Improving Circulating Tumor DNA-based Liquid Biopsy Through Understanding Molecular Regulators of Cell Free DNA Release

By

Brad Davidson

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Cancer Biology

May 10th, 2024

Nashville, Tennessee

Approved:

Alissa Weaver, M.D., Ph.D. (Chair)

Christine Lovly, M.D., Ph.D.

P. Brent Ferrell, M.D.

Ben Ho Park, M.D., Ph.D. (Advisor)

ACKNOWLEDGMENTS

The work presented herein was made possible by financial support from MICTP T32 (T32CA009592-31A1) (B.A.D.), The American Cancer Society 131356-RSG-17-160-01-CSM as well as The Breast Cancer Research Foundation, Komen Foundation, NIH CA214494, CA194024 (B.H.P.). We would also like to thank and acknowledge the support of The Canney Foundation, the Sage Patient Advocates, the Marcie and Ellen Foundation, The Eddie and Sandy Garcia Foundation, the support of Amy and Barry Baker, the support of John and Donna Hall, and the Vanderbilt-Ingram Cancer Center support grant (NIH CA068485) and Breast Cancer SPORE (NIH CA098131). And of course, schematic figures were made possible by and created with Biorender.

First, I must acknowledge my mentors in Paula Hurley, Sarah Croessmann, and Ben Park. Paula, thank you for your insightful questions in lab meeting, always having your door open to my random technical questions, and being calm and collected in all situations. Sarah/Cressy, thank you for telling me that not joining the Ben Ho Park lab would be the worst mistake of my life, for endless gossip, and for always trying to convince me that the sky isn't falling. Ben, ever since you made poop jokes in your departmental introduction seminar there was no way I wasn't going to join your lab. Thank you for caring about me personally just as much as scientifically and for your unending optimism and gratitude. I aspire to have the outlook on life that you do. You are definitely wrong about the oxford comma though.

Other mentors who made a major impact on me at Vanderbilt include people with direct impacts on my research to those who haven't seen a pipette in years. My thesis committee in Brent Ferrell, Christine Lovly, Pierre Massion, and Alissa Weaver were indispensable for keeping me on track in my research. My short stint in the Irish/Ihrie labs was also the gift that kept on giving

by giving me an understanding of high dimensional data, thinking outside the box, presentation style, and connecting me with Justine Sinnaeve and Gabrielle Rushing who springboarded me into the rare disease advocacy field. To Chip, Pui, and Kristin of the DADA2 Foundation, thank you for taking me under your wing and helping me realize what is important to me for my career post-graduation.

Thank you to all Park/Hurley Lab West members past and present – I probably won't mention you all here, but the Park/Hurley Lab environment has been special since day 1 and you all helped make it the fun place that it (usually) was to spend my PhD. Adam, we really committed to the bit with the suicide pact of joining the Park Lab, and we made it. I literally wouldn't have been able to do anything computational without you, and having someone to experience this PhD with from start to finish made it bearable. To my landlord Sarah Reed, as soon as you rotated in the Park Lab I knew I needed you to join (even though I left to go on vacation while mentoring you for your rotation - oops). You carried me on your back through the physical/emotional turmoil of the back half of my PhD, and I could not be more grateful for knowing you. To my co-cell-line-generator Riley, thank you for being the hype-woman I always needed. From starting the HR and Self-deprecation jars to baking desserts for every milestone, everyone needs a Riley in their life. To Justin/Yustin, thank you for your constant companionship, putting up with all the weird things I say, and your ability to read my emails and check my experimental results.

Finally, I have to thank my family and friends from before coming to Vanderbilt, back home in Maryland and everywhere else. Most of you had no idea what I was doing and would sometimes ask "when are you going to be done?" or "When are you coming back to Maryland?". Don't worry y'all, the wait is over. My St. Mary's people in the Bens, the Joes, Graeme, Third Left, WC33, Caroline Prison Gang, and the whole SMCM Schaeferites cohort I came up with; My

Havre de Grace people in Chris Worrell, Anthony, Chris Scott, and many more; my NIH people in Ben and Sam; Mom and Dad; you all in part got me through this, and I can't wait to see you all on a more regular basis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

BACKGROUND AND RESEARCH DIRECTIONS

Sections of this chapter have been previously published in: Davidson, B.A., Croessmann, S. & Park, B.H. The breast is yet to come: current and future utility of circulating tumour DNA in breast cancer. *Br J Cancer* 125, 780–788 (2021). https://doi.org/10.1038/s41416-021-01422-w

Liquid Biopsies of Circulating DNAs in Cancer

Precision cancer medicine, or the tailoring of treatment to target unique characteristics of a tumor, has greatly improved the outcomes of cancer patients since its inception in the 1990s[1,2]. This treatment paradigm is driven by the identification of specific genetic mutations and/or protein expression signatures in the tumor, followed by the application of therapies against those changes. Theoretically, this should improve on-tumor activity and decrease off-target toxicity as compared to non-targeted treatments. Traditionally, these tumor specific changes have been identified through direct tumor biopsy and subsequent histopathology or sequencing. Initially, targeted therapies are often very effective, but cancers almost universally eventually develop resistance through a variety of mechanisms such as losing expression of a targeted protein or developing a *de novo* mutation[3]. This implies the need to collect serial information about tumors as they progress and modify treatment to combat new tumor phenotypes. Unfortunately, repeating a tumor biopsy to obtain this information is not simple. Tumor biopsies sometimes fail to obtain usable sample, in addition to being expensive and invasive[4-6]. Tumors are also highly heterogeneous, leading to different results from biopsies of the same tumor[7]. Further, a traditional biopsy of a primary tumor will never identify the burden or contributions of potential metastatic disease.

Liquid biopsies, defined for the purpose of this dissertation as a blood test that draws on cancer biomarkers circulating in the blood of patients, represent an effective method to address

these shortcomings. Given that blood tests are relatively easy to administer and all tumor sites must connect to a blood supply, liquid biopsies are poised to complement the serial application of targeted therapies. In addition, liquid biopsies tests are increasing in scope to identify cancer at early stages, indicate treatment choice, and monitor cancer progression[8]. There are many different blood analytes being studied in these capacities, including circulating RNAs, DNAs, tumor cells, and extracellular vesicles[9].

Circulating tumor DNA (ctDNA) based biopsies are the most established of these techniques. DNA was first found to circulate in the blood in 1948[10] and in the following years a variety of conditions and physiologic states were shown to increase cfDNA in the blood, including exercise, auto-immunity, and cancer[11-13]. This discovery was followed by the finding that some of the DNA in cancer patient blood comprised mutations attributable to the cancer, indicating that some DNA in the blood is ctDNA[14]. Indeed, both total cell-free DNA (cfDNA) and ctDNA in cancer patient blood increase with cancer stage and are both poor prognostic factors[15-17]. Expectedly then, cfDNA and ctDNA concentrations in patient blood also increase with cancer stage. Interestingly, the amount and thereby detectability of both DNA populations also varies by cancer type. For example, small cell lung cancer (SCLC) tends to be detected easily by ctDNA-based tests, while glioma is very difficult to detect[18-20]. The reasons for these differences in detectability are somewhat unclear, although in the case of brain cancers is likely mediated by the blood brain barrier preventing ctDNA from entering circulation[21]. Although our current understanding of what cancers release high amounts of ctDNA and why is poor, it is obvious that ctDNA is an up-and-coming biomarker in the treatment of cancer patients.

Methods in Circulating Tumor DNA Mutation Detection

The ability of a liquid biopsy to detect ctDNA in patient blood samples can be greatly modified by the techniques used to isolate the DNA from blood, as well as the downstream techniques used to quantify its presence. The liquid biopsy field is currently moving towards using large scale -omics to detect ctDNA and away from previously used PCR-based techniques that are potentially better equipped to detect specific mutations, but at a lower throughput. In this section, I will discuss the pros and cons these various techniques in the ctDNA analysis pipeline and why various liquid biopsy labs might think about employing them.

Sample Preparation

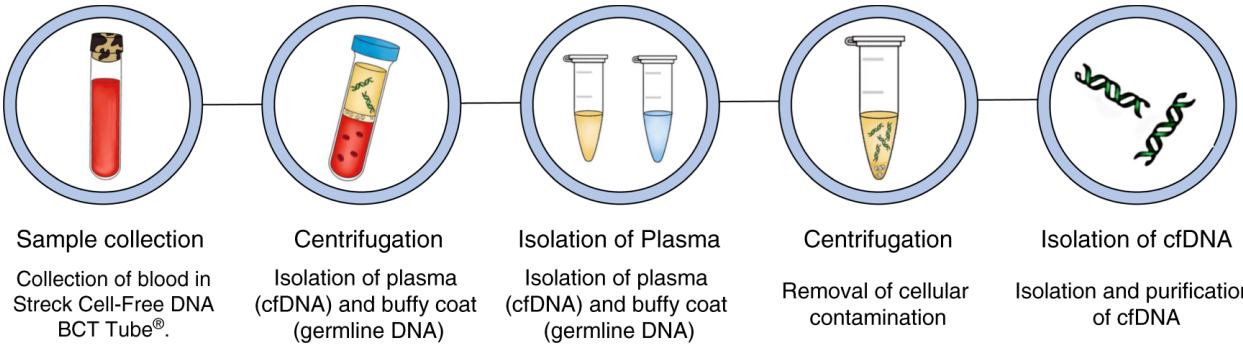


Figure 1.1. Blood processing for cell-free DNA isolation. Blood samples are collected in tubes made specifically to stabilize all cfDNA. The collected sample is then centrifuged and cfDNA-containing plasma is isolated. Remaining cellular contamination is eliminated with a final centrifugation, and cfDNA is isolated and purified through a variety of commercial kits.

The extraction and isolation of cell free DNA from plasma is an important step in the ability to utilize liquid biopsies for genetic information. An overview of general sample processing for cfDNA is shown in Fig. 1.1. Unfortunately, techniques for cfDNA isolation have not been fully standardized, leading to variance in results[22]. Improper sample processing can cause potential

contamination from the rapid lysis of white blood cells and/or the degradation of target DNAs. For example, EDTA blood collection tubes may be and have been traditionally used, but must be rapidly processed to prevent compromise of the sample. In their place, Streck Cell-Free DNA BCT collection tubes allow for stable storage blood for up to 14 days at room temperature and are ideal for circulating free DNA (cfDNA) isolation[23]. After collection, blood is centrifuged to separate the plasma, which contains most cfDNA, and the buffy coat, which contains germline DNA. Plasma is centrifuged a second time to remove cellular contaminants and processed using a circulating nucleic acid isolation kit (i.e. QIAamp® MinElute ccfDNA Kits, but many others exist potentially leading to even more variable results). The sample is then ready to be analyzed. ctDNA can also be isolated from samples such as cerebrospinal fluid, urine, and stool[24-26]. However, these tests are tailored to specific cancers related to these samples (brain, prostate/bladder, and colorectal respectively) and would not provide the systemic-level information a blood-based liquid biopsy can provide. The presence of ctDNA can be analyzed within total cfDNA isolated from these samples using a variety of methods, although most are polymerase chain reaction (PCR) or next-generations sequencing (NGS) based (Figure 1.2).

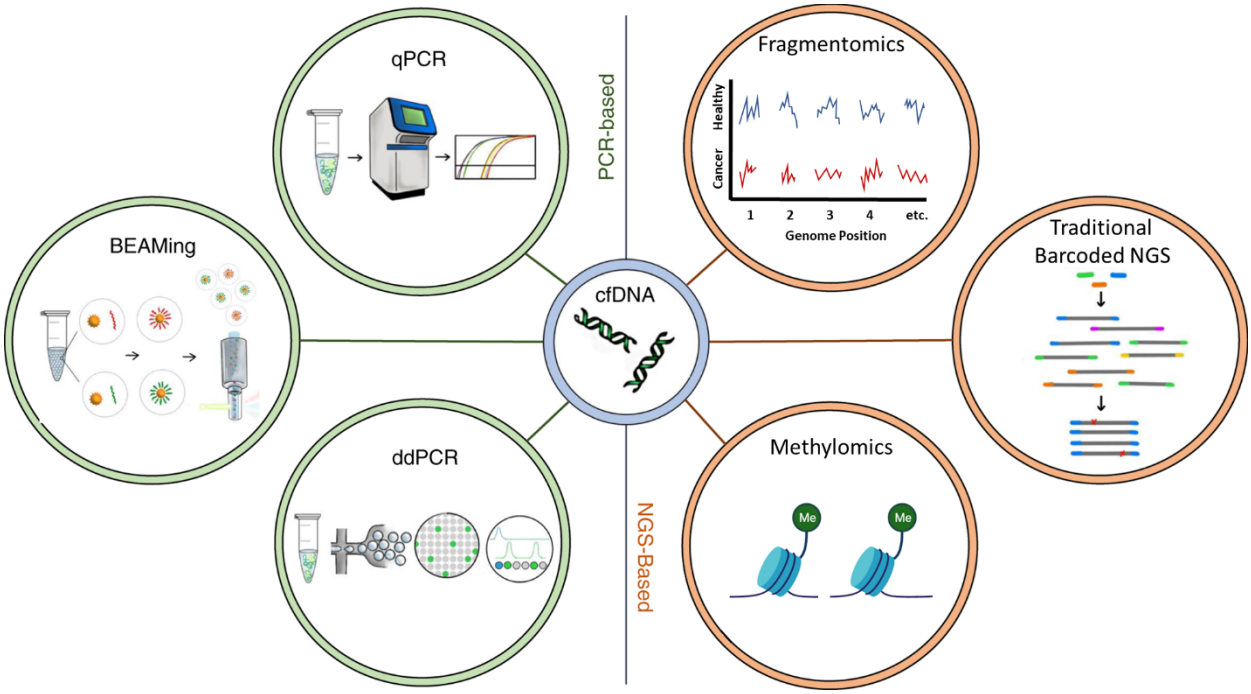


Figure 1.2. Techniques for the detection of ctDNA within the total pool of isolated cfDNA. Most ctDNA detection techniques can be divided into PCR and NGS based. PCR tests include qPCR, BEAMing, and ddPCR, all relying on the fluorescent detection of PCR-amplified DNA molecules of interest. NGS tests are numerous, but can be grouped into general categories that look for mutational changes (traditional barcoded NGS), methylation changes (methylomics), and fragmentation changes (fragmentomics).

Detection Methods: PCR-based

PCR-based techniques—including real-time quantitative PCR (qPCR), Beads, Emulsion, Amplifying and Magnetics (BEAMing) and droplet-digital PCR (ddPCR)—offer the simplest approach for the detection and quantification of ctDNA. PCR-based methods generally use target-specific probes that are designed to identify single-nucleotide variants. Thus, only known cancer-associated variants can be queried, and only a few variants can be probed at one time. Consequently, these techniques are typically applied to detect mutations that are highly prevalent or to track previously verified mutations.

qPCR is a rapid and cheap method that most often uses a quenched fluorescent probe, which is released during amplification and emits a detectable fluorescent signal. Although the specificity of qPCR is quite high, significant variation can occur between repeated runs[27], which has led to the development of new qPCR-based techniques to improve assay performance. However, these detection techniques have a relatively low threshold of mutation detection (0.1%) and cannot accurately detect extremely rare variants[28,29].

BEAMing is a digital PCR method that combines the techniques of PCR and flow cytometry[30]. Magnetic beads are tagged with bait DNA and mixed with oil, PCR reagents and target DNA to form an emulsion. Each oil droplet is capable of undergoing PCR-mediated amplification of the DNA that is bound to the bait DNA. The emulsions are then magnetically purified and opened to isolate the amplified target DNA, which is identified with hybridized fluorescent probes and enumerated by flow cytometry. BEAMing has been reported to detect 1 mutant DNA molecule among 10,000 normal molecules, which is 10× more sensitive than improved qPCR assays[31]. However, despite the improved sensitivity, BEAMing is laborious and expensive, and only a few laboratories and commercial entities are capable of efficiently employing the technology.

The most widely used of these PCR-based DNA detection techniques is ddPCR, which incorporates the methodology of both qPCR and BEAMing[32]. ddPCR first requires the creation of an emulsion of quenched fluorescent DNA probes, PCR components and sample DNA. Theoretically, the emulsion is diluted to contain ≤1 DNA molecule per oil droplet. PCR amplification removes the probe's quencher, and fluorescent droplets are quantified using flow cytometry, providing a ratio of droplets with on target DNA signal against a specific mutation. Compared to BEAMing and qPCR, ddPCR is a relatively inexpensive method and offers improved performance metrics[27,33,34].

Detection Methods: Next Generation Sequencing Based

The inner workings of next-generation sequencing (NGS) are beyond the scope of this review and have been discussed elsewhere[35-37]. However, NGS represents a massive improvement over purely PCR-based methods in that it allows for increased scalability and the detection of previously unknown mutations or other modifications. Generally, NGS can either be used to sequence many targeted sequences of interest at massive scale, or sequence across the entire genome. Within the context of ctDNA evaluation however, NGS typically inspects targeted intron/exon junctions and specific exonal regions of a subset of cancer-related genes to identify mutations. Although this approach can detect mutant alleles at frequencies as low as 0.1%, many sequencing platforms have a random error rate of the same frequency[38]. Therefore, standard versions of this technique might not be sufficient for analyzing rare, or low variant allele frequency (VAFs), ctDNA such as those in early cancer patients. However, advanced barcoding systems, improved hybrid capture techniques, and/or greater sequencing read depth have been implemented to minimize errors and maximize detection.

The first such technology to be developed was dubbed 'Safe-SeqS' for Safe-Sequencing System[39]. Additional iterations, including 'integrated Digital Error Suppression Cancer Personalized Profiling by deep sequencing' (iDES-CAPPseq)[40] and Targeted Error Correction Sequencing (TEC-seq)[20], among others, ensued. Each of these techniques utilize unique molecular identifiers to barcode individual cfDNA molecules and generate redundant sequences of the same barcoded molecule. These cfDNA molecules can then be consolidated to determine if any mutations found are bona fide mutations or whether they are due to sequencing errors, as true mutations should be present in multiple reads containing the same barcodes. The development of these technologies significantly improves the sensitivity of NGS-based

techniques while enabling the interrogation of thousands of genomic positions at once. Additionally, pharmaceutical and biotech companies are currently developing tests based on both the fragmentation of cfDNA/ctDNA as well as its methylation. Cancers are known to alter expression profiles on an epigenetic level through differential regulation of transcription factors, which can be visualized in cfDNA through altered DNA fragmentation, and through differential methylation, which can be visualized in cfDNA through altered methylation marks on released DNA[41-43]. NGS-based technologies such as DELFI (DNA evaluation of fragments for early interception) for fragmentation/fragmentomics and bisulfite sequencing for methylation can be used to test these differences in DNA isolated from human blood samples, and are gaining traction as they move through clinical trials although they are currently primarily experimental[44,45]. Although promising, the clinical validity of all of these NGS-based techniques for the detection of ctDNA still requires further evaluation from a technical standpoint, and their clinical utility—that is, whether these tests can change outcomes to help guide care for cancer patients—has yet to be demonstrated.

Current and Future Impacts of Circulating DNAs in the Cancer Clinic

The potential clinical utilities of ctDNA-based liquid biopsies have been extensively reviewed. Liquid biopsies will likely become an essential part of cancer treatment at all stages of disease[8]. At the current moment however, liquid biopsies are used primarily to detect potentially targetable mutations in late-stage disease, with some forays into early stage disease. Potential future utilities include the early detection of cancer and the detection of minimum residual disease (Figure 1.3). We will use primarily breast cancer in the following sections as a model disease to illustrate the utility of ctDNA.

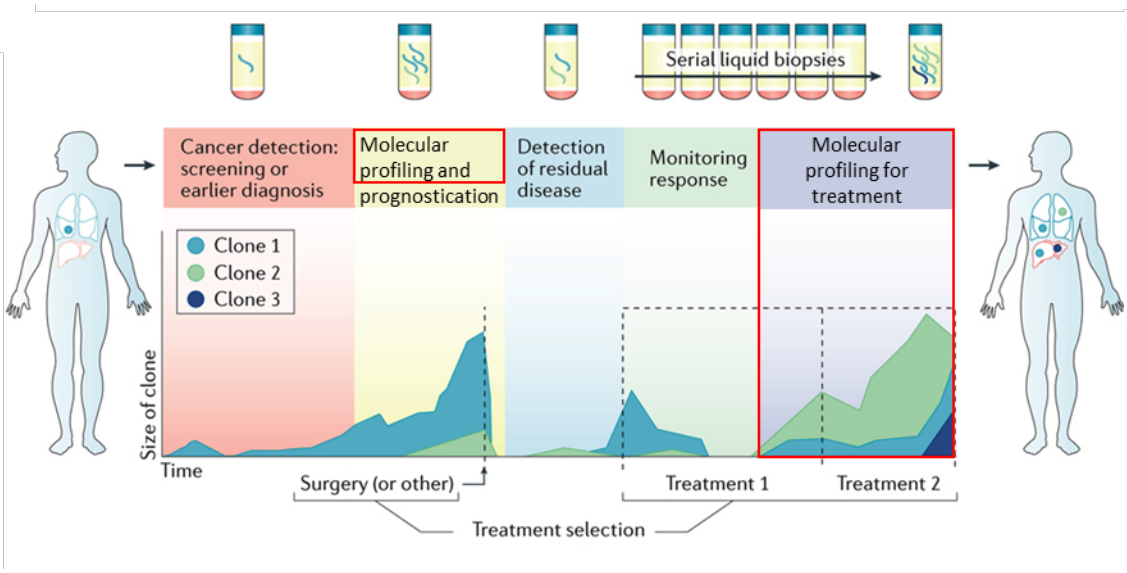


Figure 1.3. Current and Future utility of ctDNA-based liquid biopsies across a cancer patient's treatment course. Currently employed clinical uses are outlined in red. Adapted from Wan et al[8].

Current Clinical Utility: Molecular Profiling

As of 2024, most clinical guidelines suggest ctDNA-based liquid biopsy testing only in specific cancers and primarily in late-stage disease[46]. The purpose of this application is primarily to determine potentially targetable mutations to be paired with appropriate FDA approved

therapies. This modality was initially performed through single-gene companion diagnostics, such as the cobas® EGFR Mutation Test v2 and therascreen® PIK3CA RGQ PCR Kit to test for mutations in one specific gene to indicate one specific therapy. More recently, multi-gene panels were developed that can serve as companion diagnostics for multiple therapeutics across multiple cancers at once in in Guardant360® CDx and FoundationOne® Liquid CDx. To highlight one example, the cobas® EGFR Mutation Test v2 as well as both the Guardant360® and FoundationOne® tests are well validated in their ability to identify non-small cell lung cancer (NSCLC) patients who might be responsive to the mutation specific EGFR inhibitor Osimertinib. These multi-gene tests have greatly broadened the use of ctDNA in the clinic, as their ability to simultaneously detect multiple targetable mutations led to their widespread use as a way to enroll patients in target-specific clinical trials. For example, the phase III clinical trial of elacestrant in late stage breast cancer used ctDNA mutation detection to enroll patients with *ESR1* mutations and saw improved progression free survival (PFS) in these patients over the general population[47]. Falling into a similar category are basket, umbrella, and platform trials, which have begun enrolling patients into different treatment arms based on mutations detected in circulating DNA. An example is the plasmaMatch trial, which leveraged the ability of ctDNA testing to detect various mutations in a population of advanced breast cancer patients and determine what treatments might best target tumor-driving mutations. PlasmaMATCH specifically stratified patients based on *ESR1*, *HER2*, and *AKT* mutations, finding that targeted treatment in the *HER2* and *AKT* mutant populations lead to an unexpectedly high ~20% response rate[48]. Together, these trials indicate the benefit ctDNA based testing as a molecular diagnostic can, and are currently bringing, to advanced stage patients.

Further, many academic medical centers are performing multi-gene ctDNA testing simultaneously with traditional tumor biopsies at presentation as an FDA-approved 'complementary diagnostic.' This can be particularly useful as tumor biopsies sometimes fail to isolate evaluable tissue. One study of 161 patients saw that primary tumor results could be generated for only 124/161 patients, while ctDNA calls were evaluable in 151/161 patients[49], indicating the utility complementary testing can provide patients whose tumor biopsies are unanalyzable. Such complementary tests can also support the relevance of tumor biopsy-confirmed mutations, as well as potentially identify mutations contributed from metastatic sites.

Future Clinical Utility: Early Detection of Cancer

The use of ctDNA in detecting cancer at its earliest, pre-symptomatic, screening stage is being fervently studied. Most up-and-coming early detection tests are Multi-Cancer Early Detection (MCED) tests, attempting to detect any cancer from a simple blood test. Data analyzing viral DNA in EBV-driven nasopharyngeal carcinomas illustrate the difficulty in this field, suggesting the development of additional technology will be acquired to achieve the level of sensitivity and specificity necessary for a primary cancer screening test. In a prospectively enrolled clinical trial, qPCR was used to detect circulating Epstein Barr virus (EBV) DNA in the blood of asymptomatic patients before the onset of nasopharyngeal carcinoma[50]. Detection metrics were excellent in this study, with 97.1% sensitivity and 98.6% specificity. This power was due in large part to the high amount of target DNA expected to be present in each cancer cell, with 50 EBV genomes present on average in each cell which have 10 repeats of their test's target gene, suggesting that approximately 500 molecular markers would be required to achieve similar sensitivity and specificity in more traditional mutation-based assays.

To overcome this barrier, researchers have explored the inclusion of other analytes than simply ctDNA, such as protein biomarkers, into liquid biopsy. CancerSEEK is a technique developed with this approach in mind, utilizing 61 amplicons across 16 genes and 8 proteins[51]. For the detection of early-stage (stages I–III) ovarian and liver cancer, this technique achieved 99% sensitivity in a retrospective analysis. However, this technique didn't perform as successfully in all cancers, with only a 33% median sensitivity observed in breast cancers. In the prospective DETECT-A (a pan-cancer detection study performed in approximately 10,000 women), CancerSEEK in combination with PET-CT was able to increase sensitivity and specificity compared to CancerSEEK alone[52]. Importantly, the use of PET-CT allowed the investigators to identify/locate the tissue of origin from which the cancers arose. However, the results were highly variable depending on the cancer type, and cancers with standard of care screening modalities such as breast cancers were typically identified through this traditional mechanism before in blood.

Other techniques that seek to expand the amount of identifiable molecular markers in ctDNA rely on DNA methylation and fragmentation. Both targeted and untargeted studies have found that methylated gene promoters represent hundreds to thousands of changes in transformed cells and can serve as biomarkers for specific cancers[53,54]. For example, Shen et al. performed whole genome methylation analysis in a small cohort of early-stage breast cancer patients using the technique of cell-free Methylated DNA Immunoprecipitation sequencing (cfMeDIP-seq) to achieve an ~85% detection rate[55]. Currently, large clinical studies led by GRAIL Inc. (Illumina) are ongoing to determine the potential of their whole genome methylation-based MCED test for early detection across cancers. Preliminary reports indicate an overall sensitivity of 51.5% across stage I-III disease, increasing with cancer stage and again varying with cancer type[56].

Circulating DNA fragmentation has also been reported to be altered in cancer, with alterations in transcription factor footprinting and overall fragment length[57]. Based on these alterations Cristiano et al. developed an MCED named DELFI (DNA EvaLuation of Fragments for early Interception), reporting a test sensitivity of 57% to 99% across eight cancers in stage I-III disease with 98% specificity[44]. Together, these tests are potentially promising and could shift the diagnosis of cancer patients towards earlier stage disease especially in cancers without established screening tests. Unfortunately, serious concern about these MCED tests and the companies backing them has been raised by the liquid biopsy community due to poor results in many diseases. These poor results are particularly concerning as most data delineating the efficacy of MCED tests is currently retrospective and uses a case-control study design, which inflates detection values[58]. Much is yet to be done to determine the clinical relevance of these tests, but only the completion of ongoing prospective clinical trials will determine their utility.

Future Clinical Utility: Disease Prognostication

As well as investigating the use of liquid biopsy approaches for the detection of early disease, researchers have made strides towards predicting patient outcomes through ctDNA testing. This approach could stratify patients at high and low risk, allowing for additional treatment or de-escalation. The detection of mutations by ctDNA in the pre-treatment early-stage disease setting is currently the most explored pathway in this field. For example, Rothé et al. found that, in HER2+ breast cancers, a lack of detectable *PIK3CA* and *TP53* variants in patient plasma prior to neoadjuvant chemotherapy and subsequent surgery was associated with a high pathological complete response (pathCR) rate[59]. A separate study demonstrated that the prevalence of mutant *PIK3CA* ctDNA pre-surgery was associated with poor relapse-free survival and overall survival,

regardless of breast cancer subtype, corroborating the detection of ctDNA at presentation as a poor prognostic factor[60].

The prognosis of early-stage disease using ctDNA is not limited to the detection of single mutational variants in ctDNA; Amplifications, epigenetics, and a constellation of mutations have also been used to indicate outcomes. A small study of pan-breast cancer patients demonstrated that detection of the 1q21.3 amplification in cfDNA prior to neoadjuvant therapy was prognostic for relapse, with all 8 ctDNA positive patients relapsing within 3 years[61]. Additionally, studies involving ctDNA methylation determined that the detection of a specific methylated region pre-neoadjuvant chemotherapy was a poor prognostic factor, with >70% of patients with this marker relapsing within 5 years[62]. Although these studies focused on prognostication in the pre-treatment setting, other investigators have studied ctDNA after neoadjuvant therapy. Two similar studies showed that the detection of tumor variants by ddPCR after neoadjuvant therapy correlated with poorer disease free and overall survival in triple negative breast cancer[63,64]. McDonald et al. corroborated this result in a pan-breast cancer cohort using a patient-personalized next-generation sequencing panel[65]. While these studies indicate the power of ctDNA in prognosticating patient outcomes, no prospective clinical trial to date has validated the prognostic potential of ctDNA in early-stage disease. Reliable tests would be valuable here, as the identification of high-risk groups through ctDNA positivity either before or after therapy initiation could lead to beneficial treatment that could increase pathCRs.

Future Clinical Utility: Minimum Residual Disease Detection

The rate of breast cancer recurrence within 10 years of curative-intent therapy has been shown to be 36.8%[66]. The ability to accurately identify those patients in whom cancer will recur could lead to increased survival through earlier treatment and prevention. Preliminary studies by Beaver et

al. demonstrated that ddPCR could detect mutant *PIK3CA* ctDNA in some early-stage breast cancer patients after surgery as well as before[67]. In the 29 patient cohort, *PIK3CA* mutations were present in both tumor tissue and pre-surgery blood samples from 14 patients. Post-surgery, 5 of these 14 patients remained mutation-positive, as assessed by ddPCR, suggesting the presence of minimal residual disease (MRD), and raising the possibility that ctDNA abundance post-surgery could identify patients at risk of recurrence. Indeed, one patient with detectable ctDNA post-surgery had triple negative metaplastic breast cancer and died within two years of her initial diagnosis. In a more extensive study, Garcia-Murillas et al. used a similar ddPCR approach to track patient-specific ctDNA mutations in the plasma of patients with early-stage breast cancer[68]. In this study, serial post-surgery ctDNA liquid biopsy identified 12 of 15 relapses and correctly classified 96% of non-relapsing patients. In a similar retrospective study, Olsson et al. used ddPCR to demonstrate that levels of cancer-specific genetic rearrangements were associated with relapse[69]. Patient-specific panels of 4 to 6 rearrangements detected ctDNA retrospectively in 93% of patients in whom cancer recurred. Despite these positive results with ddPCR, this approach is limited by the number of genetic alterations that can be queried and the need for a priori knowledge of tumor mutations. NGS overcomes these limitations owing to its ability to query multiple loci in an unbiased way. Accordingly, studies have leveraged NGS-based approaches in conjunction with liquid biopsies for the detection of ctDNA as a surrogate for MRD. To show the strength of this method, Parsons et al. demonstrated that an NGS-based approach was capable of measuring 488 mutations with 100 times more power than a single mutation ddPCR test *in vitro*[70]. MRD was detected using a patient-specific version of this assay, and all patients in whom ctDNA was detected 1-year post treatment relapsed (6 of 6). However, many patients still relapsed within 10 years without the detection of ctDNA at 1 year, potentially

indicating the need for long-term follow up in MRD testing. In a similar approach by Coombes et al., 49 breast cancer patients were monitored using ultradeep sequencing of 16 variants[71]. ctDNA was detected in 16 of the 18 patients in whom recurrence occurred as early as 2 years prior to relapse over a 100-month follow-up period. Again however, multiple patients relapsed without ctDNA ever being detected. These studies indicate the potential for ctDNA to predict relapse but indicates that the frequency and duration of monitoring will be critical issues for prospective trials testing ctDNA as a prognostic marker for MRD and relapse. In addition, performance improvements need to be implemented to identify those patients who remain ctDNA negative but still relapse.

Future Clinical Utility: Treatment Monitoring

Serial plasma monitoring in patients with metastatic breast cancer gives clinicians the opportunity to quickly determine the response (if any) to current treatment. Multiple studies in metastatic breast cancer have demonstrated that a direct correlation exists between the levels of tumor-specific ctDNA and changes in diagnostic imaging seen in response to treatment. Although most of these studies primarily track the frequency of commonly mutated genes such as *PIK3CA* and *TP53*[72-74], other alterations have also been tracked and yielded similar results[61]. Furthermore, these studies have demonstrated a link between the persistence of mutations in the ctDNA during ongoing therapy and poor outcome. According to additional studies, ctDNA dynamics early on during treatment of late stage disease might also predict the clonal composition of the tumor upon progression. In the PALOMA-3 Study, a Phase 3 study of combination CDK4/6 inhibitor palbociclib in combination with fulvestrant, results suggested that a drop off in detection of *PIK3CA* mutations upon treatment initiation led to improved responses over time[75]. This observation was not related to the baseline level of PIK3CA ctDNA detection,

which was not predictive of outcome. Interestingly, the dynamics of ESR1 mutations on treatment were not indicative of treatment response, indicating that this paradigm of mutation tracking indicating the effectiveness of treatment might require optimization for various combinations of treatments and mutations. However, the ability to identify and track these variants early-on could significantly impact the treatment course in patients who do not respond on the ctDNA level. It is important to note, however, that such techniques are not yet useful for clinical management, as discordant responses between ctDNA and scans have been documented[76], possibly due in part, to tumor heterogeneity and the subclonal nature of metastatic cancers. For this method of surveillance to become clinically relevant, assay sensitivity, subclonal dynamics, and the emergence of resistance mutations need further study and development.

Mechanisms of Cell-Free DNA Release

In the past, most studies that have sought to improve ctDNA-based liquid biopsy have focused on improving the technologies around detection, from initial sample processing to sequencing workflows. However, the basic biogenesis and degradation of cell-free DNA is a poorly studied field full of conflicting results. In a world where ctDNA-based liquid biopsies are poised to change the face of the cancer clinic, increased knowledge about this field is currently a missed opportunity to improve detection. Here, I review the current state of the field of cell-free DNA release biology to indicate the need for increased discovery efforts in this field, especially in the light of bettering liquid biopsy tests. Major mechanisms of cell-free DNA release are highlighted in Figure 1.4.

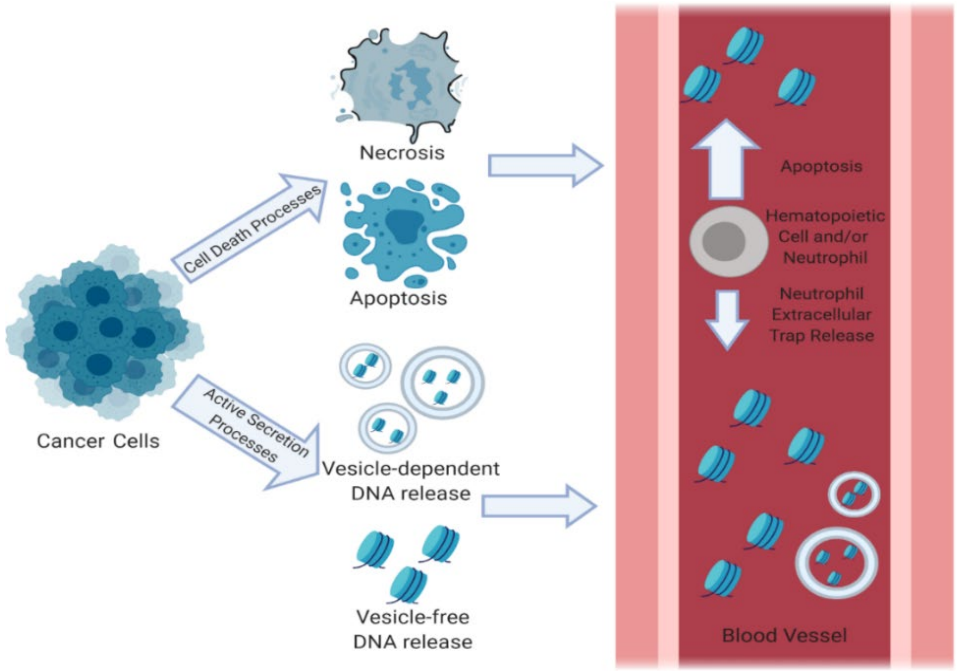


Figure 1.4. Major sources of circulating DNA. Cancer cells contribute to circulating DNA content through two major mechanisms: cell death and active secretion. While many forms of cell death might have a role in DNA release from cancer cells, the most explored are apoptosis and necrosis. Active secretion of many vesicle subtypes containing DNA as well as vesicle-free DNA may also contribute to circulating DNA. Most of the total cfDNA content is thought to come from the apoptotic death of hematopoietic cells. The extrusion of neutrophil nuclei as neutrophil extracellular traps also plays a role, but its relative impact is

Cell Death: Apoptosis and Necrosis

Apoptotic and necrotic cell death are two major mechanisms of cell-free DNA release reported in the literature, although other forms of cell death may play a role. Support for apoptosis as the major contributor to cfDNA release chiefly derives from fragmentation patterns in human blood. When testing both healthy control and cancer patient samples, the overwhelming pattern displayed is a major fragment at approximately 166bp with varying degrees of repeating fragments at 166bp increments[77-79]. This profile is characteristic of apoptosis as it implicates cleavage by Caspase Activated DNAse (CAD), an enzyme only activated by caspases during the early stages of apoptosis. CAD rapidly cleaves non-protein bound double-stranded DNA and 166bp is the approximate length of DNA capable of being bound by a single histone, protecting

the DNA from CAD[80,81]. The repeating fragments consist of partially processed chromatin linked by uncut DNA[82]. Further studies supporting apoptosis as a major mechanism of DNA release treated cells or mice with apoptosis inducing drugs and found increased cfDNA and ctDNA, especially at fragment sizes corresponding to CAD-cleaved apoptotic DNA[82-85]. This was first outlined in Jahr et al. upon treating mice with a FAS receptor agonistic antibody, resulting in an increase in cfDNA release into mouse blood at a fragmentation pattern consistent with apoptotic-released DNA[82]. Confoundingly, various studies have also shown a lack of correlation between apoptosis and cfDNA or ctDNA release[86-88]. In addition, ctDNA fragments are reported to be enriched at sizes significantly smaller and larger than the characteristic 166bp apoptotic peak. Together, these studies suggest that apoptosis is likely an important ctDNA release mechanism that leads to a major peak at 167bp, but it is not clear whether further processing of DNA or contribution of alternative pathways lead to these inconsistencies.

Necrosis has likewise been indicated as a source of cfDNA and ctDNA, and is distinguished by large fragments of >10kb. This was first shown by Jahr et al. upon treating mice with acetaminophen to induce liver necrosis, after which they found increased cfDNA in blood at extremely high sizes[82]. Some studies have also reported increases in DNA fragmentation size in cancer patients[78,89,90]. This observation is most easily explained by increases in necrosis, although these studies were observational and did not explore mechanism. Additionally, tumor necrosis has been correlated to ctDNA detection in a cohort of early stage lung cancer patients. These reports have led to an acknowledgement that necrosis may contribute to cfDNA and ctDNA content, however again confounding results exist. In xenografted mouse studies, some cell lines with less active necrosis *in vivo* release higher amounts of ctDNA than those with higher amounts of necrosis[85,91]. Additionally, the fragmentation pattern of DNA traditionally

found in human blood indicates that necrosis may not be a common factor, as large fragments are rarely seen. When these fragments are seen however, these large fragments are often called as contamination from poor blood handling leading to lysis of white blood cells, potentially hiding the role of necrosis.

Active Secretion: Internal or External to Vesicles?

Although multiple methods of active release of cfDNA have been described, most studies have focused on extracellular vesicles (EVs). EVs are broadly defined as lipid-bound particles released by cells. EVs have been classified into categories and subcategories differentiated by size, content, and mechanism of secretion. The broadest classifications are small EVs (sEVs) and large EVs (lEVs), where sEVs are less than 200nm in size and lEVs are greater than 200nm in size[92]. Vesicle subtypes within each of these categories have been shown to contain DNA. However, EV DNA content varies based on vesicle origin, model used, and isolation technique.

Exosomes are the most studied EV, defined by their formation as intralumenal vesicles inside multivesicular bodies (MVBs) and their size of <150nm[93]. Exosomes carry diverse cargo, including various DNAs, and are thought to be secreted by most cell types[93]. Specific molecular regulators can modulate the release of exosomes and their content, with Rab27a knockdown decreasing overall exosome release and senescence induction through H-ras^{V12} increasing exosomal DNA content[94-96]. Within the field of exosomal DNA there are many points of contention including DNA size and localization. Studies of exosomes released from various cell lines show broad DNA peaks across 100-2500bp, sharp peaks between 2000-5000bp and peaks greater than 10,000bp in size depending on the study[97-100]. In addition, studies of human plasma exosomes report both smaller fragments that mimic an apoptotic ladder pattern as well as sizes greater than 10,000bp[98,101]. While there is high variance, the mostly commonly displayed peak

across these studies lies between 2000-5000bp. Interestingly, these DNAs have been reported to be simultaneously internal and external or completely external depending on study. Initially these assays were carried out with DNAse treatment of intact exosomes, cleaving DNA on the surface of the vesicle while leaving intralumenal DNA intact. These studies primarily show either completely internal DNA or a mixture of internal and external DNA[97,98,102]. A subsequent study termed a "Reassessment of Exosome Composition" by Jeppessen et al., saw instead that no DNA was associated with exosomes, with both internal or external DNA being an artefact of improper vesicle isolation[103].

Many of these discrepancies might be attributable to different models of exosome release as well as exosome isolation. Mechanistic reports of DNA packaging and release have been undertaken to determine whether it is truly packaged into exosomes or whether it is simply associated with them (Figure 1.5). Yokoi et al. report that drug induced genomic instability drives an increase in the amount of exosomes released containing DNA[102]. They correlated this phenomenon with the formation of micronuclei. Early endosomes, MVBs, and autophagosomes were visualized as present at micronuclei through transmission electron microscopy. Exosomes are formed from invaginations of the early endosome, which itself can convert through endosome maturation into MVBs. Multivesicular bodies can then fuse with the plasma membrane to release their content, or fuse with autophagosomes to become amphisomes. Amphisomes can also fuse with the plasma membrane to release their content. This observational data suggests that DNA might be loaded directly into early endosomal vesicles or autophagosomes for later release in both an exosome-dependent or amphisome-dependent manner. However, using two exosome isolation methods, Jeppessen et al. never found DNA or DNA-associated proteins in exosomes[103]. DNA was only found in isolated non-vesicular fractions. Using multi-color structured illumination

microscopy, this group instead found that DNA was contained within amphisomes, but rarely colocalized with exosomes within the amphisome. Based on this, Jeppessen et al. posited that genomic DNA from collapsing micronuclei is taken up into autophagosomes which later fuse with MVBs to create amphisomes, with DNA internal to amphisomes but external to exosomes. This would lead to DNA release associated with but separate from exosomes upon amphisome fusion with the plasma membrane. While these studies agree on a potential amphisome dependent mechanism of DNA secretion, the presence and association of DNA with exosomes is debated. This discordance ultimately reflects a poor understanding of the mechanisms of active DNA release through endocytic pathways.

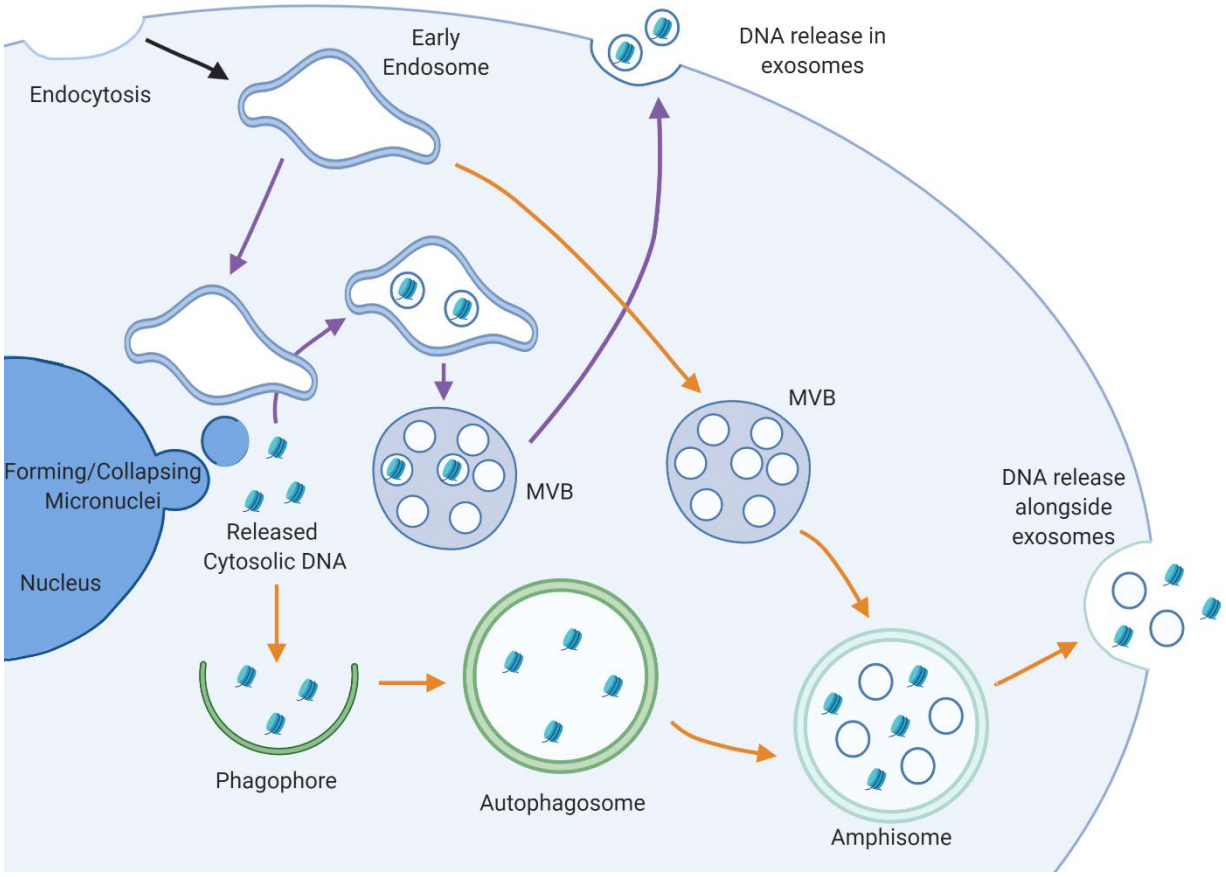


Figure 1.5. Potential mechanisms of DNA release through exosome-dependent (Purple, Yokoi et al.) and amphisome-dependent (Orange, Jeppessen et al.) pathways. Adapted from Jeppessen et al[99].

In addition, the overall contribution of "exosomal" DNA to total cfDNA and ctDNA content in humans is unclear. While Yokoi et al. delineate that less than 1% of exosomes from cancer patient blood are associated with DNA, Kahlert et al. and Fernando et al. posit that most cfDNA present in the blood is associated with exosomes[98,101,102]. Further adding to this controversy, Helmig et al. indicated that exercise induced increases in cfDNA without any change in exosomal DNA[104], and Wang et al. reported that ~50% of DNA in cancer cell conditioned media is associated with exosomes[87]. Finally, the commonly reported large size of DNA found in exosomes is in contrast with the pattern of DNA fragmentation in patient plasma. The myriad of conflicting data in this field indicates that the role of DNA release through sEVs and associated pathways is not well understood.

Larger vesicles and other processes may also contribute to active secretion of DNA from normal and cancer cells. Microvesicles are vesicles 0.2-1um in size which are shed through outward blebbing of plasma membranes and have been reported as associated with DNA[105]. Similarly, large oncosomes are vesicles between 1-10um in diameter derived specifically from cancer cells and have been reported associate with most released DNA in vitro and in vivo as large fragments exceeding 10Kbp[106]. Again, these sizes of DNA are often not tested in patient samples and are regarded as genomic DNA contamination. Together, these results indicate the potential for lEVs to carry DNA, but similar to sEVs the field is not well settled and few studies fully characterize the DNA present in a reliable way.

Unbiased genetic screening for phenotypes of interest in cancer

Functional genomic screens are powerful tools by which genotype-phenotype relationships can be directly interrogated on a large scale, potentially testing the effects of specific genes with actions ranging from drug resistance to exosome release. The earliest forms of these screens in mammalian cells relied on RNAi or cDNA libraries to respectively deplete or over-express genes of interest. However, technology progressed past these screening tools, as they are prone to off-target effects and are generally less versatile than the current workhorse of functional genomics, CRISPR-Cas9.

The CRISPR-Cas9 system consists of two components, the bacterial endonuclease Cas9 that creates double stranded breaks in DNA, and a single guide RNA (sgRNA) that guides Cas9 to its target sequence based on homology[107]. Upon inducing breakage, the DNA can either be subsequently repaired by non-homologous end joining (NHEJ) or homologous recombination (HR). In NHEJ, the original gene sequence is destroyed as this DNA repair mechanism repairs the sequence somewhat erroneously, often introducing insertions or deletions (indels) that lead to mRNA products that fall prey to nonsense mediate decay (NMD). In HR, a DNA template either supplied exogenously by the researcher or present endogenously as an un-edited DNA of similar sequence is used to faithfully repair the sequence. The rate at which these two processes happen differs across cell types and various cellular conditions[108]. This makes CRISPR-Cas9 a powerful tool for genome editing, as it can destroy the function of most genes when DNA is repaired by NHEJ, or can introduce new mutations into original sequences when supplied with exogenous template that contains a mutation of interest by HR.

In most CRISPR screens, a pool of sgRNAs is delivered through a lentiviral vector into a large pool Cas9-expressing cells at a low multiplicity of infection to achieve single gene knockout in

each cell. Cells which did not receive virus are removed using a selection agent such as puromycin. Then, a biological pressure or manipulation of interest is applied to spur the dropout or growth of specific knockout clones. NGS is used to amplify the sgRNA barcodes, which due to their lentiviral delivery have integrated into host genomes. Any skewing in favor of or away from specific barcodes can indicate their preferential positive or negative selection and indicate a gene of interest.

Commonly used CRISPR screen designs in oncologic research are outlined in Figure 1.6. Various CRISPR screen designs can yield genes that promote or inhibit growth at baseline, play roles in metastasis, control specific molecular phenotypes, promote or inhibit drug tolerance, and identify cellular rewiring due to mutations, among other uses[109]. For example, by culturing cells *in vitro* and/or *in vivo*, genes that drop out or are enriched over time can indicate powerful growth regulating genes (Fig 1.6A). One scheme often used us to identify hits in this type of screen is to screen across the entire genome *in vitro* first, then re-screen *in vivo* with a smaller pool of genes. Yamauchi et al. utilized this technique to identify 2,256 genes which were important to leukemia survival initially, further pruned their list through literature search to 470 genes, and re-screened in vivo to identify 130 high-confidence hits[110]. In addition, screens for metastasis make use of the injection of library-infected cells into the bloodstream or primary tumor locations and compare input barcodes to barcodes present at sites of metastasis to determine regulators of this process (Figure 1.6B). Scheidmann et al isolated circulating tumor cells from a human breast cancer patient, infected them with a lentiviral sgRNA library, and injected them into the mammary fat pad in order to identify genes that regulated various stages of metastasis[111]. They identified *PLK1* as a major regulator of intravasation of cancer cells from the breast into circulation and identified metastatic signatures using genes whose knockout

specifically encouraged metastasis into all organs collected as well as into specific organs. Further, CRISPR screens can be used to identify regulators of specific signaling pathways by combining CRISPR screening with reporting constructs. Fomicheva et al leveraged this technique to identify regulators of density-dependent cell-cycle arrest, a key property lost in solid cancers[112]. Using a FUCCI indicator, which causes cells to fluoresce red when arrested and green when cycling, the researchers serially selected cells by flow sorting which continued cycling even under confluent conditions and eventually identified *TRAF3* and generally non-canonical NF-kβ signaling as important regulators of density dependent proliferation. CRISPR screening can also be used to identify genes whose inactivation sensitizes cells to drug treatment (Figure 1.6D). Tiedt et al applied the combination of dabrafenib and trametinib to five library infected BRAF-mutant colorectal cancer cell lines and compared these screens to identical arms treated only with vehicle. Using this technique, they identified GRB2 as a gene shared across all 5 cell lines that was essential to resistance to this drug combination[113]. Finally, CRISPR screening can also be used to identify mutations which render cancer cells exquisitely vulnerable to perturbation of other specific processes through synthetic lethality (Figure 1.6E). For example, Gallo et al compared the representation of CRISPR barcodes between WT and CCNE1-high RPE1 cells after serial passages and identified *PKMYT1* as a gene specifically negatively selected in the CCNE1-high cells[114]. They then identified, characterized, and proposed the usage of an inhibitor of this gene as a treatment for CCNE1-amplified cancers. Many other strategies exist that leverage the power of CRISPR screening to identify mediators of cancer-related phenotypes, such as mechanisms controlling surface protein expression or CAR-T cell therapy escape[115,116], highlighting the power of this technique to identify diverse facets of molecular biology.

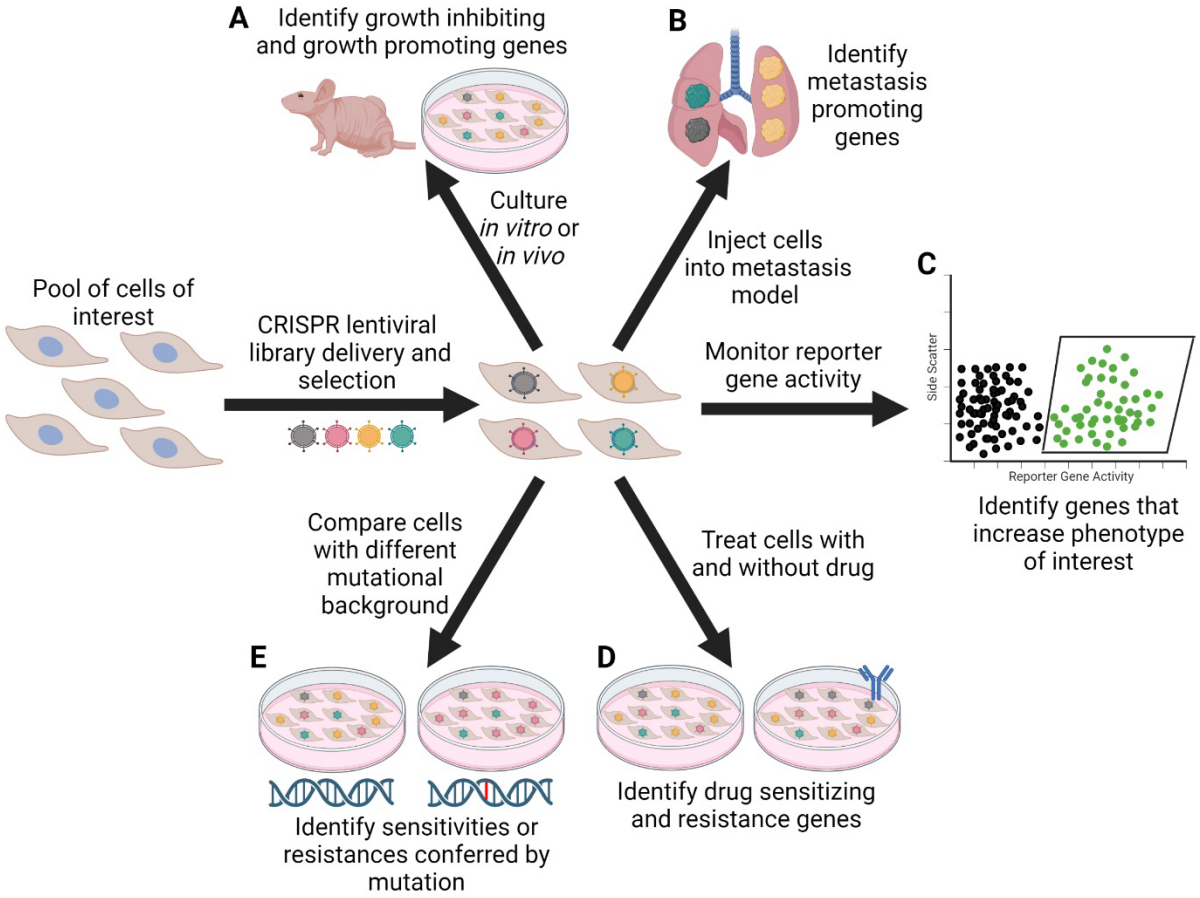


Figure 1.6. Common uses of CRISPR screening in Cancer Biology. Adapted from He et al[104]. Cells of interest are infected with various CRISPR lentiviral sgRNA libraries. Knockout takes place, aiming to knock out a single gene in a single cells across either a small subset of genes or the whole genome. After selection for cells that receive knockout, the cells are manipulated to determine phenotypes of interest. Allowing cells to grow without manipulation either in vitro or in vivo leads to the outgrowth of cells with growth-inhibiting gene knockouts and the depression of cells with growth-promoting genes knocked out (A). Injecting CRISPR cells into circulation or allowing them to metastasize from a primary tumor site to a far-away collection site selectes for cells with knockouts that antagonize metastasis (B). Cell lines with reporter backgrounds can be selected by flow-assisted cell sorting or other techniques based on background, and collected cells reveal genes that positively and negatively regulate the phenotype of the reporter (C). Comparing two cell populations allows researchers to determine growth inhibiting or promoting genes in specific circumstances. In cancer biology, this is most common when looking for genes that interact with drugs (D) or drugs that interact with mutations (E) to produce phenotypes not present in controls.

Summary and Research Objectives

Liquid biopsy is a promising technique that is already impacting clinical decision making by supplementing traditional tumor biopsy as a way to derive molecular information about various cancers. However, it has potential to provide much greater impact through avenues such as early detection and minimal disease detection. Most studies attempt to elevate liquid biopsy to the levels of sensitivity and specificity required to be used for these approaches by adding additional analytes or increasing the quality of sequencing performed as part of the assay. However, few have attempted to identify and modify regulators of the cfDNA life cycle to increase the presence of ctDNA and therefore improve liquid biopsy accuracy. We sought to address this gap in knowledge by determining the most relevant mechanisms of cell-free DNA release, of which many have been proposed, including apoptosis, necrosis, and active release.

In chapter II of this dissertation, I include the materials and methods used in this work. Chapters III and IV are data chapters primarily derived from work about to be published in *Communications Biology*. In Chapter III of this dissertation, we define a model system for the release of cell-free DNA and create and perform a novel screen for mediators of cell-free DNA release in two cell lines. In Chapter IV, we validate the hits identified in our screen using genetic and drug studies and determine that apoptotic processes are the major regulator of cell-free DNA release. Finally, in Chapter V, I discuss limitations, implications, and future directions that could be taken based on this work.

CHAPTER II

Sections of this chapter have been previously published in: Davidson, *et al*. 2024. An in vitro CRISPR screen of cell-free DNA identifies apoptosis as the primary mediator of cell-free DNA release. *Communications Biology*, provisionally accepted.

MATERIALS AND METHODS

Cell lines used

A complete list of cell lines used can be found in Table 2.1. MCF-10A, hTERT-IMEC, MCF-7, T-47D, BT474, ZR-75-1, MDA-MB-231, HCT116, and DLD1 cells were acquired from the American Type Culture Collection (ATCC). HepG2/C3A and HEK-293T cells were kindly provided by Dr. Emily Hodges (Vanderbilt University). CAL-51, Sum-185PE, MDA-MB-453, MDA-MB-468, HCC38, HCC70, HCC1143, HCC1937, and HCC1806 were provided by Dr. Brian Lehmann (Vanderbilt University Medical Center). A549, PC9, NCI-H841, NCI-H1607, and NCI-H2227 were provided by Dr. Christine Lovely (Vanderbilt University Medical Center). MCF-10A cells were grown in DMEM:F12 (1:1) (GIBCO) supplemented with 5% horse serum (Life Technologies), 20 ng/mL epidermal growth factor (EGF; Sigma-Aldrich), 10 μg/mL insulin (Life Technologies), 0.5 μg/mL hydrocortisone (Sigma-Aldrich), 0.1 μg/mL cholera toxin (Sigma-Aldrich), and 1% penicillin-streptomycin (PS; Life Technologies). hTERT-IMEC cells were grown in MCF-10A media with 2% charcoal dextran stripped fetal bovine serum (Life Technologies) in place of 5% horse serum. Sum185-PE cells were grown in MCF-10A media replacing horse serum for fetal bovine serum (FBS; Life Technologies), doubling the concentration of hydrocortisone and halving that of insulin. MCF-7, BT474, ZR-75-1, CAL-51, HepG2/C3A, HEK293T, HCT116, MDA-MB-231, MDA-MB-453, and MDA-MB-468 were grown in DMEM supplemented with 10% FBS and 1% PS. A549, T-47D, DLD-1, PC9, HCC38, HCC70, HCC1143, HCC1806, HCC1937, NCI-H841, NCI-H1607, and NCI-H2227 were all grown in RPMI supplemented with 10% FBS and 1% PS. The cell lines used were verified by STR profiling.

Table 2.1. Resource list for studies carried out in this thesis.

Reagent or Resource	Source	Identifier
Primary Antibodies		
Polyclonal Rabbit Lamin A/C	Cell Signaling Technology	2032
Monoclonal Rabbit FADD (EPR4415)	Abcam	ab108601
Polyclonal Mouse Sam68 Antibody (7-1)	Santa Cruz	sc-1238
Monoclonal Rabbit GAPDH Antibody (D16H11)	Cell Signaling Technology	5174
Polyclonal Rabbit a-Tubulin	Abcam	ab4074
Monoclonal Rabbit BCL-XL (54H6)	Cell Signaling Technology	2764
Secondary Antibodies		
Goat anti-Rabbit IgG (H+L) Alexa Fluor Plus 647	ThermoFisher	A32733
Goat anti-Rabbit IgG (H+L) Alexa Fluor Plus 488	ThermoFisher	A32731
Goat anti-Mouse IgG (H+L) Alexa Fluor 488	ThermoFisher	A11029
Chemicals, Peptides, and Recombinant Proteins		
Z-VAD-FMK	SelleckChem	S7023
TRAIL Protein, Recombinant Human	Millipore Sigma	GF092
Critical Commercial Assays		
Quant-iT dsDNA Assay Kit High Sensitivity	Invitrogen	Q33120
Quant-iT dsDNA Assay Kit Broad Range	Invitrogen	Q33130
Real-Time Glo Annexin V Apoptosis and Necrosis Assay	Promega	JA1012
Miscellaneous Commercial Kits		
QIAamp MinElute ccfDNA Midi Kit	Qiagen	55284
Quick-DNA Urine Kit	Zymo Research	D3061
D5000 ScreenTape Reagents	Agilent	5067-5589
D5000 ScreenTape	Agilent	5067-5588
Lenti-vpak packaging kit	Origene	TR30037
NE-PER Nuclear and Cytoplasmic Extration Reagents	Thermo Scientific	78833
Experimental Models: Cell Lines		
MCF-10A	ATCC	CRL-10317, CVCL_0598
MCF-7	ATCC	HTB-22, CVCL_0031
T-47D	ATCC	HTB-133, CVCL_0553
BT-474	ATCC	HTB-20, CVCL_0179
ZR-75-1	ATCC	CRL-1500, CVCL_0588
MDA-MB-231	ATCC	HTB-26, CVCL_0062
HCT116	ATCC	CCL-247, CVCL_0291
DLD-1	ATCC	CCL-221, CVCL_0248
HepG2/C3A	Dr. Emily Hodges	CRL10741, CVCL_1098
HEK-293T	Dr. Emily Hodges	CRL-3216, CVCL_0063
CAL-51	Dr. Brian Lehmann	CVCL_1110
Sum-185PE	Dr. Brian Lehmann	CVCL_5591
MDA-MB-453	Dr. Brian Lehmann	HTB-131, CVCL_0418
MDA-MB-468	Dr. Brian Lehmann	HTB-132, CVCL_0419
HCC38	Dr. Brian Lehmann	CRL-2314, CVCL_1267
HCC70	Dr. Brian Lehmann	CRL-2315, CVCL_1270
HCC1143	Dr. Brian Lehmann	CRL-2321, CVCL_1245
HCC1937	Dr. Brian Lehmann	CRL-2336, CVCL_0290
HCC1806	Dr. Brian Lehmann	CRL-2335, CVCL_1258
A549	Dr. Christine Lovly	CRM-CCL-185, CVCL_0023
PC9	Dr. Christine Lovly	CVCL_B260
NCI-H841	Dr. Christine Lovly	CRL-5845, CVCL_1595
NCI-H1607	Dr. Christine Lovly	CVCL_A467
NCI-H2227	Dr. Christine Lovly	CRL-5934, CVCL_1542
Oligonucleotides		
KHDRBS1 Sequencing Primer Forward 1: GTCATGGCTTCAGGTGAGGGTG	IDT	N/A
KHDRBS1 Sequencing Primer Reverse 1: CCTGGCCCAGCACTTAACATACA	IDT	N/A
KHDRBS1 Sequencing Primer Forward 2: GGTGCCATTTGACTTCAGAGAAGG	IDT	N/A
KHDRBS1 Sequencing Primer Reverse 2: CTGCCTTGGGTAGGTCTGGAGA	IDT	N/A
FADD Sequencing Primer Forward: CACCTCTGTCCACTCAGCAC	IDT	N/A
FADD Sequencing Primer Reverse: GCAGAACGCCACAGTGGTTGA	IDT	N/A
BCL2L1 Sequencing Primer Forward: CACAGCAGCAGTTTGGATG	IDT	N/A
BCL2L1 Sequencing Primer Reverse: CTCTGAAGCACAGGGTCAT	IDT	N/A
BCL2L1 RTPCR Primer Forward: ATTCAGTGACCTGACATCCC	IDT	N/A
BCL2L1 RTPCR Primer Reverse: TTTCCGACTGAAGAGTGAGC	IDT	N/A
PIK3CA E545K ddPCR Primer Forward: TCAAAGCAATTTGTACACGAGAT	IDT	N/A
PIK3CA E545K ddPCR Primer Reverse: ATTTTAGCACTTACCTGTGACT	IDT	N/A
PIK3CA E545K ddPCR Probe: FAM-tagged CTCTGAAATCACTAAGCAGGAGAAAGATTT	IDT	N/A
ERBB2 L755S ddPCR Primer Forward:CTGATGGGGAGAATGTGAAA	IDT	N/A
ERBB2 L755S ddPCR Primer Reverse: TCTAAGATTTCTTTGTTGGCTTTG	IDT	N/A
ERBB2 L755S ddPCR Probe: HEX-tagged CCATCAAAGTGTCGAGGGAAAACA	IDT	N/A
Recombinant DNA		
Plasmid: pLenti-FADD-mGFP-P2A-Puro	Origene	RC201805L4
Plasmid: pLenti-C-mGFP-P2A-Puro	Origene	PS100093
Plasmid: pLenti-Sam68-mGFP-P2A-Puro	Origene	RC200263L4
Lentiviral Library: Brunello Human CRISPR Knockout Pooled Library in lentiCRISPRv2	Addgene	73179-LV
sgRNAs		
Human KHDRBS1 sgRNA 1: ACCTGTCAAGCAGTATCCCA	IDT	
Human KHDRBS1 sgRNA 2: TGGCACCCCACGTCCACGAG	IDT	Hs.Cas9.KHDRBS1.1.AN
Human FADD sgRNA: TGACGTTAAATGCTGCACAC	IDT	Hs.Cas9.FADD.1.AB
Human BCL2L1 sgRNA 1: CAGGCGACGAGTTTGAACTG	IDT	
Human BCL2L1 sgRNA 2: GACCCCAGTTTACCCCATCC	IDT	
Software and Algorithms		
Graphpad Prism 9.5.0	GraphPad	https://www.graphpad.com/ scientific-software/prism/
MaGECK-VISPR v0.5.6	Git-Hub, Dr. Shirley Liu	https://bitbucket.org/liulab/mageck-vispr/src/master/
Rstudio 2022.02.1+461	Posit	https://posit.co/download/rstudio-desktop/
R v4.1.3	R-Project	https://www.r-project.org/
Data Availability		
CRISPR Screen Data	Dryad	doi:10.5061/dryad.k0p2ngfd2,

cfDNA release assays

cfDNA release assays were performed as described in Figure 2.1. Cell lines and their derivatives were plated in T75 plates at the following densities: 4 X 10^5 for MCF-10A, 7.5 X 10^5 for A549, HCT116, DLD1, MDA-MB-231, NCI-H841, and 1 X 10^6 for MDA-MB-468. 24 hrs after seeding, the media was replaced to a volume of 10mL in T75. After media change, cells were allowed to grow to ~85% confluency over the next 3 days. Upon reaching this density, media was collected and centrifuged to remove live cells, then dead cells and debris at 300 × g for 10 minutes and 2000 × g for 30 minutes, respectively. DNA was isolated from 8mLs of media from T75s through the QIAamp MinElute ccfDNA Midi Kit (Qiagen), used according to manufacturer protocols. All DNA was eluted in 25uL of provided deionized water, and concentrations were measured by fluorescence using the Quant-iT dsDNA Assay Kit High Sensitivity or Quant-iT dsDNA Assay Kit Broad Range (Invitrogen) on the GloMax Discover system (Promega). Simultaneously, cell counts present on the plate were measured by Vi-Cell BLU Cell Viability Analyzer (Beckmann-Coulter). DNA concentrations were divided by total cell concentrations to normalize the data for cell growth differences. In studies where there was a vehicle, control, or non-edited cell line, the data was normalized to these groups. Samples were analyzed for fragmentation analysis using D5000 ScreenTape on TapeStation 2200 or 4200 (Agilent) according to the manufacturer's specifications. All assays were performed in at least triplicate. All drug assays used the indicated concentration of TRAIL ligand (Millipore Sigma, GF092) or Z-VAD-FMK (Selleck Chemicals, S7203).

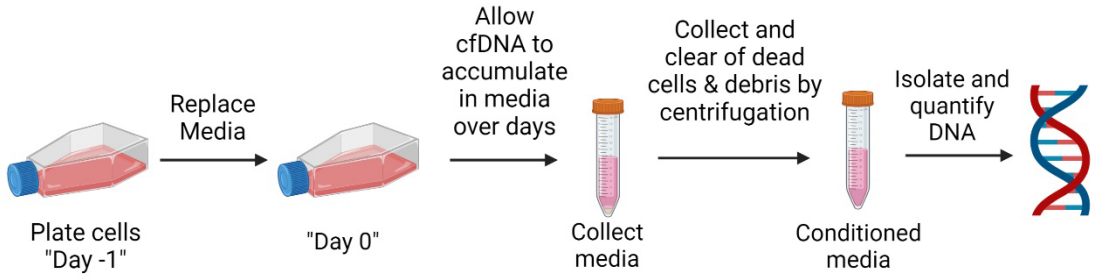


Figure 2.1. Methodology for cfDNA release assays. Cells were initially plated at cell densities described. Media is replace the following day to remove cells that did not attach. Cells are allowed to grow and release cfDNA over days, three days for most cell lines and experiments. At this timepoint media is collected and cells left on plate are counted. Collected media is cleared of dead cells and debris through serial centrifugation as described above. DNA is then isolated from conditioned media using the Qiagen MinElute ccfDNA Midi Kit, and quantified using QuantIT dsDNA assay kits. DNA release is normalized to cell counts individually for each plate.

DNA release time-course assays:

Cells were plated and assayed as described in the cfDNA Release Assays section. Analysis was performed on at least three replicates on days 1, 2, and 3. DNA was isolated from media and quantified as described in the cfDNA Release Assay section.

Released DNA degradation assay:

Cells were plated and assayed as described in the cfDNA Release Assays section, with the initial media replaced to a volume of 13mL. After incubation for 3 days, media was isolated as described above. 4 mL or media was taken for Day 0 and frozen down at -20ºC. The remaining 8 mL were placed back in the cell culture incubator in a 15mL Falcon tube and additional 4 mL collections were taken on days 3 and 7. Media was thawed simultaneously and isolated and quantified as described above.

cfDNA panel studies:

Cells were all plated at 4 X 10^5 cells per plate in three T75 cell culture flasks. The following day media was changed to remove nonadherent cells and debris. Three days later, media from all cell lines was collected, isolated, and quantified as described in the cfDNA release assays section. Fragmentation patterns were taken on an Agilent Tapestation 2200 or 4200. Representative fragmentation patterns were taken from initial panel samples or were products of assays optimized for higher seeding densities to maximize DNA release and detection.

CRISPR screening:

The Brunello Human CRISPR Knockout Pooled Library was purchased from Addgene as a lentivirus (#73179-LV). This library generally employs 4 guides per gene in the human genome, as well as 1,000 non-targeting guides. The workflow for these screens is delineated in Fig 3.8. One biological replicate was performed for each screen. First, a titering assay was performed with a small aliquot of the virus for each of cell line using either reverse infection (MCF-10A) or spinfection (A549) as outlined in the Broad Genome Perturbation Web Portal Protocols (https://portals.broadinstitute.org/gpp/public/resources/protocols). Cell lines were infected with the Brunello library at an MOI of 0.3-0.5 and a guide depth of 400X. Transduced cells were selected with puromycin (Thermo Fisher Scientific) for 3-5 days at 1 μg/mL (MCF-10A), 1.5 μg/mL (NCI-H41), or 2 μg/mL (A549). After selection, cells were maintained in culture for 28 days to eliminate any essential genes which might contaminate the pool of cfDNAs. At 25 days, cells were seeded to reach 90% confluency on day 3 and cfDNA and gDNA were extracted, respectively. At the end of each screen, at least 3×10^7 cells were collected to maintain 400X guide depth, and all media from each plate was collected. cfDNA was extracted from the media using the Quick-DNA Urine Kit (Zymo) and combined. gDNA was extracted using the QiaAMP DNA

Blood Maxi Kit (Qiagen). The sgRNA sequences were amplified and sequencing adapters were added, following the protocol outlined in the Broad Genome Perturbation Platform, using Phusion High-Fidelity PCR Master Mix (Thermo Scientific). Amplified samples were submitted and sent to the VANTAGE genomics core at Vanderbilt University Medical Center for sequencing. Analysis of read counts, β-score calculation, and p-values was performed through the MAGeCK-MLE algorithm, comparing the initial plasmid pool to the gDNA and cfDNA arms. This algorithm initially compares the presence of each gRNA barcode individually, then collapses guides against the same gene when providing β-scores, a measure of fold change. Genes were considered putative hits when the absolute value of the β-score difference between the cfDNA and gDNA portions of the screen were >0.5 for MCF-10A or >0.95 for A549 and when the gene was significantly selected ($P<0.05$ and $FDR<0.1$) in one arm of the screen but not the other or in different directions in each arm of the screen.

CRISPR gene knockout:

CRISPR gene knockout was performed by ribonuclear protein (RNP) transfection in the method recommended by Addgene (https://www.idtdna.com/pages/support/guides-and-protocols). Single-guide RNAs (sgRNA) were identified from initial Brunello CRISPR Knockout Pooled Library and commercially synthesized (IDT). Specific guides and primers used to sequence the regions where the guides cut can be found in Table 2.1. The RNP complex was assembled by incubating 1 μM Alt-R CRISPR-Cas9 sgRNA (IDT), 1 μM Alt-R S.P. HiFi Cas9 Nuclease V3 (IDT), and Cas9 PLUS Reagent (Invitrogen) with Opti-MEM (GIBCO) for 5 minutes at room temperature. Transfection complexes were formed by incubating assembled RNPs with CRISPRMAX transfection reagent in Opti-MEM for 20 minutes. Transfection complexes were plated first into 96-well plates followed by addition of cells such that the final concentration of

cells/well was 40,000 and the final concentration of RNP was 10 nM. Cells were incubated with transfection complexes in a tissue culture incubator for 48 hrs and were subsequently single cell diluted to create clonal populations. Selected clones were confirmed for targeted knockout by Sanger sequencing and immunoblot. Sanger sequencing was performed through Azenta Life Sciences. For CRISPR cell pools, single cell dilution was not performed and cells were allowed to grow to confluency, at which point protein was harvested and the cells were seeded for assay.

Immunoblot analysis:

Cells were seeded in respective normal growth media and harvested during passages for protein lysates. Cells were lysed in RIPA Lysis and Extraction Buffer (ThermoFisher, 89900) supplemented with cOmplete EDTA-free Protease Inhibitor Cocktail (Millipore Sigma, 04693159001) and PhosSTOP Phosphatase Inhibitor Cocktail (Millipore Sigma, PHOSS-RO) Tablets. Lysates were sonicated and protein concentrations were measured using the Microplate BCA Protein Assay Kit (Thermo Scientific, 23252). Samples were diluted and normalized in 4X NuPAGE LDS Sample Buffer (Invitrogen, NP0007) with 5% beta-mercaptoethanol (Aldrich) and were heated for 10 minutes at 70ºC. Protein lysates were then resolved by SDS-PAGE using NuPAGE 4 to 12 % Bis-Tris 1.0-1.5mm Mini Protein Gels and transferred onto PVDF membranes (Invitrogen, IB24002). After a 2-hour incubation at room temperature with 5% BSA in TBST blocking buffer, blots were incubated overnight at 4ºC in blocking buffer with primary antibody. Blots were washed three times in TBST before incubation with fluorescent secondary antibodies. Images were taken on the ChemiDoc MP Imaging System (BioRad). Cell fractionation was performed using the NE-PER Nuclear and Cytoplasmic Extraction Reagents in place of the above technique for cellular fractionation experiments (ThermoFisher, 78833). Antibodies used in these studies can be found in Table 2.1, and were diluted as follows: Lamin A/C (1:1000, Cell Signaling

Technologies, 2032), α-tubulin (1:1000, Abcam, ab4074), Sam68 (1:500, Santa Cruz, sc-1238), FADD (1:1000, Abcam, ab108601), GAPDH (1:1000, Cell Signaling Technologies, 5174), BCL-XL (1:1000, Cell Signaling Technologies, 2764), Goat anti-Rabbit IgG (H+L) Alexa Fluor Plus 647 (1:10,000 ThermoFisher, A32733), Goat anti-Rabbit IgG (H+L) Alexa Fluor Plus 488 (1:10,000, ThermoFisher, A32731), and Goat anti-Mouse IgG (H+L) Alexa Fluor 488 (1:10,000, ThermoFisher, A11029).

Stable over-expression and re-expression cell line generation:

Lentiviral expression vectors with CMV promoters driving GFP-tagged human Sam68 and FADD were purchased from Origene (PS100093, RC200263L4, RC201805L4). Lentiviral particles containing these vectors were isolated using the Lenti-vpak Lentiviral Packaging Kit (Origene, TR30037) as directed by the manufacturer. When ready to transduce, lentivirus was thawed rapidly at 37ºC. Cells were seeded 50,000 cells in 1mL maintenance media without any antibiotics into 6 well plates and reverse transduced with 500µL of virus per well. Control wells were seeded in the absence of virus. After 48 hours, cell lines were selected with puromycin at the following doses: MCF-10A .4µg/mL for selection and maintenance, all cells grown in DMEM were selected at 2µg/mL and maintained at 0.5 µg/mL, and all cells grown in RPMI were selected at 0.5µg/mL and maintained at 0.25µg/mL. After selection, cells were then flow sorted at the Vanderbilt Flow Cytometry Shared Resource on the FACSAria III (BD) for the top 1% of GFP expressors from the baseline cell pools and utilized in indicated experiments.

Cell growth assay:

Exponentially growing MCF-10A cells of each knockout genotype were plated at 2 X 10^3 cells per well in six-well plates. On indicated days, cells were counted using a Beckman Coulter Vi-Cell BLU Cell Viability Analyzer. All cell lines were counted in triplicate.

Baseline cell death assay:

Cells were plated at 2 x 10^4 cells per well in clear-bottomed white 96 well plates (Greiner-Bio One). The next day, the media was replaced with 100µL growth media, and 100µL 2X Detection Reagent prepared from the RealTime-Glo Annexin V Apoptosis and Necrosis Assay Kit (Promega). Plates were incubated for 24 hours and read for fluorescence and luminescence as directed by Promega on the Glomax Discovery system. Cells were then trypsinized and counted within each well using the Vi-Cell BLU Cell Viability Analyzer, and signal was normalized to the average concentration of cells in wells for each cell type. In studies where there was a vehicle, control, or non-edited cell line, the data was normalized to these groups. When used to profile the cell line panel, background resultant from the usage of different media types was subtracted prior to normalization to cell counts.

RT-PCR splicing assay:

Primers were designed that would simultaneously amplify BCL-XL and BCL-XS splice products of *BCL2L1.* Primer sequences can be found in Table S5. Control cells and Sam68 KO cells were plated at 300,000 cells per well in 6 well plates using standard MCF-10As growth media as described previously. Cells were allowed to grow to 80% confluence over 2 days. Cells were then harvested for RNA using the RNAeasy Mini Kit (Qiagen, 74104). Equivalent 1ug quantities of RNA were added into iScript cDNA Synthesis Kit (BioRad, 1708890) and converted to cDNA by

manufacturers protocols. PCRs were performed with primers described above using Phusion Hot Start II High-Fidelity Master Mix (ThermoFisher, F565L) with an annealing temperature of 63 degrees Celsius. Samples were then run on an agarose gel and visualized by UV light with GelRed (Biotium, 41003).

Digital droplet PCR:

Custom primers and probes were developed and used as indicated against the *PIK3CA* H1047R, *PIK3CA* E545K, and *ERBB2* L755S mutations. These sequences can be found in Table 2.1. For the double mutant *PIK3CA* E545K and *ERBB2* L755S assay, a primer/probe master mix was prepared, mixing stock primers and probes at 100uM each to a concentration of 18uM in 5uL per sample. For the single mutant *PIK3CA* H1047R assay, the master mix was prepared with primers instead at 50uM stock concentration diluted to 18uM in 5uL per sample. 5uL of this mix was combined with 45uL 1:2 diluted isolated cfDNA and 50uL ddPCR Supermix for Probes (No dUTP) (Biorad). This solution was distributed into cartridges and formed into droplets using the QX200 droplet generator (Biorad). PCR was performed according to the manufacturer's protocol, and results were read using the QX200 droplet reader (Biorad).

Statistics:

GraphPad Prism 9.5.0 (La Jolla, CA) was used to perform all comparisons of two groups through student's t-test or ANOVA for comparisons of three or more groups. For ANOVA, Dunnett's tests were used to compare the control group to all other groups or a Sidak's test to compare specific groups. Correlation analyses were performed in R (Posit) using the Pearson correlation. Outliers were removed from correlation analyses using the ROUT method at the most stringent Q-value of 0.1%. A *P* value of less than 0.05 was considered significant for all studies.

CHAPTER III

IDENTIFICATION OF MODULATORS OF CELL-FREE DNA RELEASE THROUGH A NOVEL CRISPR SCREENING MODALITY

Sections of this chapter have been previously published in: Davidson, *et al*. 2024. An in vitro CRISPR screen of cell-free DNA identifies apoptosis as the primary mediator of cell-free DNA release. *Communications Biology*, provisionally accepted.

Abstract

Few studies have described the release and degradation of cell-free DNA. Those that have often conflict in terms of the importance of certain mechanisms or characteristics of its release. By understanding the biogenesis of cell-free DNA (cfDNA), we can identify opportunities to increase DNA release from cancer cells and thereby increase the sensitivity of clinical ctDNA testing. Here, we perform an in-depth characterization of the properties of DNA released from cell lines *in vitro.* We initially profile the release and degradation dynamics of DNA in a small 6 cell line panel, then identify the capacity of varying cell lines to release DNA across 24 cell lines. We find that different cell lines release highly variable DNA release capacities and display variable fragmentation sizes of released DNA. We see that these fragmentation patterns generally fall into two groups, being either "left skewed" towards smaller fragments or "right skewed" towards larger fragments, typically indicative of apoptotic or necrotic/vesicular DNA respectively. These properties were cell-line intrinsic, and likely not derived from culture conditions. Further, we developed a novel CRISPR screen known as cfCRISPR that leverages the presence of CRISPR barcodes in non-nuclear DNA populations and applied this screening to identify the most relevant molecular mediators of cfDNA release in MCF-10A and A549 cell lines. Our results identify apoptotic pathway genes, including those in the TRAIL extrinsic apoptotic pathway and BCL-2 family genes, as major regulators of cfDNA release among other genes.

Introduction

Most ongoing efforts to improve the sensitivity and specificity of ctDNA-based liquid biopsy have focused on adding more analytes such as protein or RNA, increasing sequencing depth and breadth, or decreasing sequencing errors[40,51,117,118]. While these efforts show exciting progress, few studies have leveraged the basic biogenesis, degradation, and elimination of total cell-free DNA (cfDNA) and cancer-specific ctDNA to improve testing accuracy. In part, this is due to the relative lack of knowledge regarding the life cycle of ctDNA and factors that mediate its release. Past observational human studies demonstrate that the release of cfDNA and ctDNA is variable. cfDNA concentrations in blood differ significantly among individuals, cycle throughout the day, and are modified by physiologic conditions such as exercise and inflammation[20,119-121]. ctDNA levels also range from undetectable to extremely high variant allele fractions relative to total cfDNA depending on cancer type, stage, and other unknown factors[20,122]. Moreover, though the half-life of ctDNA has been widely cited as 1 to 2 hours, this originates from a single study of one colorectal cancer patient who had serial measurements before and after surgical removal of the primary tumor[123].

Apoptosis, necrosis, and active release through vesicular pathways are the most supported mechanisms of cfDNA and ctDNA release, but all are debated in the literature[124-126]. Apoptosis has long been assumed to be the primary mechanism of cfDNA release from cells based in part on the fragmentation pattern of cfDNA in blood. cfDNA from both healthy controls and cancer patients is primarily found at ~167bp in length with a repeating ladder pattern, which has been attributed to caspase-dependent cleavage[41,82,127,128]. While the role of apoptosis has been implicated in multiple *in vitro* and *in vivo* studies[82,88,129], other reports show a lack of correlation between apoptosis and cfDNA release[86-88]. In addition, ctDNA fragments are reported to be

enriched at sizes significantly smaller and larger than this peak, suggesting they are further acted on or released through additional pathways[78,130,131]. Still other studies have proposed necrosis as a major source of cfDNA release corresponding instead to DNA fragments around 10,000bp in size[82,88]. One such study showed that necrosis was correlated with ctDNA detection in a retrospective analysis of lung cancer patients[82,88,132]. Finally, vesicles released from cells have been reported to be associated with cfDNA. While most studies observe vesicle-associated DNA is >1000bp, studies have reached differing conclusions about the DNA content of various vesicle populations[94,97,98,101,102,106,133,134]. In contrast, in 2019 Jeppesen *et al.* showed small extracellular vesicles do not contain DNA, and previous results may have been due to incomplete vesicle purification[103]. One potential contributor to these discrepancies is the lack of rigorous model validation and the use of different models across studies. In addition, different systems for the isolation, purification, and quantitation of cell-free DNA have make it difficult to compare across studies. For example, various in vitro studies report discrepant DNA release over time in culture, all using different cell lines and methods of DNA isolation and detection[86-88,129]. These uncertainties regarding the origins of cfDNA underscore the need for additional research to gain new insights that could be translated for clinical care. Here in this chapter, I profile cfDNA release across many cancer cell lines, including quantity of release, rate of degradation, and fragmentation pattern, as well as perform a novel CRISPR screening modality, cfCRISPR, to identify specific molecular mediators of cfDNA release in two cell lines to address these issues.

Results:

A panel of human cell lines reveals convergent cfDNA release kinetics and divergent fragmentation patterns

To accurately analyze mechanisms of cfDNA release *in vitro*, we initially assessed six human cell lines to develop a standardized cfDNA assay. Cells were grown without media changes for the indicated days, and designated replicates harvested for cfDNA at days 1, 2 and 3. Our isolation methodology specifically retains most large and small vesicular populations, allowing us to analyze their effects on cfDNA release. Despite reported discordance in cfDNA quantitation over time[87,88,129], all cell lines demonstrated increases in the quantity of cfDNA over 3 days (Fig. 3.1A). Furthermore, in contrast to previously reported large shifts in fragmentation patterns, in some cases moving from small 167bp fragments to large >1500bp fragments over time[86], we primarily observed an increase in the concentration of fragments at the same size that was most prevalent at previous time points (Fig. 3.1B). To characterize the contribution of DNA degradation in our system, media containing cfDNA was collected at day 3, incubated in new flasks without cells, and profiled 0, 3, and 7 days after removal. The half-life of cfDNA for all cell lines was approximately 3 days, with smaller fragments appearing over the course of the assay (Fig. 3.1C, Fig. 3.1D). These results suggest that DNA is degrading *in vitro* over several days and suggests that small fragments may be derived from degradation of larger cfDNA products after cellular release.

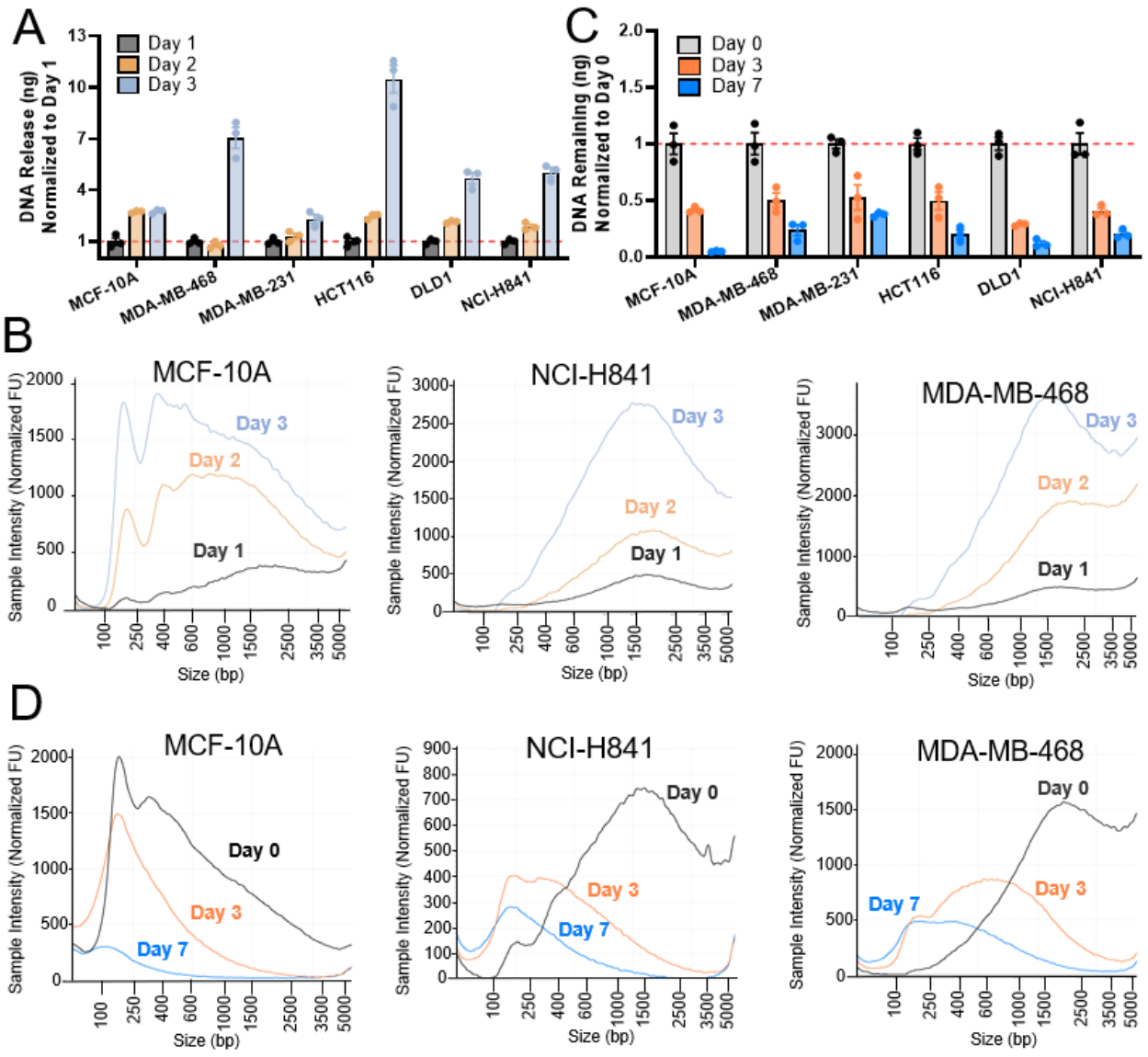


Figure 3.1. Release and degradation of cell-free DNA from a panel of cell lines. (**A**) Release of cfDNA in vitro over time. Data represent mean fold change ± SEM in absolute DNA release from Day 1 for each cell line. (**B**) Electropherograms of samples from (A), assessing changes in cfDNA release with increased cell incubation periods, were individually run at least n=3 times and representative traces are shown. (**C**) Degradation of in vitro cfDNA over time at physiologic temperatures. Data represent mean fold change ± SEM in absolute DNA quantity from Day 0 for each cell line. (**D**) Electropherograms of samples from (C), assessing the degradation of cfDNA in culture media after removal from cells, were individually run at least n=3 times and representative traces are shown.

cfDNA release properties were then assessed in a larger panel of 24 human cell lines. This panel included nontumorigenic (n=2), hepatocellular cancer (n=1), colorectal cancer (n=2), lung cancer (n=5), and breast cancer cell lines (n=14). After normalizing for variations in proliferation, there was a striking difference in the amount of cfDNA released across cancer cell lines (Fig. 3.2A). Interestingly, these trends were also observed when comparing the absolute quantities of DNA released from cells (Fig. 3.2B). Examination of cfDNA fragmentation patterns across the cell line panel revealed two distinct fragmentation patterns. Cell lines demonstrated either "left-skewed" cfDNA with major peaks at around 167bp reminiscent of patterns derived from human blood samples, or "right-skewed" cfDNA with largest peaks at sizes greater than 1000bp (Fig. 3.3). Although media conditions varied across cell lines, differences in cfDNA release did not correlate with media differences and the switching of types of serum in culture media did not lead to large changes in fragmentation pattern (Fig. 3.4). We also find that various serum types do not confound our experiments since they contain a dearth of DNA, as multiple of our cell lines were found to have almost no DNA present in their serum-enriched media (Fig. 3.1). The classification of each cell line into left- and right-skewed groups can be found in Table 3.1, and overlaid fragmentation patterns of left- vs. right-skewed cell lines can be found in Figure 3.5A. Interestingly, right-skewed cell lines show overall greater cfDNA release capacity but do not show increased proliferation or cell death (Figs 3.5B, 3.5C, 3.5D, 3.5E). Furthermore, using the Cancer Cell Line Encyclopedia and Genentech databases[135,136], the intrinsic expression of DNases across the cell lines was found to be generally very low and did not correlate with cfDNA release quantity or fragmentation pattern type (Table 3.2, 3.3). Using RNA-seq from multiple TCGA cohorts, we confirmed that expression of DNases in human tumor tissue is similarly low (Figs. 3.6A, 3.6B, 3.6C). Thus, local DNase expression in tumors does not account for the difference in fragmentation patterns between

patient cfDNA and our *in vitro* data (Fig. 3.6D, 3.6E, Table 3.2, 3.3). Taken together, these results reveal a striking cell-intrinsic diversity of cfDNA release.

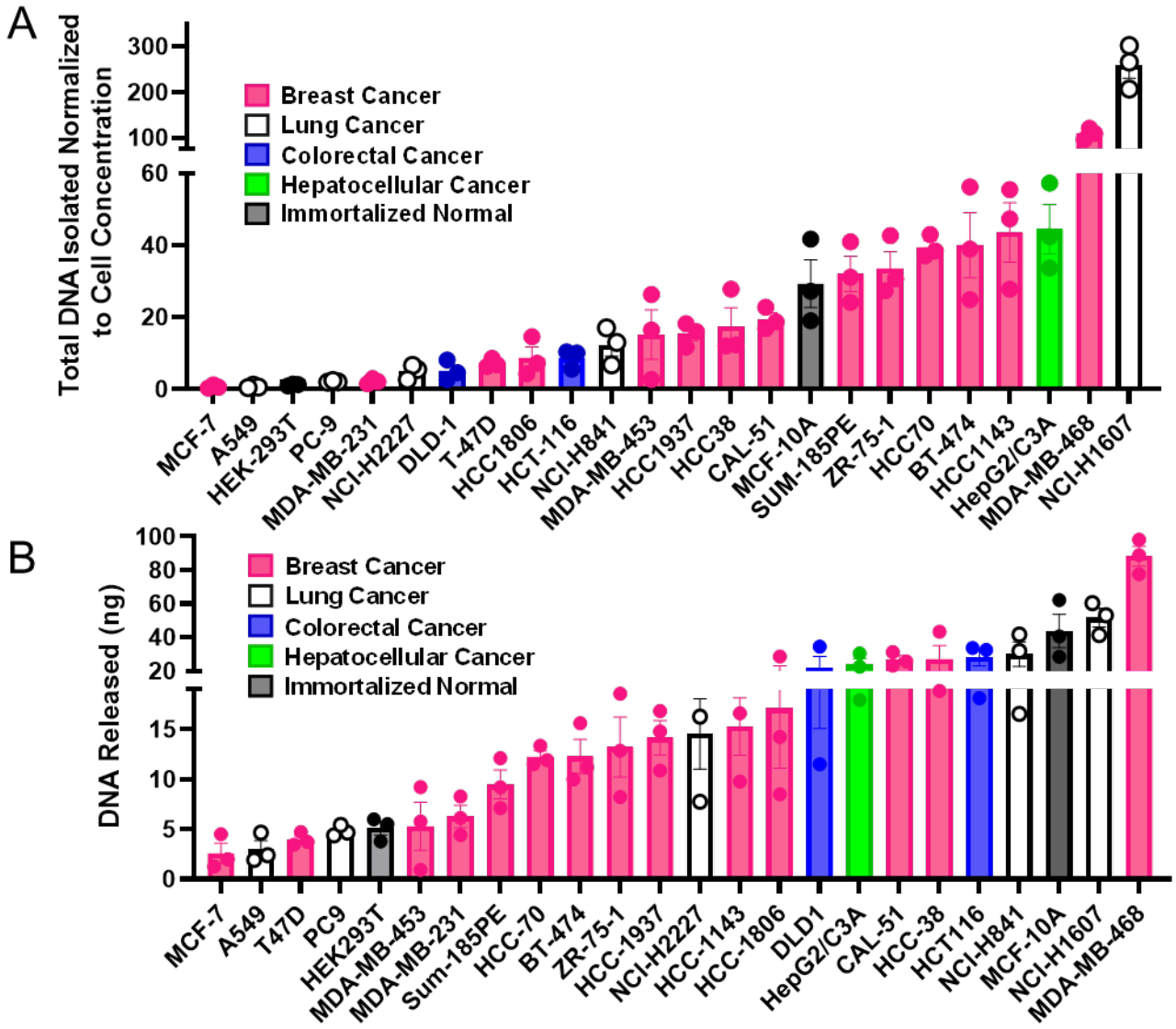


Figure 3.2. Release of quantity of cell-free DNA across a 24-cell line panel. (A) Quantification of cfDNA release from cell lines in culture. Data represent mean fold change ± SEM in DNA release normalized to cell concentration at time of collection for each cell line. (B) Quantification of raw values of cfDNA release from cell lines in culture. Data represent mean fold change ± SEM in absolute DNA release.

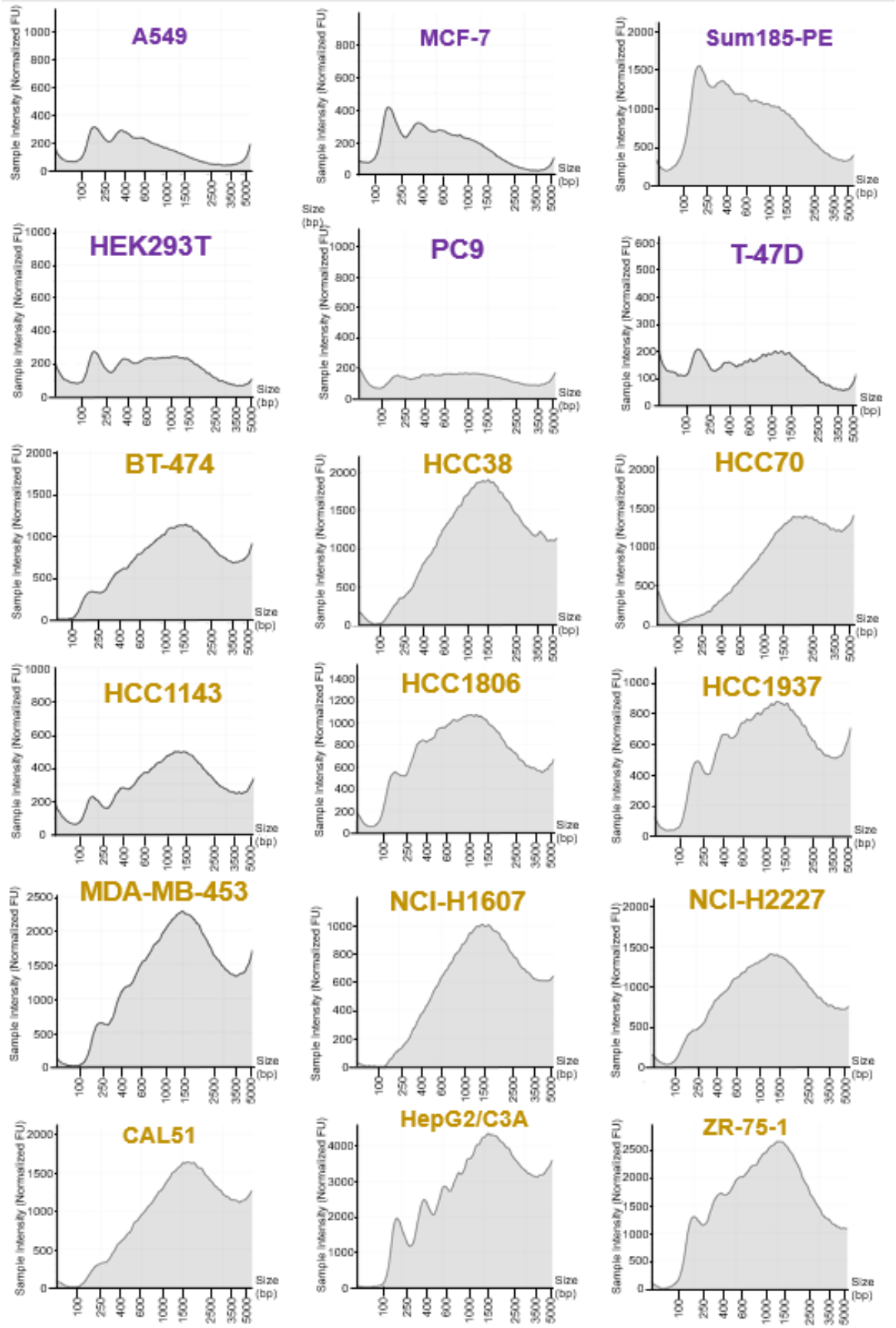

A549
MCF-7
Sum185-PE
HEK293T
PC9
T-47D
BT-474
HCC38
HCC70
HCC1143
HCC1806
HCC1937
MDA-MB-453
NCI-H1607
NCI-H2227
CAL51
HepG2/C3A
ZR-75-1
Sample Intensity (Normalized FU)
Size (bp)

Figure 3.3. Fragmentation of cell-free DNA across a 24-cell line panel. Fragmentation patterns of selected cell lines from the cfDNA panel. Cell lines with purple text are representative of cell lines with a left skew, whereas cell lines with yellow text are representative of lines with a right skew.

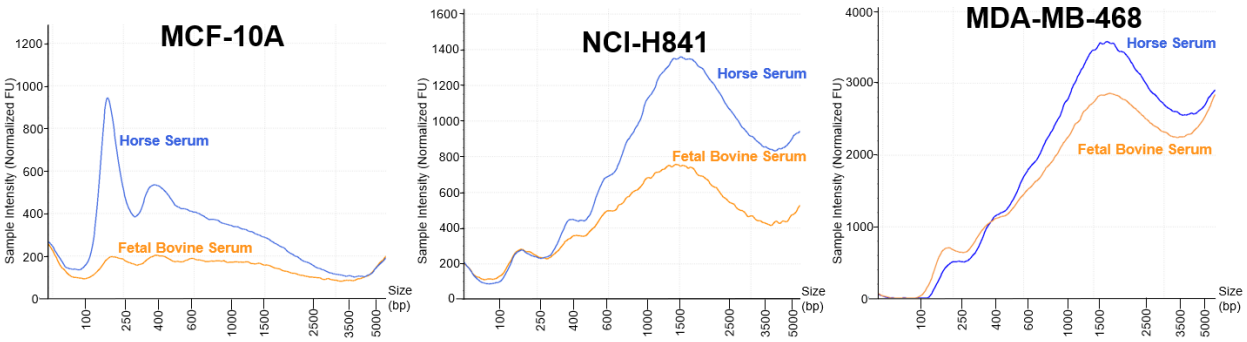


Figure 3.4. Fragmentation of cell-free DNA in response to serum switching in media. (C) Fragmentation patterns of cell lines treated with different serum conditions. Blue traces represent horse serum treatment, whereas orange represent fetal bovine serum. Electropherograms were individually run at least n=3 times.

Table 3.1. Cell lines from cfDNA panel in Figure 1E categorized by left (largest peak at ~167bp) or right skewing (largest peak at >1000bp) cfDNA release pattern.

Cell Line	Skew
A549	Left
HEK	Left
MCF-10A	Left
MCF7	Left
MDAMB231	Left
PC9	Left
Sum185	Left
T47D	Left
BT474	Right
Cal51	Right
DLD1	Right
HCC1143	Right
HCC1806	Right
HCC1937	Right
HCC38	Right
HCC70	Right
HCT116	Right
HepG2/C3A	Right
MDAMB453	Right
MDAMB468	Right
NCI-H1607	Right
NCI-H2227	Right
NCI-H841	Right
ZR75-1	Right

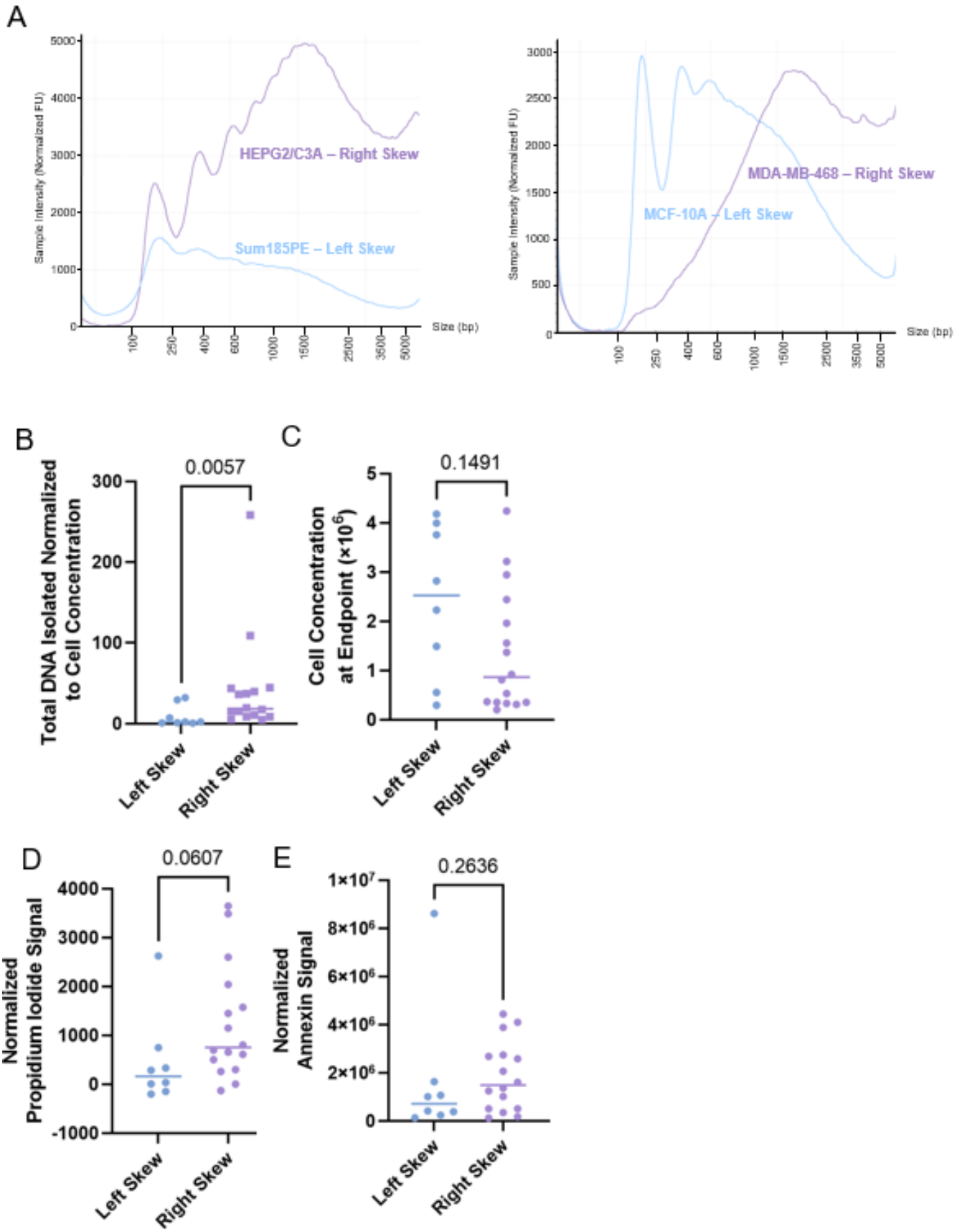

A
Sample Intensity (Normalized FU)
HEPG2/C3A – Right Skew
Sum185PE – Left Skew
Size (bp)
MDA-MB-468 – Right Skew
MCF-10A – Left Skew
B
Total DNA Isolated Normalized to Cell Concentration
0.0057
Left Skew
Right Skew
C
Cell Concentration at Endpoint (×10^6)
0.1491
D
Normalized Propidium Iodide Signal
0.0607
E
Normalized Annexin Signal
0.2636

Figure 3.5. Comparison of cell lines with left and right skewing of cell-free DNA. (**A**) Fragmentation patterns of example left-skewing and right-skewing cell lines overlapped to indicate their different fragmentation pattern. In the left panel, left-skewed Sum-185PE and right-skewed HEPG2/C3A are reproduced as an overlay from Figure 1F. In the right panel, example fragmentation patterns from left-skewed MCF-10A is overlapped with right-skewed MDA-MB-468. (**B)** Quantification of total DNA release from left-skewed and right-skewed cell lines. Data represent mean DNA release normalized to cell concentration from Figure 1E (n=3 for each cell line). Bar line is median, and Mann-Whitney U test was run due the large discrepancy in total number of left-skewed vs. right-skewed cell lines (8 vs. 16). **(C)** Quantification of proliferation from left-skewed and right-skewed cell lines. Data represent mean cell concentration after seeding at equivalent starting concentrations used to normalized cfDNA release in Figure 1E (n=3 for each cell line). Bar line is median, and Mann-Whitney U test was run due the large discrepancy in total number of left-skewed vs. right-skewed cell lines. **(D)** Quantification of Annexin V signal from left-skewed and right-skewed cell lines. Data represent mean Annexin V normalized to cell concentration (n=3 for each cell line). Bar line is median, and Mann-Whitney U test was run due the large discrepancy in total number of left-skewed vs. right-skewed cell lines (8 vs. 16). **(E)** Quantification of Propidium Iodide signal from left-skewed and right-skewed cell lines. Data represent mean Annexin V normalized to cell concentration (n=3 for each cell line). Bar line is median, and Mann-Whitney U test was run due the large discrepancy in total number of left-skewed vs. right-skewed cell lines (8 vs. 16).

Table 3.2. Correlation of expression (TPM) from CCLE of DNASE family members with cell free DNA release and skew in various cell lines.
Outlier refers to MDA-MB-468 Absolute and Normalized cfDNA Release (ROUT method, Q=.1%).

Cell Line	DNASE1	DNASE2	DNASE2B	DNASE1L1	DNASE1L2	DNASE1L3	SERPINB1	DFFA	DFFB	ENDOG	GAPDH	Absolute cfDNA Release (ng)	Normalized cfDNA Release (ng normalized to cell counts)	Skew
A549	2	39	0.1	12	0.7		90	28	4	4	4394	3.019934633	0.755928569	Left
MCF-7	10	178		30	10		50	38	8	23	2938	2.619988753	0.697733356	Left
PC-9	4	89		8	2		38	32	4	10	3870	4.84573802	2.177859784	Left
T-47D	14	106		11	6		7	24	5	17	916	3.983076596	7.241957448	Left
ZR-75-1	27	59		20	9	0.4	18	19	9	20	4212	12.27538128	33.6311816	Left
BT-474	10	56	3	8	2		8	29	8	20	2280	13.2140569	40.04259668	Right
CAL-51	11	86		13	3	0.1	84	53	11	10	2777	26.69795177	19.48755604	Right
HCC-1143	13	0.9		14	4	0.1	61	37	10	3	1561	15.26194486	43.60555673	Right
HCC-1937	9	182		10	3	0.1	43	37	4	22	4201	14.1421484	15.37190043	Right
HCC-38	7	87		28	3		86	27	3	13	1395	27.11418837	17.43677709	Right
HCC-70	6	140	0.2	34	4		163	29	7	6	2441	12.24975254	39.51533077	Right
HCT116	4	36	0.6	4	0.7		80	72	23	5	3811	28.05450656	8.726129569	Right
MDA-MB-231	3	43		13	0.8	0.1	135	34	6	13	3904	6.276845511	2.225831741	Right
MDA-MB-453	10	101	2	72	4		14	26	15	16	2942	5.329388922	15.22682549	Right
MDA-MB-468	6	207	0.1	25	2	0.1	143	25	8	18	2218	88.19726179	108.8855084	Right
NCI-H2227	21	30	1	35	4		25	36	15	21	1434	14.51336673	4.928138107	Right
NCI-H841	4	109		23	4		15	63	13	19	1443	30.0278627	12.30650111	Right
Expression vs. Absolute cfDNA Release														
R-Squared Value	0.02286	0.1608		0.0004436	0.06263		0.1616	0.017	0.02461	0.0063	0.05641			
R-Squared Value without Outlier	0.00451	0.01947		0.0273	0.0548		0.00449	0.469	0.1956	0.0142	0.08538			
P-Value	0.5624	0.1107		0.9361	0.3326		0.1097	0.616	0.5476	0.7625	0.3587			
P-Value without Outlier	0.8047	0.6063		0.5409	0.3829		0.8052	0.003	0.0863	0.6604	0.2721			
Expression vs. Normalized cfDNA Release														
R-Value	0.00411	0.1296		0.003744	0.005		0.1418	0.073	0.00149	0.0019	0.04926			
R-Value without Outlier	0.1106	0.02281		0.0003556	0.01545		0.002197	0.039	3.3E-05	0.0237	0.05187			
P-Value	0.8069	0.1558		0.8155	0.7874		0.1363	0.295	0.8831	0.8698	0.3919			
P-Value without Outlier	0.2082	0.5766		0.9447	0.6465		0.8631	0.462	0.9832	0.569	0.3963			
Left vs. Right Skew t-test														
P-value	0.5116	0.6563		0.4217	0.3928		0.8088	0.41	0.0997	0.8409	0.5957			

Table 3.3. Correlation of expression (TPM) from Genentech dataset of DNASE family members with cell free DNA release and skew in various cell lines.
Outlier refers to MDA-MB-468 absolute and normalized cfDNA Release (ROUT method, Q=.1%)

Cell Line	DNASE1	DNASE2	DNASE2B	DNASE1L1	DNASE1L2	DNASE1L3	SERPINB1	DFFA	DFFB	ENDOG	GAPDH	Absolute cfDNA Release (ng)	Normalized cfDNA Release (ng normalized to cell counts)	Skew
A549	4	34		17	1	0.1	69	41	5	23	4844	3.019934633	0.755928569	Left
MCF-7	9	86	0.1	21	5		34	43	7	19	4060	2.619988753	0.697733356	Left
T-47D	5	139		11	0.8		9	37	2	27	1890	3.983076596	7.241957448	Left
MCF-10A	2	82		23	0.6		72	47	11	12	5185	43.76018449	29.36925133	Left
CAL-51	7	62		21	4		48	79	11	38	4728	26.69795177	19.48755604	Right
HCC-38	2	48		19	0.7		81	30	2	6	2502	27.11418837	17.43677709	Right
HCC-70	3	116		34	2		102	31	5	9	3574	12.24975254	39.51533077	Right
HCT116	4	43	0.2	6	1		31	92	18	18	4007	28.05450656	8.726129569	Right
HCC-1143	3	0.4	0.3	18	0.4		87	45	4	4	3522	15.26194486	43.60555673	Right
MDA-MB-453	9	166	3	87	3		42	42	12	17	3244	5.329388922	15.22682549	Right
MDA_MB-468	7	209		44	3		122	34	9	32	2879	88.19726179	108.8855084	Right
ZR-75-1	6	137	0.9	12	0.9		25	25	4	8	6344	12.27538128	33.6311816	Right
DLD1	3	57		13	2		163	62	13	11	5509	21.82381981	5.147127314	Right
Expression vs. Absolute cfDNA Release														
R-Value	0.004488	0.1289		0.009851	0.00245		0.2049	0.0032	0.0816	0.0858	0.0097			
R-Value without Outlier	0.3052	0.1249		0.06512	0.08142		0.07643	0.1843	0.1892	0.0209	0.0668			
P-Value	0.8279	0.2283		0.747	0.8724		0.1204	0.8541	0.3441	0.3315	0.7489			
P-Value without Outlier	0.0625	0.2598		0.4234	0.3686		0.3844	0.1637	0.1576	0.6543	0.4175			
Expression vs. Normalized cfDNA Release														
R-Value	0.003844	0.2748		0.07636	0.001129		0.1515	0.0999	0.0054	0.019	0.0361			
R-Value without Outlier	0.1082	0.00013		0.007845	0.1195		0.01989	0.0911	0.0705	0.2674	0.0089			
P-Value	0.8405	0.0659		0.3608	0.9132		0.1886	0.2929	0.8113	0.6532	0.5343			
P-Value without Outlier	0.2964	0.9723		0.7843	0.2711		0.662	0.3404	0.4041	0.0851	0.7708			
Left vs. Right Skew t-test														
P-value	0.944	0.8356		0.4441	0.967		0.2365	0.5831	0.4315	0.51	0.9613			

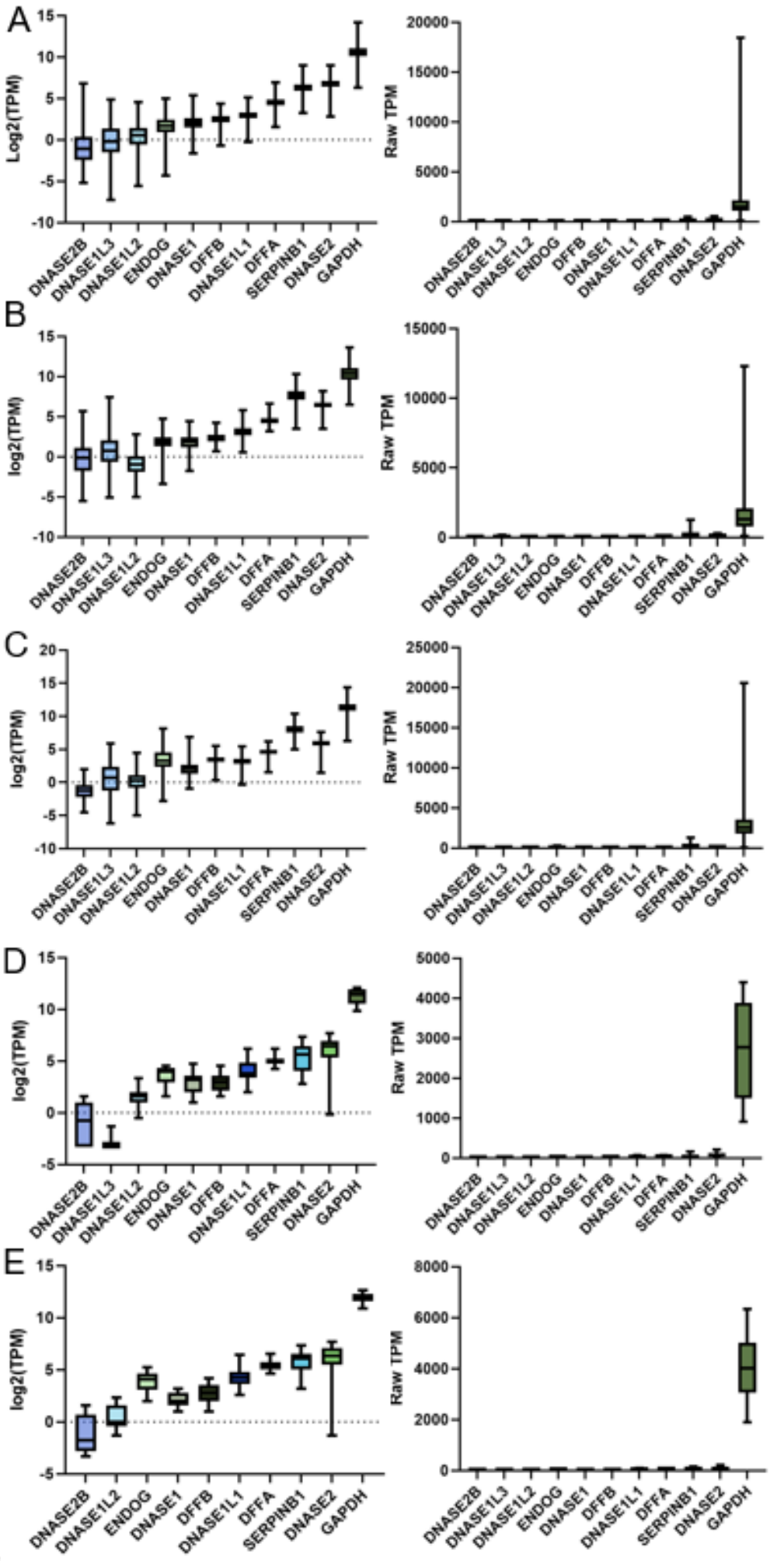

A
Log2(TPM)
Raw TPM
B
log2(TPM)
Raw TPM
C
log2(TPM)
Raw TPM
D
log2(TPM)
Raw TPM
E
log2(TPM)
Raw TPM
DNASE2B
DNASE1L3
DNASE1L2
ENDOG
DNASE1
DFFB
DNASE1L1
DFFA
SERPINB1
DNASE2
GAPDH

Figure 3.6. Expression of DNAses by TCGA and cell line cohorts. (A) Raw TPM and Log2 transformed TPMs of the expression of DNAses in the TCGA breast cancer cohort (n=1230). **(B)** Raw TPM and Log2 transformed TPMs of the expression of DNAses in the TCGA lung adenocarcinoma cohort (n=600) **(C)** Raw TPM and Log2 transformed TPMs of the expression of DNAses in the TCGA colorectal cancer cohort (n=522). **(D**) Raw TPM and Log2 transformed TPMs of the expression of DNAses in the Cancer Cell Line atlas including all available cell lines used in this study (n=6-17) **(E)** Raw TPM and Log2 transformed TPMs of the expression of DNAses in the Genetech RNA-seq database including all available cell lines used in this study (n=1-13).

cfCRISPR is a genome-wide cfDNA CRISPR-Cas9 screen that identifies putative modulators of cfDNA release.

To identify regulators of cfDNA release, a novel CRISPR screening strategy was developed. MCF-10A was initially utilized for this technique due to its release of high amounts of relative and absolute cfDNA compared to other profiled cell lines (Figure 3.2), as well as being a mostly diploid cell line. Additionally, MCF-10A cells display a left-skewed cfDNA fragmentation pattern reminiscent of that found in cfDNA from human plasma samples, including healthy controls, healthy controls with spiked-in cancer cell line cfDNA, and human cancer patients[79] (Fig. 3.7A). We reasoned that if cfDNA is shed equivalently across a cell's genome, then an individual integrated lentiviral sgRNA in that cell's genome would similarly be equally shed as cfDNA into the media. If a gene knock-out affected the rate of cfDNA release, then the relative ratio of cfDNA to cellular genomic DNA (gDNA) for that particular sgRNA lentiviral vector would be skewed. To confirm cfDNA is shed equivalently across a cell's genome, representative loci in MCF-10A and MCF-7 cfDNA were quantified using ddPCR. Equivalent representation of heterozygous mutations in *PIK3CA* and *ERBB2* previously knocked-in to these lines relative to their known genomic copy number[137] were detected for both cell lines (Fig. 3.7B), indicating that various genomic loci are likely evenly represented in cfDNA.

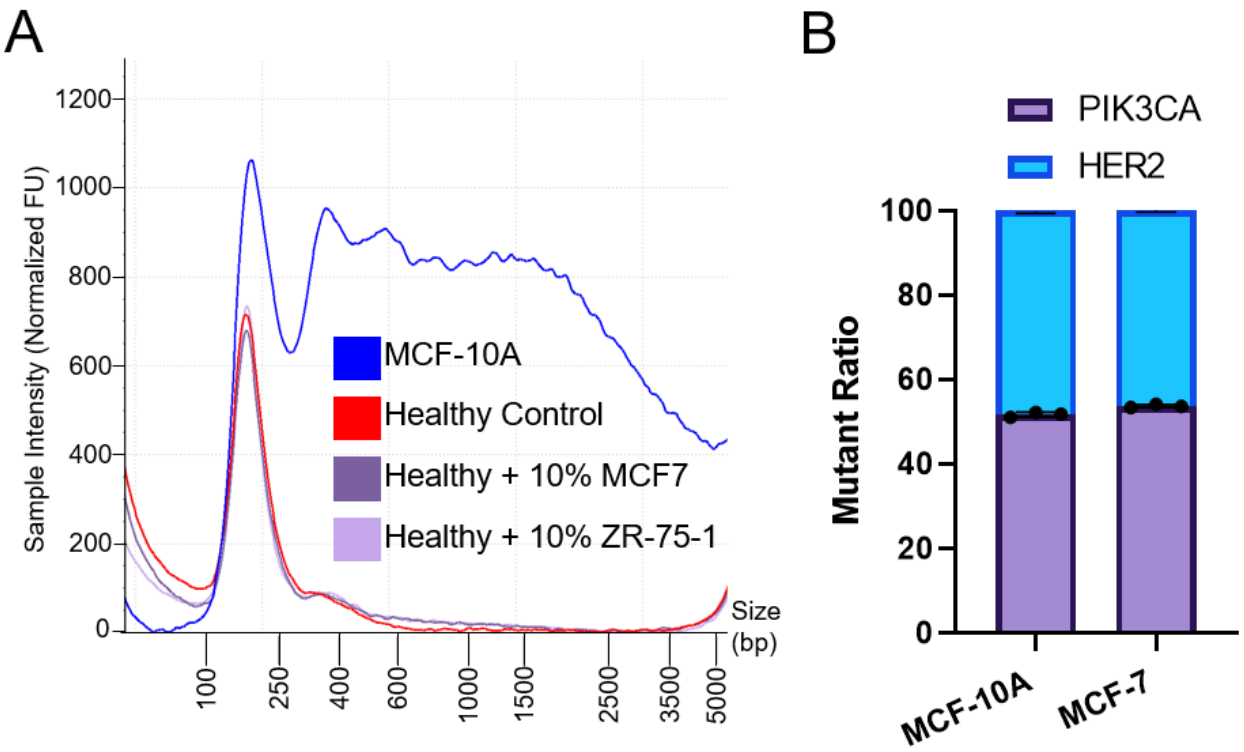


Figure 3.7. Fragmentation of cell-free DNA in response to serum switching in media. (C) Fragmentation patterns of cell lines treated with different serum conditions. Blue traces represent horse serum treatment, whereas orange represent fetal bovine serum. Electropherograms were individually run at least n=3 times.

For the first cfCRISPR screen, MCF-10A cells were infected with the Brunello Human CRISPR Knockout Pooled Library platform[138,139]. This library contains 76,441 sgRNAs targeting 19,114 genes. Infection led to a polyclonal population of cells that each contain 1 lentiviral particle and therefore one gene knockout. After puromycin selecting to remove cells that did not receive lentivirus, cells were passaged for 28 days prior to analysis to ensure the removal of essential genes to prevent false positives. At this timepoint, gDNA and cfDNA were isolated from polyclonal pool of cells adherent to the plate and the culture media, respectively. The relative representation of each gene's barcodes was compared between the starting plasmid library and each endpoint DNA (gDNA or cfDNA) population using the MAGeCK-MLE algorithm[140] (Fig. 3.8). An output

of this algorithm is the β-score, a representation of fold change between the starting plasmid library and either the endpoint gDNA or cfDNA arms of the screen. By identifying genes which were discordant in their β-score between the gDNA and cfDNA arms of the screen, we were able to identify genes which might regulate cfDNA release (Fig. 3.9A). For example, a gene whose knockout causes an increase in cell growth will be more represented in the gDNA population and therefore yield a high gDNA β-score. However, if that same gene's cfDNA β-score is negative, the knockout causes an under-representation in the cfDNA pool. The discrepancy between representation of this gene in the gDNA and cfDNA arms of the screen makes it a candidate modulator of cfDNA release. Genes which do not display a discrepancy between their cfDNA and gDNA β-score can be found along the Z-axis in Figure 3.9A and are not likely candidate regulators of cfDNA release.

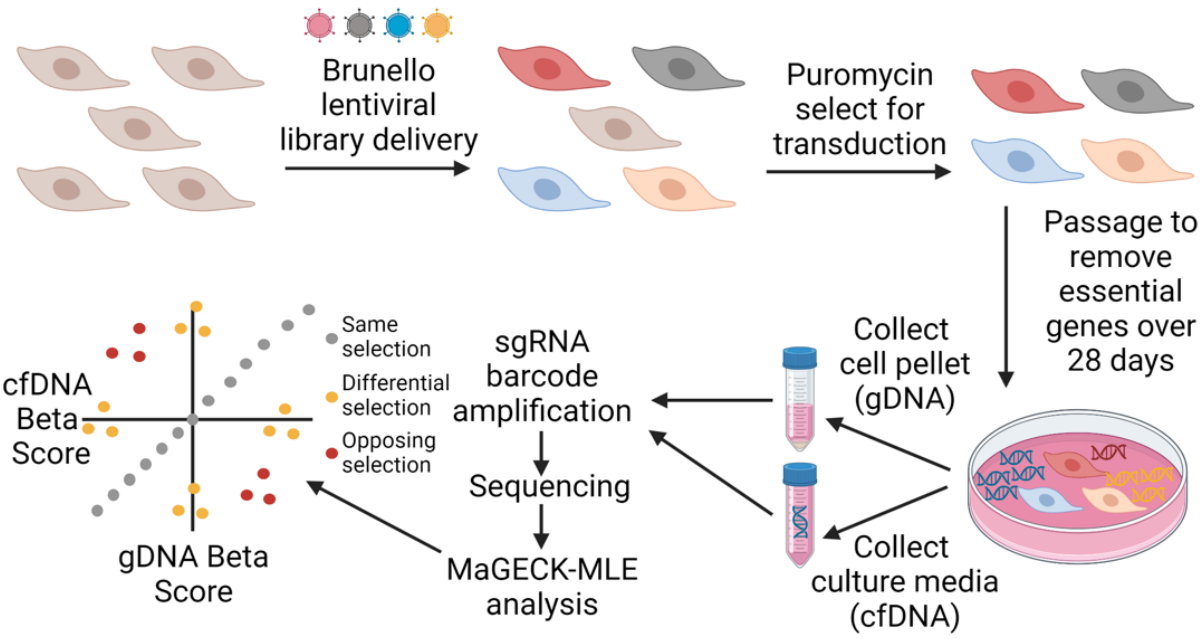


Figure 3.8. A CRISPR Screen for mediators of cell-free DNA Release. Cells are infected with the Brunello lentiviral sgRNA library at a low MOI (Multiplicity of Infection). Cells were then selected with puromycin to ensure infection in present cells. Cells were passaged for 28 days, then cfDNA and gDNA were collected and compared by MaGECK-MLE analysis of sequenced of PCR-amplified sgRNA barcode regions.

Putative genes of interest with β-scores discrepant between the cfDNA and gDNA arms of the screen were grouped based on literature-defined categories (Fig 3.9A, Table 3.4). Many of the hits are well-known members of the TRAIL extrinsic-apoptotic cell-death pathway, including *TNFRSF10A*, *TNFRSF10B*, *CASP8*, and *FADD*. Multiple RNA binding, putative RNA binding, or proteins known to interact with these proteins were also identified, including *KHDRBS1*, *RBMX2*, and *CELF1*. The remaining hits could not be easily grouped or categorized, as confirmed by PANTHER-based gene ontology analysis[141]. Based on biological process, molecular function, or cellular component, only genes involved in extrinsic apoptotic pathways were identified as enriched pathways (Figure Fig 3.9B). These data suggest that regulators of cfDNA release include multiple pro-apoptotic genes in the TRAIL pathway as well as molecules with more cryptic roles.

To test if similar hits would be yielded by orthogonal screens in different cell lines, we performed a second genome-wide CRISPR screen using the low cfDNA releasing human lung cancer cell line, A549. We performed this screen with two different timepoints – a late timepoint taken at 28 days post infection in accordance with the MCF-10A screen (Figure 3.9C), and an earlier timepoint at 14 days (Figure 3.9D). *BCL2L1* and *MCL1* were identified as significant candidates and both are members of the BCL-2 family in the late timepoint. (Table 3.5). No other significant gene families were identified by gene ontology (Figure 3.7D). *BCL2L1* was also one of the strongest hits in the early timepoint screen (Table 3.6). However, many other pathways were also implicated in the early screen aside from apoptosis, including mevalonate synthesis, metaphase/anaphase transition regulation, and trafficking of cellular components to the multivesicular body (Figure 3.7E, 3.7F).

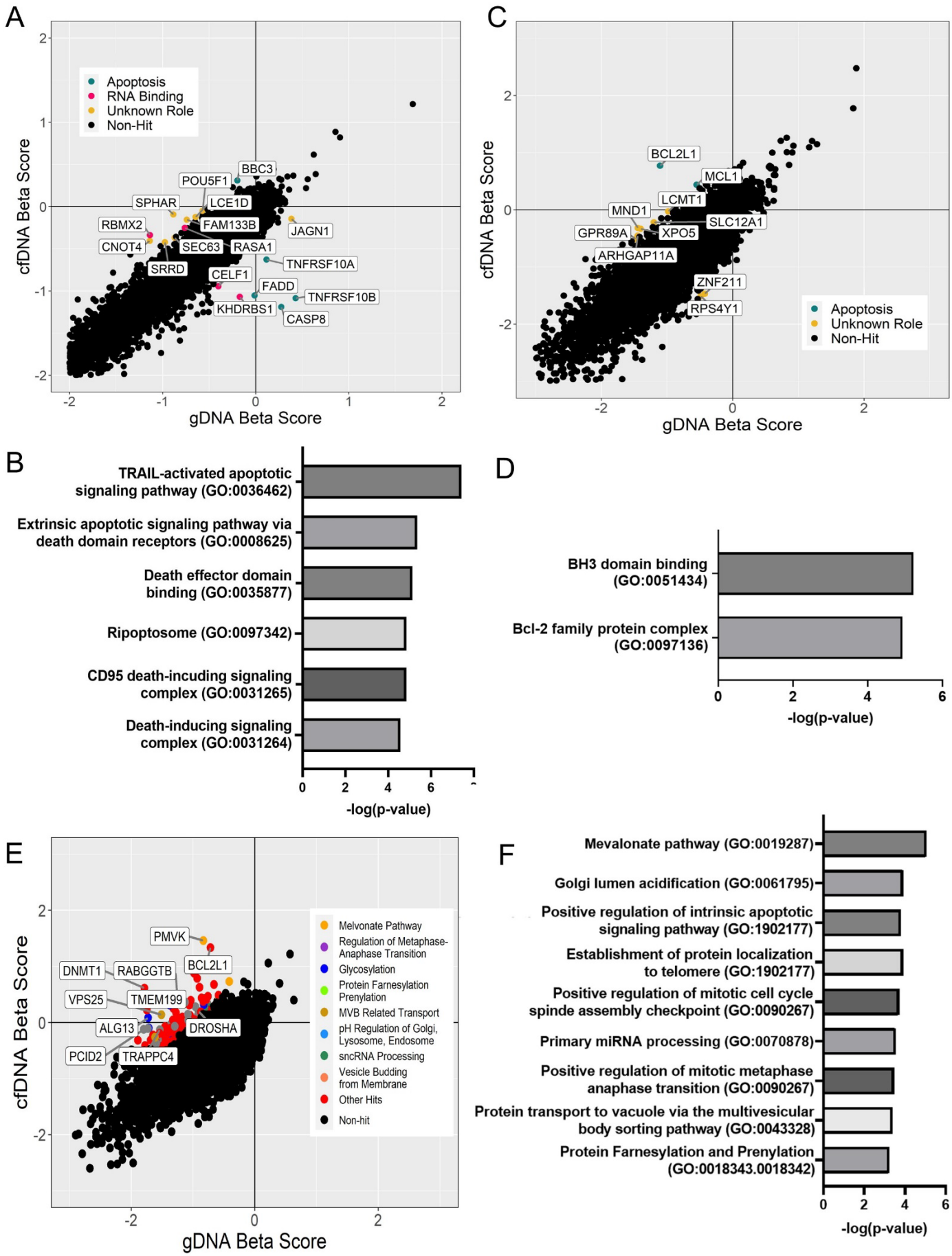

A
Apoptosis
RNA Binding
Unknown Role
Non-Hit
BBC3
POU5F1
SPHAR
LCE1D
RBMX2
FAM133B
JAGN1
CNOT4
SEC63
RASA1
TNFRSF10A
SRRD
CELF1
FADD
TNFRSF10B
KHDRBS1
CASP8
cfDNA Beta Score
gDNA Beta Score
C
BCL2L1
MCL1
LCMT1
MND1
SLC12A1
GPR89A
XPO5
ARHGAP11A
ZNF211
RPS4Y1
Apoptosis
Unknown Role
Non-Hit
cfDNA Beta Score
gDNA Beta Score
B
TRAIL-activated apoptotic signaling pathway (GO:0036462)
Extrinsic apoptotic signaling pathway via death domain receptors (GO:0008625)
Death effector domain binding (GO:0035877)
Ripoptosome (GO:0097342)
CD95 death-incuding signaling complex (GO:0031265)
Death-inducing signaling complex (GO:0031264)
-log(p-value)
D
BH3 domain binding (GO:0051434)
Bcl-2 family protein complex (GO:0097136)
-log(p-value)
E
PMVK
DNMT1
RABGGTB
BCL2L1
VPS25
TMEM199
ALG13
DROSHA
PCID2
TRAPPC4
Melvonate Pathway
Regulation of Metaphase-Anaphase Transition
Glycosylation
Protein Farnesylation Prenylation
MVB Related Transport
pH Regulation of Golgi, Lysosome, Endosome
sncRNA Processing
Vesicle Budding from Membrane
Other Hits
Non-hit
cfDNA Beta Score
gDNA Beta Score
F
Mevalonate pathway (GO:0019287)
Golgi lumen acidification (GO:0061795)
Positive regulation of intrinsic apoptotic signaling pathway (GO:1902177)
Establishment of protein localization to telomere (GO:1902177)
Positive regulation of mitotic cell cycle spinde assembly checkpoint (GO:0090267)
Primary miRNA processing (GO:0070878)
Positive regulation of mitotic metaphase anaphase transition (GO:0090267)
Protein transport to vacuole via the multivesicular body sorting pathway (GO:0043328)
Protein Farnesylation and Prenylation (GO:0018343.0018342)
-log(p-value)

Figure 3.9. Results of CRISPR screens for mediators of cfDNA Release in MCF-10A and A549 cells. (A) Genes plotted by their β-scores in both cfDNA and gDNA arms of MCF-10A CRISPR screen. Putative hits are highlighted and grouped by literature-based shared functions. (B) Gene ontology of genes from the MCF-10A CRIPSR screen determined as putative hits. Data are -log p-values derived from PANTHER gene ontology of and include enriched pathways in biological process, molecular function, and cellular component. (C) Genes plotted by their β-scores in both cfDNA and gDNA arms of A549 28-day late screen. Putative hits are highlighted and grouped by literature-based shared functions. (D) Gene ontology of genes from the A549 28 day late CRIPSR screen determined as putative hits. Data are -log p-values derived from PANTHER gene ontology of and include enriched pathways in biological process, molecular function, and cellular component. (E) Genes plotted by their β-scores in both cfDNA and gDNA arms of A549 14-day early screen. Putative hits are highlighted and grouped by literature-based shared functions. (F) Gene ontology of genes from the A549 14 day early CRIPSR screen determined as putative hits. Data are -log p-values derived from PANTHER gene ontology of and include enriched pathways in biological process.

Table 3.4. Commonalities exhibited by hits derived from MCF-10A CRISPR screen for regulators of cfDNA release at 0.5 Beta difference cutoff.

Gene	Beta Difference	cfDNA Beta Score	gDNA Beta Score	Expected Change in cfDNA Release when Deleted	Literature Based Classification
TNFRSF10B (DR5)	1.51496	-1.086	0.4286	Decrease	Apoptosis
CASP8	1.46545	-1.19	0.2751	Decrease	Apoptosis
FADD	1.04145	-1.054	-0.013	Decrease	Apoptosis
KHDRBS1 (Sam68)	0.89919	-1.069	-0.17	Decrease	RNA Binding
RBMX2	0.79647	-0.339	-1.135	Increase	RNA Binding
SPHAR	0.79035	-0.093	-0.884	Increase	Unknown Role
TNFRSF10A (DR4)	0.74404	-0.628	0.1162	Decrease	Apoptosis

CNOT4	0.73251	-0.405	-1.138	Increase	Unknown Role
FAM133B	0.58493	-0.155	-0.74	Increase	Unknown Role
SRRD	0.55264	-0.423	-0.975	Increase	Unknown Role
CELF1	0.54444	-0.945	-0.401	Decrease	RNA Binding
JAGN1	0.52982	-0.145	0.3845	Decrease	Unknown Role
POU5F1 (OCT3/4)	0.52218	-0.049	-0.571	Increase	Unknown Role
LCE1D	0.52111	-0.124	-0.645	Increase	Unknown Role
RASA1	0.50798	-0.253	-0.761	Increase	RNA Binding
BBC3	0.50591	0.3093	-0.197	Increase	Apoptosis
SEC63	0.50157	-0.368	-0.87	Increase	Unknown Role

Table 3.5. Commonalities exhibited by hits derived from A549 CRISPR screen for regulators of cfDNA release at 0.95 Beta difference cutoff.

Gene	Beta Difference	cfDNA Beta Score	gDNA Beta Score	Expected Change in cfDNA Release when Deleted	Literature Based Classification
BCL2L1	1.87473	0.77053	-1.1042	Increase	Apoptosis
MND1	1.12401	-0.31819	-1.4422	Increase	Unknown Role
XPO5	1.06902	-0.33288	-1.4019	Increase	Unknown Role
ZNF211	1.044	-1.4691	-0.4251	Decrease	Unknown Role
RPS4Y1	1.02604	-1.5113	-0.4853	Decrease	Unknown Role
ARHGAP11A	0.99075	-0.45425	-1.445	Increase	Unknown Role
MCL1	0.98674	0.43931	-0.5474	Increase	Apoptosis
SLC12A1	0.97971	-0.21949	-1.1992	Increase	Unknown Role
GPR89A	0.96582	-0.51168	-1.4775	Increase	Unknown Role
LCMT1	0.952065	-0.028315	-0.9804	Increase	Unknown Role

Table 3.6. Literature-based commonalities exhibited by hits derived from 14-day timepoint A549 CRISPR screen for regulators of cfDNA release at 0.95 Beta difference cutoff.

Gene	Beta Difference	cfDNA Beta Score	gDNA Beta Score	Literature Based Classification
DNMT1	2.39892	0.61322	-1.7857	Other Hits
PMVK	2.29263	1.4601	-0.83253	Mevalonate pathway
BCL2L1	2.05159	1.3347	-0.71689	Other Hits
CCDC84	1.96885	0.22235	-1.7465	Other Hits
GGPS1	1.94511	0.39081	-1.5543	Other Hits
CMPK1	1.92894	0.35604	-1.5729	Other Hits
UHRF1	1.85902	0.87776	-0.98126	Other Hits
ALG13	1.809953	0.083153	-1.7268	Glycosylation
CDIPT	1.77237	0.99387	-0.7785	Other Hits
RABGGTB	1.73649	0.48159	-1.2549	Protein Farnesylation/Prenylation
G6PD	1.71928	0.7883	-0.93098	Other Hits
DOLK	1.68815	0.30965	-1.3785	Glycosylation
PCID2	1.67681	-0.12199	-1.7988	Regulation of Metaphase-Anaphase Transition
MVK	1.6736	0.4367	-1.2369	Mevalonate pathway
VPS25	1.65236	0.14006	-1.5123	MVB Related Transport
ANAPC11	1.62389	-0.10711	-1.731	Regulation of Metaphase-Anaphase Transition
ALG2	1.614484	-0.099116	-1.7136	Glycosylation
POLR3H	1.61433	-0.61857	-2.2329	Other Hits
SRF	1.60913	0.35533	-1.2538	Other Hits
SNF8	1.59134	1.0384	-0.55294	MVB Related Transport
ARL2	1.59061	0.26781	-1.3228	Other Hits
HSPD1	1.57464	-0.31656	-1.8912	Other Hits
CCT4	1.57055	-0.62405	-2.1946	Other Hits
COASY	1.56985	0.25205	-1.3178	Other Hits
ZNF407	1.52	0.62692	-0.89308	Other Hits
UBA1	1.50925	-0.50305	-2.0123	Other Hits
TAF6	1.50685	-0.48785	-1.9947	Other Hits
HMGCS1	1.50055	-0.77065	-2.2712	Other Hits
TMEM199	1.49223	0.36533	-1.1269	pH Regulation of Membrane Bound Organelles
PFDN5	1.4513	0.1772	-1.2741	Other Hits
PCNA	1.44116	-0.68784	-2.129	Other Hits
C19orf52	1.4295	-1.1125	-2.542	Other Hits
PCYT1A	1.42832	0.39282	-1.0355	Other Hits

TIMM10	1.42449	-0.22321	-1.6477	Other Hits
TRAPPC4	1.41023	-0.12807	-1.5383	Vesicle Budding from Membrane
TOMM70A	1.37498	-0.07752	-1.4525	Other Hits
SACM1L	1.37157	-0.31323	-1.6848	Other Hits
NUBP1	1.37012	0.04462	-1.3255	Other Hits
SRP14	1.36844	0.30994	-1.0585	Other Hits
POLR3B	1.36358	-0.41332	-1.7769	Other Hits
SOD1	1.3556	-0.923	-2.2786	Other Hits
VPS28	1.3518	-0.2623	-1.6141	MVB Related Transport
FEN1	1.351767	0.032767	-1.319	Other Hits
TXN	1.34908	-0.19232	-1.5414	Other Hits
TP53RK	1.33771	-0.77109	-2.1088	Other Hits
ENY2	1.33657	0.67529	-0.66128	Other Hits
INCENP	1.328107	-0.029893	-1.358	Regulation of Metaphase-Anaphase Transition
IARS2	1.31606	-0.13464	-1.4507	Other Hits
HSPE1	1.3046	-1.1349	-2.4395	Other Hits
SUDS3	1.29574	0.43456	-0.86118	Other Hits
DTYMK	1.29435	-0.12585	-1.4202	Other Hits
RANGAP1	1.291	-0.6948	-1.9858	Other Hits
UBL5	1.28042	-0.50748	-1.7879	Other Hits
PRMT1	1.27895	-0.54995	-1.8289	Other Hits
MSTO1	1.278602	0.041702	-1.2369	Other Hits
ATP1A1	1.267213	0.096213	-1.171	Other Hits
YRDC	1.26202	-0.71828	-1.9803	Other Hits
TNPO1	1.25905	-0.10315	-1.3622	Other Hits
H3F3A	1.25842	-0.34068	-1.5991	Other Hits
GPR89B	1.25251	-0.31119	-1.5637	pH Regulation of Membrane Bound Organelles
TBCB	1.24309	0.42427	-0.81882	Other Hits
FNTA	1.23778	0.27983	-0.95795	Protein Farnesylation/Prenylation
ATP6V0B	1.237622	-0.0088778	-1.2465	pH Regulation of Membrane Bound Organelles
TRAPPC1	1.22756	0.14566	-1.0819	Vesicle Budding from Membrane
TRAPPC3	1.22571	-0.06869	-1.2944	Vesicle Budding from Membrane
DAD1	1.224	-0.2133	-1.4373	Other Hits
PRMT5	1.2223	-1.5587	-2.781	Other Hits
GLMN	1.22203	0.46514	-0.75689	Other Hits
CCT5	1.22054	-0.79986	-2.0204	Other Hits
TRAPPC11	1.22045	-0.37235	-1.5928	Vesicle Budding from Membrane
SCAP	1.216884	0.023584	-1.1933	Other Hits
THOC3	1.21169	-0.47711	-1.6888	Other Hits

TMX2	1.21129	0.37583	-0.83546	Other Hits
RPP21	1.21074	-0.41136	-1.6221	Other Hits
SFPQ	1.2063	-1.2325	-2.4388	Other Hits
CFLAR	1.20538	-0.25212	-1.4575	Other Hits
RFC3	1.20516	-0.17584	-1.381	Other Hits
EIF2B5	1.204	-1.2073	-2.4113	Other Hits
PPP4C	1.20011	-0.24329	-1.4434	Other Hits
EIF5A	1.19976	-0.25094	-1.4507	Other Hits
ANKRD49	1.1972	-0.4816	-1.6788	Other Hits
UPF2	1.1933	0.40691	-0.78639	Other Hits
POLR2A	1.18407	-0.77063	-1.9547	Other Hits
FDX1L	1.17747	-0.73683	-1.9143	Other Hits
TBCE	1.17213	0.12623	-1.0459	Other Hits
PPP2R2A	1.168394	0.039694	-1.1287	Other Hits
MFN2	1.16724	-0.51116	-1.6784	Other Hits
SMG1	1.165753	0.013853	-1.1519	Other Hits
PHB2	1.16136	-0.98814	-2.1495	Other Hits
CDC23	1.15512	-0.21948	-1.3746	Regulation of Metaphase-Anaphase Transition
ARMC7	1.15286	-0.93294	-2.0858	Other Hits
SBDS	1.1488	-0.8291	-1.9779	Other Hits
RFT1	1.14297	0.31725	-0.82572	Glycosylation
POLR3C	1.14172	-0.49978	-1.6415	Other Hits
MVD	1.14095	0.7306	-0.41035	Mevalonate pathway
VPS13D	1.1384	-0.5115	-1.6499	Other Hits
ZNHIT2	1.1369	-1.3263	-2.4632	Other Hits
PSMG3	1.13628	-0.53132	-1.6676	Other Hits
RBBP4	1.13626	-0.12784	-1.2641	Other Hits
RBM8A	1.13519	-0.52251	-1.6577	Other Hits
CAPZB	1.13496	-0.32114	-1.4561	Other Hits
CHMP2A	1.12751	-0.27049	-1.398	MVB Related Transport
WDR4	1.125491	0.024791	-1.1007	Other Hits
ZNF207	1.123815	0.043015	-1.0808	Regulation of Metaphase-Anaphase Transition
EEF1A1	1.1201	-1.5304	-2.6505	Other Hits
LEMD2	1.116291	0.054991	-1.0613	Nuclear Envelope Organization
DHFR	1.113581	-0.030619	-1.1442	Other Hits
TLN1	1.113448	0.069448	-1.044	Other Hits
RABGGTA	1.10989	-0.11161	-1.2215	Protein Farnesylation/Prenylation
SHFM1	1.106007	0.073407	-1.0326	Other Hits
EIF1	1.10576	-0.48984	-1.5956	Other Hits

DROSHA	1.10447	0.19568	-0.90879	sncRNA Processing
METTL14	1.104071	0.078471	-1.0256	Other Hits
TBCC	1.103	-0.3308	-1.4338	Other Hits
ATP6AP2	1.09978	0.33556	-0.76422	pH Regulation of Membrane Bound Organelles
SF1	1.096841	-0.099459	-1.1963	Other Hits
DGCR8	1.09394	-0.02606	-1.12	sncRNA Processing
TOMM22	1.09319	-0.52961	-1.6228	Other Hits
CCT6A	1.09264	-0.89616	-1.9888	Other Hits
POLR1C	1.09124	-0.97596	-2.0672	Other Hits
GINS2	1.08754	-0.67756	-1.7651	Other Hits
IGBP1	1.08751	-0.70059	-1.7881	Other Hits
METTL3	1.08683	0.27229	-0.81454	sncRNA Processing
CHURC1-FNTB	1.085668	-0.011732	-1.0974	Other Hits
NUPL1	1.078225	0.053225	-1.025	Other Hits
DPAGT1	1.07817	-0.26913	-1.3473	Glycosylation
NUP35	1.077837	-0.088063	-1.1659	Other Hits
NDC80	1.07674	-0.66576	-1.7425	Other Hits
NARFL	1.07597	-0.33233	-1.4083	Other Hits
PKMYT1	1.07483	-0.42267	-1.4975	Other Hits
SBNO1	1.07306	-0.39094	-1.464	Other Hits
GNB1L	1.07268	0.48747	-0.58521	Other Hits
CIAO1	1.071448	-0.098852	-1.1703	Other Hits
DUT	1.07012	-0.71298	-1.7831	Other Hits
TTC27	1.06057	-0.22693	-1.2875	Other Hits
IPO13	1.06008	-0.45542	-1.5155	Other Hits
SSFA2	1.05985	0.18662	-0.87323	Other Hits
RPL35	1.0539	-1.1508	-2.2047	Other Hits
FNTB	1.04533	0.12054	-0.92479	Protein Farnesylation/Prenylation
SRP54	1.04455	-0.17925	-1.2238	Other Hits
CDC16	1.04366	-0.25084	-1.2945	Regulation of Metaphase-Anaphase Transition
CHMP7	1.04044	0.45652	-0.58392	MVB Related Transport
SMARCA2	1.03945	-0.52905	-1.5685	Other Hits
MBTPS1	1.03701	-0.16599	-1.203	Other Hits
ALDOA	1.03613	-0.48367	-1.5198	Other Hits
NDNL2	1.03453	-0.48627	-1.5208	Other Hits
ATP6V0C	1.03385	-0.36265	-1.3965	pH Regulation of Membrane Bound Organelles
LRRC37A	1.0268	-1.1122	-2.139	Other Hits
TRMT5	1.026669	0.050239	-0.97643	Other Hits

SPC24	1.0247	-0.9063	-1.931	Other Hits
TTC1	1.02101	-0.18579	-1.2068	Other Hits
CENPO	1.01368	-0.46932	-1.483	Other Hits
MCL1	1.0132	0.23454	-0.77866	Other Hits
PFDN2	1.01032	-0.80218	-1.8125	Other Hits
HNRNPU	1.00904	-0.13556	-1.1446	Other Hits
DICER1	1.00717	0.26377	-0.7434	sncRNA Processing
CPSF2	1.00398	-0.19742	-1.2014	Other Hits
HARS	1.0031	-0.4865	-1.4896	Other Hits
BUD31	1.00254	-0.83086	-1.8334	Other Hits
RCC1	1.00233	0.06049	-0.94184	Nuclear Envelope Organization
DNAJC11	1.00225	-0.65215	-1.6544	Other Hits
RPN2	1.00121	-0.32969	-1.3309	Other Hits

Discussion

This section explores the characteristics of cell-free DNA release in a panel of cancer and non-cancerous cell lines and leverages this knowledge to select cell lines for CRISPR screening for mechanisms of cell-free DNA release. Through this panel we were able to separate cell lines into two groups, those which have a left-skewed cell-free DNA release with a major peak at around 167bp and those which display a right-skewed cell-free DNA release pattern with a major peak greater than 1000bp in size. To our knowledge, this is the largest evaluated panel for cell-free DNA release, and the first to also profile fragmentation pattern. We were unable to correlate cancer status, cancer type, DNAse expression, or media conditions with this skewing phenomenon, indicating it is a cell intrinsic process with unknown regulators. Previous research has indicated that these large, right skewed fragments might be associated with vesicular DNA or necrotic DNA, whereas the small fragments seen in the left skewed cell lines are traditionally thought to be resultant from apoptotic DNA release[82,106,133]. This concept will be explored in Chapter IV through the creation of cell lines with gene knockouts for hits of interest, which were primarily in apoptotic pathways. Another possibility is that the DNA in the left-skewed cell lines

is degraded at a rapid rate towards small fragments. However, we find that in our culture system, although degradation does create a left shift when starting with larger fragments, even 50% degradation requires at least 3 days of incubation. This implies if degradation was the cause of left skewing that we would see a longer trail of larger fragments that had yet to degrade. We also find that DNA tends to increase in quantity released over days in culture within its particular fragmentation pattern without deviating into the other fragmentation pattern at any point, contrary to the idea of left skewed DNA being simply degraded DNA as well as previous results from other groups showing shifts in fragmentation patterns released from cell lines over time[86-88,129]. Together, this cell line panel is the most in-depth analysis of *in vitro* cell-free DNA release yet conducted and shows a striking diversity in quantity and fragmentation pattern of DNA released by various cell lines.

Visualizing the cell-free DNA release of various cell lines allowed us to select cell lines with varying characteristics for CRISPR screening, broadening the impact of any pathways with shared hits. We decided to screen in two left shifted cell lines as they best mimic that found in human blood, one high releasing in MCF-10A and one low releasing in A549. Given that MCF-10A are also a non-tumorigenic epithelial cell line and A549 are a tumorigenic lung cancer line, we expected to see major differences in genes that were hits, but expected any overlapping pathways to be highly relevant to the release of cell-free DNA in most cellular systems. In the late timepoint of both screens, we found that apoptotic processes were strong hits. Particularly, the MCF-10As displayed multiple hits in the extrinsic-apoptotic TRAIL pathway, whereas A549s displayed multiple hits in the intrinsic apoptotic pathway. Cancer cell lines often downregulate external cell death receptors[142,143], so it is unsurprising to find a lack of extrinsic cell death pathways as hits in the A549 screen. Additionally, MCF-10As have been shown to

lowly express the gene product BCL-XL of BCL2L1, the main hit in the A549 screens, indicating why this discrepancy occurred[144].

In the early timepoint of the A549 screen, many non-apoptotic genes were found that will not be explored in the rest of this thesis. However, many of them posit intriguing questions. Multiple groups of genes hits in this screen regulate processes that if damaged would cause increased DNA damage, including the establishment of protein localization to telomeres, positive regulation of cell-cycle spindle assembly checkpoints, and positive regulation of the metaphase/anaphase transition. It is possible that disruption of these processes leads to increased DNA damage and thereby increased formation of micronuclei which upon collapsing lead to increased cytosolic DNA, which has previously been shown to be trafficked out of cells through vesicular pathways[94,103]. Other hits in this screen were involved in vesicle trafficking, vesicle pH balance, and the loading of cargo into the multivesicular body. Defects in this trafficking could either lead to further accumulation of DNA in the cytoplasm, or could lead to trafficking away from traditional disposal sites such as the lysosome and towards extracellular release. However, multiple ribosomal genes were hits in this arm of the screen, indicating that many of these extra hits may simply be essential to cell function and were called as hits more as a contaminant than as a molecule actively controlling DNA release. However, these hits in DNA damage and vesicle trafficking pathways remain interesting for future studies, as little is known about predictors of DNA release from cancers. Future research directions might inquire if increased DNA damage leads to increased DNA release both *in vitro* and in humans, and might also determine the effect of vesicle release inhibition or activation on the release of DNA. Together, these results outline multiple possible groups of molecules that regulate cell-free DNA release that can be mined for future study.

In chapter IV, we will investigate apoptosis as a major regulator of cfDNA release, as hits in this pathway were present across all screens and additionally had high hit strength. CRISPR screens can provide false positive results due to stochastic effects, so rigorous validation through single gene knockouts is commonly needed to determine both the voracity and mechanism of any phenotype identified. To this end, Chapter IV will primarily focus on genes of interest identified by our CRISPR screens in apoptotic pathways, such as *FADD* and *BCL2L1*, and their genetic manipulation to confirm the role of these genes in cfDNA release.

CHAPTER IV

CONFIRMATION OF CELL DEATH AS A MAJOR MODULATOR OF CELL-FREE DNA RELEASE

Sections of this chapter have been previously published in: Davidson, *et al*. 2024. An in vitro CRISPR screen of cell-free DNA identifies apoptosis as the primary mediator of cell-free DNA release. *Communications Biology*, provisionally accepted.

Abstract

Three major cellular processes have been proposed as the primary pathway of cfDNA release in apoptosis, necrosis, and active release through the vesicle pathway. Our previous CRISPR screens revealed regulators of apoptotic pathways across cell lines and timepoints as strong hits for the modification of cell-free DNA release. The TRAIL pathway and Sam68 were major hits in the MCF-10A screen, and through work in this chapter were confirmed as major positive regulators of cfDNA release through their apoptotic roles. Similarly, BCL2-family members were highlighted in the A549 screen, and work in this chapter confirmed *BCL2L1* as a major negative regulator of cfDNA release through its apoptotic roles. Manipulations of these genes through genetic knockout, over-expression, and drug studies typically did not change the intrinsic fragmentation pattern skewing of DNA released from respective cell lines, indicating that the notion posited in previous studies that apoptotic DNA release represents only the 167bp peak and vesicle-related DNA represents only the >1000bp peak may be incorrect. Instead, we propose that cell-free DNA released through apoptotic pathways can be of either of these sizes, and that most cfDNA release previously identified in *in vitro* is apoptotic. In addition, these results indicate that the DNA fragmentation pattern found in human blood may not be intrinsically apoptotic, but instead the degradation of various DNAs to small fragments through circulating DNAses.

Introduction

Apoptosis is a well-studied process of programmed cell death characterized morphologically by cell shrinkage, chromatin condensation and membrane blebbing[145]. After initiation of his process through either cell-intrinsic or cell-extrinsic means, enzymes known as caspases cleave proteins and other macromolecules internal to the cell. Membrane blebs containing intact organelles are released from the cell membrane and are phagocytosed by macrophages and degraded. Eventually, degradation of cellular structures causes the cell membrane to finally rupture as the cell dies.

Apoptosis is brought on primarily through two canonical pathways: an intrinsic mitochondrial driven mechanism and an extrinsic receptor-triggered mechanism[145]. The TRAIL pathway is a prototypical example of an extrinsic receptor triggered mechanism, although others such as the FAS and TNFα pathways are also widely studied pathway members. TRAIL signaling is initiated when the TRAIL ligand binds to death receptors TRAIL-R1/DR4 or TRAIL-R2/DR5, resulting in receptor trimerization and the formation of the death-inducing signaling complex (DISC). The protein FADD is then recruited to the complex and can further recruit pro-caspase 8 and pro-caspase 10, which auto-hydrolyze into activated caspase enzymes. These caspases then go on to cleave pro-caspase 3 and Bid into caspase 3 and tBid. Caspase 3 continues to cleave various macromolecules including other caspases, proteins, and DNA, leading to cell death. Meanwhile, tBid links the extrinsic pathway to the intrinsic pathway by cleaving inhibitors of the pro-apoptotic mitochondrial outer membrane permeabilization (MOMP) initiators Bax and Bak[146]. The intrinsic apoptotic pathway is driven primarily through changes in the transcription of MOMP inhibiting proteins such as the BCL-2 family members BCL-2, BCL-XL, BCL-W, and MCL1 and upregulation of MOMP promoting proteins[145]. For example, when a cell has

extensive DNA damage, the tumor suppressor protein p53 senses the damage and causes the transcription of pro-apoptotic regulators such as Bax and can down-regulate the transcription of anti-apoptotic proteins such as BCL-2[147,148]. The permeabilization of the mitochondria leads to the release of cytochrome c, an irreversible step in apoptosis, which through a signaling cascade leads to activation of caspase 9 and caspase 3 and cell death through processes similar to those seen in extrinsic apoptosis[145].

Proteins whose known effects don't directly implicate apoptotic regulation can still modify apoptotic processes by altering the expression of apoptosis initiating or inhibiting proteins. Sam68 is one such protein with multifaceted roles in cytoplasmic growth signaling pathways, transcriptional regulation, DNA damage repair, and alternative splicing. Sam68 was first identified as a phosphorylation target of c-Src during mitosis, and was later described as part of multiple cytoplasmic signaling pathways downstream of this phosphorylation event, including forming part of a complex that may regulate Ras activation during insulin signaling[149-151]. In addition, it acts at the transcriptional level to repress or increase expression of various genes through interactions with various transcription factors in the nucleus. For example, its binding to AR-V7 in prostate cancer cells spurs transcription of its downstream targets[152]. Nuclear Sam68 has also been shown to be essential for PARylation of DNA, a critical step in multiple DNA damage repair pathways, and its depletion leads to apoptotic cell death after genotoxic stress due to an inability to repair DNA damage[153,154]. Sam68's most clear link to apoptosis however is through its splicing functions, where it is known to funnel gene products such as the *BCL2L1* gene product BCL-X towards anti-apoptotic spliceforms, in this case being BCL-XL, over the pro-apoptotic BCL-XS[155].

In this section, proteins found in the MCF-10A and A549 screen that have implicated roles in apoptosis will be studied for their potential effects on cfDNA release. This analysis will focus on the pro-apoptotic TRAIL pathway mediator FADD and the potentially pro-apoptotic splicing protein Sam68 in MCF-10A cells, while in A549 cells the *BCL2L1* anti-apoptotic gene product BCL-XL will be explored. Previously known roles of each of these potential cfDNA modulators in apoptosis can be found in Figure 4.1. Through genetic and pharmacologic manipulations of these genes and others, we found that cfDNA is almost entirely composed of apoptosis derived DNA. This is in spite of the fact that most DNA derived from the cell lines in our panel is of large fragment sizes >1000bp, which are traditionally thought to originate from non-apoptotic sources such as necrosis or vesicle-associated release. These results indicate that most DNA released from cells is apoptotic, and that we can modify DNA release from cells in a way that could be clinically meaningful by manipulating apoptotic processes. In addition, these results might indicate a reason for the extreme variability seen between and within cancer patients with regards to the ctDNA content in their bloodstreams, as if more or less apoptotic cell death is taking place in the tumor due to various stimuli such as drug treatment or inactivation of apoptotic mediators it would lead to more or less DNA release into the blood.

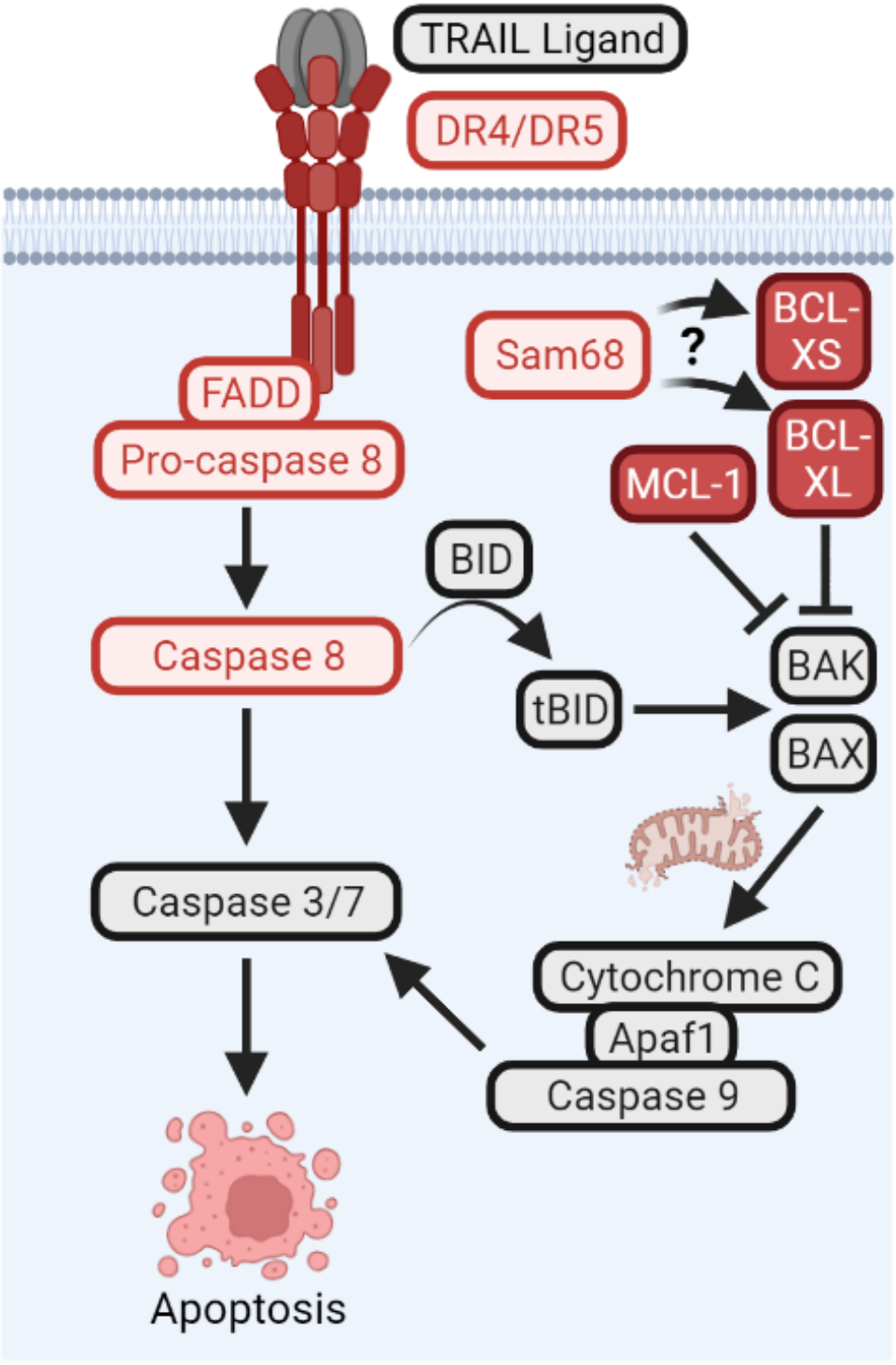


Figure 4.1. Role of CRISPR screen hits in apoptotic pathways. Apoptosis-related hits in the MCF-10A screen were primarily pro-apoptotic genes in TRAIL extrinsic apoptotic pathway, outlined in light red. The strongest hits in the late A549 screen were anti-apoptotic genes in related apoptotic pathways as outlined in dark red including BCL-2 family members. Sam68 was the next strongest hit in MCF-10A screen after the TRAIL pathway. This gene has multifaceted roles in splicing, DNA damage signaling, and transcription, but has been proposed to splice the *BCL2L1* gene product in other systems towards the anti-apoptotic BCL-XL and away from the pro-apoptotic BCL-XS.

Results

Confirmation of Sam68 and FADD as mediators of cell-free DNA release

To validate that the genes identified in our CRISPR screen were involved in cfDNA release, *KHDRBS1* (Sam68) and *FADD* (FADD) were selected for generation of CRISPR-mediated gene knockouts in MCF-10A cells. FADD is a central hub in the TRAIL pathway, as its recruitment by the trimerization of TRAIL receptors provides the scaffolding to activate caspase 8 cleavage and downstream apoptotic signaling[156]. Sam68 has a more cryptic role given its multifaceted functions in splicing, transcriptional/post-transcriptional gene regulation, and DNA damage response[154,157,158], but was chosen as the top candidate among our putative RNA binding genes. Cell lines generated included four knockout cell lines using two distinct sgRNAs for *KHDRBS1* (Sam68 KO1/KO2 and KO3/KO4, respectively) and two knockout cell lines using one sgRNA for *FADD* (FADD KO1/KO2). Untreated cell lines ("parental") and CRISPR-targeted single cell clones that resulted in wild-type ("targeted wild-type"; TWT) were used as controls. Parental cells account for the natural phenotype of the cell line, while TWT cells control for the transfection process and off-target effects. Complete knockout in Sam68 KO and FADD KO lines was confirmed by sequencing and immunoblot (Figs. 4.1A, 4.1B, 4.1C, 4.1D). Our CRISPR screen analysis (Fig. 3.9A) predicts the directional effect of knockout on cfDNA release – hits below the Z-axis will likely lead to decreases in cfDNA release when knocked out, whereas those above the Z-axis will likely lead to increases. Sam68/KHDRBS1 and FADD are positioned below this Z-axis. Their low position on the y-axis represents negative selection in the cfDNA arm of our screen while its near zero β-score in the gDNA arm represented on the x-axis indicates a lack of selection. Therefore, we would expect that knockout of these genes would lead to decreased cfDNA release as they were not selected against cellularly, but less of their DNA was found in culture media than expected. Indeed, after grouping cell lines based on their genotype, analysis of cfDNA release

from our knockout cell lines confirmed that knockout of Sam68 or FADD lead to a significant decrease in cfDNA release as expected, with FADD KOs demonstrating a greater than 75% decrease (Fig. 4.3A). DNA fragmentation analysis displayed a decrease in cfDNA release across all fragment sizes (Fig. 4.3B). Re-expression of GFP-tagged Sam68 and FADD proteins in their respective knockout cell lines was able to fully rescue cfDNA release, while maintaining the cell line's innate left-skewed fragmentation pattern. Overexpression of each protein in the parental or TWT MCF-10A control cell lines concordantly led to a significant increase in cfDNA release. Though this cell line is left-skewed at baseline, overexpression of both proteins led to a dominant peak at ~400bp (Figs. 4.3C, 4.3D, Fig. 4.3E, 4.3F).

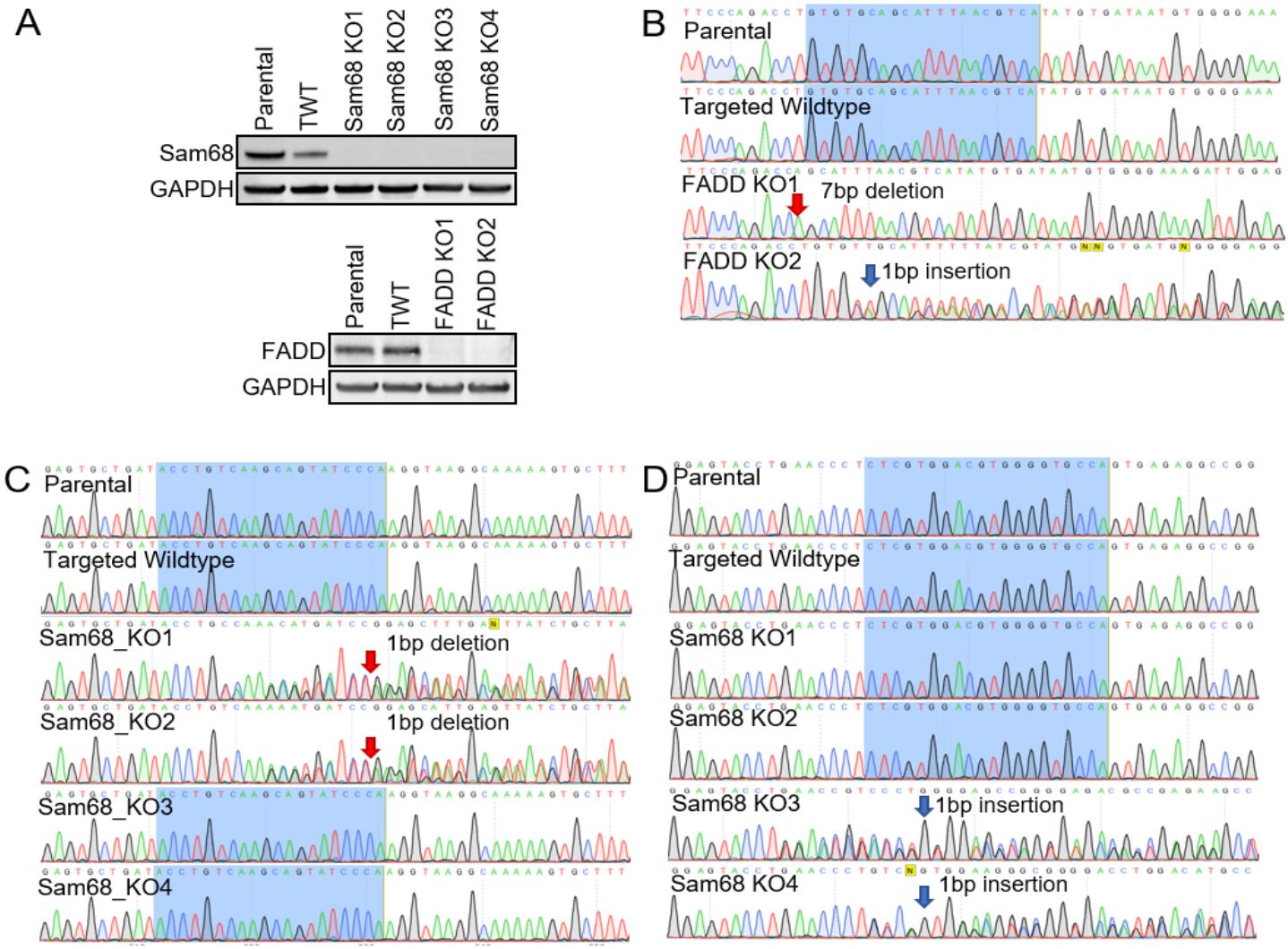


Figure 4.2. Validation of Sam68 and FADD knockouts in MCF-10A. (A) Immunoblot analysis of Sam68 and FADD after CRISPR-mediated knockout (KO) in the MCF-10A background. TWT = Targeted Wild-Type. (B) Sequence of FADD clones at the cut-site of FADD_sgRNA_1, with sgRNA sequence highlighted in blue. (C) Sequence of Sam68 clones at the cut-site of Sam68_sgRNA_1, with sgRNA sequence highlighted in blue. (D) Sequence of Sam68 clones at the cut-site of Sam68_sgRNA_2, with sgRNA sequence highlighted in blue.

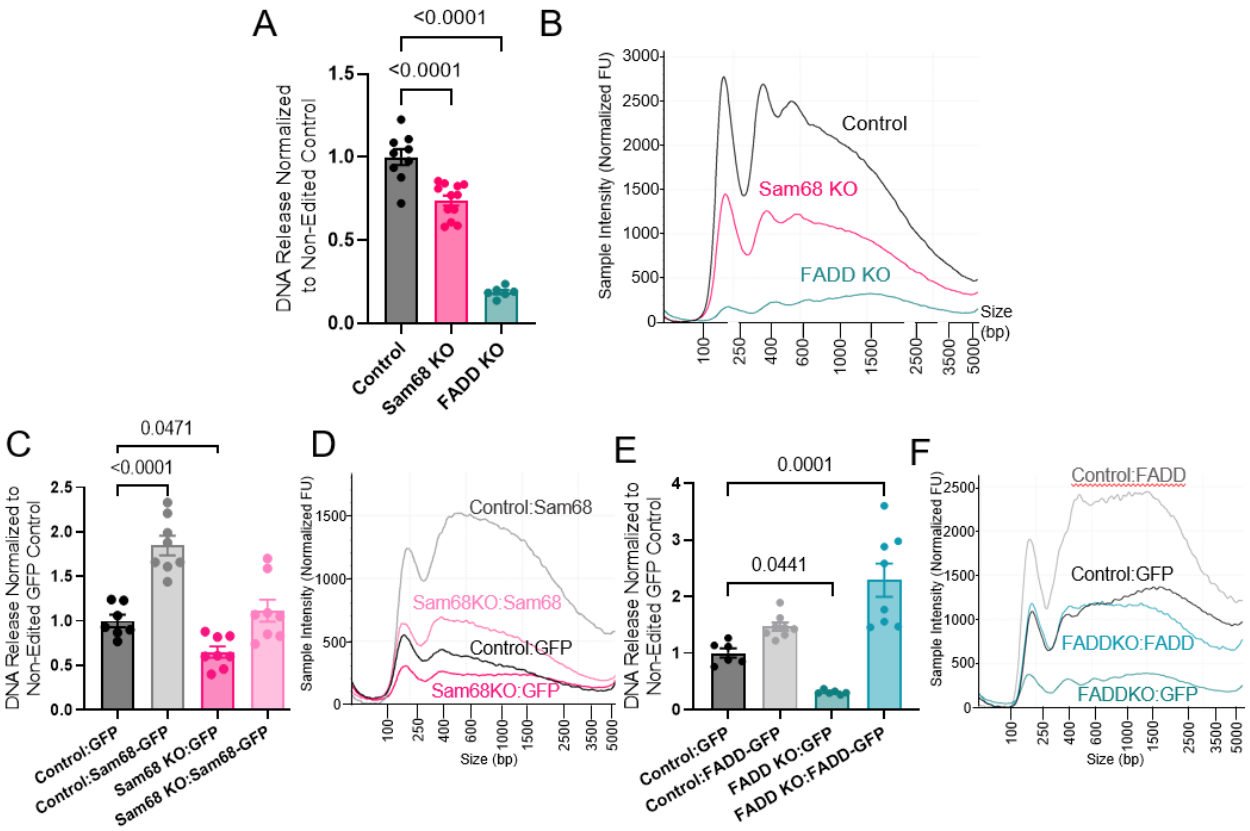


Figure 4.3. Confirmation of Sam68 and FADD as regulators of cfDNA release in MCF-10A background. (A) Quantification of DNA release from MCF-10A KO cell lines in culture. Parental cells and TWT cells were averaged and labeled control. All four Sam68 KO and both FADD KO cell lines are respectively grouped. Data represent mean fold change ± SEM internally normalized to cell concentration for each cell line and then normalized to control; n=3 for each cell line before combining. (B) Fragmentation pattern of DNA released from MCF-10A cell lines in culture. Electropherograms were individually run at least n=3 times and representative traces were selected. (C) Quantification of DNA release from Sam68 KO MCF-10A cell lines rescued by Sam68-GFP overexpression. Parental and TWT were grouped and labeled control, and two Sam68 KO cell lines were grouped. Data represent mean fold change ± SEM internally normalized to cell concentration for each cell line and then normalized to control; n=3-4 for each cell line before combining. (D) Fragmentation pattern of DNA released from MCF-10A Sam68 KO lines and rescues in culture. Electropherograms were individually run at least n=3 times and representative traces were selected. (E) Quantification of DNA release from FADD KO MCF-10A cell lines rescued by FADD-GFP overexpression. Parental and TWT were grouped and labeled control, and two FADD KO cell lines were grouped. Data represent mean fold change ± SEM internally normalized to cell concentration for each cell line and then normalized to control; n=3-4 for each cell line before combining. (F) Fragmentation pattern of DNA released from MCF-10A FADD KO lines and rescues in culture. Electropherograms were individually run at least n=3 times and representative traces were selected.

Further characterization of the knockout cell lines revealed an increased rate of growth when compared to the non-edited controls (Fig. 4.4A). To determine whether this was due to increased cell proliferation or decreased cell death, cells were analyzed for cell death markers. Interestingly, all knockout lines demonstrated decreased early apoptotic cell death initiation through Annexin V labeling and decreased membrane permeability with propidium iodide (PI) compared to the control lines (Fig. 4.4B, C). These changes indicate that differential cell growth is due, at least in part, to alterations in apoptotic pathways.

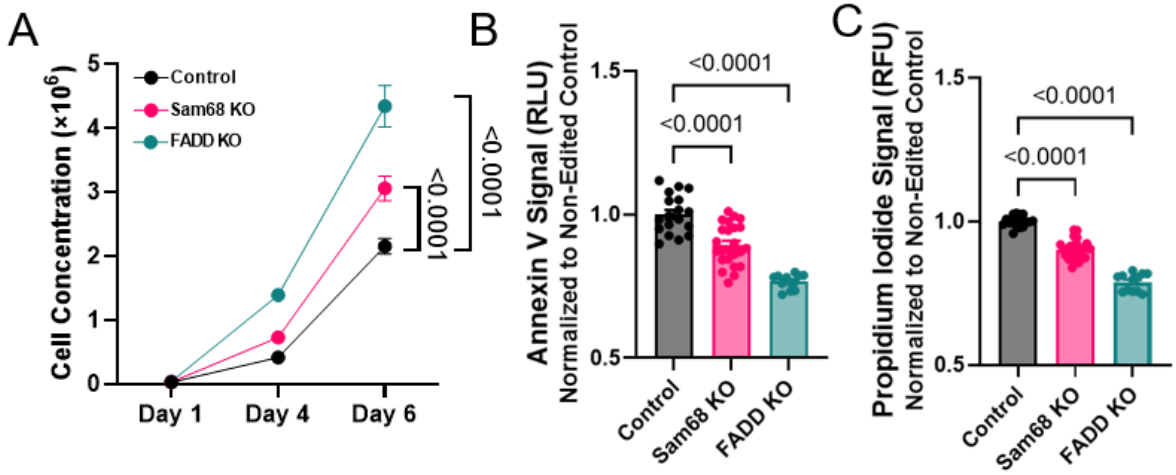


Figure 4.4. Effects of Sam68 and FADD KO on growth and death. (A) Cell growth assay of KO MCF-10A cell lines. Parental and TWT were grouped and labeled control. All four Sam68 KO cell lines and both FADD KO cell lines are respectively grouped. Data represent mean cell concentration ± SEM; n=3 for each cell line before combining. (B) Annexin V and (C) Propidium Iodide (PI) assay of Sam68 and FADD KO MCF-10A cell lines. Cell lines are grouped as in (A). Data represent mean signal (RFU for PI; RLU for Annexin) ± SEM internally normalized to cell concentration for each cell line and then normalized to control; n=6 for each cell line before combining, statistics were ANOVA with Dunnett's multiple comparisons test.

To determine if the screen-identified mediators of apoptosis affect cfDNA release in human cancer cells, GFP-tagged Sam68 and FADD were over-expressed in five different cancer cell lines. Over-expression increased cfDNA release across all lines, with the most pronounced increases in the FADD over-expressing cell lines (Figure 4.5A). Over-expression of Sam68 or FADD resulted in significantly increased Annexin V signal in most lines, indicating increased propensity for apoptosis (Figure 4.5B). Membrane permeability via PI was not measured due to fluorescence overlap with GFP. cfDNA increases were consistent with control fragmentation patterns, revealing an unaltered skew with increased amplitude (Figure 4.5C).

The role of FADD as a mediator of the extrinsic apoptosis pathway is well-defined in the literature, but the role of Sam68 in cell death is less understood. A previous group demonstrated Sam68 interacts with the TNFα receptor complex to promote apoptosis and NF-κB activation[159], suggesting Sam68 might also have a direct role in signaling at the TNFR superfamily member TRAIL receptor[160-163]. To determine the potential pathway overlap of Sam68 and FADD, we performed a double knockout of FADD and Sam68 in the MCF-10A background (Fig. 4.6A). We saw that double knockout cells did not display further decreased cfDNA release compared to FADD only mutants, indicating that these two genes likely are involved in the same pathway (Fig 4.6B). To further define the interplay between FADD and Sam68, we overexpressed FADD and Sam68 in opposing knockout cell lines. In doing so, we found that FADD overexpression in a Sam68 knockout led to an increase in cfDNA release, while Sam68 overexpression in a FADD knockout background did not lead to rescue (Figure 4.6C). This implies that FADD is required for Sam68's impact on cfDNA release, but Sam68 is not required for FADD's function. In terms of evidence for a direct physical interaction between these at the TRAIL receptor, previous literature suggests that Sam68 can be both a nuclear protein and a cytoplasmic protein[154,164,165].

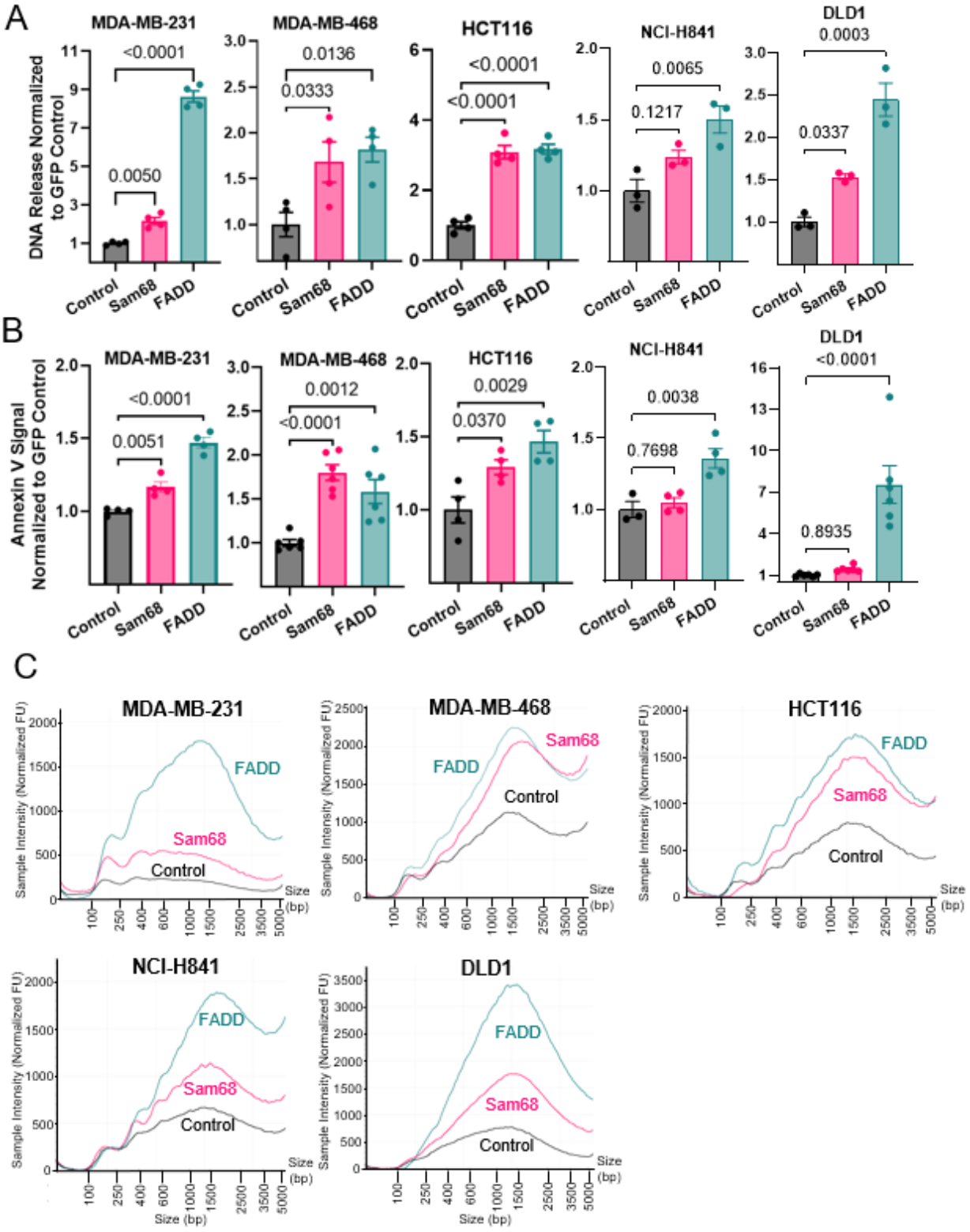
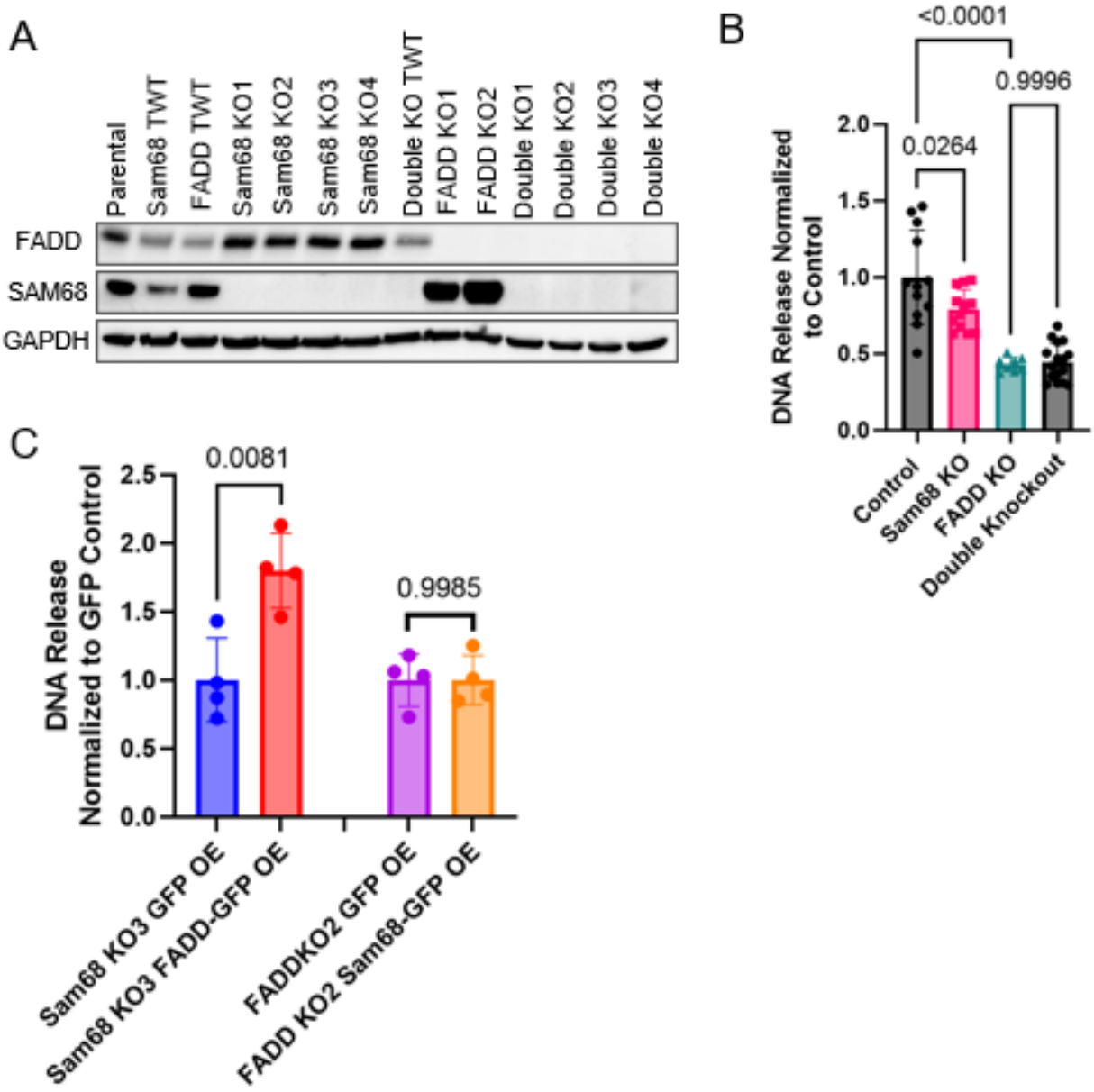

A
MDA-MB-231
MDA-MB-468
HCT116
NCI-H841
DLD1
DNA Release Normalized to GFP Control
<0.0001
0.0050
0.0136
0.0333
<0.0001
<0.0001
0.0065
0.1217
0.0003
0.0337
Control
Sam68
FADD
B
Annexin V Signal Normalized to GFP Control
<0.0001
0.0051
0.0012
<0.0001
0.0029
0.0370
0.0038
0.7698
<0.0001
0.8935
C
Sample Intensity (Normalized FU)
Size (bp)
100
250
400
600
1000
1500
2500
3500
5000

Figure 4.5. Overexpression of Sam68 and FADD in various cancer cell lines. (A) Quantification of DNA release from 5 cancer cell lines with overexpression of GFP-tagged Sam68 or FADD. Data represent mean fold change ± SEM in DNA release normalized to cell concentration for each cell line, then overall to GFP control; n=4. (B) Quantification of Annexin V signal from five cancer cell lines with overexpression of GFP-tagged Sam68 and FADD. Data represent mean fold change ± SEM in RLU signal normalized to cell concentration at collection for each cell line, then overall to GFP control; n=4. (C) Fragmentation pattern of five cancer cell lines with overexpression of GFP-tagged Sam68 and FADD. Electropherograms were individually run at least n=3 times and representative traces were selected.



Supplementary Fig. 4.6. Manipulation of Sam68 and FADD in MCF-10A background indicates shared pathway. (**A**) Immunoblot analysis of Sam68 and FADD after CRISPR-mediated double knockout in the MCF-10A background. (**B**) Quantification of DNA release from MCF-10A KO and double KO cell lines in culture. Parental cells and TWT cells were averaged and labeled control. All four Sam68 KO, both FADD KO, and all four double knockout lines initially derived from Sam68 KO3 and Sam68 KO4 are respectively grouped. Data represent mean fold change ± SEM internally normalized to cell concentration for each cell line and then normalized to parental MCF-10A; n=4 for each cell line before combining. (**C**) Quantification of DNA release from MCF-10A KO and over-expression lines. Parental cells and TWT cells were averaged and labeled control. One Sam68 and one FADD KO line were infected with either GFP-only lentivirus, or a GFP-tagged version of the opposing protein. Data represent mean fold change ± SEM internally normalized to cell concentration for each cell line and then normalized to parental MCF-10A; n=3 for each cell line before combining.

Within our models, nuclear/cytoplasmic fractionation of endogenous Sam68 and the localization of an overexpressed Sam68-GFP fusion protein revealed Sam68 is exclusively a nuclear protein and is likely not involved in receptor activation with FADD in our models (Fig. 4.7A, Fig 4.7B). These results indicate that Sam68 plays a role in apoptotic regulation and FADD is required for this function, but the mechanistic link between these two genes in our model system remains unknown. Previous studies have shown that Sam68 can splice the BCL-X mRNA product of the *BCL2L1* gene, and depletion of Sam68 leads to accumulation of the anti-apoptotic product BCL-XL spliceoform over the pro apoptotic BCL-XS[155,166]. Given that Sam68 knockout MCF-10A cells display decreased apoptosis, we hypothesized that Sam68 may be mediating the splicing of this apoptosis-related protein. However, we were unable to observe differences in the expression of these products at the mRNA level by RT-PCR (Fig 4.7C). Further delineation of Sam68's role in apoptosis and the TRAIL pathway will be the focus of future study.

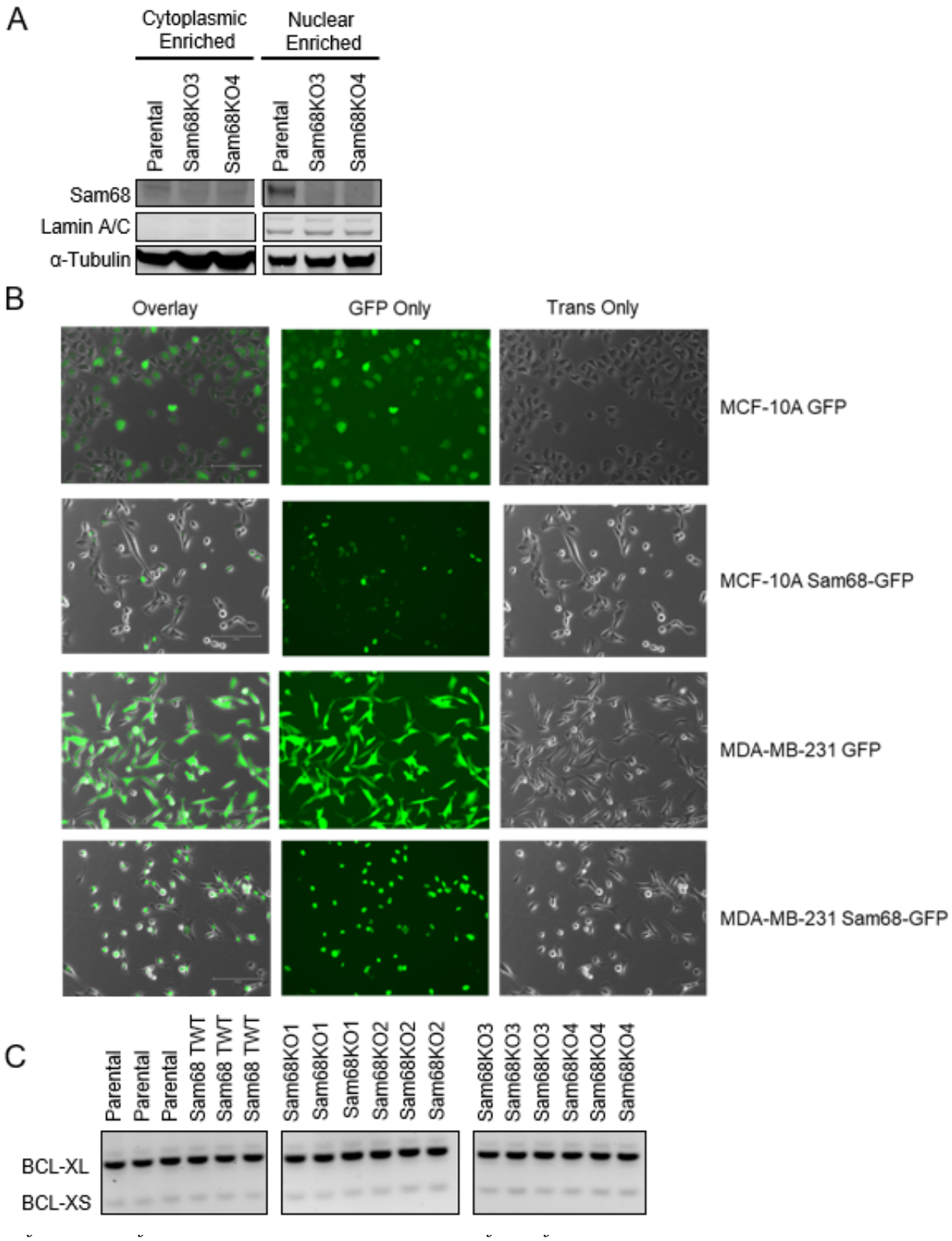


Figure 4.7. Role of Sam68 in MCF-10A Background. (A) Immunoblot analysis of Sam68 in nuclear and cytoplasmic enriched fractions of MCF-10A Sam68 KO cells. (B) Localization of GFP or GFP-tagged Sam68 in MCF10A and MDA-MB-231 cells. Scale bar is 150uM. (C) Splicing of BCL-X into BCL-XL and BCL-XS in MCF-10A Sam68 KO cell lines.

Confirmation of TRAIL as a mediator of cell-free DNA release

Given the top cfCRISPR screen candidates were primarily in the TRAIL pathway, the knockouts were next evaluated to determine if they were resistant to TRAIL-induced apoptotic cell death. Both Sam68 and FADD knockout lines were resistant to TRAIL-induced apoptosis, with FADD knockouts demonstrating complete resistance, and Sam68 showing partial resistance (Fig. 4.8A). In turn, TRAIL administration led to increased cfDNA release in the non-edited controls as well as a muted increase in the partially resistant Sam68 knockout lines, but did not alter release in FADD knockouts (Fig. 4.8B). We also performed a similar analysis in the MDA-MB-468 breast cancer cell line, finding that Sam68 or FADD knockout alone did not lead to changes in cfDNA release. However, upon TRAIL administration Sam68 and FADD knockout lines displayed decreased cfDNA release compared to treated wild-type cells, further validating Sam68 and FADD as mediators of the TRAIL pathway and cfDNA release (Fig. 4.9A, 4.9B).

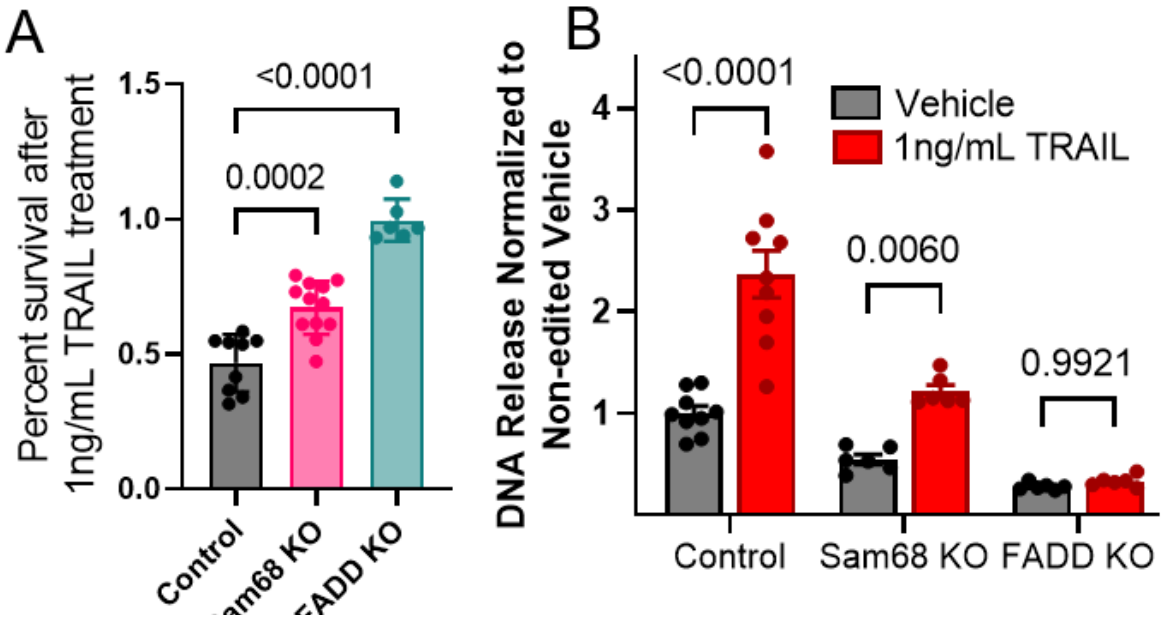


Figure 4.8. Effects of pharmacologic apoptotic pathway modulation on MCF-10A KO cells. (A) Cytotoxic assay of MCF-10A KO panel treated with 1ng/mL TRAIL ligand. Parental and TWT were grouped and labeled control. All four Sam68 mutant cell lines and both FADD mutant cell lines were grouped, respectively. Data represent mean percent survival ± SEM as normalized to vehicle of each cell line condition; n=3 for each cell line before combining. (B) Quantification of DNA release from MCF-10A KO panel treated with 1ng/mL TRAIL ligand. Parental and TWT were grouped and labeled control. All four Sam68 mutant cell lines and both FADD mutant cell lines were respectively grouped. Data represent mean fold change ± SEM internally normalized to cell concentration for each cell line and then normalized to control; n= 3 for each cell line before combining.

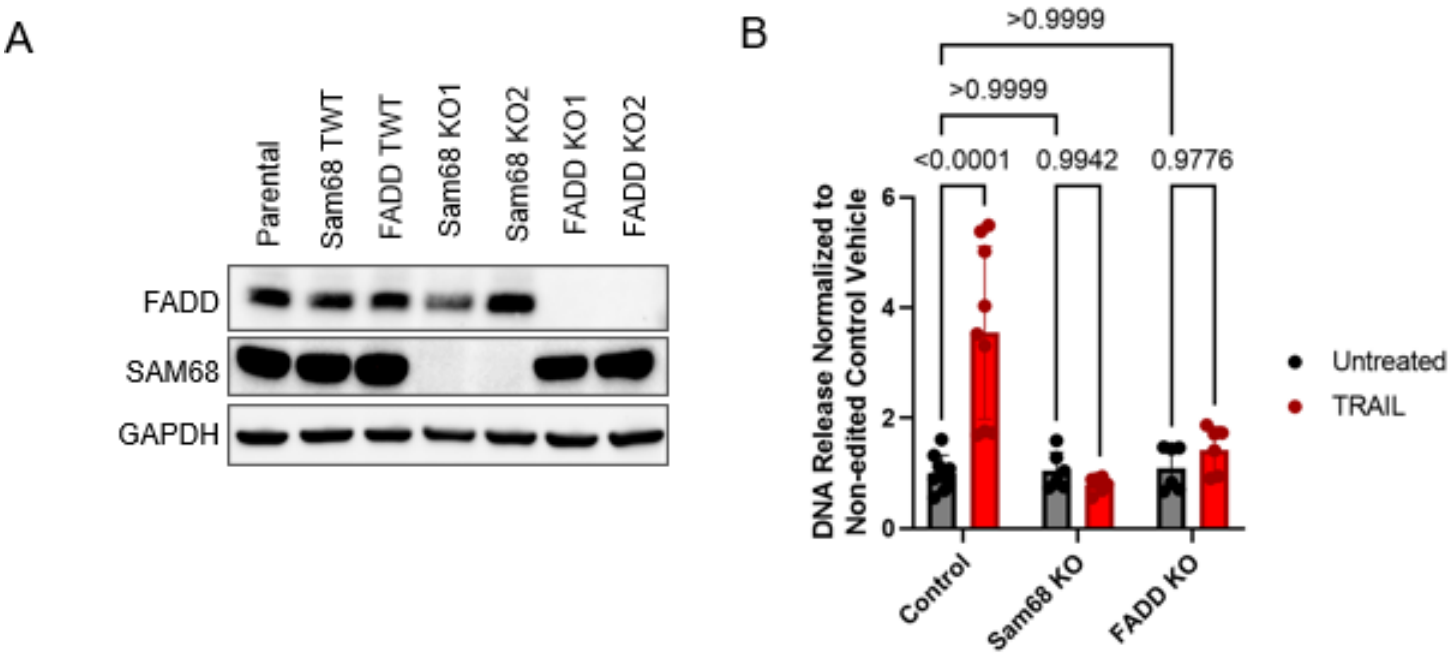


Figure 4.9. Manipulation of MCF-10A screen hits in non-MCF-10A backgrounds. (A) Immunoblot analysis of Sam68 and FADD after CRISPR-mediated double knockout in the MDA-MB-468 background. **(B)** Quantification of DNA release from MDA-MB-468 KO cell lines in culture. Parental cells and TWT cells were averaged and labeled control. Both Sam68 KO and FADD KO knockout lines initially derived are respectively grouped. Data represent mean fold change ± SEM internally normalized to cell concentration for each cell line and then normalized to parental MDA-MB-468; n=3 for each cell line before combining.

To determine the generalizability of TRAIL to initiate cfDNA release through apoptotic mechanisms, we also validated its effects on the same cancer cell lines used to validate Sam68 and FADD. In these cell lines, a dose dependent increase in cfDNA release was observed (Fig. 4.10A). These gains in cfDNA release were associated with increases in both Annexin V and PI assays, suggesting these responses were mediated by increased apoptosis (Fig. 4.10B, 4.10C). Again, this manipulation did not modify the inherent fragmentation patterns of the cell lines (Fig. 4.10D).

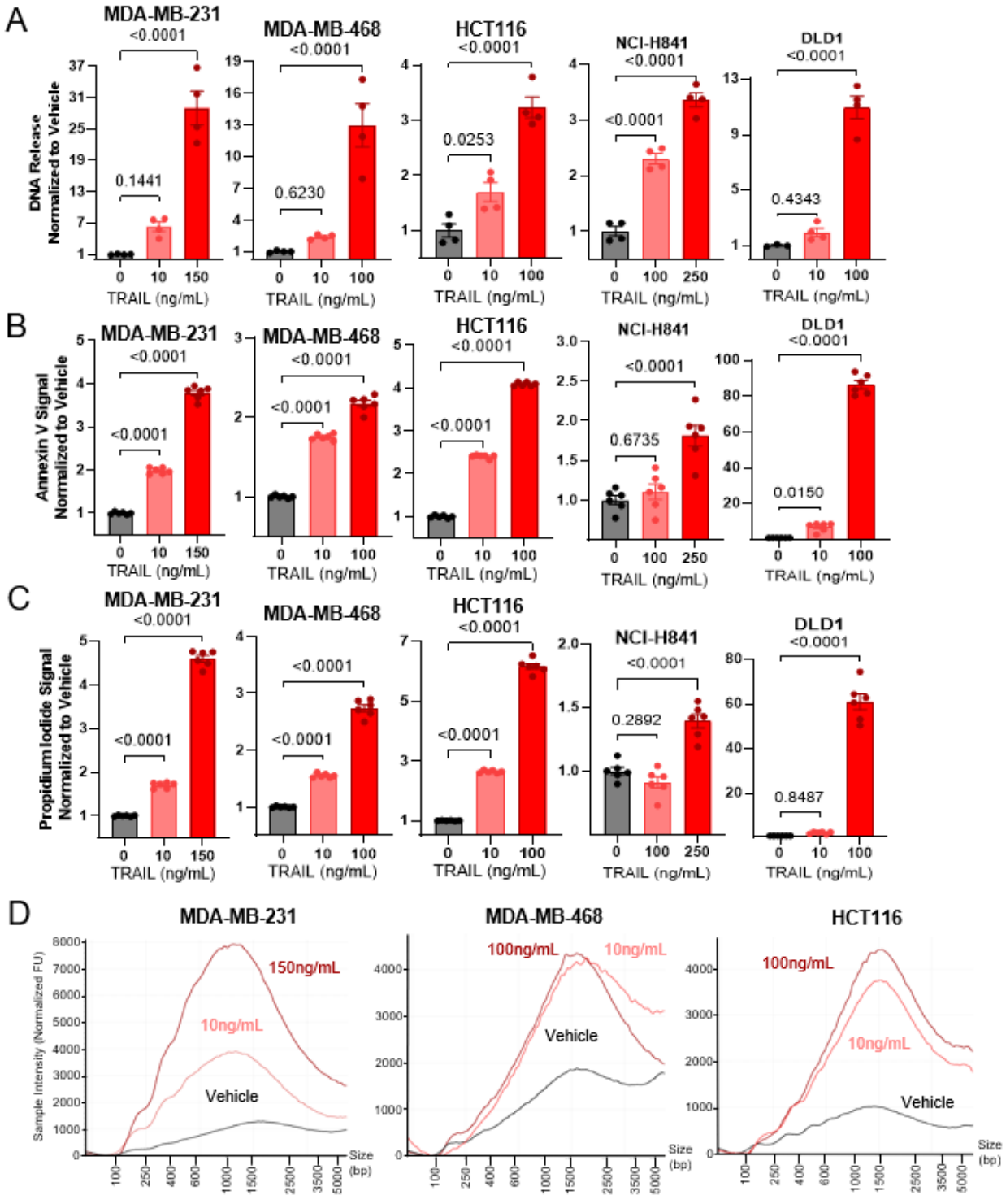


Figure 4.10. TRAIL treatment of various cancer cell lines. (A) Quantification of DNA release from DLD1 cancer cell lines treated with TRAIL ligand. Data represent mean fold change ± SEM in DNA release normalized to cell concentration at collection for each cell line then overall to GFP control; n=4. (B) Quantification of Annexin V signal from cancer cell lines treated with TRAIL Ligand. Data represent mean fold change ± SEM in signal (RLU) normalized to cell concentration at collection for each cell line, then overall to vehicle control; n=6. (C) Quantification of Propidium Iodide signal from cancer cell lines treated with TRAIL Ligand. Data represent mean fold change ± SEM in signal (RFU) normalized to cell concentration at collection for each cell line, then overall to vehicle control; n=6.All statistics were ANOVA with Dunnett's multiple comparison test at endpoint. (D) Fragmentation pattern of cancer cell lines treated with TRAIL ligand. Electropherograms were individually run at least n=3 times and representative traces were selected.

Confirmation of BCL2L1 as a mediator of cell-free DNA release

BCL2L1 and *MCL1* were identified as significant hits in the A549 CRISPR screen at the late timepoint and are both members of the BCL-2 family. *BCL2L1* was also the third strongest hit in the early timepoint (Table 5). BCL-XL, one of the gene products of BCL2L1, as well as MCL-1, are anti-apoptotic multi-domain members of the BCL-2 family that regulate apoptosis by preventing mitochondrial membrane permeabilization, and their simultaneous targeting is considered synthetically lethal in CD34+ hematopoietic stem cells[167,168]. To test whether these genes positively regulated cfDNA release, four BCL2L1 knockout cell lines were generated using two sgRNAs (Fig. 4.11A, 4.11B, 4.11C). Interestingly and unexpectedly, cfDNA release increased but the fragmentation pattern of the knockout lines was skewed towards larger fragments (Fig. 4.12A, 4.12B). This change was concomitant with increases in baseline apoptotic indices (Fig 4.12 C, 4.12D), indicating that these effects were likely through apoptosis. Knockout of *BCL2L1* was also performed in two cell lines without single cell dilution (Fig. 4.13A). This CRISPR "knockdown" in a bulk population led to an increase in cfDNA release in two cell lines tested (Fig. 4.13B) akin to the *BCL2L1* knock out in A549 cells, further corroborating BCL2L1 as a modifier of cfDNA release.

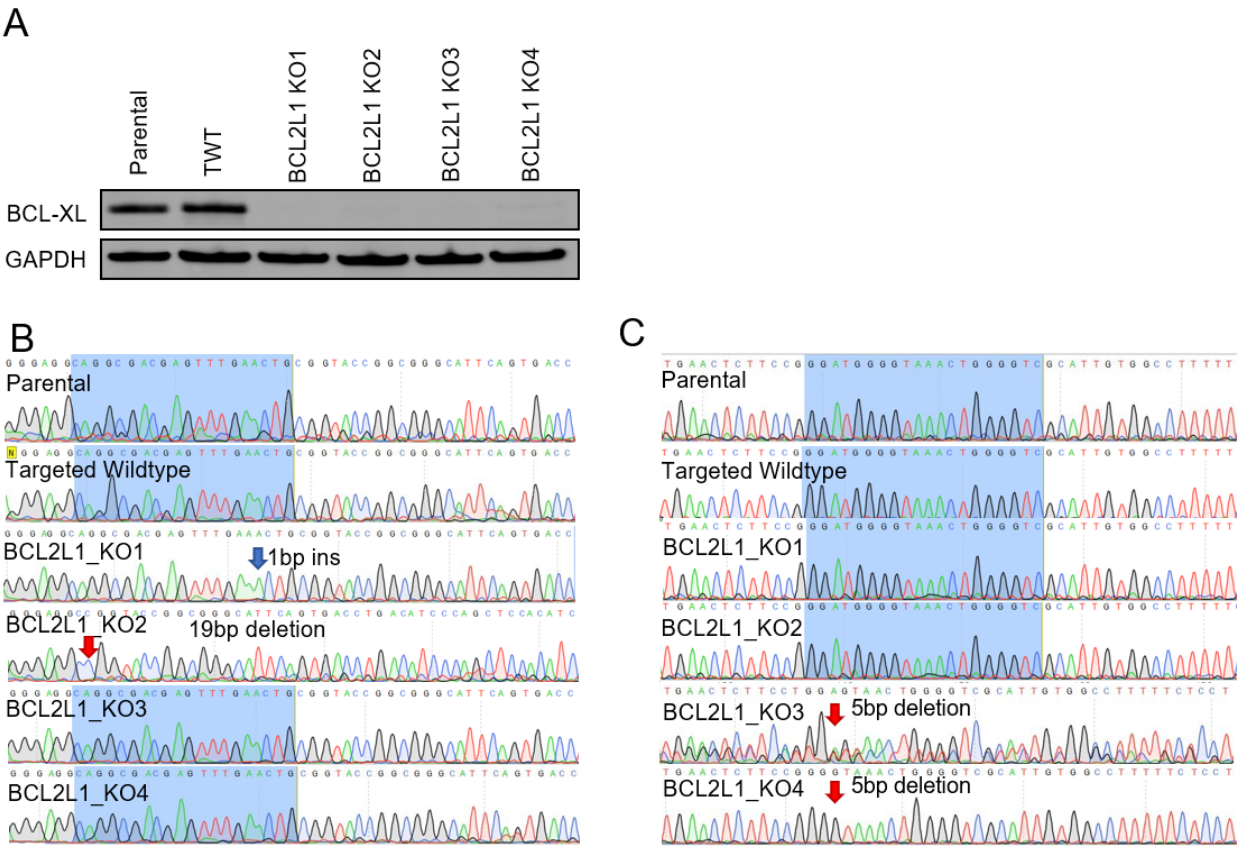


Figure 4.11. Validation of BCL-XL knockout in MCF-10A. (**A**) Immunoblot analysis of BCL-XL after CRISPR-mediated knockout (KO) in the A549 Background. TWT = Targeted Wild-Type. (**B**) Sequence of BCL2L1 clones at the cut-site of BCL2L1_sgRNA_1, with sgRNA sequence highlighted in blue. (**C**) Sequence of BCL2L1 clones at the cut-site of BCL2L1_sgRNA_2, with sgRNA sequence highlighted in blue.

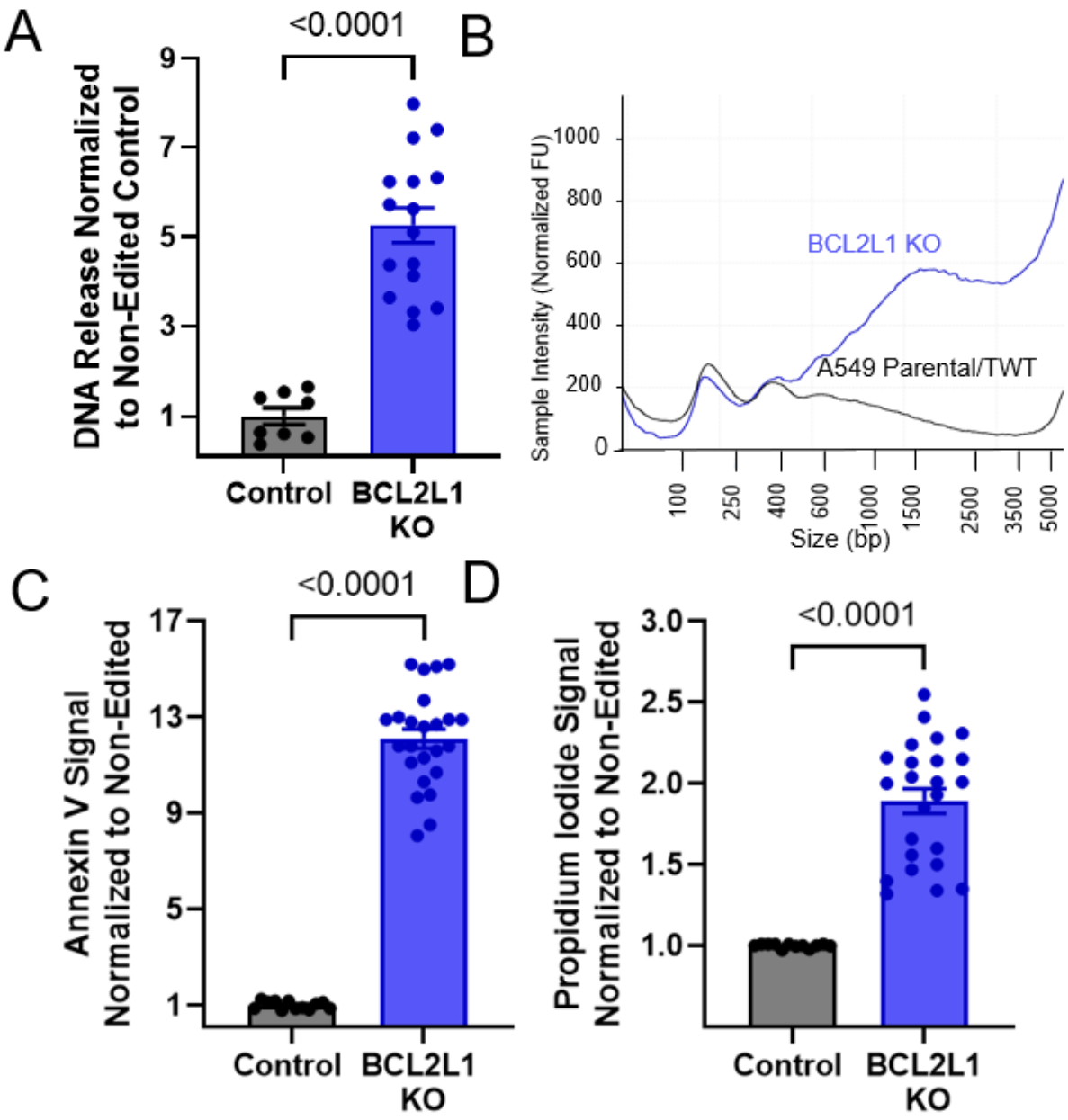


Figure 4.12. Confirmation of BCL-XL as a regulator of cfDNA release in A549 background. (**A**) Quantification of DNA release from A549 KO cell lines in culture. Parental cells and TWT cells were averaged and labeled control. All four *BCL2L1* KO cell lines were grouped. Data represent mean fold change ± SEM in DNA release normalized to cell concentration for each cell line then normalized to control; n=4 for each cell line before combining. (**B**) Electropherograms were individually run at least n=3 times and representative traces were selected. (**C**) Annexin V and (**D**) Propidium Iodide assay on all generated A549 cell lines. Parental cells and TWT cells are combined. All four *BCL2L1* KO cell lines were grouped. Data represent mean signal (RFU for PI or RLU for Annexin) ± SEM normalized to cell concentration then overall to Control; n=6 for each cell line before combining.

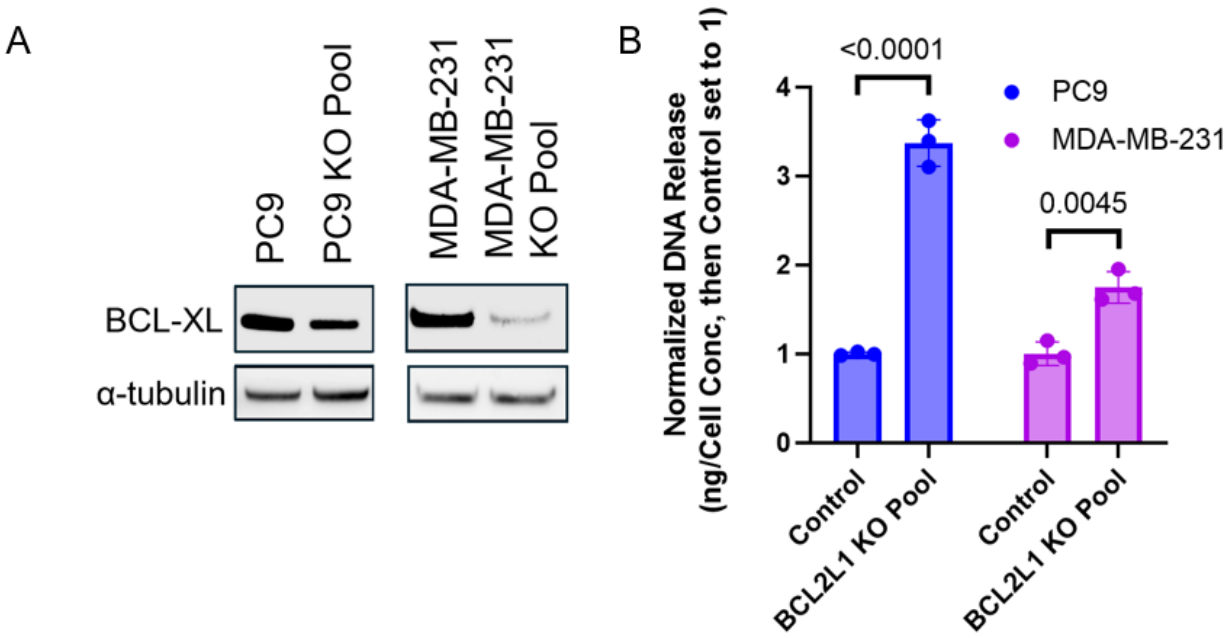


Figure 4.13. Manipulation of *BCL2L1* in additional cell lines. (A) Immunoblot analysis of BCL-XL in CRISPR knockout pooled cell lines in the PC9 and MDA-MB-231 cell lines. **(B)** Quantification of DNA release from pooled KO cell lines in culture. Control cells are parental cells treated only with Cas9. Data represent mean fold change ± SEM internally normalized to cell concentration for each cell line and then normalized to control for each line; n=3 for each cell line or cell pool.

Confirmation of apoptosis as a broad mediator of cell-free DNA release

To further elucidate the relationship between cfDNA release and apoptosis, Annexin V and PI signals were next quantified across the 24-cell line panel and compared with cfDNA release. There was a significant correlation between cfDNA release and both baseline apoptotic indices (Fig. 4.14A, 4.14B). In addition, cell lines were treated with the pan-caspase inhibitor ZVAD-FM-K, which has been known to block all forms of apoptosis, in order to determine the degree to which DNA released from cells is apoptotic in nature. Treatment with ZVAD-FM-K decreased cfDNA release across all cancer cell lines tested by nearly 75%, but did not alter fragmentation pattern skewing, suggesting again that the majority of DNA release from cells is apoptotic, and apoptotic release does not consistently lead to the putative apoptotic fragmentation pattern of 167bp with a continuing ladder pattern (Fig. 4.14C, 4.14D). When applying this drug to the MCF-10A cell line, nearly all cfDNA release was abolished, and even the very low releasing FADD KO had its cfDNA release further diminished (Fig. 4.14E). Together, these results suggest that generally apoptosis is the major mechanism of cell-free DNA release over all previously proposed mechanisms, and the prescribed fragmentation pattern of apoptotic DNA as 167bp has been inaccurate, or at least not accurate in all cases of apoptotic processes.

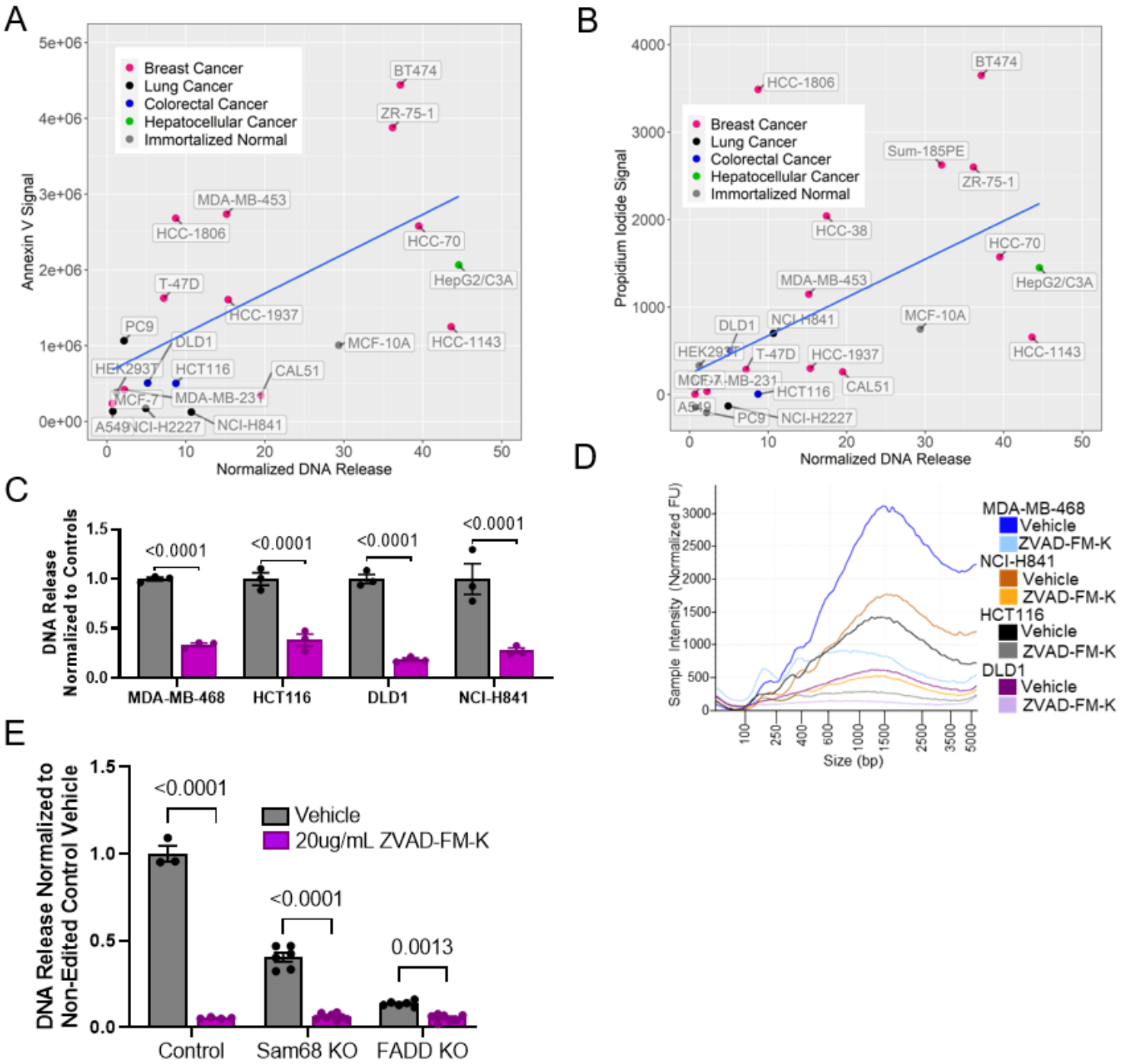


Figure 4.14. Apoptotic processes regulate cfDNA release across all cell types tested. (A) Quantification of DNA release from A549 KO cell lines in culture. Parental cells and TWT cells were averaged and labeled control. All four *BCL2L1* KO cell lines were grouped. Data represent mean fold change ± SEM in DNA release normalized to cell concentration for each cell line then normalized to control; n=4 for each cell line before combining. **(B)** Electropherograms were individually run at least n=3 times and representative traces were selected. **(C)** Annexin V and **(D)** Propidium Iodide (PI) assay on all generated A549 cell lines. Parental cells and TWT cells are combined. All four *BCL2L1* KO cell lines were grouped. Data represent mean signal (RFU for PI or RLU for Annexin) ± SEM normalized to cell concentration then overall to Control; n=6 for each cell line before combining. **(E)** Quantification of DNA release from MCF-10A KO cell lines treated with 20ug/mL ZVAD-FM-K. Parental and TWT were grouped and labeled control. Two Sam68 mutant cell lines and both FADD mutant cell lines were respectively grouped. Data represent mean fold change ± SEM in DNA release internally normalized to cell concentration for each cell line and then normalized to control vehicle; n=3 for each cell line before combining.

Discussion

This section explored a subset of potential cfDNA-modulating hits in the apoptotic pathway collected from the CRISPR screens for cfDNA release performed in Chapter III. While previous research has indicated through mechanistic approaches that apoptosis was likely a major regulator of cell free DNA release[82], multiple reports also showed a lack of correlation between apoptosis and cfDNA or ctDNA release[86-88]. Here, using genetic and pharmacologic manipulations of both specific and general mediators of apoptosis in various cell models with different cfDNA release characteristics, we unequivocally confirm that cfDNA release is mainly controlled by apoptosis. All three gene knockout groups generated to confirm our CRISPR screens for cfDNA release in Sam68, FADD, and BCL-XL, showed expected positive or negative modifications in cfDNA release. All three of these genes play direct or indirect roles in apoptosis. The over-expression of Sam68 and FADD and subsequent increases seen in cfDNA release and pro-apoptotic phenotypes across multiple cancer cell lines indicates the generalizability of these molecules as positive regulators of cfDNA release through apoptosis. Meanwhile, the knockout of anti-apoptotic BCL-XL lead to increases in cfDNA release, indicating its negative regulation of the process of cfDNA release, with its generalizability confirmed through knockdown in two additional cell lines.

In addition to these specific molecular regulators of apoptosis, we showed that generalized apoptotic processes lead to increased cfDNA release. Higher propensity to undergo apoptosis at baseline, measured through Annexin V and PI stainings, lead to a positive relationship with cfDNA release – a phenotype that has been been somewhat inconsistent across studies[86-88]. Here, we find this trend where others did not likely due to the large number of individual datapoints used from various cell lines. In addition, these previous studies had attempted to

correlate DNA release to apoptosis induced during drug studies[87,88]. Treating the cells with various drugs could lead to increased variance, especially since not all cell-death inducing or blocking drugs are equivalently effective across cell lines. Further, we also blocked apoptosis with the pan-caspase inhibitor ZVAD-FM-K in a variety of cell lines. While there was a variety of responses depending on cell line, all cell lines decreased their DNA release by greater than 50%. MCF-10A displayed an extremely strong response to caspase inhibition, with cfDNA release being almost completely abrogated. The cancer cell lines tested did not display such a strong decreases in cfDNA release, although most displayed a roughly 75% decrease. These large reductions indicate that while apoptosis may not be the only process of cfDNA release in cancer cells especially, it is the major release mechanism. This discrepancy may be attributable to cancer-specific or associated phenotypes that could potentially predispose to cfDNA release through alternative mechanisms such as vesicle production and DNA damage.

Perhaps the most interesting finding regarding the administration of various pro-and anti-apoptotic stimuli was the general fragmentation pattern of DNA released from cell lines. When left shifted cell lines such as MCF-10A decreased their DNA release due to genetic deletion of pro-apoptotic genes, their fragmentation pattern decreased overall in amplitude. However, upon pro-apoptotic over-expression of Sam68 and FADD in this cell line, the fragment lengths released increased across the spectrum of sizes measured, including at the expected left skewed 167 peak, but also at larger sizes. When the anti-apoptotic BCL-XL protein was knocked out in left skewed, low-releasing A549 lung cancer cell lines, cfDNA was released primarily at larger fragment sizes than the typical left-skewed pattern of this cell line. In all right-skewed cancer cell lines treated with pro- or anti-apoptotic stimuli, fragmentation patterns didn't change in skewing, only in amplitude. These large DNA sizes were previously attributed to necrosis or

vesicular DNA[86,97,98,133]. However, these results indicate that these large DNAs are apoptotic in nature. This is in line with Jeppessen et al., who indicated that little to no cfDNA was contained within small vesicles and that DNA of large fragment sizes similar to those we now see as apoptotic co-purify with non-vesicular fractions of improved small extracellular vesicle prerparations[103]. Together, our studies indicate that DNA fragments found in extracellular vesicle preparations have been apoptotic fragments sometimes stuck to the outside of the extracellular vesicles. This finding also challenges the canonical fragmentation pattern of apoptotic DNA being 167bp[82]. Instead, this work implies that DNA is not always released from cells at the smaller "apoptotic" sizes seen in patient blood samples, but rather may undergo post-release processing in the blood that results in smaller sizes. Together, these results show that specific and general apoptotic stimuli control cfDNA release to a much higher level than any other process yet discovered, and challenges the assumptions previously made about fragmentation patterning of cfDNA released both in cell culture and in living organisms.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

Sections of this chapter have been previously published in: Davidson, *et al*. 2024. An in vitro CRISPR screen of cell-free DNA identifies apoptosis as the primary mediator of cell-free DNA release. *Communications Biology*, provisionally accepted.

Conclusions & Future Directions

cfDNA is now routinely used in the clinical setting to help guide the diagnosis and treatment of patients in many distinct areas of medicine. In particular, for oncology ctDNA-based liquid biopsies have shown great promise as tools for guiding cancer treatment but continue to have limitations in sensitivity and negative predictive value[169]. Most attempts to address these shortcomings have evaluated the use of additional analytes or applied advanced sequencing bioinformatic techniques, although a recent study has also implicated pre-treatment with inhibitors of DNA degradation[40,44,51,56,117,118,170,171]. However, the mechanisms of how cfDNA is liberated into circulation has not been proposed as a potential means of increasing cfDNA release and therefore sensitivity of liquid biopsies. Herein, we present genetic evidence that apoptotic pathways are a major regulator of cfDNA release using novel cfDNA-based CRISPR screens, single gene knockouts, overexpression rescue studies, and drug treatments across multiple human cell lines. In contrast to previous literature suggesting roles for vesicles as well as necrosis in cfDNA/ctDNA release[98,100,101,106,133,134], our screens did not implicate these processes. Our DNA isolation methods purposefully retain most large and small vesicle populations, and the lack of vesicle-related genes seen in both screens indicates a minor or null role for vesicle populations on cfDNA release. Importantly, we instead show that cfDNA released from human cells is primarily derived from apoptotic pathways. Even Sam68, a known RNA binding protein, mediated its effects on cfDNA release in part through apoptotic pathways as measured by changes in Annexin V and

other biologic parameters. Interestingly, our screens in the MCF-10A and A549 cell lines revealed hits in different parts of the apoptotic pathway, with *FADD* in the MCF-10A screen being a mediator of extrinsic cell death pathways and *BCL2L1* in the A549 screen being an inhibitor of intrinsic apoptosis. The identification of these cell line-specific hits reflects the biological differences between the two cell lines. A549 cells are known to be resistant to TRAIL ligand, and MCF-10As display very low expression of BCL-XL, the apoptosis-inhibiting spliceoform of *BCL2L1*[144,172,173]. Thus, it would have been surprising to see these hits or other similar ones identified across both screens. The convergence of non-overlapping hits from both screens onto the same overarching biological process of apoptosis confirms the validity of our findings and is not uncommon in studies CRISPR screening in multiple cell types[113,174]. Taken together, these screens support our findings that apoptosis is a major mediator of cfDNA release. However, hits we generated but did not explore (Tables 3.4, 3.5, 3.6) could represent additional pathways that regulate cfDNA release. Though we did not explore many of these hits, due in part to the relative obscurity of many of the genes identified, they may represent exciting targets for future study. Studies of cfDNA not related to specific hits found by our screens that are still critical to fully understanding the basic biology of cfDNA release include the evaluation of cfDNA release after modification of vesicular release and DNA damage repair pathways.

We also found that the fragmentation pattern of DNA released through these apoptotic pathways can contain larger sizes traditionally assumed to be released through vesicular or necrotic pathways, rather than the expected 167bp size expected following apoptotic DNA cleavage. This was visualized in multiple right-skewed cell lines increasing cfDNA quantities at sizes >1000bp after apoptosis induction through Sam68, FADD, and TRAIL treatment. In addition, the left-skewed A549 began skewing rightwards upon loss of anti-apoptotic *BCL2L1*, and the left-skewed

MCF-10A began skewing more rightwards upon overexpression of pro-apoptotic Sam68 and FADD. This suggests that apoptotic DNA is not always released from cells at the smaller "apoptotic" sizes seen in patient blood samples and assumed in the literature, but rather may undergo post-release processing in the blood that results in smaller "left-skewed" sizes akin to those found in our cell-free DNA degradation experiments.

In addition, we present a new cfDNA CRISPR screening modality in cfCRISPR that may prove useful for other applications. Traditionally, CRISPR screens require the isolation of DNA from tens of millions of cells at any given timepoint. In analyzing cfDNA, we saw that most genes were not discrepant between genomic and cell-free DNA in their representation. This provides possibilities for future screens to serially identify genes of interest over time using sequential harvesting of culture media containing cfDNA instead of or in addition to gDNA. In theory this also opens the door to use CRISPR screens to study other DNA-related processes, such as DNA trafficking into multi-vesicular bodies by harvesting DNA from non-nuclear populations[103]. Thus, in this work we not only improve our understanding of cfDNA biology, but also provide a new platform to perform genome wide CRISPR screens in non-nuclear DNA populations.

While our results come from *in vitro* models, we believe they may have implications for *in vivo* and patient settings, to be assessed in future studies. The finding that we can increase or decrease cfDNA and ctDNA by specific gene knock out *in vitro* raises provocative questions with clinical implications. For example, if a cancer patient has a high variant allele fraction for a specific mutation with ctDNA testing, does this necessarily represent the predominant clonal population or a subclone that is prone to higher cfDNA release due to inactivation or activation of certain genes? In addition, although recombinant TRAIL and agonist TRAIL receptor therapies did not demonstrate efficacy in clinical trials, there is the possibility that these drugs, as well as many

other agents with distinct functions, could be repurposed as "ctDNA adjuvant diagnostics". Although such drugs may increase both ctDNA and total cfDNA release from normal cells, this may still provide meaningful benefit since the sensitivity and negative predictive value of liquid biopsies are often limited by the lack of total cfDNA, and not necessarily by the relatively low amount of ctDNA to total cfDNA. Indeed, some current commercial ctDNA assays state a technical lower limit of detection well below one mutant molecule per 100,000 wildtype cfDNA molecules (0.001% allele fraction)[175]. However, 100,000 cfDNA molecules is equivalent to approximately 300 ng of cfDNA (one haploid genome = 3 picograms of DNA), and the amount of cfDNA obtained for clinical testing is often far below this threshold to achieve such sensitivity and therefore false negatives remain problematic. Increasing total cfDNA, along with ctDNA, could solve this current unmet need, and allow for highly sensitivity and specific testing that may guide clinical care with precision.

Concluding Thoughts

The work presented in this dissertation presents evidence for multiple paradigm shifts in the fields of basic cfDNA biology and clinical liquid biopsy testing. By studying the cfDNA release properties of a large cell line panel, we identified that cell lines release cfDNA in various amounts and sizes in a manner caused by intrinsic biological properties of specific lines. However, we identified that the main release mechanism of DNA across these lines is apoptotic, no matter the cfDNA release characteristics of the cell line. This goes against previous literature which labeled cfDNA released as larger or smaller fragments as released by different mechanisms. Together, these insights highlight the need for further study as to what characteristics of individual cell lines lead to these divergent cfDNA release properties. For example, delineating how apoptosis can lead to DNA release at different fragmentation sizes in

different cell lines will be a key question for future study. In addition, the development of cfCRISPR as a CRISPR screening platform that can be used to screen for changes in DNA related phenotypes, such as DNA release or DNA packaging into various vesicular bodies, could lead to myriad future discoveries in various fields including exosome biology, DNA damage, and cfDNA release. In addition, the ability of specific biological and chemical manipulations to modify cfDNA release from cells in culture could help shift the field of liquid biopsy away from attempting to improve the sensitivity and specificity of liquid biopsy diagnostics on the back end through improved detection technologies towards technologies which increase DNA input into assays in the first place. Finally, this work could be extended into experiments attempting to explain in part why some cancer patients have extremely high and extremely low ctDNA burdens, as varying propensities of their cancers towards apoptosis or other release factors could be at play. Together, this work answers timely questions in the field of cfDNA release biology while presenting multiple avenues for further investigation into cfDNA release *in vitro* and in vivo. Continued study in this field will support efforts to improve ctDNA-based liquid biopsies for cancer that could someday greatly impact cancer diagnosis, prognosis, and treatment paradigms.

REFERENCES

1 Doroshow, D. B. & Doroshow, J. H. Genomics and the History of Precision Oncology. *Surg Oncol Clin N Am* **29**, 35-49 (2020).

2 Meisel, J. L., Venur, V. A., Gnant, M. & Carey, L. Evolution of Targeted Therapy in Breast Cancer: Where Precision Medicine Began. *Am Soc Clin Oncol Educ Book* **38**, 78-86 (2018).

3 Aldea, M. *et al.* Overcoming Resistance to Tumor-Targeted and Immune-Targeted Therapies. *Cancer Discov* **11**, 874-899 (2021).

4 Overman, M. J. *et al.* Use of research biopsies in clinical trials: are risks and benefits adequately discussed? *J Clin Oncol* **31**, 17-22 (2013).

5 Dowlati, A. *et al.* Sequential tumor biopsies in early phase clinical trials of anticancer agents for pharmacodynamic evaluation. *Clin Cancer Res* **7**, 2971-2976 (2001).

6 Gomez-Roca, C. A. *et al.* Sequential research-related biopsies in phase I trials: acceptance, feasibility and safety. *Ann Oncol* **23**, 1301-1306 (2012).

7 Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883-892 (2012).

8 Wan, J. C. M. *et al.* Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* **17**, 223-238 (2017).

9 Ignatiadis, M., Sledge, G. W. & Jeffrey, S. S. Liquid biopsy enters the clinic - implementation issues and future challenges. *Nat Rev Clin Oncol* **18**, 297-312 (2021).

10 Mandel, P. & Metais, P. [Nuclear Acids In Human Blood Plasma]. *C R Seances Soc Biol Fil* **142**, 241-243 (1948).

11 Leon, S. A., Shapiro, B., Sklaroff, D. M. & Yaros, M. J. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res* **37**, 646-650 (1977).

12 Tan, E. M. & Kunkel, H. G. Characteristics of a soluble nuclear antigen precipitating with sera of patients with systemic lupus erythematosus. *J Immunol* **96**, 464-471 (1966).

13 Neuberger, E. W. I. *et al.* Physical activity specifically evokes release of cell-free DNA from granulocytes thereby affecting liquid biopsy. *Clin Epigenetics* **14**, 29 (2022).

14 Anker, P., Lyautey, J., Lederrey, C. & Stroun, M. Circulating nucleic acids in plasma or serum. *Clin Chim Acta* **313**, 143-146 (2001).

15 Fan, G. *et al.* Prognostic value of circulating tumor DNA in patients with colon cancer: Systematic review. *PLoS One* **12**, e0171991 (2017).

16 Pietrasz, D. *et al.* Plasma Circulating Tumor DNA in Pancreatic Cancer Patients Is a Prognostic Marker. *Clin Cancer Res* **23**, 116-123 (2017).

17 Fernandez-Garcia, D. *et al.* Plasma cell-free DNA (cfDNA) as a predictive and prognostic marker in patients with metastatic breast cancer. *Breast Cancer Res* **21**, 149 (2019).

18 Zhang, Y. *et al.* Pan-cancer circulating tumor DNA detection in over 10,000 Chinese patients. *Nat Commun* **12**, 11 (2021).

19 Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, 224ra224 (2014).

20 Phallen, J. *et al.* Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med* **9** (2017).

21 De Mattos-Arruda, L. *et al.* Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat Commun* **6**, 8839 (2015).
22 Lone, S. N. *et al.* Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments. *Mol Cancer* **21**, 79 (2022).
23 Toro, P. V. *et al.* Comparison of cell stabilizing blood collection tubes for circulating plasma tumor DNA. *Clin Biochem* **48**, 993-998 (2015).
24 Takousis, P. *et al.* A standardised methodology for the extraction and quantification of cell-free DNA in cerebrospinal fluid and application to evaluation of Alzheimer's disease and brain cancers. *N Biotechnol* **72**, 97-106 (2022).
25 Casadio, V. *et al.* Cell-Free DNA Integrity Analysis in Urine Samples. *J Vis Exp* 10.3791/55049 (2017).
26 Augustus, E. *et al.* The art of obtaining a high yield of cell-free DNA from urine. *PLoS One* **15**, e0231058 (2020).
27 Hindson, C. M. *et al.* Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat Methods* **10**, 1003-1005 (2013).
28 Miyazawa, H. *et al.* Peptide nucleic acid-locked nucleic acid polymerase chain reaction clamp-based detection test for gefitinib-refractory T790M epidermal growth factor receptor mutation. *Cancer Sci* **99**, 595-600 (2008).
29 Wang, H. *et al.* Allele-specific, non-extendable primer blocker PCR (AS-NEPB-PCR) for DNA mutation detection in cancer. *J Mol Diagn* **15**, 62-69 (2013).
30 Diehl, F. *et al.* BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat Methods* **3**, 551-559 (2006).
31 Taniguchi, K. *et al.* Quantitative detection of EGFR mutations in circulating tumor DNA derived from lung adenocarcinomas. *Clin Cancer Res* **17**, 7808-7815 (2011).
32 Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* **83**, 8604-8610 (2011).
33 O'Leary, B. *et al.* Comparison of BEAMing and Droplet Digital PCR for Circulating Tumor DNA Analysis. *Clin Chem* **65**, 1405-1413 (2019).
34 Taylor, S. C., Laperriere, G. & Germain, H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Sci Rep* **7**, 2409 (2017).
35 El Achi, H., Khoury, J. D. & Loghavi, S. Liquid Biopsy by Next-Generation Sequencing: a Multimodality Test for Management of Cancer. *Curr Hematol Malig Rep* **14**, 358-367 (2019).
36 Chen, M. & Zhao, H. Next-generation sequencing in liquid biopsy: cancer screening and early detection. *Hum Genomics* **13**, 34 (2019).
37 Alekseyev, Y. O. *et al.* A Next-Generation Sequencing Primer-How Does It Work and What Can It Do? *Acad Pathol* **5**, 2374289518766521 (2018).
38 Glenn, T. C. Field guide to next-generation DNA sequencers. *Mol Ecol Resour* **11**, 759-769 (2011).
39 Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**, 9530-9535 (2011).
40 Newman, A. M. *et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol* **34**, 547-555 (2016).

41 Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164**, 57-68 (2016).
42 Galardi, F. *et al.* Cell-Free DNA-Methylation-Based Methods and Applications in Oncology. *Biomolecules* **10** (2020).
43 Feinberg, A. P. Phenotypic plasticity and the epigenetics of human disease. *Nature* **447**, 433-440 (2007).
44 Cristiano, S. *et al.* Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385-389 (2019).
45 Liu, M. C. *et al.* Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol* **31**, 745-759 (2020).
46 Garcia-Pardo, M., Makarem, M., Li, J. J. N., Kelly, D. & Leighl, N. B. Integrating circulating-free DNA (cfDNA) analysis into clinical practice: opportunities and challenges. *Br J Cancer* **127**, 592-602 (2022).
47 Bidard, F. C. *et al.* Elacestrant (oral selective estrogen receptor degrader) Versus Standard Endocrine Therapy for Estrogen Receptor-Positive, Human Epidermal Growth Factor Receptor 2-Negative Advanced Breast Cancer: Results From the Randomized Phase III EMERALD Trial. *J Clin Oncol* **40**, 3246-3256 (2022).
48 Turner, N. C. *et al.* Circulating tumour DNA analysis to direct therapy in advanced breast cancer (plasmaMATCH): a multicentre, multicohort, phase 2a, platform trial. *Lancet Oncol* **21**, 1296-1308 (2020).
49 Razavi, P. *et al.* High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. *Nat Med* **25**, 1928-1937 (2019).
50 Chan, K. C. A. *et al.* Analysis of Plasma Epstein-Barr Virus DNA to Screen for Nasopharyngeal Cancer. *N Engl J Med* **377**, 513-522 (2017).
51 Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926-930 (2018).
52 Lennon, A. M. *et al.* Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. *Science* **369** (2020).
53 de Almeida, B. P., Apolonio, J. D., Binnie, A. & Castelo-Branco, P. Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers. *BMC Cancer* **19**, 219 (2019).
54 Xu, Z., Sandler, D. P. & Taylor, J. A. Blood DNA Methylation and Breast Cancer: A Prospective Case-Cohort Analysis in the Sister Study. *J Natl Cancer Inst* **112**, 87-94 (2020).
55 Shen, S. Y. *et al.* Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* **563**, 579-583 (2018).
56 Klein, E. A. *et al.* Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann Oncol* **32**, 1167-1177 (2021).
57 Thierry, A. R. Circulating DNA fragmentomics and cancer screening. *Cell Genom* **3**, 100242 (2023).
58 Fernandez-Uriarte, A., Pons-Belda, O. D. & Diamandis, E. P. Cancer Screening Companies Are Rapidly Proliferating: Are They Ready for Business? *Cancer Epidemiol Biomarkers Prev* **31**, 1146-1150 (2022).

59 Rothe, F. *et al.* Circulating Tumor DNA in HER2-Amplified Breast Cancer: A Translational Research Substudy of the NeoALTTO Phase III Trial. *Clin Cancer Res* **25**, 3581-3588 (2019).
60 Oshiro, C. *et al.* PIK3CA mutations in serum DNA are predictive of recurrence in primary breast cancer patients. *Breast Cancer Res Treat* **150**, 299-307 (2015).
61 Goh, J. Y. *et al.* Chromosome 1q21.3 amplification is a trackable biomarker and actionable target for breast cancer recurrence. *Nat Med* **23**, 1319-1330 (2017).
62 Widschwendter, M. *et al.* Methylation patterns in serum DNA for early identification of disseminated breast cancer. *Genome Med* **9**, 115 (2017).
63 Riva, F. *et al.* Patient-Specific Circulating Tumor DNA Detection during Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. *Clin Chem* **63**, 691-699 (2017).
64 Cavallone, L. *et al.* Prognostic and predictive value of circulating tumor DNA during neoadjuvant chemotherapy for triple negative breast cancer. *Sci Rep* **10**, 14704 (2020).
65 McDonald, B. R. *et al.* Personalized circulating tumor DNA analysis to detect residual disease after neoadjuvant therapy in breast cancer. *Sci Transl Med* **11** (2019).
66 Cheng, L. *et al.* Hazard of recurrence among women after primary breast cancer treatment--a 10-year follow-up using data from SEER-Medicare. *Cancer Epidemiol Biomarkers Prev* **21**, 800-809 (2012).
67 Beaver, J. A. *et al.* Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clin Cancer Res* **20**, 2643-2650 (2014).
68 Garcia-Murillas, I. *et al.* Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci Transl Med* **7**, 302ra133 (2015).
69 Olsson, E. *et al.* Serial monitoring of circulating tumor DNA in patients with primary breast cancer for detection of occult metastatic disease. *EMBO Mol Med* **7**, 1034-1047 (2015).
70 Parsons, H. A. *et al.* Sensitive Detection of Minimal Residual Disease in Patients Treated for Early-Stage Breast Cancer. *Clin Cancer Res* **26**, 2556-2564 (2020).
71 Coombes, R. C. *et al.* Personalized Detection of Circulating Tumor DNA Antedates Breast Cancer Metastatic Recurrence. *Clin Cancer Res* **25**, 4255-4263 (2019).
72 Dawson, S. J. *et al.* Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* **368**, 1199-1209 (2013).
73 Kodahl, A. R. *et al.* Correlation between circulating cell-free PIK3CA tumor DNA levels and treatment response in patients with PIK3CA-mutated metastatic breast cancer. *Mol Oncol* **12**, 925-935 (2018).
74 Liang, D. H. *et al.* Cell-free DNA as a molecular tool for monitoring disease progression and response to therapy in breast cancer patients. *Breast Cancer Res Treat* **155**, 139-149 (2016).
75 O'Leary, B. *et al.* Early circulating tumor DNA dynamics and clonal selection with palbociclib and fulvestrant for breast cancer. *Nat Commun* **9**, 896 (2018).
76 Frenel, J. S. *et al.* Serial Next-Generation Sequencing of Circulating Cell-Free DNA Evaluating Tumor Clone Response To Molecularly Targeted Drug Administration. *Clin Cancer Res* **21**, 4586-4596 (2015).
77 Mouliere, F. *et al.* Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* **10** (2018).
78 Jiang, P. *et al.* Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A* **112**, E1317-1325 (2015).

79 Lapin, M. *et al.* Fragment size and level of cell-free DNA provide prognostic information in patients with advanced pancreatic cancer. *J Transl Med* **16**, 300 (2018).
80 Sanchez-Osuna, M. *et al.* Caspase-activated DNase is necessary and sufficient for oligonucleosomal DNA breakdown, but not for chromatin disassembly during caspase-dependent apoptosis of LN-18 glioblastoma cells. *J Biol Chem* **289**, 18752-18769 (2014).
81 Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett* **589**, 2914-2922 (2015).
82 Jahr, S. *et al.* DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* **61**, 1659-1665 (2001).
83 Choi, J. J., Reich, C. F., 3rd & Pisetsky, D. S. Release of DNA from dead and dying lymphocyte and monocyte cell lines in vitro. *Scand J Immunol* **60**, 159-166 (2004).
84 Marques, J. F., Junqueira-Neto, S., Pinheiro, J., Machado, J. C. & Costa, J. L. Induction of apoptosis increases sensitivity to detect cancer mutations in plasma. *Eur J Cancer* **127**, 130-138 (2020).
85 Panagopoulou, M. *et al.* Circulating cell-free DNA release in vitro: kinetics, size profiling, and cancer-related gene methylation. *J Cell Physiol* **234**, 14079-14089 (2019).
86 Bronkhorst, A. J. *et al.* Characterization of the cell-free DNA released by cultured cancer cells. *Biochim Biophys Acta* **1863**, 157-165 (2016).
87 Wang, W. *et al.* Characterization of the release and biological significance of cell-free DNA from breast cancer cell lines. *Oncotarget* **8**, 43180-43191 (2017).
88 Rostami, A. *et al.* Senescence, Necrosis, and Apoptosis Govern Circulating Cell-free DNA Release Kinetics. *Cell Rep* **31**, 107830 (2020).
89 Wang, B. G. *et al.* Increased plasma DNA integrity in cancer patients. *Cancer Res* **63**, 3966-3968 (2003).
90 Umetani, N. *et al.* Increased integrity of free circulating DNA in sera of patients with colorectal or periampullary cancer: direct quantitative PCR for ALU repeats. *Clin Chem* **52**, 1062-1069 (2006).
91 Gasparello, J. *et al.* Liquid biopsy in mice bearing colorectal carcinoma xenografts: gateways regulating the levels of circulating tumor DNA (ctDNA) and miRNA (ctmiRNA). *J Exp Clin Cancer Res* **37**, 124 (2018).
92 Tkach, M. & Thery, C. Communication by Extracellular Vesicles: Where We Are and Where We Need to Go. *Cell* **164**, 1226-1232 (2016).
93 McAndrews, K. M. & Kalluri, R. Mechanisms associated with biogenesis of exosomes in cancer. *Mol Cancer* **18**, 52 (2019).
94 Takahashi, A. *et al.* Exosomes maintain cellular homeostasis by excreting harmful DNA from cells. *Nat Commun* **8**, 15287 (2017).
95 Lee, T. H. *et al.* Oncogenic ras-driven cancer cell vesiculation leads to emission of double-stranded DNA capable of interacting with target cells. *Biochem Biophys Res Commun* **451**, 295-301 (2014).
96 Ostrowski, M. *et al.* Rab27a and Rab27b control different steps of the exosome secretion pathway. *Nat Cell Biol* **12**, 19-30; sup pp 11-13 (2010).
97 Thakur, B. K. *et al.* Double-stranded DNA in exosomes: a novel biomarker in cancer detection. *Cell Res* **24**, 766-769 (2014).

98 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).
99 Montermini, L. *et al.* Inhibition of oncogenic epidermal growth factor receptor kinase triggers release of exosome-like extracellular vesicles and impacts their phosphoprotein and DNA content. *J Biol Chem* **290**, 24534-24546 (2015).
100 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).
101 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).
102 Yokoi, A. *et al.* Mechanisms of nuclear content loading to exosomes. *Sci Adv* **5**, eaax8849 (2019).
103 Jeppesen, D. K. *et al.* Reassessment of Exosome Composition. *Cell* **177**, 428-445 e418 (2019).
104 Helmig, S., Fruhbeis, C., Kramer-Albers, E. M., Simon, P. & Tug, S. Release of bulk cell free DNA during physical exercise occurs independent of extracellular vesicles. *Eur J Appl Physiol* **115**, 2271-2280 (2015).
105 Lazaro-Ibanez, E. *et al.* Different gDNA content in the subpopulations of prostate cancer extracellular vesicles: apoptotic bodies, microvesicles, and exosomes. *Prostate* **74**, 1379-1390 (2014).
106 Vagner, T. *et al.* Large extracellular vesicles carry most of the tumour DNA circulating in prostate cancer patient plasma. *J Extracell Vesicles* **7**, 1505403 (2018).
107 Asmamaw, M. & Zawdie, B. Mechanism and Applications of CRISPR/Cas-9-Mediated Genome Editing. *Biologics* **15**, 353-361 (2021).
108 Her, J. & Bunting, S. F. How cells ensure correct repair of DNA double-strand breaks. *J Biol Chem* **293**, 10502-10511 (2018).
109 He, C. *et al.* CRISPR screen in cancer: status quo and future perspectives. *Am J Cancer Res* **11**, 1031-1050 (2021).
110 Yamauchi, T. *et al.* Genome-wide CRISPR-Cas9 Screen Identifies Leukemia-Specific Dependence on a Pre-mRNA Metabolic Pathway Regulated by DCPS. *Cancer Cell* **33**, 386-400 e385 (2018).
111 Scheidmann, M. C. *et al.* An In Vivo CRISPR Screen Identifies Stepwise Genetic Dependencies of Metastatic Progression. *Cancer Res* **82**, 681-694 (2022).
112 Fomicheva, M. & Macara, I. G. Genome-wide CRISPR screen identifies noncanonical NF-kappaB signaling as a regulator of density-dependent proliferation. *Elife* **9** (2020).
113 Tiedt, R. *et al.* Integrated CRISPR screening and drug profiling identifies combination opportunities for EGFR, ALK, and BRAF/MEK inhibitors. *Cell Rep* **42**, 112297 (2023).
114 Gallo, D. *et al.* CCNE1 amplification is synthetic lethal with PKMYT1 kinase inhibition. *Nature* **604**, 749-756 (2022).
115 Burr, M. L. *et al.* An Evolutionarily Conserved Function of Polycomb Silences the MHC Class I Antigen Presentation Pathway and Enables Immune Evasion in Cancer. *Cancer Cell* **36**, 385-401 e388 (2019).
116 Wang, D. *et al.* CRISPR Screening of CAR T Cells and Cancer Stem Cells Reveals Critical Dependencies for Cell-Based Therapies. *Cancer Discov* **11**, 1192-1211 (2021).

117 Zviran, A. *et al.* Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring. *Nat Med* **26**, 1114-1124 (2020).
118 Gydush, G. *et al.* Massively parallel enrichment of low-frequency alleles enables duplex sequencing at low depth. *Nat Biomed Eng* **6**, 257-266 (2022).
119 Breitbach, S., Tug, S. & Simon, P. Circulating cell-free DNA: an up-coming molecular marker in exercise physiology. *Sports Med* **42**, 565-586 (2012).
120 Duvvuri, B. & Lood, C. Cell-Free DNA as a Biomarker in Autoimmune Rheumatic Diseases. *Front Immunol* **10**, 502 (2019).
121 Madsen, A. T., Hojbjerg, J. A., Sorensen, B. S. & Winther-Larsen, A. Day-to-day and within-day biological variation of cell-free DNA. *EBioMedicine* **49**, 284-290 (2019).
122 Kato, S. *et al.* Analysis of Circulating Tumor DNA and Clinical Correlates in Patients with Esophageal, Gastroesophageal Junction, and Gastric Adenocarcinoma. *Clin Cancer Res* **24**, 6248-6256 (2018).
123 Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics. *Nat Med* **14**, 985-990 (2008).
124 Kustanovich, A., Schwartz, R., Peretz, T. & Grinshpun, A. Life and death of circulating cell-free DNA. *Cancer Biol Ther* **20**, 1057-1067 (2019).
125 Grabuschnig, S. *et al.* Putative Origins of Cell-Free DNA in Humans: A Review of Active and Passive Nucleic Acid Release Mechanisms. *Int J Mol Sci* **21** (2020).
126 Bronkhorst, A. J., Ungerer, V. & Holdenrieder, S. The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol Detect Quantif* **17**, 100087 (2019).
127 Lo, Y. M. *et al.* Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med* **2**, 61ra91 (2010).
128 Thierry, A. R. *et al.* Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic Acids Res* **38**, 6159-6175 (2010).
129 Ungerer, V., Bronkhorst, A. J., Van den Ackerveken, P., Herzog, M. & Holdenrieder, S. Serial profiling of cell-free DNA and nucleosome histone modifications in cell cultures. *Sci Rep* **11**, 9460 (2021).
130 Markus, H. *et al.* Refined characterization of circulating tumor DNA through biological feature integration. *Sci Rep* **12**, 1928 (2022).
131 Mouliere, F. *et al.* High fragmentation characterizes tumour-derived circulating DNA. *PLoS One* **6**, e23418 (2011).
132 Abbosh, C. *et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446-451 (2017).
133 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).
134 Jin, Y. *et al.* DNA in serum extracellular vesicles is stable under different storage conditions. *BMC Cancer* **16**, 753 (2016).
135 Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607 (2012).
136 Klijn, C. *et al.* A comprehensive transcriptional portrait of human cancer cell lines. *Nat Biotechnol* **33**, 306-312 (2015).
137 Zabransky, D. J. *et al.* HER2 missense mutations have distinct effects on oncogenic signaling and migration. *Proc Natl Acad Sci U S A* **112**, E6205-6214 (2015).

138 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016).
139 Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9**, 5416 (2018).
140 Li, W. *et al.* Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol* **16**, 281 (2015).
141 Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* **8**, 1551-1566 (2013).
142 Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
143 Fulda, S. Evasion of apoptosis as a cellular stress response in cancer. *Int J Cell Biol* **2010**, 370835 (2010).
144 Carne Trecesson, S. *et al.* BCL-X(L) directly modulates RAS signalling to favour cancer cell stemness. *Nat Commun* **8**, 1123 (2017).
145 Elmore, S. Apoptosis: a review of programmed cell death. *Toxicol Pathol* **35**, 495-516 (2007).
146 Falschlehner, C., Emmerich, C. H., Gerlach, B. & Walczak, H. TRAIL signalling: decisions between life and death. *Int J Biochem Cell Biol* **39**, 1462-1475 (2007).
147 Ozaki, T. & Nakagawara, A. Role of p53 in Cell Death and Human Cancers. *Cancers (Basel)* **3**, 994-1013 (2011).
148 Hemann, M. T. & Lowe, S. W. The p53-Bcl-2 connection. *Cell Death Differ* **13**, 1256-1259 (2006).
149 Taylor, S. J. & Shalloway, D. An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis. *Nature* **368**, 867-871 (1994).
150 Fumagalli, S., Totty, N. F., Hsuan, J. J. & Courtneidge, S. A. A target for Src in mitosis. *Nature* **368**, 871-874 (1994).
151 Najib, S. & Sanchez-Margalet, V. Sam68 associates with the SH3 domains of Grb2 recruiting GAP to the Grb2-SOS complex in insulin receptor signaling. *J Cell Biochem* **86**, 99-106 (2002).
152 Stockley, J. *et al.* The RNA-binding protein Sam68 regulates expression and transcription function of the androgen receptor splice variant AR-V7. *Sci Rep* **5**, 13426 (2015).
153 Fu, K. *et al.* Sam68/KHDRBS1-dependent NF-kappaB activation confers radioprotection to the colon epithelium in gamma-irradiated mice. *Elife* **5** (2016).
154 Fu, K. *et al.* Sam68/KHDRBS1 is critical for colon tumorigenesis by regulating genotoxic stress-induced NF-kappaB activation. *Elife* **5** (2016).
155 Paronetto, M. P., Achsel, T., Massiello, A., Chalfant, C. E. & Sette, C. The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *J Cell Biol* **176**, 929-939 (2007).
156 Schneider, P. *et al.* TRAIL receptors 1 (DR4) and 2 (DR5) signal FADD-dependent apoptosis and activate NF-kappaB. *Immunity* **7**, 831-836 (1997).
157 Sun, X. *et al.* Sam68 Is Required for DNA Damage Responses via Regulating Poly(ADP-ribosyl)ation. *PLoS Biol* **14**, e1002543 (2016).
158 Frisone, P. *et al.* SAM68: Signal Transduction and RNA Metabolism in Human Cancer. *Biomed Res Int* **2015**, 528954 (2015).
159 Ramakrishnan, P. & Baltimore, D. Sam68 is required for both NF-kappaB activation and apoptosis signaling by the TNF receptor. *Mol Cell* **43**, 167-179 (2011).

160 Lafont, E. *et al.* The linear ubiquitin chain assembly complex regulates TRAIL-induced gene activation and cell death. *EMBO J* **36**, 1147-1166 (2017).
161 Walczak, H. Death receptor-ligand systems in cancer, cell death, and inflammation. *Cold Spring Harb Perspect Biol* **5**, a008698 (2013).
162 Jin, Z. & El-Deiry, W. S. Distinct signaling pathways in TRAIL- versus tumor necrosis factor-induced apoptosis. *Mol Cell Biol* **26**, 8136-8148 (2006).
163 Holoch, P. A. & Griffith, T. S. TNF-related apoptosis-inducing ligand (TRAIL): a new path to anti-cancer therapies. *Eur J Pharmacol* **625**, 63-72 (2009).
164 Sanchez-Margalet, V. & Najib, S. Sam68 is a docking protein linking GAP and PI3K in insulin receptor signaling. *Mol Cell Endocrinol* **183**, 113-121 (2001).
165 Li, Z. *et al.* Sam68 expression and cytoplasmic localization is correlated with lymph node metastasis as well as prognosis in patients with early-stage cervical cancer. *Ann Oncol* **23**, 638-646 (2012).
166 Bielli, P. *et al.* The transcription factor FBI-1 inhibits SAM68-mediated BCL-X alternative splicing and apoptosis. *EMBO Rep* **15**, 419-427 (2014).
167 Bohler, S. *et al.* Inhibition of the anti-apoptotic protein MCL-1 severely suppresses human hematopoiesis. *Haematologica* **106**, 3136-3148 (2021).
168 Popgeorgiev, N., Jabbour, L. & Gillet, G. Subcellular Localization and Dynamics of the Bcl-2 Family of Proteins. *Front Cell Dev Biol* **6**, 13 (2018).
169 Dang, D. K. & Park, B. H. Circulating tumor DNA: current challenges for clinical utility. *J Clin Invest* **132** (2022).
170 Foda, Z. H. *et al.* Detecting Liver Cancer Using Cell-Free DNA Fragmentomes. *Cancer Discov* **13**, 616-631 (2023).
171 Martin-Alonso, C. *et al.* Priming agents transiently reduce the clearance of cell-free DNA to improve liquid biopsies. *Science* **383**, eadf2341 (2024).
172 Zhuang, H. *et al.* Down-regulation of HSP27 sensitizes TRAIL-resistant tumor cell to TRAIL-induced apoptosis. *Lung Cancer* **68**, 27-38 (2010).
173 Li, X. *et al.* Reversal of the Apoptotic Resistance of Non-Small-Cell Lung Carcinoma towards TRAIL by Natural Product Toosendanin. *Sci Rep* **7**, 42748 (2017).
174 Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**, 1515-1526 (2015).
175 Cohen, J. D. *et al.* Detection of low-frequency DNA variants by targeted sequencing of the Watson and Crick strands. *Nat Biotechnol* **39**, 1220-1227 (2021).