating Exosomes. *Cold Spring Harb. Symp. Quant. Biol.* **81**, 275–280 (2016).
- Kahlert, C. et al. Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. *J. Biol. Chem.* **289**, 3869–3875 (2014).
- Zhang, H. et al. Identification of distinct nanoparticles and subsets of extracellular vesicles by asymmetric flow field-flow fractionation. *Nat. Cell Biol.* **20**, 332–343 (2018).
- Jeppesen, D. K. et al. Reassessment of exosome composition. *Cell* **177**, 428–445.e18 (2019).
- Elzanowska, J., Semira, C. & Costa-Silva, B. DNA in extracellular vesicles: biological and clinical aspects. *Mol. Oncol.* **15**, 1701–1714 (2021).
- Thery, C. et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society

- for Extracellular Vesicles and update of the MISEV2014 guidelines. *J. Extracell. Vesicles* **7**, 1535750 (2018).
13. Lotvall, J. et al. Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. *J. Extracell. Vesicles* **3**, 26913 (2014).
14. Lazaro-Ibanez, E. et al. DNA analysis of low- and high-density fractions defines heterogeneous subpopulations of small extracellular vesicles based on their DNA cargo and topology. *J. Extracell. Vesicles* **8**, 1656993 (2019).
15. Elkon, R. & Agami, R. Characterization of noncoding regulatory DNA in the human genome. *Nat. Biotechnol.* **35**, 732–746 (2017).
16. Levings, D., Shaw, K. E. & Lacher, S. E. Genomic resources for dissecting the role of non-protein coding variation in gene-environment interactions. *Toxicology* **441**, 152505 (2020).
17. Ueki, R. et al. DNA aptamer assemblies as fibroblast growth factor mimics and their application in stem cell culture. *Chem. Commun.* **55**, 2672–2675 (2019).
18. Ueki, R., Atsuta, S., Ueki, A. & Sando, S. Nongenetic reprogramming of the ligand specificity of growth factor receptors by bispecific DNA aptamers. *J. Am. Chem. Soc.* **139**, 6554–6557 (2017).
19. Ramaswamy, V. et al. DNA aptamer assembly as a vascular endothelial growth factor receptor agonist. *Nucleic Acid Ther.* **25**, 227–234 (2015).
20. Song, Y. et al. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* **509**, 91–95 (2014).
21. Yang, S., Xia, J., Yang, Z., Xu, M. & Li, S. Lung cancer molecular mutations and abnormal glycosylation as biomarkers for early diagnosis. *Cancer Treat. Res. Commun.* **27**, 100311 (2021).
22. Mu, D. et al. Genomic amplification and oncogenic properties of the KCNK9 potassium channel gene. *Cancer Cell* **3**, 297–302 (2003).
23. Xu, J. et al. ATP11B inhibits breast cancer metastasis in a mouse model by suppressing externalization of nonapoptotic phosphatidylserine. *J. Clin. Invest.* **132**, e149473 (2022).
24. Abad, E. & Lyakhovich, A. Movement of mitochondria with mutant DNA through extracellular vesicles helps cancer cells acquire chemoresistance. *ChemMedChem* **17**, e202100642 (2022).
25. Sansone, P. et al. Packaging and transfer of mitochondrial DNA via exosomes regulate escape from dormancy in hormonal therapy-resistant breast cancer. *Proc. Natl. Acad. Sci. USA* **114**, E9066–E9075 (2017).
26. Ouyang, H. et al. Immortal human pancreatic duct epithelial cell lines with near normal genotype and phenotype. *Am. J. Pathol.* **157**, 1623–1631 (2000).
27. Gilles, M. E. et al. Personalized RNA medicine for pancreatic cancer. *Clin. Cancer Res.* **24**, 1734–1747 (2018).
28. Tanton, H. et al. F(1)F(0)-ATP synthase inhibitory factor 1 in the normal pancreas and in pancreatic ductal adenocarcinoma: effects on bioenergetics, invasion and proliferation. *Front. Physiol.* **9**, 833 (2018).
29. Zhu, G., Zhou, L., Liu, H., Shan, Y. & Zhang, X. MicroRNA-224 promotes pancreatic cancer cell proliferation and migration by targeting the TXNIP-mediated HIF1alpha pathway. *Cell. Physiol. Biochem.* **48**, 1735–1746 (2018).
30. Thery, C., Amigorena, S., Raposo, G. & Clayton, A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr. Protoc. Cell Biol.* **Chapter 3**, Unit 3.22 (2006).
31. Peinado, H. et al. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nat. Med.* **18**, 883–891 (2012).
32. Boelens, M. C. et al. Exosome transfer from stromal to breast cancer cells regulates therapy resistance pathways. *Cell* **159**, 499–513 (2014).
33. Genschmer, K. R. et al. Activated PMN exosomes: pathogenic entities causing matrix destruction and disease in the lung. *Cell* **176**, 113–26.e15 (2019).
34. Wei, Y. et al. Pyruvate kinase type M2 promotes tumour cell exosome release via phosphorylating synaptosome-associated protein 23. *Nat. Commun.* **8**, 14041 (2017).
35. Zhang, Y. et al. Pericytes control vascular stability and auditory spiral ganglion neuron survival. *eLife* **12**, e83486 (2023).
36. Xie, X., Guo, L. W. & Kent, C. K. miR548ai antagonism attenuates exosome-induced endothelial cell dysfunction. *Cell Death Discov.* **7**, 318 (2021).
37. Wang, G. et al. A nanounit strategy reverses immune suppression of exosomal PD-L1 and is associated with enhanced ferroptosis. *Nat. Commun.* **12**, 5733 (2021).
38. Kendig, K. I. et al. Sentieon DNaseq variant calling workflow demonstrates strong computational performance and accuracy. *Front. Genet.* **10**, 736 (2019).
39. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
40. Birolo, G. & Telatin, A. BamToCov: an efficient toolkit for sequence coverage calculations. *Bioinformatics* **38**, 2617–2618 (2022).
41. Robinson, J. T., Thorvaldsdottir, H., Turner, D. & Mesirov, J. P. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics* **39**, btac830 (2023).
42. Chen, S. et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92–100 (2024).
43. Fairley, S., Lowy-Gallego, E., Perry, E. & Flicek, P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res.* **48**, D941–D947 (2020).
44. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

Acknowledgements

This work was supported by funds from the Ryan Foundation (Greenwood Village, CO) to M.R.F. We thank the Electron Microscopy Core and the Nanomaterials characterization Core at the University of Nebraska, Medical Center for their assistance. The transmission electron microscopy images in Supplementary Fig. 2 were taken at the University of Nebraska, Lincoln. For that reason, we acknowledge the Light and Electron microscopy (LEMC) supported by NCIBC (NHI grant # P20 GM113126) and the UNL CryoEM Facility part of the Nebraska Center for Biotechnology.

Author contributions

Project conceptualization and experiments design: M.R.F., A.A.O.; supervision: M.R.F., A.A.O.; data acquisition: A.A.O., W.T., G.K., C.J., D.C., N.F.; scanning electron microscopy in the initial manuscript, A.D. and J.D.; sequencing and bioinformatics: W.T.; data analysis: A.A.O., W.T., and M.R.F.; manuscript writing: A.A.O.; review/editing: A.A.O., W.T., M.R.F., D.R.W.; funding acquisition: M.R.F.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-07567-1>.

Correspondence and requests for materials should be addressed to Appolinaire A. Olou or M. Rohan Fernando.

Peer review information *Communications Biology* thanks Nuno Bastos and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Christina Karlsson Rosenthal.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025