

AD _____

AWARD NUMBER: W81XWH-10-1-0563

TITLE: Mitochondrial DNA Biomarker Discovery and Validation for the Detection, Prognosis and Treatment of Prostate Cancer

PRINCIPAL INVESTIGATOR: Jason H. Bielas, Ph.D.

RECIPIENT: Fred Hutchinson Cancer Research Center
Seattle, WA 98109-1024

REPORT DATE: August 2013

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release; distribution is unlimited.

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE			<i>Form Approved</i> <i>OMB No. 0704-0188</i>		
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE CE * * • dGEFH		2. REPORT TYPE Final		3. DATES COVERED 19 Jul 2010 - 18 Feb 2013	
4. TITLE AND SUBTITLE Mitochondrial DNA Biomarker Discovery and Validation for the Detection, Prognosis and Treatment of Prostate Cancer			5a. CONTRACT NUMBER AAA		
			5b. GRANT NUMBER Y1FY PEFEEI HA		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Jason H. Bielas, Ph.D. E-Mail: jbielas@fhcrc.org			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Fred Hutchinson Cancer Research Center 1100 Fairview Ave N, J6-300 Seattle, WA 98109-4433			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The mitochondrial genome offers excellent potential as a DNA-based biomarker for cancer detection, as most tumors are composed of cells that each share thousands of identical somatic mtDNA point mutations, a phenomenon called homoplasmy. Here we propose the use of the sensitive mutation detection assays to test whether the frequency of circulating homoplasmic mtDNA tumor mutations in patients with prostate cancer would be a specific and sensitive marker of prostatic tumor therapeutic response, progression, and recurrence. Prostate cancer presents an ideal tumor model to test this new technology, as the frequency of prostate gland biopsy, homoplasmic mtDNA mutations, recurrence following radical prostatectomy are high.					
15. SUBJECT TERMS Biomarker Discovery, Prostate Cancer, Cancer Genetics, Mitochondria, Survivorship, DNA Mutation, Circulating tumor cells (CTC's)					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)
			UU	44	

Table of Contents

	<u>Page</u>
1. Introduction	1
2. Keywords	1
3. Overall Project Summary	1
4. Key Research Accomplishments	5
5. Conclusion	5
6. Publications, Abstracts, and Presentations	6
7. Inventions, Patents and Licenses	6
8. Reportable Outcomes	7
9. Other Achievements	7
10. References	7
11. Appendices	9
12. Table & Figures	10

INTRODUCTION:

We urgently require new markers that are predictive of the biological behavior of tumors to guide the types and aggressiveness of therapy. We have proposed to address this challenge by the development of a new technology that will exploit mitochondrial DNA mutations as novel biomarkers for tumor progression, therapeutic response, and cancer recurrence. Our hypothesis is that an ultrasensitive measurement of CTC and ctmtDNA prevalence, marked by homoplasmic mtDNA mutations identical to those in the primary tumor, will serve as early independent prognostic indicators of tumor stage, therapeutic response, progression and recurrence. We have two specific aims. In Aim 1, we will determine the rate and types of somatic mtDNA mutations in prostatic cancers. In Aim 2, we will establish whether the prevalence of circulating tumor mtDNA can serve as a sensitive marker of clinical stage, progression, and recurrence.

KEYWORDS:

- Biomarker Discovery
- Prostate Cancer
- Cancer Genetics
- Mitochondria
- Survivorship
- DNA Mutation
- Circulating tumor cells (CTCs)
- Mutation detection

OVERALL PROJECT SUMMARY:

Mutations in mitochondrial DNA (mtDNA) lead to a diverse collection of diseases that are challenging to diagnose and treat. Each human cell has hundreds to thousands of mitochondrial genomes and disease-associated mtDNA mutations are homoplasmic in nature, i.e. the identical mutation is present in a preponderance of mitochondria within a tissue (Chatterjee et al., 2006; Taylor and Turnbull, 2005). Although the precise mechanisms of mtDNA mutation accumulation in disease pathogenesis remain elusive, we have documented multiple homoplasmic mutations from prostate tumor samples. Indicative of their involvement in tumorigenesis and their potential utility as prognostic and predictive markers, mtDNA mutations identified in prostatic cancers by our group (Table 1) and previously by others (Petros et al., 2005) are predominantly nonsynonymous. However, in our dataset, no statistical correlations were found between mtDNA mutation and clinical significance with respect to Gleason score, PSA level, clinical stage, recurrence, therapeutic response, or progression.

Mutations in prostate tumors: discovery and monitoring

We have successfully identified multiple homoplasmic mutations from prostate tumor samples (Table 1). To ensure accurate identification of homoplasmic tumor mtDNA mutations, patient-matched normal peripheral blood cells and multiple pure prostatic carcinoma cells isolated with laser capture microdissection (LCM) from surgically resected tumors (radical prostatectomies) were collected from patients who have given informed consent and received no prior treatment. The entire mitochondrial genome was sequenced in prostatic cancer, adjacent normal tissue, and blood samples from each patient first by PCR amplifying the mtDNA with 28 pairs of primers, previously described (Taylor et al., 2001). Clonally expanded mtDNA mutations were scored only when the sequence of the tissue samples differed from that of the patient-matched normal peripheral blood cells. All regions with detected mutations were reamplified and sequenced to rule out the possibility of the mutations being produced by polymerase errors during the PCR or sequencing processes. In addition, to guard against the sample mix-up and contamination that has confounded many mtDNA mutation studies (Salas et al., 2005), we compared each patient's sequences to the revised Cambridge Reference Sequence (rCRS) to confirm they shared common polymorphisms.

To track these mutations using the Random Mutation Capture (RMC) methodology, we set out to expand RMC coverage to other mutational targets in accordance with our statement of proposed work. We first identified robust restriction enzyme recognition sequences that can be accurately monitored for mutation. Digestion of mtDNA was performed with the required restriction enzymes (Table 1) and the efficiency of digestion was monitored by QPCR with primers that flank the restriction sites. However, suboptimal restriction enzyme efficiencies were observed, limiting our ability to detect point mutations, as compared to our resolution with TaqI. Alternatively, given that mtDNA deletions are found in many tumors (Kulawiec et al., 2010)(Table 1), an adaptation of the RMC technology would allow us to utilize these mutations to track CTCs and ctmtDNA. As such, we developed a highly sensitive tool, termed Digital Deletion Detection (3D), for detection, quantification, and characterization of rare (< 1 per 10^8 molecules) deletion events in mtDNA (Taylor et al., 2013). 3D utilizes droplet digital PCR for direct enumeration of deletion mutations and provides several process pathways for characterization of the diversity of deletion sizes and breakpoints within the sample population. Single molecule compartmentalization minimizes amplification bias, allowing simultaneous analysis of a large range of deletion products and more accurate sequencing analysis. The assay can be readily adapted to interrogate multiple diverse regions within the genome.

3D assay design

Digital Deletion Detection (3D) is an extremely sensitive tool for the absolute quantification and characterization of rare deletion molecules. The basic strategy behind 3D is a three-step process: enrich, amplify, and analyze. The first step, based on methods developed previously by Bielas and colleagues, enriches for deletion-bearing molecules and improves mutant specificity (Bielas and Loeb, 2005; Vermulst et al., 2008a). This step consists of targeted endonucleolytic digestion of templates to selectively digest wild-type (WT) molecules, thus allowing the preferential PCR amplification of molecules bearing an appropriate deletion (Figure 1A). After digestion, the DNA molecules are sequestered into homogenous 1 nl water-in-oil emulsion droplets and subjected to normal PCR amplification (Figure 1B). The concentration of molecules within the droplets is adjusted such that most droplets contain no mutant genomes, while a small fraction contains only one. Thus a single well in the reaction actually consists of many thousand single molecule reaction chambers. This process allows each captured deletion to be amplified without introducing many of the PCR artifacts and biases that are common to bulk amplification reactions (i.e. template switching and preferential amplification of short templates).

Following amplification, the deletions can be analyzed via two process pathways. In the quantification pathway, high resolution quantification of deletions is accomplished through the use of droplet digital PCR (ddPCR) (Pinheiro et al., 2012). With the inclusion of TaqMan reporter chemistry, droplets bearing amplified templates are readily distinguished by their fluorescence amplitude using a cytometry system. Because the droplet volumes are highly uniform, Poisson statistics can be applied to calculate the average number of deletion-bearing molecules per droplet and the absolute concentration of mutant molecules determined with high precision and accuracy (Pinheiro et al., 2012). Alternatively, in the characterization pathway, droplets are disrupted and amplicons recovered. The deletions can then be directly sequenced using high-throughput or 'next generation' sequencing, or cloned for use in Sanger sequencing or other downstream applications.

3D sensitivity and recovery

Using the quantification process pathway of 3D, we measured the absolute deletion frequency within a region spanning the ND1/ND2 genes in mitochondrial DNA isolated from human epithelial cells in tissue culture. We measured the deletion frequency to be 1.6 ± 0.4 deletions per ten million genomes (or 1.6×10^{-7} per genome) (Figure 2). We next asked whether 3D was able to fully recover all of the deletions within a sample over a broad range of deletion frequencies. To address this we performed a series of reconstruction experiments. First, a plasmid harboring a fragment of mtDNA containing a known deletion in the ND1/ND2 region was mixed at a constant concentration (3 copies/ μ l) against increasingly higher levels of genomic mtDNA (up to 2.5×10^6 copies/ μ l). We then performed 3D analysis to determine if the low concentration of the control molecules could be accurately quantified in the presence of increasing concentrations of background DNA (Figure 2). This

reconstruction demonstrated accurate quantification of target molecules across a range of frequencies spanning eight orders of magnitude, with sensitive recovery at frequencies as low as 1×10^{-7} per genome. Because we reached the endogenous deletion frequency of the background DNA, we were unable to test lower frequencies in the reconstruction experiment.

Capturing and analyzing sample complexity

Analysis of fluorescence amplitudes of the three control plasmids following ddPCR revealed that under the current conditions, a given template will yield an average droplet fluorescence intensity inversely proportional to the template size (Figure 3). When the three control templates were combined, this effect led to a striking multimodal distribution in the fluorescence amplitudes (Figure 3A). More generally, we found that the sample heterogeneity is reflected in the distribution of fluorescence amplitudes (Figure 3C). Thus, the average amplitude and distribution of the droplet fluorescence can be used to predict deletion sizes and complexity (e.g. presence of a single, clonal deletion vs. a heterogeneous population of multiple deletions). This discovery led to the development of QuantiSize (Laurie et al., 2013), which combines quantification and size determination in a single ddPCR experiment. With standard ddPCR reagent concentrations, DNA amplification is eventually limited by the availability of dNTPs and inhibited by the presence of pyrophosphate (Hori et al., 2007; Xiao et al., 2004). This means that the amplification of long DNA templates consumes more dNTPs and generates more pyrophosphate, resulting in fewer products than short templates at the endpoint of a standard reaction. Because the final number of products generated within a droplet determines its level of fluorescence, the measured fluorescence amplitude of droplets containing short templates will be greater than that of droplets containing long templates. The QuantiSize assay exploits this fact to generate an equation relating fluorescence amplitude to amplicon size by using measurements of known size standards. The equation describing the relationship between fluorescence amplitude and amplicon size can be used to calculate the size of any unknown ddPCR template that shares common primer and probe binding sites with the size standards. Creating size standards that have primer and probe binding sites in common with DNA samples can be accomplished in a number of ways including cloning sample DNA into a vector and appending adapter sequences to both the sample DNA and size standards (Zhang et al., 2003).

We created a set of size standards applicable to Illumina NGS libraries containing inserts ranging from 25 to 1000 base pairs flanked by adapter sequences compatible with the Illumina MiSeq platform. A pair of primers and a fluorescent TaqMan probe were designed to hybridize to the adapter sequences such that the length of each amplicon is 160 base pairs plus the length of the insert. As the primers and probe are specific to the MiSeq adapter sequences, only the adapter-ligated molecules that are amplifiable on the MiSeq flow cell will be quantified.

A ddPCR experiment was performed with the aforementioned size standards in separate wells of a 96-well plate. Droplets containing the target (positive) increased in fluorescence following amplification of the target whereas droplets lacking the target (negative) remained at the background level of fluorescence (Figure 4A). The distribution of droplet amplitudes is consistent across most amplicon lengths, but the 760 and 860 bp amplicons show a broader distribution of amplitudes (Figure 4B). An inverse, linear correlation between amplicon size and mean fluorescence amplitude was observed ($R^2=0.99436$) (Figure 4C). The equation describing this correlation allows for the calculation of amplicon size given a measured fluorescence amplitude. The slope of this equation provides a measure of the difference in mean fluorescence amplitude that is expected with a given difference in amplicon size. Maximizing the magnitude of this slope maximizes the resolution of size standards, which is advantageous for the purpose of determining the length of unknown amplicons more accurately. The size standards used for QuantiSize are highly analogous to the standards used in gel and capillary electrophoresis. The size of unknown DNA can be determined by visually comparing the fluorescence amplitude of the size references to that of the unknown DNA or by entering the fluorescence amplitude value into the equation describing the relationship between average fluorescence amplitude and amplicon size for the size standards.

The droplet reader software counts positive and negative droplets by using a threshold of fluorescence between the well-defined populations of high and low fluorescence amplitude droplets. For one particular TaqMan probe tested, the fluorescence amplitude of droplets containing amplicons larger than 660 bp was too low to reliably discriminate between positive and negative droplets when templates are amplified with a one-minute elongation time. When this is the case, the average fluorescence amplitude for these amplicons cannot be calculated. Increasing the elongation time to two minutes increases the fluorescence amplitude of all droplets containing amplifiable template (Figure 5). This enables the acquisition of accurate concentration and fluorescence amplitude data for longer templates, but the slope of the relationship between amplicon size and fluorescence amplitude is decreased (from $m = -11.66$ to $m = -9.12$), which decreases the ability to resolve small differences in amplicon size (Figure 5). Decreasing the elongation time to 30 seconds increases the resolution of the relationship between amplicon size and fluorescence amplitude, but prevents targets longer than 460 bp from amplifying to the point that they fluoresce detectably above the background fluorescence (Figure 5). This is likely due to the fact that longer products require more time for complete polymerization of nascent strands to occur. Thus, there is a tradeoff between the resolution and range of QuantiSize, though the assay can be easily adjusted to fit particular experimental needs.

3D Summary

In order to adequately detect *de novo* mtDNA deletions and trace the frequency dynamics, an assay is needed that can enrich for and directly quantify extremely rare deletion events. Current approaches to analyzing mtDNA deletions include Southern blotting (DiMauro and Hirano, 1993), direct sequencing (Ameur et al., 2011; Kato et al., 2011; Sequeira et al., 2012; Spelbrink et al., 2000), and PCR amplification (Kraytsberg et al., 2008). Sequencing of deletions via cloning is laborious, time-consuming, prone to cloning artifacts, and allows only the most abundant deletion types to be analyzed (Supplementary Notes 3 and 4). Massively-parallel or 'next generation' sequencing is rapidly becoming a preferred means for high-throughput screening of individual DNA molecules. As an example, Illumina, Inc. (San Diego, CA) offers systems that generate from 17 million (MiSeq®) up to 3 billion simultaneous sequencing reads per run (HiSeq®)(Liu et al., 2012). However, given a relatively short read length of less than 150 bp and the fact that the majority of the reads will be off-target, this remains insufficient to adequately resolve mtDNA deletions that occur at frequencies of less than one in a million genomes. Even assuming no off-target reads, the MiSeq® instrument would still only yield about one deletion in ten runs. It is therefore critical that a selection step be performed to limit the number of off-target reads and to enrich for deletion-bearing molecules.

PCR-based methods, including long-distance PCR and real-time quantitative PCR, are among the most frequently employed methods for both selection and amplification of deletions (Chabi et al., 2003; Cortopassi and Arnheim, 1990; He et al., 2002; Kraytsberg et al., 2008). Generally speaking, these assays distinguish wild-type from deleted genomes through exploitation of differences in amplicon fragment lengths and amplification efficiencies. Given that they do not select for deleted molecules prior to amplification, one of the main drawbacks is high background signal from contaminating wild-type molecules, limiting the effective sensitivity. Furthermore, these bulk PCR assays tend to introduce a number of additional artifacts arising from preferential amplification of small templates (allelic preference), introduction of false deletions through template jumping, and other PCR errors (Kraytsberg and Khrapko, 2005). Real-time quantitative PCR (qPCR) can be quite sensitive, but its reliance on relative differences in crossing thresholds rather than direct quantification makes it more suitable for measuring fold changes rather than absolute deletion frequencies (Chabi et al., 2003; He et al., 2002). Digital PCR methods, including long single molecule PCR (long smPCR) (Guo et al., 2010; Kraytsberg and Khrapko, 2005) and the random mutation capture assay developed for mtDNA deletions (deletion RMC)(Vermulst et al., 2008a; Vermulst et al., 2008b) achieve direct quantification through the use of single molecule partitioning in 96-well plates. Partitioning additionally serves to minimize artifacts of template jumping and allelic preference that are common in bulk PCR reactions (Kraytsberg and Khrapko, 2005). Despite these advantages, this approach becomes laborious and costly when using the wells of a multi-well plate as the partition, and only a handful of the most common deletions within a sample are yielded.

The digital deletion detection (3D) assay shows a marked improvement in specificity, sensitivity, and accuracy over other available methods. This is achieved via a three-step process of selection, amplification, and characterization (i.e. quantification or sequencing). As with deletion RMC, high specificity for deletion-bearing molecules is achieved through the destruction of WT template molecules by restriction endonuclease, thereby selecting for and enriching mutant molecules prior to amplification. Following enrichment, partitioning for digital PCR amplification is performed through the generation of up to 20,000 droplet partitions, the equivalent of over 200 96-well plates, within a single reaction well. Quantification is greatly facilitated through the use of TaqMan reporter probes and cytometry, which allows for rapid enumeration of all partitions that contain an amplifiable template and direct quantification of all deletions within a sample.

Mutations in tumor mtDNA that do not disrupt restriction sites

The majority of detected homoplasmic mutations (Table 1) did not mutate at a known restriction enzyme recognition site. As such, to achieve our project's goals, it was necessary to develop a novel mutation detection assay (see 2012 updated SOW) to test whether the frequency of circulating homoplasmic mtDNA tumor mutations in patients with prostate cancer would be a specific and sensitive marker of therapeutic response, progression, and recurrence. As such, we began the development of a novel DNA-based technology, termed Rolling Cypher Seq, which also exploits somatic mtDNA mutations for the early detection of disease. By exploiting advances in Next Generation Sequencing (NGS) technologies, Rolling Cypher Seq is expected to permit the enumeration of any mutation with unparalleled sensitivity, regardless of its location in the genome.

The identification of rare somatic mutations that are present in a small fraction of DNA templates is essential for DNA-based early detection methodologies. Although massively parallel sequencing instruments are, *in principle*, well-suited for this task, the error rates associated with NGS are too high to allow confident identification of rare variants. For example, the error rates vary from ~1% (Nazarian et al., 2010; Quail et al., 2008) to ~0.05% (Gore et al., 2011; He et al., 2010) of bases sequenced with the commonly used Illumina sequencing instruments. To address this limitation, we have developed a novel method for detecting rare mutations (< 1 mutant base pair among 10^8 wild-type nucleotides) in any target DNA molecule. Our method utilizes rolling circle amplification (RCA) on a generated library of vectors, each containing unique double-stranded barcode pairs (cyphers). Primers used in the RCA step are designed to selectively amplify the DNA molecule of interest. Since RCA copies from the same circular template molecule with each cycle, it circumvents the clonal amplification of polymerase errors observed in successive PCR cycles. Moreover, unique cyphers (Figure 6A) flanking each copy of the target molecule will allow us to deconvolute the NGS data and accurately distinguish between polymerase error artifacts and true mutations (Figure 6B). Over the last year we have been able to successfully demonstrate the utility of our method to eliminate background-sequencing errors (Figure 7). This monumental advance in sequencing resolution puts us in a position to enumerate CTC (circulating tumor cells) and ctmtDNA (circulating tumor mtDNA) in prostate cancer patients using any mutated base pair. However, RCA enrichment would still be required to monitor for these rare circulating mutant variants in blood. Unfortunately, while we are making headway, our enrichment strategy is still being optimized; thus we have been unable to enumerate these endpoints thus far. However, we have institution support to continue our assay development, we hope to complete the stated project goals once our RCA enrichment technology development is complete.

KEY RESEARCH ACCOMPLISHMENTS:

- Detected and characterized homoplasmic mutations in prostate cancer
- Developed ultra sensitive methods enumerate deletions and point mutations in DNA

CONCLUSION:

We have been able to successfully demonstrate the utility of new mutation assessment method to eliminate background-sequencing errors. This monumental advance in sequencing resolution puts us in a position to

enumerate CTC (circulating tumor cells) and ctmtDNA (circulating tumor mtDNA) in prostate cancer patients using any mutated base pair. However, mtDNA enrichment is still be required to monitor for these rare circulating mutant variants in blood. Unfortunately, while we are making headway, our enrichment strategy is still being optimized; thus we have been unable to enumerate these endpoints thus far. However, we have institution support to continue our assay development, and hope to complete the stated project goals once our RCA enrichment technology development is complete. The use of the sensitive mutation detection assays to test whether the frequency of circulating homoplasmic mtDNA tumor mutations in patients with prostate cancer is expected to be a specific and sensitive marker of prostatic tumor therapeutic response, progression, and recurrence.

PUBLICATIONS, ABSTRACTS, AND PRESENTATIONS:

Peer-Reviewed Scientific Journals:

Simultaneous digital quantification and fluorescence-based size characterization of massively parallel sequencing libraries. Laurie MT, Bertout JA, Taylor SD, Burton JN, Shendure JA, Bielas JH. *Biotechniques*. 2013 Aug; (2):61-7. (See Appendix 1)

Targeted enrichment and high-resolution digital profiling of DNA deletions in mitochondria. Taylor SD, Ericson NG, Burton JN, Prolla TA, Silber JR, Shendure J, Bielas JH. *Ageing Cell*. 2013 Aug; [Epub ahead of print]. (See Appendix 2)

Presentations:

- 08/2013 Mitochondrial DNA Mutagenesis: Insight into Human Aging, Carcinogenesis, and Novel Anticancer Therapies, Ellison Medical Foundation Colloquium on the Biology of Aging, Woods Hole, MA
- 05/2013 Mechanism and Clinical Utility of Nuclear and Mitochondrial DNA Mutations in Cancer, Biochemistry and Molecular Biology, Faculty of Medicine, Dalhousie University, Halifax, NS, Canada
- 03/2013 Mechanism and Clinical Utility of Nuclear and Mitochondrial DNA Mutations in Cancer, Irish Association for Cancer Research Annual Meeting, Dublin, Ireland
- 12/2012 Digital Detection of Rare Mutations, First Annual Droplet Digital PCR User Meet, Boston, MA
- 10/2012 Digital Detection of Rare Mutations and Tumor Infiltrating T Cells: Clinical Application in Cancer and Disease, Digital PCR Applicationas and Advances, Cambridge Healthtech Institute, San Diego, CA
- 10/2012 Mechanism and Clinical Utility of Nuclear and Mitochondrial DNA Mutations in Cancer, External Scientific Advisory Board Meeting, Fred Hutchison and University of Washington Cancer Consortium, Seattle, WA
- 09/2012 Metabolism and Mitochondrial Mutagenesis in Human Colorectal Cancer, Metabolism and Metabolites Symposium, Fred Hutchinson Cancer Research Center, Seattle, WA
- 07/2012 Nuclear and Mitochondrial DNA Mutations: Mechanisms and Disease, Meeting of the Outstanding New Environmental Scientist (ONES) Grantee Forum, NIEHS, Research Triangle Park, NC

INVENTIONS, PATENTS AND LICENSES:

12-032, Provisional Applications [61/654,236](#) filed 6/1/2012 and [61/783,815](#) filed on 3/14/2013 (the priority patent applications) and the subsequent international patent application [PCT/US2013/043158](#) filed 5/29/2013 (“Compositions and Methods for Detecting Rare Nucleic Acid Molecule Mutations”) (the patents claiming digital PCR for quantitation of DNA for sequencing) all include the following statement:

“STATEMENT OF GOVERNMENT INTEREST”

This was funded in part by **U.S. Department of Defense/Congressionally Directed Medical Research Programs Grant No. W81XWH-10-1-0563** and by National Institute of Environmental Health Sciences R01 Grant ES019319. The government has certain rights in this invention.”

REPORTABLE OUTCOMES:

- Detected and characterized homoplasmic mutations in prostate cancer
- Developed ultra sensitive methods enumerate deletions and point mutations in DNA

Both the above-mentioned outcomes should permit the measure of CTC and ctmtDNA prevalence, to advance the understanding, prevention, diagnosis, prognosis, treatment of prostate cancer.

OTHER ACHIEVEMENTS:

Nothing to Report.

REFERENCES:

- Ameur, A., Stewart, J. B., Freyer, C., Hagström, E., Ingman, M., Larsson, N.-G., and Gyllensten, U. (2011). Ultra-Deep Sequencing of Mouse Mitochondrial DNA: Mutational Patterns and Their Origins. *PLoS Genet* 7.
- Bielas, J. H., and Loeb, L. A. (2005). Quantification of random genomic mutations. *Nat Methods* 2, 285-290.
- Chabi, B., Mousson de Camaret, B., Duborjal, H., Issartel, J.-P., and Stepien, G. (2003). Quantification of Mitochondrial DNA Deletion, Depletion, and Overreplication: Application to Diagnosis. *Clinical Chemistry* 49, 1309-1317.
- Chatterjee, A., Mambo, E., and Sidransky, D. (2006). Mitochondrial DNA mutations in human cancer. *Oncogene* 25, 4663-4674.
- Cortopassi, G. A., and Arnheim, N. (1990). Detection of a specific mitochondrial DNA deletion in tissues of older humans. *Nucleic Acids Res* 18, 6927-6933.
- DiMauro, S., and Hirano, M. (1993). Mitochondrial DNA Deletion Syndromes. In *GeneReviews*, R.A. Pagon, T.D. Bird, C.R. Dolan, and K. Stephens, eds. (Seattle (WA)).
- Gore, A., Li, Z., Fung, H. L., Young, J. E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M. A., Kiskinis, E., *et al.* (2011). Somatic coding mutations in human induced pluripotent stem cells. *Nature* 471, 63-67.
- Guo, X., Kudryavtseva, E., Bodyak, N., Nicholas, A., Dombrovsky, I., Yang, D., Kravtsov, Y., Simon, D. K., and Khrapko, K. (2010). Mitochondrial DNA deletions in mice in men: substantia nigra is much less affected in the mouse. *Biochim Biophys Acta* 1797, 1159-1162.
- He, L., Chinnery, P. F., Durham, S. E., Blakely, E. L., Wardell, T. M., Borthwick, G. M., Taylor, R. W., and Turnbull, D. M. (2002). Detection and quantification of mitochondrial DNA deletions in individual cells by real-time PCR. *Nucleic Acids Res* 30, e68.
- He, Y., Wu, J., Dressman, D. C., Iacobuzio-Donahue, C., Markowitz, S. D., Velculescu, V. E., Diaz, L. A., Jr., Kinzler, K. W., Vogelstein, B., and Papadopoulos, N. (2010). Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature* 464, 610-614.
- Hori, M., Fukano, H., and Suzuki, Y. (2007). Uniform amplification of multiple DNAs by emulsion PCR. *Biochemical and biophysical research communications* 352, 323-328.

- Kato, M., Nakamura, M., Ichiba, M., Tomiyasu, A., Shimo, H., Higuchi, I., Ueno, S.-i., and Sano, A. (2011). Mitochondrial DNA deletion mutations in patients with neuropsychiatric symptoms. *Neuroscience Research* 69, 331-336.
- Kraytsberg, Y., and Khrapko, K. (2005). Single-molecule PCR: an artifact-free PCR approach for the analysis of somatic mutations. *Expert review of molecular diagnostics* 5, 809-815.
- Kraytsberg, Y., Nicholas, A., Caro, P., and Khrapko, K. (2008). Single molecule PCR in mtDNA mutational analysis: Genuine mutations vs. damage bypass-derived artifacts. *Methods* 46, 269-273.
- Kulawiec, M., Salk, J. J., Ericson, N. G., Wanagat, J., and Bielas, J. H. (2010). Generation, function, and prognostic utility of somatic mitochondrial DNA mutations in cancer. *Environmental and molecular mutagenesis* 51, 427-439.
- Laurie, M. T., Bertout, J. A., Taylor, S. D., Burton, J. N., Shendure, J. A., and Bielas, J. H. (2013). Simultaneous digital quantification and fluorescence-based size characterization of massively parallel sequencing libraries. *BioTechniques* 55, 61-67.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology* 2012, 11.
- Nazarian, R., Shi, H., Wang, Q., Kong, X., Koya, R. C., Lee, H., Chen, Z., Lee, M. K., Attar, N., Sazegar, H., *et al.* (2010). Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* 468, 973-977.
- Petros, J. A., Baumann, A. K., Ruiz-Pesini, E., Amin, M. B., Sun, C. Q., Hall, J., Lim, S., Issa, M. M., Flanders, W. D., Hosseini, S. H., *et al.* (2005). mtDNA mutations increase tumorigenicity in prostate cancer. *Proc Natl Acad Sci U S A* 102, 719-724.
- Pinheiro, L. B., Coleman, V. A., Hindson, C. M., Herrmann, J., Hindson, B. J., Bhat, S., and Emslie, K. R. (2012). Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal Chem* 84, 1003-1011.
- Quail, M. A., Kozarewa, I., Smith, F., Scally, A., Stephens, P. J., Durbin, R., Swerdlow, H., and Turner, D. J. (2008). A large genome center's improvements to the Illumina sequencing system. *Nat Methods* 5, 1005-1010.
- Salas, A., Yao, Y. G., Macaulay, V., Vega, A., Carracedo, A., and Bandelt, H. J. (2005). A critical reassessment of the role of mitochondria in tumorigenesis. *PLoS Med* 2, e296.
- Sequeira, A., Martin, M. V., Rollins, B., Moon, E. A., Bunney, W. E., Macciardi, F., Lupoli, S., Smith, E. N., Kelsoe, J., Magnan, C. N., *et al.* (2012). Mitochondrial mutations and polymorphisms in psychiatric disorders. *Frontiers in genetics* 3, 103.
- Spelbrink, J. N., Toivonen, J. M., Hakkaart, G. A., Kurkela, J. M., Cooper, H. M., Lehtinen, S. K., Lecrenier, N., Back, J. W., Speijer, D., Foury, F., and Jacobs, H. T. (2000). In vivo functional analysis of the human mitochondrial DNA polymerase POLG expressed in cultured human cells. *J Biol Chem* 275, 24818-24828.
- Taylor, R. W., Taylor, G. A., Durham, S. E., and Turnbull, D. M. (2001). The determination of complete human mitochondrial DNA sequences in single cells: implications for the study of somatic mitochondrial DNA point mutations. *Nucleic Acids Res* 29, E74-74.
- Taylor, R. W., and Turnbull, D. M. (2005). Mitochondrial DNA mutations in human disease. *Nat Rev Genet* 6, 389-402.
- Taylor, S. D., Ericson, N. G., Burton, J. N., Prolla, T. A., Silber, J. R., Shendure, J., and Bielas, J. H. (2013). Targeted enrichment and high-resolution digital profiling of mitochondrial DNA deletions in human brain. *Aging cell*.
- Vermulst, M., Bielas, J. H., and Loeb, L. A. (2008a). Quantification of random mutations in the mitochondrial genome. *Methods* 46, 263-268.
- Vermulst, M., Wanagat, J., Kujoth, G. C., Bielas, J. H., Rabinovitch, P. S., Prolla, T. A., and Loeb, L. A. (2008b). DNA deletions and clonal mutations drive premature aging in mitochondrial mutator mice. *Nat Genet* 40, 392-394.
- Xiao, M., Phong, A., Lum, K. L., Greene, R. A., Buzby, P. R., and Kwok, P. Y. (2004). Role of excess inorganic pyrophosphate in primer-extension genotyping assays. *Genome research* 14, 1749-1755.
- Zhang, Y., Zhang, D., Li, W., Chen, J., Peng, Y., and Cao, W. (2003). A novel real-time quantitative PCR method using attached universal template probe. *Nucleic acids research* 31, e123.

APPENDICES (PREVIOUSLY UPLOADED):

Appendix 1: Simultaneous digital quantification and fluorescence-based size characterization of massively parallel sequencing libraries. Laurie MT, Bertout JA, Taylor SD, Burton JN, Shendure JA, Bielas JH. *Biotechniques*. 2013 Aug; (2):61-7.

Appendix 2: Targeted enrichment and high-resolution digital profiling of DNA deletions in mitochondria. Taylor SD, Ericson NG, Burton JN, Prolla TA, Silber JR, Shendure J, Bielas JH. *Aging Cell*. 2013 Aug; [Epub ahead of print].

Appendix 3: Jason Bielas, PhD - Curriculum Vitae

TABLE & FIGURES:

Listed in Order Below

Figure 1

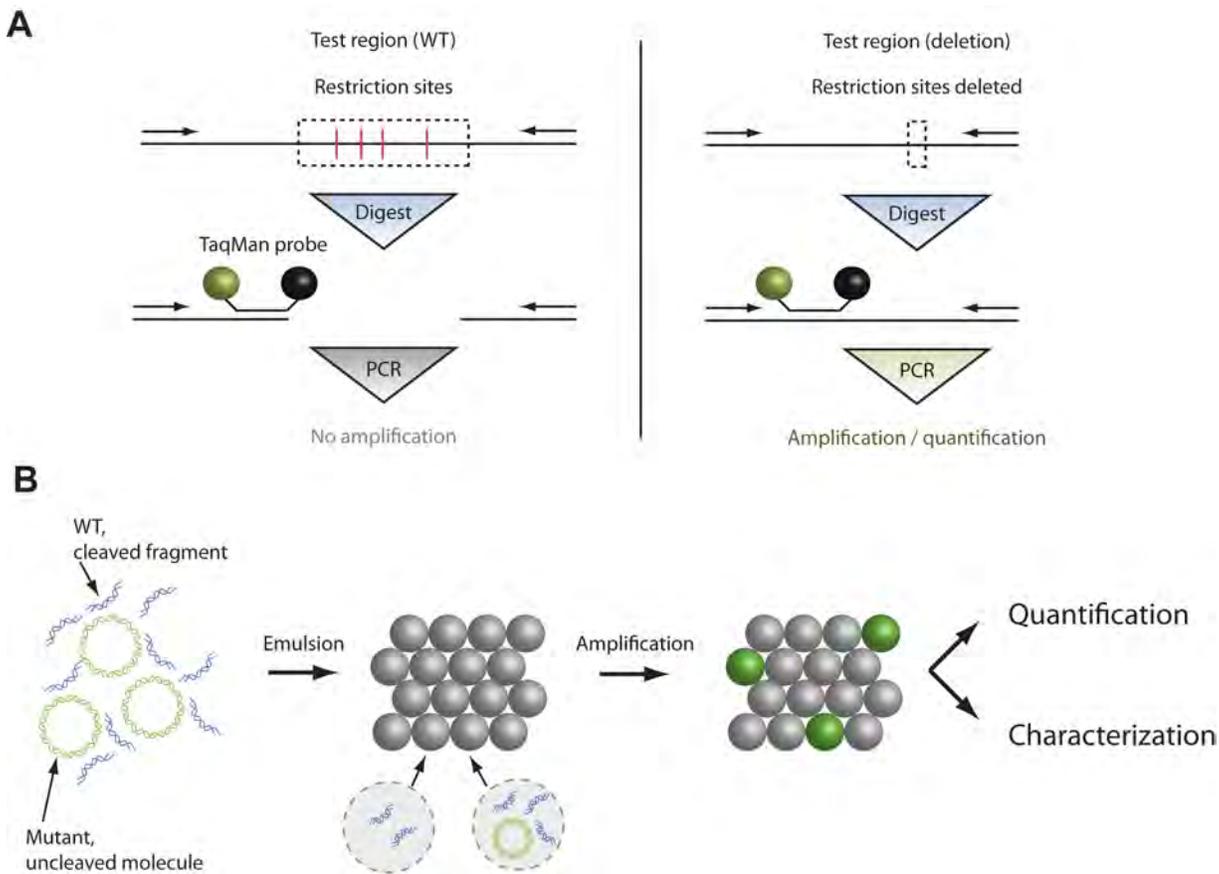


Figure 1. Overview of Digital Deletion Detection (3D). (A) Enrichment of deletion-bearing molecules. WT molecules harbor endonuclease recognition sites within the target region. Upon digestion, the target is cleaved, making the WT molecule unsuitable as a template for PCR amplification. In contrast, mutant molecules that harbor deletions and remove the restriction recognition sites are resistant to digestion. These molecules serve as templates for PCR amplification. The presence of the TaqMan® hydrolysis probe allows for the detection and enumeration of each molecule in the sample bearing the appropriate deletion. (B) Mutant target molecules (depicted as an idealized, unbroken circular mitochondrial chromosome) are individually sequestered into 1 nL water-in-oil droplets along with TaqMan® PCR chemistry and target-specific TaqMan® probes. Droplets are thermally cycled. Because the average number of molecules per droplet is less than one, positive droplets (green droplets) represent individual reaction vessels for single-molecule quantitative PCR amplification. Droplets are individually scanned and scored as positive or negative, thus providing a digital quantification of all deletion-bearing molecules within the sample. Alternatively, droplets can be disrupted and the amplification products subjected to physical characterization, for example cloning, sequencing, or other applications.

Figure 2

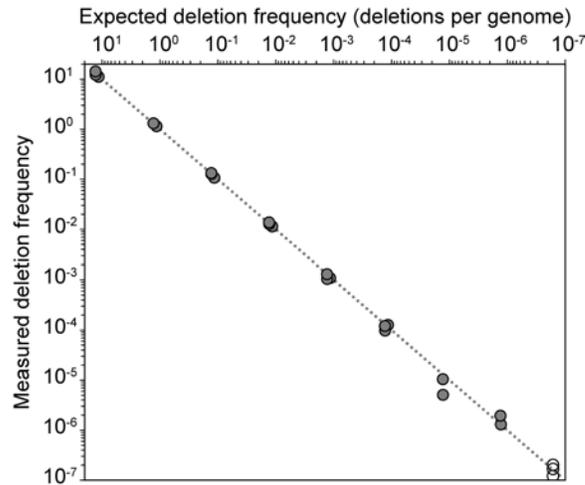


Figure 2. Sensitivity and recovery. 3D was performed on *TaqI*-digested HCT 116 mtDNA using primers and probes for the human ND1/ND2 site to give the endogenous deletion frequency (empty circles). Reconstruction experiments were performed by spiking in 3 molecules μL^{-1} of a control plasmid bearing a portion of the human mitochondrial genome with a known deletion ($3534\Delta 997$) into a serial dilution series of *TaqI*-digested HCT 116 mtDNA (filled circles). The predicted deletion frequency is plotted against the measured deletion frequency. Each data point represents an individual experiment. The reconstruction data were fit to $y = x$ (dotted line) with a correlation coefficient $R^2 = 0.9942$.

Figure 3

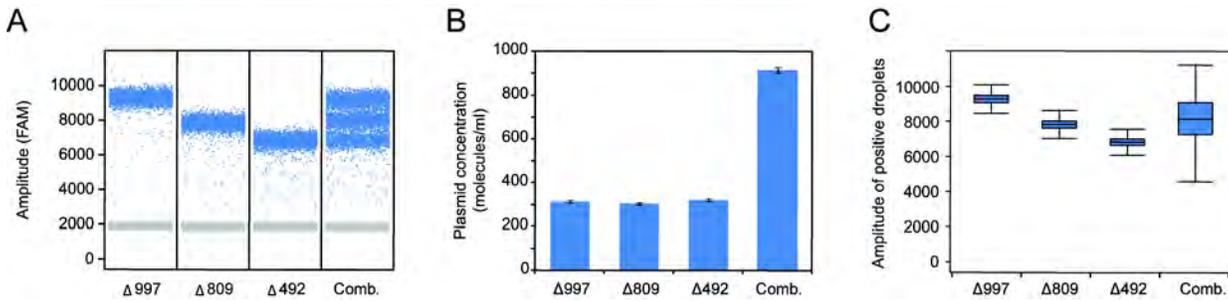


Figure 3. Effects of sample heterogeneity on 3D analysis. (A) Three plasmid controls (3534 Δ 997, 3719 Δ 809, and 3871 Δ 492) were diluted to an expected concentration of 300 molecules μL^{-1} template and subjected to 3D analysis, either individually or combined. Blue dots represent droplets whose amplitudes are above the threshold ('positives'), while gray droplets are those whose amplitudes are below the threshold value ('negatives'). (B) Measured deletion concentration for individual and combined templates. Error bars indicate the Poisson 95% confidence intervals for each concentration determination. (C) Box and whisker plot showing the distribution of positive droplets for each template. When used as a template in PCR, each plasmid yields different size fragments (185 bp, 372 bp, and 686 bp, respectively). There is an inverse relationship between average fluorescence amplitude and template length, as well as a relationship between the sample complexity and the breadth of the distribution of positive droplets.

Figure 4

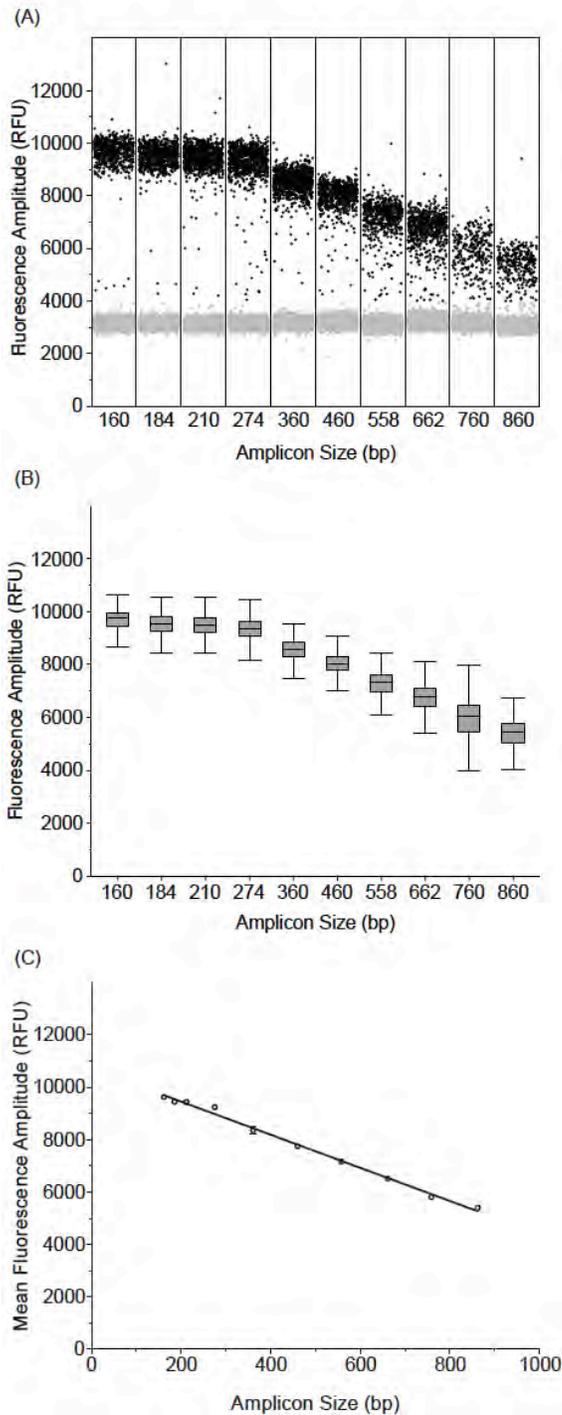


Figure 4. ddPCR amplification of 10 size standards designed for use with the QuantiSize assay. All size standards were amplified in parallel with standard reagent and thermal cycling conditions. (A) Scatter plot of fluorescence amplitude of individual droplets for each size standard. Droplets whose fluorescence amplitude is above a specified threshold (“positives”) are shown in black and droplets with fluorescence amplitude below the threshold (“negatives”) are shown in grey. (B) Box-and-whisker plots showing distribution of fluorescence amplitudes of positive droplets. Horizontal bars mark the mean fluorescence amplitude, boxes mark the interquartile range, and whiskers mark the 95% confidence interval. (C) Plot of mean fluorescence amplitude \pm SEM versus amplicon size showing a linear correlation ($R^2=0.9943$).

Figure 5

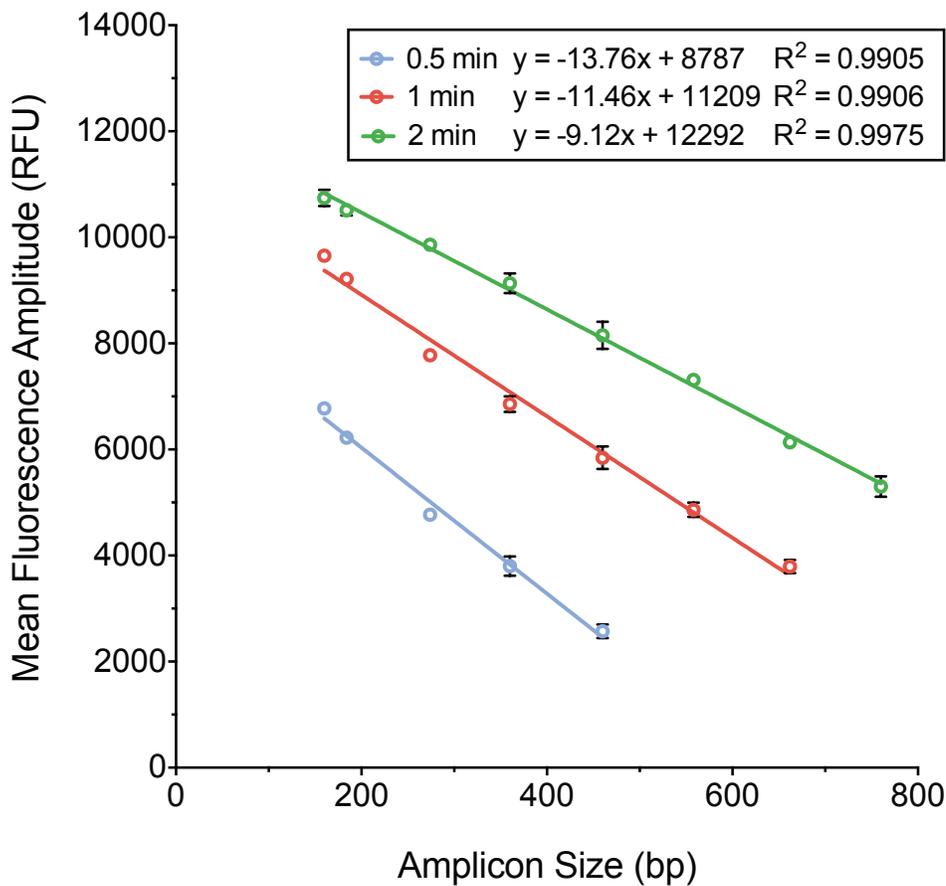


Figure 5. Effect of ddPCR elongation time on the relationship between fluorescence amplitude \pm SEM and amplicon size. Three ddPCR experiments were carried out with the same size standards using 0.5, 1, and 2 minute elongation times during droplet thermal cycling. With a 0.5 minute elongation time (blue), the slope of the regression line relating fluorescence amplitude to amplicon size was -13.760 ($R^2=0.9905$). With a 1 minute elongation time (red), the slope was -11.460 ($R^2=0.9906$). With a 2 minute elongation time (green), the slope was -9.123 ($R^2=0.9975$). As the magnitude of the slope of the relationship between fluorescence amplitude and amplicon size increases, so does the ability to accurately resolve small differences in amplicon size. Larger templates require longer elongation times for positive droplets to fluoresce discernibly above the background level of droplet fluorescence.

Figure 6

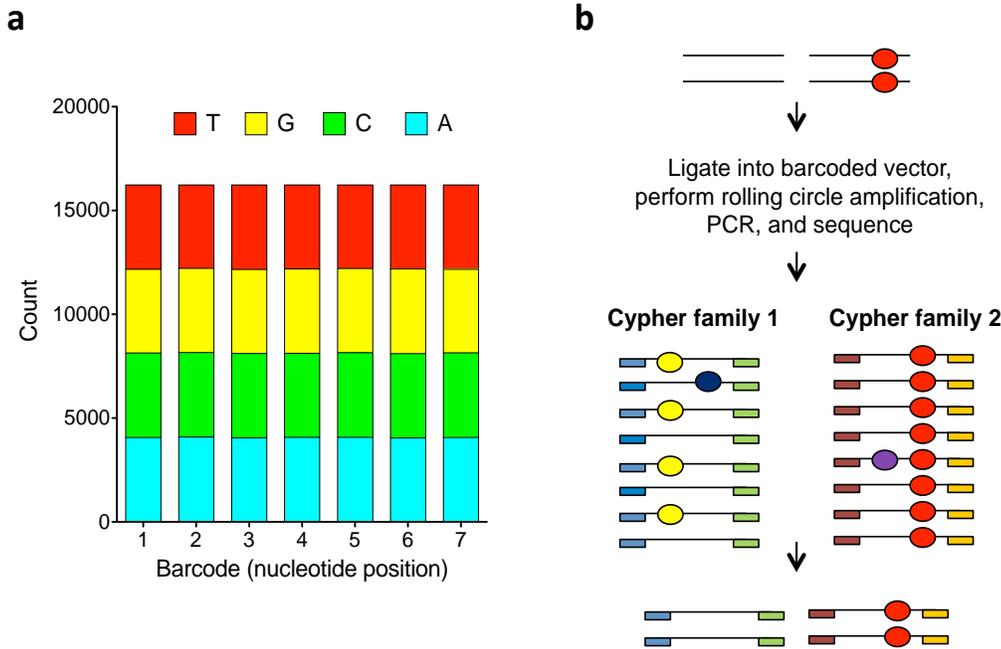


Figure 6 (a) Data generated from 5.1 million reads in a single NGS run on our MiSeq® demonstrates optimal coverage and diversity at the upstream seven base pair cypher in our vector library (b) Rolling Cypher Seq eliminates errors introduced during library preparation and sequencing. By ligating our target into our vector containing double-stranded cyphers, grouping all reads with identical cyphers and their reverse complements into families and creating a consensus sequence, we can computationally eliminate errors introduced during library preparation (yellow circles) and during sequencing (blue and purple circles). Only mutations that are present in all reads (red circles) from the same cypher and its reverse complement will be counted as true.

Figure 7

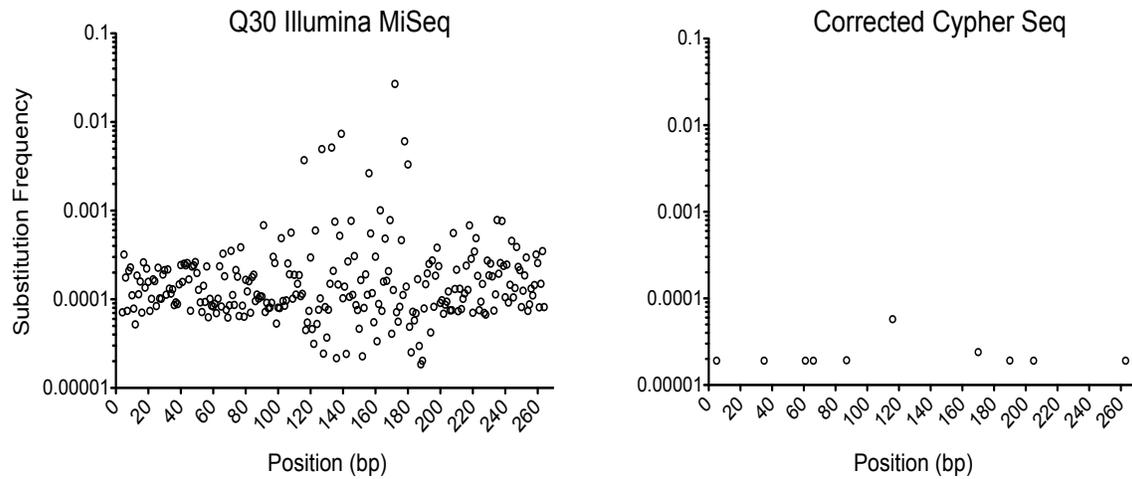


Figure 7 Exon 4 of TP53 was TOPO-cloned, transformed into E-coli, mini-prepped, and sequenced by Sanger sequencing. Wild-type Exon 4 was then ligated into a library of Cypher Seq vectors and sequenced on the Illumina MiSeq instrument with a depth of over a million. Sequences were then compared to wild-type TP53 sequence and substitutions were plotted before (left panel) and after correction (right panel) with Cypher Seq. Remaining base substitutions most likely reflect errors introduced during replication in E-coli prior to ligation into the barcoded vectors.

Table 1: Clonal mtDNA mutations identified in patient-matched tissue and blood.

Patient ID	Gleason Score ^a	Gene ^b	DNA ^c	Protein ^d	Tumor (%) ^e	Normal (%) ^f	Blood (%) ^g	Restriction Site ^h
23300B	3+3	ND5	13913T>C	L526P	50	N.D.	N.D.	Hpy188I
22871B	3+3	ATPase6	9038T>C	M171T	50	N.D.	N.D.	FatI, CviAII, NlaIII, HpyCH4V
		D-loop	16092T>C		40	70	80	
23388H	3+3							
23529W	3+3	16S rRNA	2107G>A		100	N.D.	N.D.	MnII
23570B	3+3	ND4	11256A>G	Y166C	N.D.	60	N.D.	
23481P	3+3							
23002C	3+4							
23159R	3+4	D-loop	16145G>A		100	20	10	
22896V	3+4							
23204T	3+4							
23378W	3+4	ND1	3643G>A	V113I	60	N.D.	N.D.	Hpy166II
23390H	3+4							
23569O	3+4	ND3	10228T>C	L57S	100	N.D.	N.D.	
23036D	4+3	ND5	13525G>A	E397K	100	N.D.	60	TaqI
23171D	4+3	D-loop	313delCCCCGCTTCT		50	N.D.	50	AcI, BslI, FauI
		ND2	4752T>C	S95P	100	N.D.	50	
		D-loop	16093T>C		100	60	80	
22962M	4+3	D-loop	309delC		50	N.D.	N.D.	
22959H	4+3							
23253G	4+3	Cyt b	15750T>C	L335P	100	N.D.	N.D.	
22927H	4+3	D-loop	251G>A		70	N.D.	N.D.	
		Cyt b	14846G>A	G34S	100	N.D.	N.D.	
22888L	4+3	D-loop	16027T>C		100	N.D.	40	
19575Y	3+3							
22951A	3+3							
22949P	3+3	tRNA E	14724G>A		50%	N.D.	N.D.	
23127O	3+3							
22904L	3+3	ND1	3982G>A	A226T	80%	N.D.	N.D.	CviKI-1
		Cyt b	15345T>C	L200S	70%	N.D.	N.D.	HpyCH4V
22880H	3+3	D-loop	234A>G		100%	40%	40%	
		COXIII	9942G>A	D246N	50%	N.D.	N.D.	
23024K	3+3							
23270S	3+3							
22905A	3+4							
19416E	3+4	Cyt b	14774insC	frameshift	100%	N.D.	N.D.	
22860F	3+4	D-loop	523delAC		100%	N.D.	N.D.	
		12S rRNA	1282G>A		80%	N.D.	N.D.	
		COXII	8269G>A		100%	20%	30%	SfcI
		ND3	10320A>G	I88V	100%	N.D.	N.D.	
23027B	3+4	COXI	6131A>G		60%	N.D.	N.D.	Hpy188I
		COXI	6910C>T	A336V	80%	N.D.	N.D.	TseI, ApeKI, Fnu4HI, SfcI
22870S	3+4	D-loop	16171G>A		60%	N.D.	N.D.	
22975M	3+4							
23201A	3+4	ND3	10264T>C	I69T	70%	N.D.	N.D.	Tsp509I
23028W	3+4	12S rRNA	1547insT		100%	70%	50%	
23266N	3+4							
23312L	3+4							
23168H	4+3	ND5	12473T>C	I46T	80%	N.D.	N.D.	
23183S	4+5	ND5	13480G>A	G382stop	80%	N.D.	N.D.	
		Cyt b	14798T>C	F18L	50%	N.D.	N.D.	
23461T	4+3 partially 5	12S rRNA	1440G>A		N.D.	50	N.D.	DdeI

^a Tumors were graded based on the Gleason Grading System, with the first number indicating the grade of the majority (>50%) of the tumor (on a scale from 1-5), and the second number signifying the grade of the minority (<50%, but >5%) of the tumor.

^b The region or gene in the mitochondrial genome where the mutation was detected by Sanger sequencing

^c The mutation is indicated by the base position in mtDNA, followed by the type of change. A T to C substitution at position 1000 would be described as 1000T>C, while a deletion of a G at position 500 would be 500delG.

^d Amino acid change resulting from the mutation, indicated by the original amino acid followed by the position of the residue, and then the resulting amino acid.

^e Prevalence of mutation in LCM cancerous prostate tissue as a percentage, based on Sanger sequencing chromatogram reads.

N.D. indicates that the mutation was not detectable via Sanger sequencing.

^f Prevalence of mutation in LCM normal prostate tissue as a percentage, based on Sanger sequencing chromatogram reads.

^g Prevalence of mutation in blood DNA as a percentage, based on Sanger sequencing chromatogram reads.

^h If the mutation disrupts a restriction site, the corresponding restriction enzymes are listed.

Reports

Simultaneous digital quantification and fluorescence-based size characterization of massively parallel sequencing libraries

Matthew T. Laurie¹, Jessica A. Bertout¹, Sean D. Taylor¹, Joshua N. Burton², Jay A. Shendure², and Jason H. Bielas^{1,3,4}
¹*Translational Research Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA,*
²*Department of Genome Sciences, University of Washington, Seattle, WA,* ³*Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, WA,* and ⁴*Department of Pathology, University of Washington, Seattle, WA*

BioTechniques 55:61-67 (August 2013) doi 10.2144/000114063

Keywords: droplet digital PCR; absolute quantification; next-generation sequencing; massively parallel sequencing; library preparation; quality control; size determination

Due to the high cost of failed runs and suboptimal data yields, quantification and determination of fragment size range are crucial steps in the library preparation process for massively parallel sequencing (or next-generation sequencing). Current library quality control methods commonly involve quantification using real-time quantitative PCR and size determination using gel or capillary electrophoresis. These methods are laborious and subject to a number of significant limitations that can make library calibration unreliable. Herein, we propose and test an alternative method for quality control of sequencing libraries using droplet digital PCR (ddPCR). By exploiting a correlation we have discovered between droplet fluorescence and amplicon size, we achieve the joint quantification and size determination of target DNA with a single ddPCR assay. We demonstrate the accuracy and precision of applying this method to the preparation of sequencing libraries.

Massively-parallel next-generation sequencing (NGS) technology is rapidly revolutionizing the fields of genomics, molecular diagnostics, and personalized medicine through the increasingly efficient and economical generation of unprecedented volumes of data (1–7). A common characteristic of the various commercially available NGS technologies is the need to load a precise number of viable DNA library molecules onto the instrument to optimize the yield of data from an individual sequencing experiment (8–11). Performing a sequencing run with either too many or too few library molecules results in compromised data yields or completely failed sequencing runs that waste sample, expensive reagents, user time, and instrument time. Similarly, if library molecules are not the appropriate length to fully utilize the capabilities of the sequencing platform, fewer bases can be sequenced in an NGS run and the throughput is wasted. Thus, accurate quantification and size determination of library DNA is essential for achieving optimal data yield and maximizing a laboratory's efficiency and sequencing throughput.

Protocols for the preparation of NGS libraries include quality control steps to validate the size and concentration of amplifiable library molecules (i.e., molecules properly ligated to NGS adapter sequences) before committing to a sequencing run. Manufacturers typically recommend quantification with quantitative real-time PCR (qPCR) and size determination with gel or capillary electrophoresis. It is also possible to enumerate library DNA using UV spectrophotometry, the Quant-iT PicoGreen assay, or the Agilent BioAnalyzer, but these methods are not ideal because they quantify amplifiable and non-amplifiable molecules equally (12–14). These methods are also only capable of measuring mass per volume, which must be converted to copy number using an estimated average size of library molecules which can introduce further error (15). Although qPCR is widely considered the best option for library quantification, there are considerable drawbacks to the method, including amplification biases due to template size and GC-content as well as the need for a standard curve to estimate the absolute quantity of DNA (16). Creating a standard curve

for each sample to be analyzed is a difficult and uncertain process that leads to inaccuracies in measurements of absolute target quantity (15,17). When intercalating dyes are used for quantification, the concentration reading can include non-amplifiable DNA as dyes measure dsDNA indiscriminately. Because of these potential inaccuracies, some NGS platform manufacturers recommend performing titration runs on their instrument to determine the proper loading amount. The high cost of reagents and the length of NGS runs make this an expensive and time-consuming step.

We have developed a new assay capable of concurrently measuring the absolute concentration and length of unknown amplifiable DNA templates, making it well suited for quality control of NGS libraries. The assay, which we have termed *QuantiSize*, is based on the previously validated droplet digital PCR (ddPCR) absolute quantification system (18,19) and adds the ability to calculate the size of target DNA by exploiting a linear correlation we have discovered between the fluorescence amplitude of ddPCR droplets and the size of amplicons within them.

Method summary:

QuantiSize allows for the determination of absolute quantity and size distribution of target DNA molecules in a single experiment. This assay exploits a correlation, reported herein, between the length of an amplified DNA molecule and the fluorescence amplitude produced in droplet digital PCR (ddPCR), to allow the user to calculate the size of unknown DNA. As ddPCR simultaneously measures the concentration of target DNA, the user can accurately determine the target population size and quantity in a single step.

As a quantification method, ddPCR has demonstrated greater precision and sensitivity than real-time PCR (18). We demonstrate that QuantiSize accurately measures the size and concentration of target DNA simultaneously while avoiding the limitations of other quantification systems and, we highlight the utility of this assay for preparation of NGS libraries.

Materials and methods

Purification of DNA size standards

An exACTGene 50 bp DNA Ladder (Fisher, Waltham, MA) and 1 kb Plus DNA Ladder (Fisher) were run on a 1.0% UltraPure Low-Melting Point Agarose (Invitrogen, Carlsbad, CA) electrophoresis gel and the 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, and 1000 bp ladder bands were manually excised. The DNA in these gel slices was purified using the QIAcube automated gel extraction protocol with the QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany). The size and purity of all DNA fragments were verified by gel electrophoresis.

Ligation of DNA size standards to adapter sequences

The DNA fragments were ligated to the following sequences containing the Nextera version 1 adapters (Epicenter Biotechnologies, Madison, WI):

Adapter 1:

5'-CAAGCAGAAGACGGCATAACGAGATTTCGCCT-TACGGTCTGCCTTGCCAGCCCGCTCAGAGATGTGTATAAGAGACAG[index 1]CCC-3'

Adapter 2:

5'-GGG[index 2]CTGTCTCTTATACACATCTCTGATGGC-GCGAGGGAGGCGCGATCTAGTGTAGATCTCGGTGGTCGCCGTATCATT-3'

All size standard fragments were ligated to adapters with the following 7 bp indices:

Index 1: 5'-TACCTCT-3'

Index 2: 5'-ACACATT-3'

Ligations were carried out in 20 μ L reactions containing 1 μ L T4 DNA Ligase and 2 μ L 10 \times T4 DNA Ligase Buffer (New England BioLabs, Ipswich, MA). The ligation reactions were incubated at room temperature for 2 h. The ligated DNA was purified with a phenol/chloroform/isoamyl alcohol extraction. All samples were sent to the Fred Hutchinson Cancer Research Center ABI capillary sequencing facility to verify that the correct insert had been ligated to the adapters in each sample.

Library preparation for Illumina MiSeq

Samples of the plasmid pET-23a were sheared to an average size of 150 bp using the Covaris S220 Ultrasonicator (Covaris, Woburn, MA). The sheared DNA was run on a 1.0% UltraPure Low-Melting Point Agarose gel and a gel slice corresponding to ~100–200 bp was manually excised and purified using the QIAcube gel extraction protocol. The sheared DNA was blunted and phosphorylated using the Quick Blunting Kit (New England BioLabs) and purified with a phenol/chloroform/isoamyl alcohol extraction.

Eight samples of sheared DNA were ligated to adapter sequences with unique 7 bp indices in separate 20 μ L reactions containing 1 μ L T4 DNA Ligase and 2 μ L 10 \times T4 DNA Ligase Buffer. The ligation reactions were incubated at room temperature for 2 h. All ligations were purified with a phenol/chloroform/isoamyl alcohol extraction.

All ligated libraries were amplified with 20 cycles of standard PCR using the following primer sequences:

Primer 1:

5'-AATGATACGGCGACCACCGA-3'

Primer 2:

5'-CAAGCAGAAGACGGCATAACGA-3'

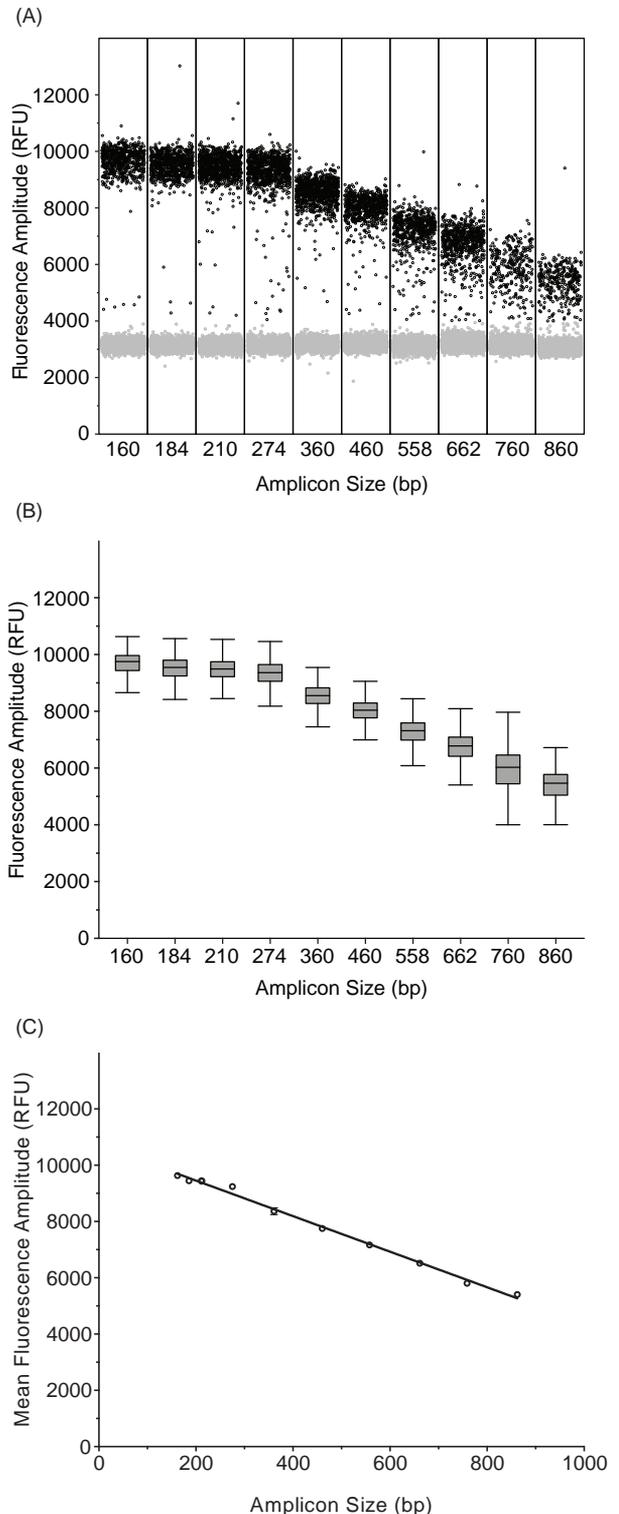


Figure 1. ddPCR amplification of 10 size standards designed for use with the QuantiSize assay. All size standards were amplified in parallel with standard reagent and thermal cycling conditions. (A) Scatter plot of fluorescence amplitude of individual droplets for each size standard. Droplets whose fluorescence amplitude is above a specified threshold (“positives”) are shown in black and droplets with fluorescence amplitude below the threshold (“negatives”) are shown in gray. (B) Box-and-whisker plots showing distribution of fluorescence amplitudes of positive droplets. Horizontal bars mark the mean fluorescence amplitude, boxes mark the interquartile range, and whiskers mark the 95% confidence interval. (C) Plot of mean fluorescence amplitude \pm SEM vs. amplicon size showing a linear correlation ($R^2 = 0.9943$).

Figure 2

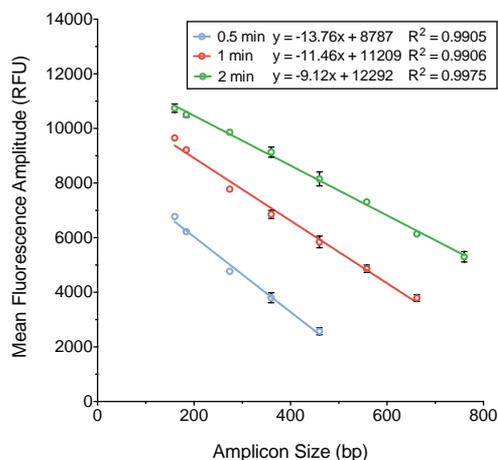


Figure 2. Effect of ddPCR elongation time on the relationship between fluorescence amplitude \pm SEM and amplicon size. Three ddPCR experiments were carried out with the same size standards using 0.5, 1, and 2 min elongation times during droplet thermal cycling. With a 0.5 min elongation time (blue), the slope of the regression line relating fluorescence amplitude to amplicon size was -13.760 ($R^2 = 0.9905$). With a 1 min elongation time (red), the slope was -11.460 ($R^2 = 0.9906$). With a 2 min elongation time (green), the slope was -9.123 ($R^2 = 0.9975$). As the magnitude of the slope of the relationship between fluorescence amplitude and amplicon size increases, so does the ability to accurately resolve small differences in amplicon size. Larger templates require longer elongation times for positive droplets to fluoresce discernibly above the background level of droplet fluorescence.

The amplified libraries were quantified using the ddPCR system (Bio-Rad, Hercules, CA) and the QuantiT PicoGreen assay (Invitrogen). In the ddPCR experiment, the libraries were run in parallel with adapter-ligated size standards to allow for the estimation of library size distribution. The measured concentrations of the 8 differently indexed libraries were used to dilute and combine the libraries in a molar ratio of 100:50:10:1 with 2 libraries at each concentration. The combination of libraries was denatured and diluted in preparation for loading onto the MiSeq flow cell (Illumina, San Diego, CA) as per the Illumina protocol.

Genomic DNA was purified from the human colon cancer cell line HCT 116 using the QIAcube automated purification protocol with the DNeasy Kit (QIAGEN). The Nextera XT DNA Sample Preparation Kit (Illumina) was used to generate a MiSeq compatible library from the HCT 116 DNA. The optional bead-based normalization step in the Nextera XT protocol was omitted and the library was instead normalized by quantification with ddPCR and the volume was adjusted to 2 nM as required by the MiSeq loading protocol. The 2 nM library was denatured and further diluted as per the manufacturer's guidelines. The standard phi X control library (Illumina) was spiked into the denatured HCT 116 library at 5% by volume. The library and phi X mixture were then loaded into a MiSeq 300-Cycle v2 Reagent Kit (Illumina).

TaqMan probe and primer design

The following primers (Invitrogen) and TaqMan probe (Applied Biosystems, Foster City, CA) were designed to hybridize to the Nextera adapter sequences:

Primer 1:

5'-GCGACCACCGAGATCTACAC-3'

Primer 2:

5'-AGCAGAAGACGGCATAACGAG-3'

Probe:

5'-FAM-CTGT+CT+CT+TA+TA+CA+CATC-IBFQ-3'

The '+' indicates that the previous base is a locked nucleic acid (LNA) base.

Droplet digital PCR

All quantified DNA libraries and size standards were prepared for droplet PCR in 25 μ L reactions containing 2 \times ddPCR Master Mix (Bio-Rad), 250 nM TaqMan probe, 900 nM each of the appropriate flanking primers, and \sim 10,000 copies of target DNA. Emulsified 1 nL reaction droplets were made by adding 20 μ L of each reaction mixture to the sample wells of a droplet generator DG8 cartridge (Bio-Rad) and 70 μ L ddPCR Droplet Generation Oil (Bio-Rad) to the oil wells of the cartridge for use in the QX100 Droplet Generator (Bio-Rad). Forty microliters of the generated droplet emulsions were transferred to Twin.tec semi-skirted 96-well PCR plates (Eppendorf, Hamburg, Germany), which were then heat sealed with pierceable foil sheets. To amplify the target DNA, the droplet emulsions were thermally cycled using the following protocol: initial denaturation at 95°C for 10 min, followed by 40 cycles of 94°C for 30 s and 60°C for 1 min. The fluorescence of each thermally cycled droplet was measured using the QX100 Droplet Reader. All measurements were performed in triplicate.

Data analysis

The equation of the line fitting the correlation between amplicon size and fluorescence amplitude for the size standards was generated using Microsoft Excel (Redmond, WA) and applied to the measured fluorescence amplitude of each sequencing library to calculate amplicon size. The .fastq data files produced by the MiSeq were imported to Sequencher (Gene Codes, Ann Arbor, MI) and aligned to the pET-23a plasmid sequence to generate a sequence alignment/map file (SAM). A perl script was used to count the length of each read pair by retrieving the number corresponding to the "TLEN" field of the SAM file. Only library molecules for which both paired-end reads passed the quality filter were included in the analysis.

Results and discussion

The ability of the QuantiSize assay to combine quantification and size determination in a single ddPCR experiment is derived from a correlation that exists between the fluorescence amplitude of droplets and the size of amplicons within them. With standard ddPCR reagent concentrations, DNA amplification is eventually limited by the availability of dNTPs and inhibited by the presence of pyrophosphate (19,20); thus long DNA templates, which consume more dNTPs and generate more pyrophosphate, will produce fewer products than short templates at the end point of a standard reaction. Because the final number of products generated within a droplet determines its level of fluorescence, the measured fluorescence amplitude of droplets containing short templates will be greater than that of droplets containing long templates. The QuantiSize assay exploits this fact to generate an equation relating fluorescence amplitude to amplicon size by using measurements of known size standards. The equation describing the relationship between fluorescence amplitude and amplicon size can be used to calculate the size of any unknown ddPCR template that shares common primer and probe binding sites with the size standards. Creating size standards that have primer and probe binding sites in common with DNA samples can be accomplished in a number of ways including cloning sample DNA into a vector and appending adapter sequences to both the sample DNA and size standards (21).

We created a set of size standards applicable to Illumina NGS libraries containing inserts ranging from 25 to 1000 base pairs flanked by adapter sequences compatible with the Illumina MiSeq platform. A pair of primers and a fluorescent TaqMan probe were designed to hybridize to the adapter sequences such that the length of each amplicon is 160 bp plus the length of the insert. As primers and probe are specific to the MiSeq adapter sequences, only adapter-ligated molecules that will be amplifiable on the MiSeq flow cell will be quantified.

A ddPCR experiment was performed with the aforementioned size standards in separate wells of a 96-well plate. Droplets containing the target (positive) increased in fluorescence following amplification of the target whereas droplets lacking the target (negative) remained at the background

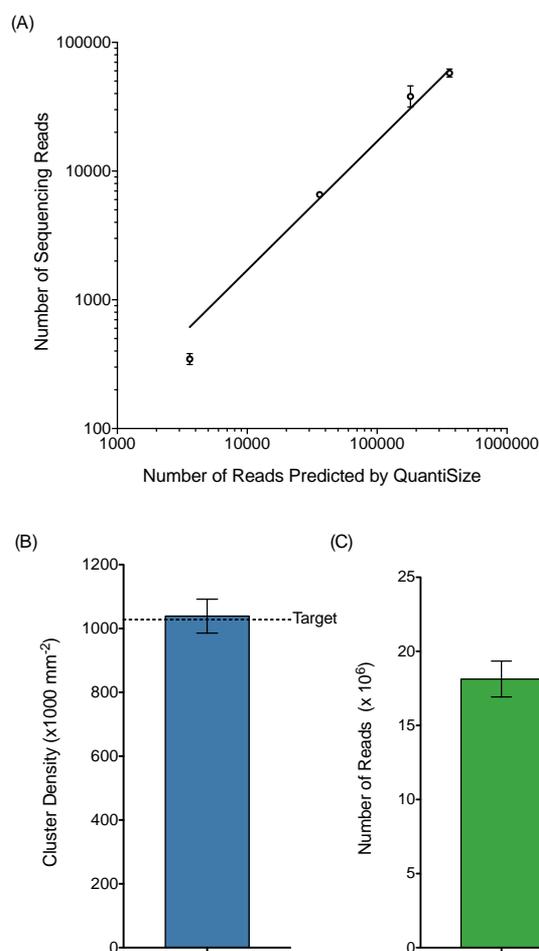


Figure 3. Cluster density and number of sequencing reads \pm SEM across multiple sequencing runs performed using QuantiSize. (A) Eight uniquely indexed libraries were loaded onto the MiSeq with two libraries at each concentration. The libraries were loaded in a concentration ratio of 100:50:10:1, based on ddPCR measurements. Due to the binding kinetics of library molecules on the MiSeq flow cell, the number of reads generated by the MiSeq is expected to be a fraction of the number of library molecules loaded. The relative numbers of MiSeq reads for each library closely correspond to the relative numbers of molecules loaded according to ddPCR measurements ($R^2 = 0.9693$). (B) Mean cluster density (\pm SEM) resulting from three separate sequencing runs using Nextera-prepared samples normalized based on QuantiSize measurements. The target cluster density (represented by a horizontal dashed line) was 1.028×10^6 clusters/mm² including a 5% phi X control and the observed mean cluster density from the 3 runs was $1.039 \pm 0.053 \times 10^6$ cluster/mm². (C) Mean number of reads (\pm SEM) resulting from three separate sequencing runs. The observed mean number of reads was $1.813 \pm 0.070 \times 10^7$.

level of fluorescence (Figure 1A). The distribution of droplet amplitudes is consistent across most amplicon lengths, but the 760 and 860 bp amplicons show a broader distribution of amplitudes (Figure 1B). An inverse, linear correlation between amplicon size and mean fluorescence amplitude was observed ($R^2 = 0.99436$) (Figure 1C). The equation describing this correlation allows for the calculation of amplicon size given a measured fluorescence amplitude. The slope of this equation provides a measure of the difference in mean fluorescence amplitude that is expected with a given difference in amplicon size. Maximizing the magnitude of this slope maximizes the resolution of size standards, which is advantageous for the purpose of determining the length of unknown amplicons more accurately. The size standards used for QuantiSize are highly analogous to the standards used in gel and capillary electrophoresis. The size of unknown DNA can be determined by visually comparing the fluorescence amplitude of the size references to that of the unknown DNA or by entering the fluores-

cence amplitude value into the equation describing the relationship between average fluorescence amplitude and amplicon size for the size standards.

The droplet reader software counts positive and negative droplets by using a threshold of fluorescence between the well-defined populations of high and low fluorescence amplitude droplets. One particular TaqMan probe displayed a fluorescence amplitude for droplets containing amplicons larger than 660 bp that is too low to reliably discriminate between positive and negative droplets when templates are amplified with a 1 min elongation time. When this is the case, the average fluorescence amplitude for these amplicons cannot be calculated. Increasing the elongation time to two minutes increases the fluorescence amplitude of all droplets containing amplifiable template (Figure 2). This enables the acquisition of accurate concentration and fluorescence amplitude data for longer templates, but the slope of the relationship between amplicon size and fluorescence amplitude is decreased (from $m = -11.66$ to $m = -9.12$), which decreases the ability to resolve small differences in amplicon size (Figure 2). Decreasing the elongation time to 30 s increases the resolution of the relationship between amplicon size and fluorescence amplitude, but prevents targets longer than 460 bp from amplifying to the point that they fluoresce detectably above the background fluorescence (Figure 2). This is likely due to the fact that longer products require more time for complete polymerization of nascent strands to occur. Thus, there is a tradeoff between the resolution and range of QuantiSize, though the assay can be easily adjusted to fit particular experimental needs.

To validate the use of the QuantiSize assay for the sizing and quantification steps in library preparation for NGS, we compared the quantity and size distribution of library DNA predicted by the QuantiSize assay to the quantity and size distribution of reads generated by the Illumina MiSeq platform. Eight uniquely indexed test libraries were generated by ligating DNA sheared to an average size of 150 bp onto MiSeq compatible adapter sequences similar to those used to create the aforementioned size standards. The libraries were run in individual wells of a ddPCR experiment alongside the set of size standards. Using the concentrations measured by ddPCR, the 8 uniquely indexed libraries were diluted and combined in a molar ratio of 100:50:10:1 with 2 libraries at each concentration. The observed number of MiSeq reads containing each index was compared with the expected number of copies of each uniquely indexed library loaded onto the MiSeq. The observed ratio of the number of reads containing each index very closely matched the expected ratio of 100:50:10:1 that was measured by ddPCR and the correlation between the expected and actual number of library molecules gave an R^2 value of 0.9693 (Figure 3A).

The Nextera XT DNA Sample Preparation Kit was used to prepare a sequencing library from genomic DNA extracted from the human colon cancer cell line HCT 116. In lieu of the optional bead normalization step in the Nextera XT protocol, the concentration and size distribution of the library were measured with QuantiSize and the library concentration was adjusted to the proper concentration based on this measurement. The process of quantifying, normalizing, denaturing, and loading the library onto the MiSeq was repeated three times to demonstrate the precision of QuantiSize for predicting cluster density. The target cluster density for each MiSeq run was 1.0×10^6 clusters/mm² (1.028×10^6 clusters/mm² including the phi X control DNA). The mean cluster density \pm SEM obtained was $1.039 \pm 0.053 \times 10^6$ clusters/mm² (Figure 3B). The mean number of reads \pm SEM obtained was $1.813 \pm 0.070 \times 10^7$ (Figure 3C). There are several potential sources of error in the MiSeq sample loading process including the pipetting of small volumes, variability in the efficiency of the denaturation reaction, variability of flow cell surface area, and user error. These factors may account for some of the observed variance in cluster density. However, even with the potential error caused by these factors, the target cluster density was achieved with high precision using the QuantiSize method.

The equation relating amplicon size and fluorescence amplitude can be applied either to the average (mean or median) fluorescence amplitude of a sample or to the fluorescence amplitude of individual droplets. Applying

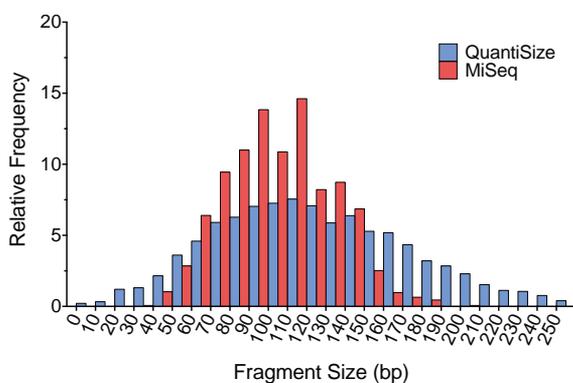


Figure 4. Comparison of library molecule size distribution. The QuantiSize assay was performed on a DNA library prepared for the MiSeq in order to predict the distribution of library molecule sizes. The DNA library was amplified in parallel with a set of size standards using the same primers and TaqMan probe, allowing us to estimate the expected amplicon size within each individual droplet. The resulting size distribution is shown in blue. The actual size distribution was determined through paired-end sequencing on the Illumina MiSeq system (shown in red). Both histograms show the relative frequency of measured molecule sizes in ten base pair bins. The size distribution measured by QuantiSize is naturally wider than the distribution measured by the MiSeq due to the inherent variance in droplet amplitude that occurs even with amplicons of the same length. The DNA library was amplified in parallel with a set of size standards using the same primers and TaqMan probe, allowing us to estimate the expected amplicon size within each individual droplet.

the equation to individual droplets allows for a more detailed analysis of the distribution of product sizes present in a sample. We applied the equation generated by the adapter-ligated size standards to the fluorescence amplitude of individual droplets containing library DNA to calculate the expected amplicon size within each droplet and compared the distribution of sizes to the distribution of read sizes measured by the MiSeq (Figure 4). The frequency distribution shows a high degree of overlap and a common center point for the estimation made using the QuantiSize assay and the observations from the MiSeq. As depicted in Figure 2A, there is an inherent variance in droplet amplitude that occurs even within a completely homogeneous sample of amplicon lengths. This variance likely accounts for the wider distribution of product sizes estimated by ddPCR than were observed in the MiSeq data. Size determination with QuantiSize provides a detailed calculation of the distribution of fragment sizes present in a sample, whereas gel or capillary electrophoresis provide only a range of sizes.

The QuantiSize assay demonstrates accuracy, reliability, and flexibility through the strength of the correlation between fluorescence amplitude and amplicon size in ddPCR experiments, and the ease with which the assay can be adjusted to fit specific experimental needs. Applying the QuantiSize assay to NGS library preparation avoids the limitations of other independent quantification and size determination methods and has the potential to increase the average yield of usable data generated from sequencing runs, thereby increasing the efficiency and throughput. The ability to determine the absolute quantity and the detailed size distribution of target DNA with a single experiment will be useful for a broad range of applications that require the quantification and sizing of target DNA.

Acknowledgments

We would like to thank Nolan Ericson and Mariola Kulawiec for helpful discussion and critical reading of our manuscript. This work was supported by an Ellison Medical Foundation New Scholar award (AG-NS-0577-09), an Outstanding New Environmental Scientist Award (ONES) (R01) from the National Institute of Environmental Health Sciences (R01ES019319), and a grant (W81XWH-10-1-0563) from the

Congressionally Directed Medical Research Programs/U.S. Department of Defense. This paper is subject to the NIH Public Access Policy

Competing interests

The authors declare no competing interests.

References

- Didelot, X., R. Bowden, D.J. Wilson, T.E. Peto, and D.W. Crook. 2012. Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* 13:601-612.
- Biesecker, L.G., W. Burke, I. Kohane, S.E. Plon, and R. Zimmern. 2012. Next-generation sequencing in the clinic: are we ready? *Nat. Rev. Genet.* 13:818-824.
- Martin, J.A. and Z. Wang. 2011. Next-generation transcriptome assembly. *Nat. Rev. Genet.* 12:671-682.
- Voelkerding, K.V., S.A. Dames, and J.D. Durtschi. 2009. Next-generation sequencing: from basic research to diagnostics. *Clin. Chem.* 55:641-658.
- Su, Z., B. Ning, H. Fang, H. Hong, R. Perkins, W. Tong, and L. Shi. 2011. Next-generation sequencing and its applications in molecular diagnostics. *Expert Rev. Mol. Diagn.* 11:333-343.
- Meyerson, M., S. Gabriel, and G. Getz. 2010. Advances in understanding cancer genomes through second-generation sequencing. *Nat. Rev. Genet.* 11:685-696.
- Zhang, J., R. Chiodini, A. Badr, and G. Zhang. 2011. The impact of next-generation sequencing on genomics. *J Genet Genomics.* 38:95-109.
- Liu, L., Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu, and M. Law. 2012. Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2012:251364.
- Loman, N.J., R.V. Misra, T.J. Dallman, C. Constantinidou, S.E. Gharbia, J. Wain, and M.J. Pallen. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* 30:434-439.
- Glenn, T.C. 2011. Field guide to next-generation DNA sequencers. *Molecular ecology resources* 11:759-769.
- Quail, M.A., M. Smith, P. Coupland, T.D. Otto, S.R. Harris, T.R. Connor, A. Bertoni, H.P. Swerdlow, and Y. Gu. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:341.
- Linnarsson, S. 2010. Recent advances in DNA sequencing methods - general principles of sample preparation. *Exp. Cell Res.* 316:1339-1343.
- Meyer, M., A.W. Briggs, T. Maricic, B. Hober, B. Hoffner, J. Krause, A. Wehmann, S. Paabo, and M. Hofreiter. 2008. From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Res.* 36:e5.
- Buehler, B., H.H. Hogrefe, G. Scott, H. Ravi, C. Pabon-Pena, S. O'Brien, R. Formosa, and S. Happe. 2010. Rapid quantification of DNA libraries for next-generation sequencing. *Methods* 50:S15-S18.
- White, R.A., 3rd, P.C. Blainey, H.C. Fan, and S.R. Quake. 2009. Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC Genomics* 10:116.
- Valasek, M.A. and J.J. Repa. 2005. The power of real-time PCR. *Adv. Physiol. Educ.* 29:151-159.
- Yun, J.J., L.E. Heisler, I.I. Hwang, O. Wilkins, S.K. Lau, M. Hycza, B. Jayabalasingham, J. Jin, et al. 2006. Genomic DNA functions as a universal external standard in quantitative real-time PCR. *Nucleic Acids Res.* 34:e85.
- Hindson, B.J., K.D. Ness, D.A. Masquelier, P. Belgrader, N.J. Heredia, A.J. Makarewicz, I.J. Bright, M.Y. Lucero, et al. 2011. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* 83:8604-8610.
- Hori, M., H. Fukano, and Y. Suzuki. 2007. Uniform amplification of multiple DNAs by emulsion PCR. *Biochem. Biophys. Res. Commun.* 352:323-328.
- Xiao, M., A. Phong, K.L. Lum, R.A. Greene, P.R. Buzby, and P.Y. Kwok. 2004. Role of excess inorganic pyrophosphate in primer-extension genotyping assays. *Genome Res.* 14:1749-1755.
- Zhang, Y., D. Zhang, W. Li, J. Chen, Y. Peng, and W. Cao. 2003. A novel real-time quantitative PCR method using attached universal template probe. *Nucleic Acids Res.* 31:e123.

Received 02 May 2013; accepted 01 July 2013.

Address correspondence to Jason H. Bielas, Fred Hutchinson Cancer Research Center, Seattle, WA. E-mail: jbielas@fhcr.org

To purchase reprints of this article, contact: biotechniques@fosterprinting.com

Targeted enrichment and high-resolution digital profiling of mitochondrial DNA deletions in human brain

Sean D. Taylor,¹ Nolan G. Ericson,¹ Joshua N. Burton,² Tomas A. Prolla,³ John R. Silber,⁴ Jay Shendure² and Jason H. Bielas^{1,5,6}

¹Translational Research Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave, Seattle, 98109, WA, USA

²Department of Genome Sciences, University of Washington, 3720 15th Ave NE, Seattle, 98195WA, USA

³Department of Medical Genetics, University of Wisconsin-Madison, 425-G Henry Mall, Madison, 53706 WI, USA

⁴Neurological Surgery, University of Washington Medical Center, 1959 NE Pacific St, Seattle, 98195WA, USA

⁵Human Biology Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave, Seattle, 98109 WA, USA

⁶Department of Pathology, University of Washington Medical Center, 1959 NE Pacific St, Seattle, 98195 WA, USA

Summary

Due largely to the inability to accurately quantify and characterize *de novo* deletion events, the mechanisms underpinning the pathogenic expansion of mtDNA deletions in aging and neuromuscular disorders remain poorly understood. Here, we outline and validate a new tool termed 'Digital Deletion Detection' (3D) that allows for high-resolution analysis of rare deletions occurring at frequencies as low as 1×10^{-8} . 3D is a three-step process that includes targeted enrichment for deletion-bearing molecules, single-molecule partitioning of genomes into thousands of droplets for direct quantification via droplet digital PCR, and breakpoint characterization using massively parallel sequencing. Using 3D, we interrogated over 8 billion mitochondrial genomes to analyze the age-related dynamics of mtDNA deletions in human brain tissue. We demonstrate that the total deletion load increases with age, while the total number and diversity of unique deletions remain constant. Our data provide support for the hypothesis that expansion of pre-existing mutations is the primary factor contributing to age-related accumulation of mtDNA deletions.

Key words: aging; genome instability; mitochondrial disease; mitochondrial DNA; next-generation sequencing; rare deletion detection

Introduction

The human mitochondrial genome is a small (16.5 kb) circular DNA molecule that is present in multiple copies per cell (between 1000 and 10 000 copies depending on the cell type) (Berdanier & Everts, 2001). This small genome is densely packed with 13 structural genes that

encode the major catalytic components of the core complexes involved in oxidative phosphorylation (OXPHOS), as well as 22 tRNAs and 2 rRNAs that are essential for mitochondrial protein synthesis (Scheffler, 2008). Because of the density of the gene structure, deletions in mitochondrial DNA (mtDNA) tend to affect multiple genes, including several essential tRNAs.

Accumulated mitochondrial deletions are known to cause a number of neuromuscular disorders, including Kearns–Sayre syndrome, progressive external ophthalmoplegia, and Pearson syndrome (Chinnery, 1993; Berdanier & Everts, 2001; Greaves *et al.*, 2012). These diseases are typically (but not exclusively) associated with a 4977-bp 'common' deletion between np 8482 and np 13 460. Additionally, an increasing number of associations are being discovered between mtDNA and cancer. Cancer-associated deletions tend to be smaller (< 1 kb) than those associated with neuromuscular disorders (Lee *et al.*, 2010). Whereas accumulation of large deletions leads to mitochondrial dysfunction and apoptosis, it is thought that small deletions may confer milder phenotypes that can promote tumor cell proliferation, drug resistance, and malignancy. Finally, accumulation of mtDNA deletions in postmitotic tissue (e.g., brain, heart, and skeletal muscle) is thought to be an important driving force in both physiological and accelerated aging (Cortopassi & Arnheim, 1990; Meissner *et al.*, 2008; Vermulst *et al.*, 2008b; Khrapko & Vijg, 2009).

In neuromuscular disorders, cancer and aging, the pathological mtDNA deletions appear to be somatically acquired (Meissner, 2007; Meissner *et al.*, 2008). Furthermore, individual mitochondrial mutations must expand above a threshold intracellular frequency, typically 60–90% of a cell's mtDNA, before it reaches phenotypic expression (Vermulst *et al.*, 2012). Thus, the etiology of mitochondrial deletion diseases necessarily involves two distinct processes: the somatic generation of the deletion (s) and their subsequent expansion to phenotypic levels. However, neither of these processes is well understood (Krishnan *et al.*, 2008; de Grey, 2009; Song *et al.*, 2011). One of the key difficulties is a lack of sensitive assays to detect *de novo* deletions, which in normal tissue may be lower than 1 deletion per million genomes, and track the kinetics of their initial selection. Current assays lack the sensitivity to capture these rare events without first amplifying the target sites, typically via PCR (Cortopassi & Arnheim, 1990; He *et al.*, 2002; Chabi *et al.*, 2003; Kraytsberg *et al.*, 2008). This practice is subject to introduction of numerous artifacts, is biased toward amplification of large products, and often only allows detection of a subset of deletions that have already undergone some level of expansion. The increasingly large body of work devoted to elucidating the mechanisms by which somatically acquired deletions undergo intra- and/or intercellular expansion serves to underscore the need for more sensitive tools to study this important phenomenon (Cortopassi & Arnheim, 1990; Coller *et al.*, 2001; Foury *et al.*, 2004; Durham *et al.*, 2006; Krishnan *et al.*, 2008; Fukui & Moraes, 2009; Kato *et al.*, 2011; Payne *et al.*, 2011; Song *et al.*, 2011; Freyer *et al.*, 2012; Vermulst *et al.*, 2012).

To more sensitively characterize the formation and expansion of mitochondrial deletions, we have developed a new procedure for quantitative analysis of rare deletion events. This assay, termed 'Digital Deletion Detection' (3D), allows us to directly quantify and characterize site-specific rare mitochondrial deletions that occur at frequencies as low

Correspondence

Jason H. Bielas, Translational Research Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, PO Box 19024, Mail Stop M5-A864, Seattle, WA 98109-1024, USA. Tel.: 206-667-3170; fax: 206-667-2537;

✉ e-mail: jbielas@fhcrc.org

Accepted for publication 21 July 2013

Dispatch: 12.8.13	Journal: ACEL	CE: Krithika P.
Author Received	No. of pages: 10	PI: Sumathi
B		
Manuscript No. 1 2 1 4 6		
Journal Name		
ACEL		

as 1 deletion per 100 million genomes. We demonstrate that 3D is accurate over a broad dynamic range and is capable of detecting both specific and random deletion events within a targeted region of the mitochondrial genome. We have successfully used 3D to study accumulation of clonal and random mitochondrial deletions in human brain tissue with respect to age, allowing a high-resolution analysis of deletion dynamics in aging tissue.

Results

Assay design

Digital Deletion Detection (3D) is an extremely sensitive tool for the absolute quantification and characterization of rare deletion molecules. The basic strategy behind 3D is a three-step process: enrich, amplify, and analyze. The first step, based on methods developed previously by Bielas and colleagues, enriches for deletion-bearing molecules and improves mutant specificity (Bielas & Loeb, 2005; Vermulst *et al.*, 2008a). This step consists of targeted endonucleolytic digestion of templates to selectively digest wild-type (WT) molecules, thus allowing the preferential PCR amplification of molecules bearing an appropriate deletion (Fig. 1A). After digestion, the DNA molecules are sequestered into homogenous 1 nL water-in-oil emulsion droplets and subjected to normal PCR amplification (Fig. 1B). The concentration of molecules within the droplets is adjusted such that most droplets contain no mutant genomes, while a small fraction contains only one. Thus, a single well in the

reaction actually consists of many thousand single-molecule reaction chambers. This process allows each captured deletion to be amplified without bias and without introducing many of the PCR artifacts that are common to bulk amplification reactions (i.e., template switching and preferential amplification of short templates).

Following amplification, the deletions can be analyzed via two process pathways. In the quantification pathway, high-resolution quantification of deletions is accomplished through the use of droplet digital PCR (ddPCR) (Pinheiro *et al.*, 2012). With the inclusion of TaqMan reporter chemistry, droplets bearing amplified templates are readily distinguished by their fluorescence amplitude using a cytometry system. Because the droplet volumes are highly uniform, Poisson statistics can be applied to calculate the average number of deletion-bearing molecules per droplet and the absolute concentration of mutant molecules determined with high precision and accuracy (Pinheiro *et al.*, 2012). Alternatively, in the characterization pathway, droplets are disrupted and amplicons recovered. The deletions can then be directly sequenced using high-throughput or 'next-generation' sequencing or cloned for use in Sanger sequencing or other downstream applications.

Sensitivity and recovery

Using the quantification process pathway of 3D (Fig. 1), we measured the absolute deletion frequency within a region spanning the ND1/ND2 genes in mitochondrial DNA isolated from human epithelial cells in tissue culture. We measured the deletion frequency to be 1.6 ± 0.4 deletions

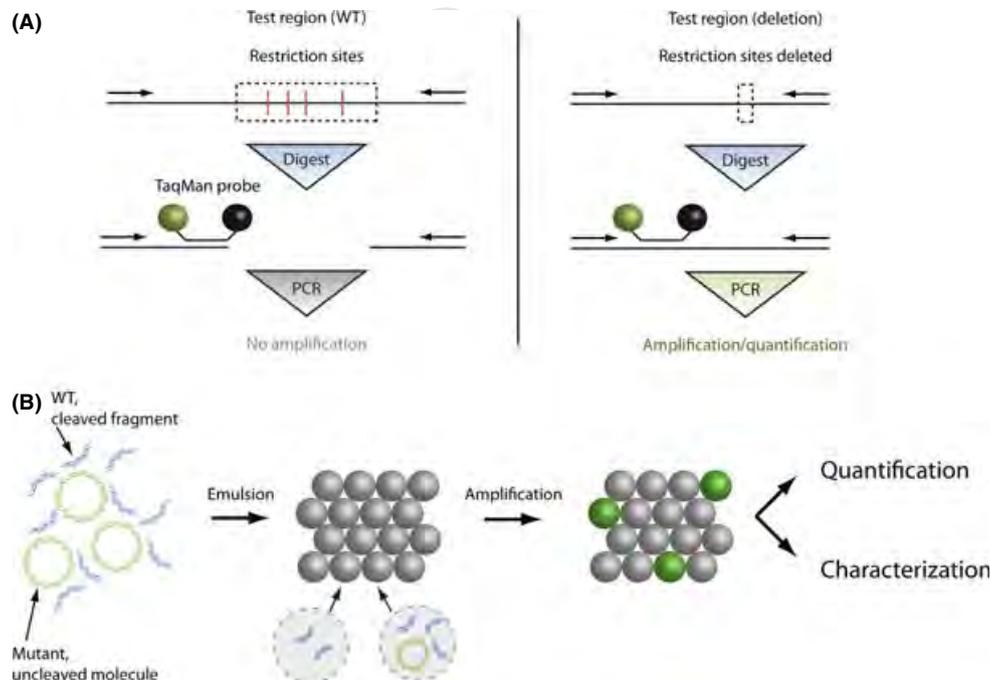


Fig. 1 Overview of Digital Deletion Detection (3D). (A) Enrichment of deletion-bearing molecules. WT molecules harbor endonuclease recognition sites within the target region. Upon digestion, the target is cleaved, making the WT molecule unsuitable as a template for PCR amplification. In contrast, mutant molecules that harbor deletions and remove the restriction recognition sites are resistant to digestion. These molecules serve as templates for PCR amplification. The presence of the TaqMan® hydrolysis probe allows for the detection and enumeration of each molecule in the sample bearing the appropriate deletion. (B) Mutant target molecules (depicted as an idealized, unbroken circular mitochondrial chromosome) are individually sequestered into 1 nL water-in-oil droplets along with TaqMan® PCR chemistry and target-specific TaqMan® probes. Droplets are thermally cycled. Because the average number of molecules per droplet is less than one, positive droplets (green droplets) represent individual reaction vessels for single-molecule quantitative PCR amplification. Droplets are individually scanned and scored as positive or negative, thus providing a digital quantification of all deletion-bearing molecules within the sample. Alternatively, droplets can be disrupted and the amplification products subjected to physical characterization, for example cloning, sequencing, or other applications.

per ten million genomes (or 1.6×10^{-7} per genome) (Fig. 2). We next asked whether 3D was able to fully recover all of the deletions within a sample over a broad range of deletion frequencies. To address this, we performed a series of reconstruction experiments. First, a plasmid harboring a fragment of mtDNA containing a known deletion in the ND1/ND2 region was mixed at a constant concentration ($3 \text{ copies } \mu\text{L}^{-1}$) against increasingly higher levels of genomic mtDNA (up to $2.5 \times 10^6 \text{ copies } \mu\text{L}^{-1}$). We then performed 3D analysis to determine whether the small concentration of the control molecules could be accurately quantified in the presence of increasing concentrations of background DNA (Fig. 2). This reconstruction demonstrated accurate quantification of target molecules across a range of frequencies spanning eight orders of magnitude, with sensitive recovery at frequencies as low as 1×10^{-7} per genome.

Because we reached the endogenous deletion frequency of the background DNA, we were unable to test lower frequencies in the reconstruction experiment. To determine whether we could detect even rarer events, we applied 3D to mtDNA isolated from muscle samples of mice, choosing a site encompassing the light chain origin of replication (Supplementary Note 4). Because deletion of this site would severely impede the ability of the genome to replicate, we expected the deletion frequency at this site to be extremely low. 3D analysis revealed a deletion frequency of $1.3 \pm 0.4 \times 10^{-8}$ per genome (Fig. S3).

Capturing and analyzing sample complexity

Next we characterized the ability of 3D to perform accurate quantification of the deletion frequency when applied to a heterogeneous population of deletions. To this end, we obtained three control plasmids, each containing an mtDNA fragment harboring a unique deletion from the minor arc of the human mitochondrial genome (3534 Δ 997,

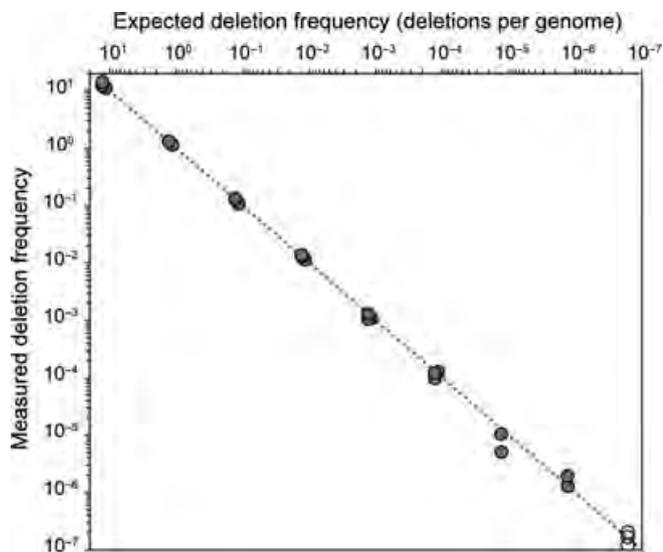


Fig. 2 Sensitivity and recovery. 3D was performed on *TaqI*-digested HCT 116 mtDNA using primers and probes for the human ND1/ND2 site to give the endogenous deletion frequency (empty circles). Reconstruction experiments were performed by spiking in $3 \text{ molecules } \mu\text{L}^{-1}$ of a control plasmid bearing a portion of the human mitochondrial genome with a known deletion (3534 Δ 997) into a serial dilution series of *TaqI*-digested HCT 116 mtDNA (filled circles). The predicted deletion frequency is plotted against the measured deletion frequency. Each data point represents an individual experiment. The reconstruction data were fit to $y = x$ (dotted line) with a correlation coefficient $R^2 = 0.9942$.

3719 Δ 809, and 3871 Δ 492). We subjected equal amounts ($300 \text{ molecules } \mu\text{L}^{-1}$) of each control plasmid to 3D analysis, either separately or combined into a single reaction, to determine whether 3D could accurately report the known concentration of a mixture of target molecules (Fig. 3A). 3D quantification of the individual plasmids yielded concentrations of 313 ± 6 , 304 ± 6 , and $322 \pm 6 \text{ molecules } \mu\text{L}^{-1}$, respectively (Fig. 3B). Quantification of the combined reaction yielded a concentration of $915 \pm 12 \text{ molecules } \mu\text{L}^{-1}$. These values match the expected concentrations within the limits of uncertainty due to the stochastic effect associated with sampling of a dilute solution (Pinheiro *et al.*, 2012).

Analysis of fluorescence amplitudes of the three control plasmids following ddPCR revealed that under the current conditions, a given template will yield an average droplet fluorescence intensity inversely proportional to the template size (Fig. 3C). When the three control templates were combined, this effect led to a striking multimodal distribution in the fluorescence amplitudes (Fig. 3A). More generally, we found that the sample heterogeneity is reflected in the distribution of fluorescence amplitudes (Fig. 3C). Thus, the average amplitude and distribution of the droplet fluorescence can be used to predict deletion sizes and complexity (e.g., presence of a single, clonal deletion vs. a heterogeneous population of multiple deletions).

Deletion dynamics in aging postmitotic tissue

While it is known that mtDNA deletions accumulate to relatively high levels in aged, postmitotic tissue in humans (Cortopassi & Arnheim, 1990), very little is known about the underlying dynamics. Specifically, as a tissue ages and accumulates deletions, it is unknown whether this increased deletion load arises through clonal expansion of an existing pool of mtDNA deletions (early acquisition), continual accumulation of new mutations (late acquisition), or an equilibrium of both processes (Khrapko, 2011). With 3D, we can now begin to directly assess these longitudinal changes. We used 3D to characterize deletions with respect to age at two regions of the mitochondrial genome from a collection of human brain tissue (Fig. 4). Using the quantification process pathway of 3D, we found that the total deletion frequency increases with age at both sites (Figs 5A and S4). The common deletion was found to gain in frequency from $1.91 \pm 0.15 \times 10^{-6}$ per genome at age 15 years to levels as high as $6.36 \pm 0.20 \times 10^{-4}$ per genome by age 80 years, an increase of over 300-fold (Table 1). These levels and accumulation rates are in agreement with previously published results (Meissner *et al.*, 2008). At the ND1/ND2 site, the absolute levels of accumulation also increased, but were generally lower than at the common deletion site. Deletion frequencies ranged from $1.9 \pm 0.5 \times 10^{-7}$ per genome to $5.25 \pm 0.22 \times 10^{-6}$ per genome, about a 25-fold increase over the same age span (Table 1). Interestingly, the increase in deletion frequency at the ND1/ND2 site showed a stronger correlation with age than the common deletion site ($R^2 = 0.812$ vs. 0.453 , respectively) (Fig. 5A).

To determine whether the increases in deletion frequency at these sites were due to expansion of existing deletions or acquisition and accumulation of new deletions, we sought to measure the ratio of unique to total deletions as a function of age. To accomplish this, emulsion droplets for a subset of patients ($n = 11$) were disrupted and the enriched mutant fragments recovered. We then performed high-throughput massively parallel sequencing analysis on each collection of amplified targets. In this way, we were able to directly profile the entire population of amplified deletion fragments at high resolution. From these data, we were able to determine the total number of unique deletion events present per sampled patient, which was then normalized

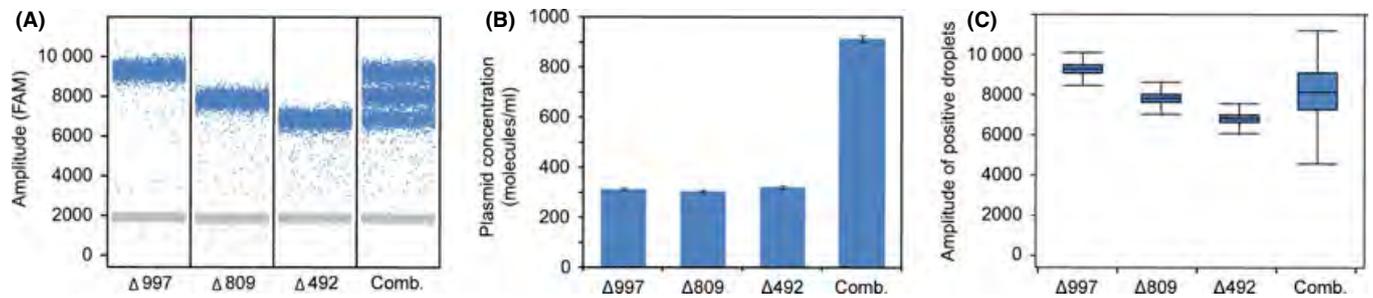


Fig. 3 Effects of sample heterogeneity on 3D analysis. (A) Three plasmid controls (3534Δ997, 3719Δ809, and 3871Δ492) were diluted to an expected concentration of 300 molecules μL^{-1} template and subjected to 3D analysis, either individually or combined. Blue dots represent droplets whose amplitudes are above the threshold ('positives'), while gray droplets are those whose amplitudes are below the threshold value ('negatives'). (B) Measured deletion concentration for individual and combined templates. Error bars indicate the Poisson 95% confidence intervals for each concentration determination. (C) Box and whisker plot showing the distribution of positive droplets for each template. When used as a template in PCR, each plasmid yields different size fragments (185 bp, 372 bp, and 686 bp, respectively). There is an inverse relationship between average fluorescence amplitude and template length, as well as a relationship between the sample complexity and the breadth of the distribution of positive droplets.

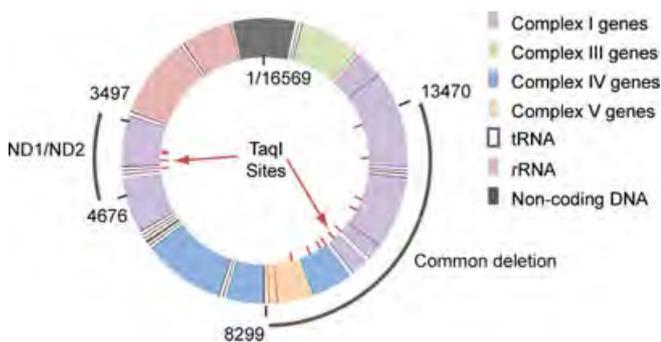


Fig. 4 Deletion sites for 3D analysis of brain mitochondrial DNA. Probe and primer sets were designed to detect deletions in two regions of the human mitochondrial genome. The first region is defined by a primer set that flanks np 8299–13470 and is designed to detect variants of the common deletion. The second primer set flanks np 3497–4676, spanning the junction between the ND1 and ND2 genes in the minor arc. The common deletion primer set flanks ten *TaqI* sites, while the ND1/ND2 primer set flanks four *TaqI* sites.

against the total number of deletions in the sample (Table 1, Fig. 5B). Linear regression analysis showed no significant correlation between the ratio of unique to total deletions and age at either site ($P = 0.120$ and $P = 0.150$ for the ND1/ND2 and common deletion sites, respectively). To ensure that our data are not influenced by sampling or processing artifacts, we analyzed a number of parameters, including the total number of genomes isolated and screened, the number of droplets used in ddPCR analysis, and site saturation effects (Data S1, Supplementary Note 5). Analysis of these parameters indicates that our data are free from any such confounding effects that might artificially skew our results (Figs S4, S5, and S6).

We next analyzed how the diversity within the pool of deletions might change with respect to age. Analysis of the amplitude distribution of positive droplets from ddPCR predicts that there is low heterogeneity at the common deletion site and high degree of heterogeneity at the ND1/ND2 site (Fig. S7). However, at both sites, the diversity does not appear to change with age. These findings were confirmed through breakpoint analysis of the sequenced deletions. Each unique deletion was individually analyzed and characterized by deletion length and relative frequency in the deletion pool (Fig. 6, Data S2). At the common deletion site, we observed a single dominant deletion in every case,

which contributed to over 90% of the deletion load (Figs 6 and S8). Although several minor variants are present in each patient, most generally contributed $< 0.5\%$ of the total deletion burden. At the ND1/ND2 site, however, there is a broad but fairly uniform distribution of deletion sizes within the ND1/ND2 deletion space across individuals of all ages (Fig. 6, bottom panel). The bulk of the deletion load was typically comprised of deletions, which individually contributed between 1 and 10% to the total deletion burden (Figs 6 and S8). The data indicate no major shift in the size distribution of deletions as well as the relative pools of high- and low-frequency deletions with age (Figs 6 and S7).

Finally, we examined the average frequency of individual deletions with respect to age. This was found by taking the ratio of the deletion frequency and the total number of unique deletions, a value we term the expansion index, which is then normalized against the youngest time point for clarity. A decrease in the normalized expansion index with respect to time denotes that deletions are being selected against, while an increase suggests positive selective pressure. We found that at both sites, the expansion ratio increases significantly with age (Fig. 5C). Concomitant with a static spectrum of deletion diversity with age, we conclude that expansion of a pre-existing set of deletions may be one of the primary drivers of age-related increases in deletion frequency.

Discussion

To adequately detect *de novo* mtDNA deletions and trace the frequency dynamics, an assay is needed that can enrich for and directly quantify extremely rare deletion events. Current approaches to analyzing mtDNA deletions include Southern blotting (DiMauro & Hirano, 1993), direct sequencing (Spellbrink *et al.*, 2000; Ameur *et al.*, 2011; Kato *et al.*, 2011; Sequeira *et al.*, 2012), and PCR amplification (Kraytsberg *et al.*, 2008). Sequencing of deletions via cloning is laborious, time-consuming, prone to cloning artifacts and allows only the most abundant deletion types to be analyzed (Supplementary Notes 3 and 4). Massively parallel or 'next-generation' sequencing is rapidly becoming a preferred means for high-throughput screening of individual DNA molecules. As an example, Illumina, Inc. (San Diego, CA, USA) offers systems that generate from 17 million (MiSeq[®]) up to 3 billion simultaneous sequencing reads per run (HiSeq[®]) (Liu *et al.*, 2012). However, given a relatively short read length of < 150 bp and the fact that the majority of the reads will be off-target, this remains insufficient to adequately

Table 1 Frequencies of mitochondrial deletion events in human brain. The error of duplicate measurements is indicated as the standard error of the mean (SEM)

ID	Age	Measured deletion frequency ($\times 10^{-7}$)				Unique deletions (per 1000 total)	
		ND1/ND2	SEM	Common	SEM	ND1/ND2	Common
P01†	28	11.3	1.1	141.5	4.2	65.2	5.0
P02†	28	5.2	0.1	38.7	4.2	224.0	9.8
P03†	43	20.0	1.7	6355.2	61.0	126.1	0.5
P04†	30	3.3	2.9	335.8	314.6	227.0	0.8
P05†	38	11.7	1.2	1883.5	556.8	115.3	1.3
P06†	38	31.1	0.1	3687.8	28.6	69.8	0.9
P07†	39	13.0	3.7	1276.0	19.7	139.7	0.7
P08†	43	23.9	0.1	1550.5	5.1	77.5	1.0
P09†	46	20.3	1.2	2174.5	16.2	157.3	1.9
P10†	54	35.3	0.4	1996.7	115.5	84.3	2.7
P11†	64	46.0	1.5	4856.7	162.3	67.6	0.8
P12	43	5.0	0.1	96.0	3.1		
P13	37	13.1	0.2	1389.2	14.8		
P14	19	7.3	1.1	285.8	13.3		
P15	32	19.1	1.7	470.0	5.7		
P16	15	1.9	0.5	19.1	1.5		
P17	45	16.0	0.2	3739.3	154.1		
P18	26	8.4	2.0	961.2	42.3		
P19	80	52.5	2.2	3084.8	20.0		
P20	78	48.0	6.3	3596.7	5.6		
P21	71	36.5	4.1	5069.5	130.5		

†Used in NGS analysis.

resolve mtDNA deletions that occur at frequencies of less than one in a million genomes. Even assuming no off-target reads, the MiSeq[®] instrument would still only yield about one deletion in ten runs. It is therefore critical that a selection step be performed to limit the number of off-target reads and to enrich for deletion-bearing molecules.

PCR-based methods, including long-distance PCR and real-time quantitative PCR, are among the most frequently employed methods for both selection and amplification of deletions (Cortopassi & Arnheim, 1990; He *et al.*, 2002; Chabi *et al.*, 2003; Kraysberg *et al.*, 2008). Generally speaking, these assays distinguish wild-type from deleted genomes through exploiting differences in amplicon fragment lengths and amplification efficiencies. Given that they do not select for deleted molecules prior to amplification, one of the main drawbacks is high background signal from contaminating wild-type molecules, thus limiting the effective sensitivity. Furthermore, these bulk PCR assays tend to introduce a number of additional artifacts arising from preferential amplification of small templates (allelic preference), introduction of false deletions through template jumping, and other PCR errors (Kraysberg & Khrapko, 2005). Real-time quantitative PCR (qPCR) can be quite sensitive, but its reliance on relative differences in crossing thresholds rather than direct quantification makes it more suitable for measuring fold changes rather than absolute deletion frequencies (He *et al.*, 2002; Chabi *et al.*, 2003). Digital PCR methods, including long single-molecule PCR (long smPCR) (Kraysberg & Khrapko, 2005; Guo *et al.*, 2010) and the random mutation capture assay developed for mtDNA deletions (deletion RMC) (Vermulst *et al.*, 2008a,b) achieve direct quantification through the use of single-molecule partitioning in 96-well plates. Partitioning additionally serves to minimize artifacts of template jumping and allelic preference that are common in bulk PCRs (Kraysberg & Khrapko, 2005). Despite these advantages, this approach becomes laborious and costly when using the wells of a multiwell plate

as the partition and yields only a handful of the most common deletions within a sample.

The Digital Deletion Detection (3D) assay shows a marked improvement in specificity, sensitivity, and accuracy over other available methods. This is achieved via a three-step process of selection, amplification, and characterization (i.e., quantification or sequencing). As with deletion RMC, high specificity for deletion-bearing molecules is achieved through the destruction of WT template molecules by restriction endonuclease, thereby selecting for and enriching mutant molecules prior to amplification. Following enrichment, partitioning for digital PCR amplification is performed through the generation of up to 20 000 droplet partitions, the equivalent of over 200 96-well plates, within a single reaction well. Quantification is greatly facilitated through the use of TaqMan reporter probes and cytometry, which allows for rapid enumeration of all partitions that contain an amplifiable template and direct quantification of all deletions within a sample.

One of the unique advancements of the 3D assay is the wealth of single-molecule information that is obtained from cytometric analysis of the droplet partitions. In other mtDNA deletion detection assays, hundreds of wells must be screened to yield a handful of successful amplifications. The corresponding template molecules can only be characterized through the additional steps of gel electrophoresis or sequencing. This process will tend to oversample large clonal deletions and thus may not yield a true representation of the biological diversity of deletions present (see Supplemental Note 3). In contrast, 3D provides an opportunity to robustly screen tens of thousands of droplet partitions, yielding hundreds of positive reactions and allowing analysis of a more complete set of deletions in the sample. Moreover, the demonstrated inverse relationship between template size and the endpoint fluorescent intensity of the droplet partitions (Fig. 3C) can be exploited to reveal information regarding the size and homogeneity of the templates in the

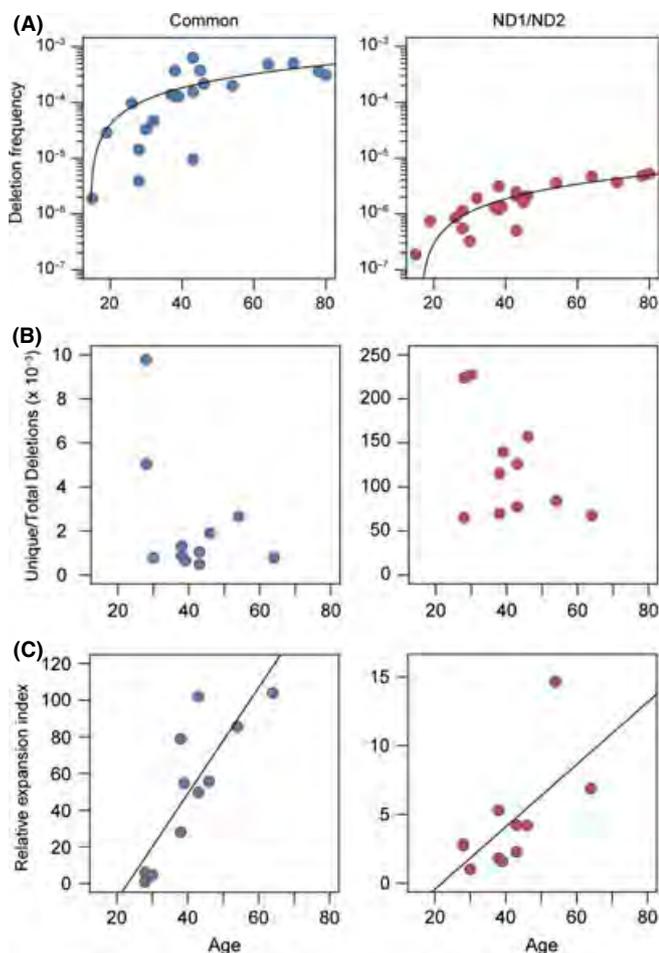


Fig. 5 3D analysis of deletion frequency in aged human brain tissue. (A) Total deletion frequency at each site is plotted against age. Deletion frequency shows a positive correlation with age at both the common deletion site ($R^2 = 0.453$, $P = 0.0008$) and the ND1/ND2 site ($R^2 = 0.812$, $P = 3 \times 10^{-8}$). The linear regression (transformed) is shown on a log-scale plot. (B) Deletion profiling was performed on a subset of patients to examine the diversity of deletions present in the deletion pool. The number of unique deletions is shown normalized against the total number of deletions for each patient in the subset, respectively. At the common deletion site, patients showed a range of 0.5–10 unique per 1000 deletions. At the ND1/ND2 site, the deletion diversity was much higher, ranging from 65 to 227 per thousand. Linear regression analysis showed no significant correlation between the unique-to-total deletion ratio and age at either site ($P = 0.15$ and $P = 0.12$ for the common and ND1/ND2 deletion sites, respectively). (C) The relative expansion index for each patient in the subset was found by taking the ratio of total deletion frequency over the number of unique deletions normalized against the youngest time point. This value gives an estimate of the average frequency of individual deletions for each patient relative to the youngest time point (i.e., the average individual deletion frequency). Linear regression showed a positive correlation for both the common deletion site ($R^2 = 0.697$, $P = 0.001$) and the ND1/ND2 site ($R^2 = 0.421$, $P = 0.03$).

sample. By analyzing the amplitude distribution of positive droplets, we were able to accurately predict whether the deletion population consisted of a few clonal expansions or a large collection of random deletions (Figs 3 and S7). In this way, cytometric analysis of the partitions could be used to gather information about the size spectrum of deletion templates in the sample without being subject to the biases inherent in individual cloning or the costs of deep sequencing. We believe that with further development, this relationship could potentially be exploited to

open new possibilities for 'next-generation' PCR technology that can dynamically sort and collect specific amplification products, similar to fluorescence-activated cell sorting with flow cytometry.

Another advantage of the 3D assay is its ability to adjust the search parameters to measure many different target deletion sets. This is achieved by defining the target deletion space through careful choice of primer locations and the restriction enzyme. This is an important advantage over many existing methods in that random deletions within a target region can be analyzed without knowing the precise breakpoints of the target deletion set *a priori*. It is noteworthy that we were able to measure the deletion loads at both sites simultaneously, given that the minor arc deletion frequency was up to 100-fold less than the major arc. In many other assays, this information would be lost to the dominant signal of the clonal expansions. Importantly, the assay is also neutral with regard to random (i.e., steady-state temporal deletions that occur at low frequency) vs. clonal events (i.e., deletions that have expanded out of the steady-state pool and that occur at relatively high frequency): the assay will detect all deletions that fall within the defined deletion space. Thus, our assay is able to account for gain or loss of steady-state temporal deletions as well as clonal expansions.

Finally, by coupling NGS with the other steps in 3D, we are able to perform high-resolution characterization of millions of breakpoints within a single sequencing run. To demonstrate the utility and sensitivity of this assay, we analyzed deletion loads within the mitochondrial genome of human brain samples. For example, at the ND1/ND2 site, we interrogated over 8 billion mitochondrial genomes and identified over 100 000 genomes with a deletion within our target region. At that site, we were able to characterize 430 individual unique deletions with an average sequencing coverage of 78-fold. Furthermore, based on the specific sequencing coverage, we were able to distinguish between clonally expanded and random, 'steady-state temporal' deletions. To our knowledge, no other assay has demonstrated the capability of identifying and analyzing such a large deletion set with comparable resolution.

Digital Deletion analysis allows for unbiased, high-resolution analysis of the full spectrum of deletions within the target site. With this tool, we can better analyze the mechanics and kinetics of deletion acquisition and expansion in aging tissue. Accumulation of mtDNA deletions, particularly in postmitotic tissue, is an important cause of human pathology and aging (Cortopassi & Arnheim, 1990; Meissner *et al.*, 2008; Vermulst *et al.*, 2008b; Khrapko & Vijg, 2009). While it is known that deletions can accumulate through a process of clonal expansion of a pre-existing pool of deletions, it is unclear whether this or an accelerated rate of *de novo* deletions is the primary driving force behind age-related deletion accumulation (Khrapko, 2011). Previous studies using mathematical simulations of cell division or analysis of the distribution of deletions in tissues conclude that many mtDNA mutations may have an early origin and have been subsequently expanded (Brierley *et al.*, 1998; Elson *et al.*, 2001; Khrapko *et al.*, 2003, 2004; Payne *et al.*, 2011). However, work from some of the same groups has also leads to the opposite conclusion that mtDNA deletions may be of late origin (Nicholas *et al.*, 2009). To address this issue, we used 3D to characterize the absolute deletion frequency and deletion spectrum of aging brain tissue at two regions of the mitochondrial genome. We found that the total deletion load increases, but that the ratio of unique to total deletions did not change from younger to older tissue. Furthermore, we observed little change in the size distribution of deletions as well as the relative pools of high- and low-frequency deletions indicating a fairly static spectrum of diversity. An important caveat is that in the present work, we are not actually tracing the dynamics of specific deletions with time, but are rather harvesting snapshots of the deletion burden across several individuals.

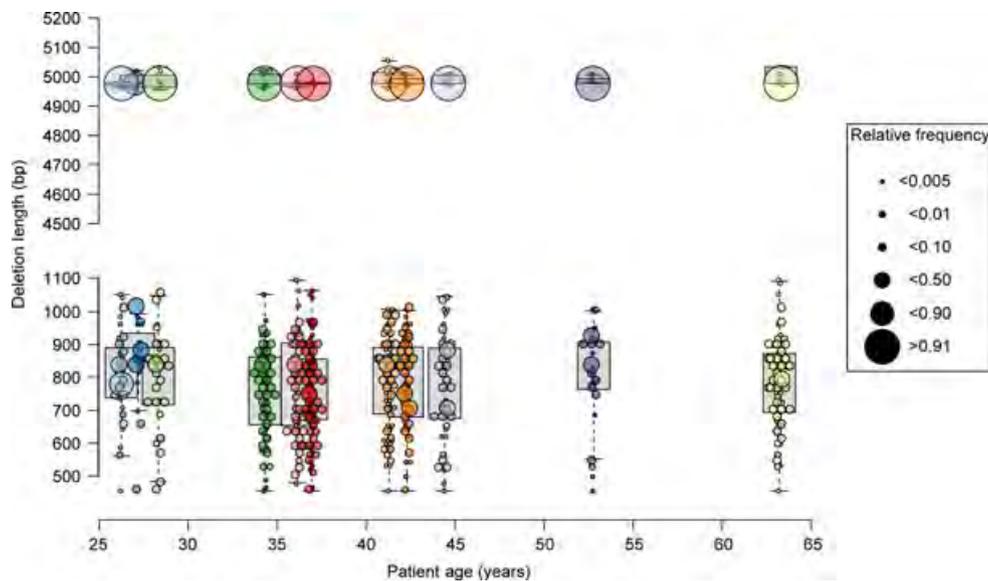


Fig. 6 High-resolution analysis of deletion dynamics. Density dot plots showing the length distribution (y-axis) and relative frequency (point size) of all unique deletions per patient in the sequenced subset (x-axis, plotted as age). Box plots and whiskers plots (gray) in the background show the 95% confidence interval of the unweighted length distribution. The common deletion site is shown on the top plot, and the ND1/ND2 deletion site is shown on the bottom plot.

Thus, we cannot rule out the contribution of newly acquired deletions to later time points. This is particularly true in the case of the common deletion where the dominance of a single-deletion species at this site makes it impossible to determine whether we are observing clonal expansion or rapid re-accumulation of the same deletion. However, at the ND1/ND2 locus, we were able to recover a large diversity of deletions without such site saturation (Figs 6 and S6). Thus, within the time frame analyzed (aged 28–80 years), our data support the hypothesis that expansion rather than generation of new deletions dominates the age-related increase in deletion load.

The fact that early mutations are allowed to accumulate to significant levels may be interpreted as evidence for some sort of selective pressure. Precisely what that pressure is, however, remains unclear. Our data show uniform random distribution of deletion lengths at the ND1/ND2 site across all ages. The absence of a shift in the diversity toward accumulation of larger deletions argues against the hypothesis that smaller mtDNA molecules possess a replicative advantage in postmitotic cells (Wallace, 1989; Fukui & Moraes, 2009). Our data are not inconsistent with *in silico* experiments that predict that clonal expansion can result from random genetic drift without the aid of selection (Coller *et al.*, 2001; Elson *et al.*, 2001). While this model has been somewhat validated for point mutations (Durham *et al.*, 2006), other selective mechanisms for deletions cannot be ruled out (de Grey, 2009). 3D will allow us to perform longitudinal studies that can trace the kinetics of clonal expansion of real deletions that will allow us to better test the *in silico* models with data from living cells.

The 3D/NGS data demonstrate that we now have the technology to perform high-resolution analysis and detailed characterization of extremely rare deletion events. Importantly, it also provides the means to begin to use mtDNA deletions as biomarkers for disease. Although mtDNA deletions accumulate readily in skeletal muscle and brain tissue, they exist at extremely low levels in blood and other rapidly proliferating tissue (DiMauro & Hirano, 1993). This has been a great hindrance to the development of blood-based biomarker assays that could be used for noninvasive screening and early detection of mitochondrial deletion

diseases. Digital Deletion Detection provides an important new tool that will allow researchers to better study the mechanisms of deletion formation, their mechanisms of expansion, and their role in the etiology of aging and disease.

Experimental procedures

Human brain tissue

Human histologically normal brain obtained from informed patients was obtained from the tissue depository of the Department of Neurological Surgery at the University of Washington. Tissue and demographic information was obtained in accord with an IRB-approved protocol (Table 1).

DNA isolation

To obtain whole DNA from human brain tissue, tissue samples (50–250 mg) were immersed in 5 mL homogenization medium (0.32 M sucrose, 1 mM EDTA, 10 mM Tris-HCl, pH 7.8) and disrupted with a glass Dounce-type homogenizer. The homogenate was transferred to a 15-mL tube and centrifuged at 4000 g. The pellet was resuspended in 3 mL lysis buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 20 mM EDTA, 1% SDS, and 0.2 mg mL⁻¹ proteinase K) and incubated at 55 °C for 3 h. DNA was isolated by phenol-chloroform extraction followed by isopropanol precipitation.

Endonucleolytic enrichment of mtDNA deletions

Rare deletion-bearing molecules were selectively enriched through endonucleolytic destruction of wild-type target sites. First, a 400 μ L digestion reaction mixture was prepared containing 10 μ g of genomic DNA, 8 μ L (800 U) of *TaqI* (New England Biolabs), and *TaqI* reaction buffer (Fermentas

). The reaction mixture was divided into $4 \times 100 \mu\text{L}$ reactions and incubated at 65 °C for 4–6 h. An additional 200 U of *TaqI*

was added to each reaction every hour. After each *TaqI* addition, samples were thoroughly mixed and briefly centrifuged to ensure efficient digestion. Following the digestion procedure, the reactions were recombined, extracted once with phenol/chloroform/isoamyl alcohol (25:24:1, v/v), precipitated by ethanol, and resuspended in 1 mM Tris, pH 8.

TaqMan probe and primer design

The following primer/probe sets were used with human total DNA for mtDNA deletion detection. Control site: 5'-CTA AAA ATA TTA AAC ACA AAC TAC CAC CTA CCTC-3' (forward primer), 5'-GTT CAT TTT GGT TCT CAG GGT TTG TTA TAA-3' (reverse primer), and 5'-6FAM- CCT CAC CAA AGC CCA TA-MGB-3' (probe). ND1/ND2 site: 5'-CGC CAC ATC TAC CAT CACC-3' (forward primer), 5'-GAT TAT GGA TGC GGT TGC TT-3' (reverse primer), and 5'-6FAM-TTG ATG GCA GCT TCT GT-MGB-3' (probe). Common deletion site: 5'-TAC CCC CTC TAG AGC CCA CT-3' (forward primer), 5'-GAG GAA AGG TAT TCC TGC TAA TGCT-3' (reverse primer), and 5'-6FAM-TGG CCC ACC ATA ATT-MGB-3' (probe).

Droplet digital PCR

The final concentration of digested DNA was adjusted to yield less than ~3500 positive molecules per μL , which is within the range of linearity for the Poisson calculation (Pinheiro *et al.*, 2012). Reaction mixtures (25 μL) contained ddPCR Master Mix (Bio-Rad), 250 nM TaqMan probe, and 1–2 μL of digested DNA (0–2 μg total). Appropriate flanking primers were added at either 900 nM or 45 nM for the quantification and sequencing process pathways, respectively (see Supplementary Notes 1 and 2). Reaction droplets were made by applying 20 μL of each reaction mixture to a droplet generator DG8 cartridge (Bio-Rad) for use in the QX100 Droplet Generator (Bio-Rad). Following droplet generation, 38 μL of the droplet emulsion was carefully transferred to a Twin.tec semi-skirted 96-well PCR plate (Eppendorf), which was then heat-sealed with a pierceable foil sheet. To amplify the fragments, thermal cycling was carried out using the following protocol: initial denaturation step at 95 °C for 10 min, followed by 40 cycles of 94 °C for 30 s, and 63.5 °C for 4 min. The thermally cycled droplets were either (i) analyzed by flow cytometry for fluorescence analysis and quantification of deletion frequencies (see Methods S1) or (ii) disrupted and the PCR products recovered and sequenced in order to verify deletions and characterize the deletion sites (see Methods S1). All experiments were performed in triplicate.

Analysis of fluorescence amplitude and quantification of deletions

Following normal thermal cycling, droplets were individually scanned using the QX100™ Droplet Digital™ PCR system (Bio-Rad). Positive (deletion-bearing) and negative droplets were distinguished on the basis of fluorescence amplitude using a global threshold. The number of mutant genomes per droplet was calculated automatically by the accompanying software (QuantaSoft, Bio-Rad) using Poisson statistics as described elsewhere (Hindson *et al.*, 2011). Quantification of deletion frequency requires ddPCR amplification using two primer sets. The first primer set flanks the test region and measures the concentration of deletion-bearing molecules. The second primer set flanks a distant region in the genome that bears no restriction recognition sites. This second or control set measures the concentration of all mtDNA genomes. Because *de novo* deletions are so rare, reactions using the different primer sets

must be run using different dilutions of the digested DNA, and the results normalized against the mass of total DNA in the reaction. Deletion frequency is calculated by taking the ratio of the normalized concentrations of deletion-bearing mtDNA molecules to the total mtDNA molecules screened. Reactions that yielded < 10 positive droplets per well were scored conservatively as having no positives above background (Pinheiro *et al.*, 2012).

Library preparation and Illumina sequencing

Human ND1/ND2 ddPCR amplification products were subjected to template conversion as described in Methods S1. Reactions were cleaned using the ZR-96 Clean and concentrator-5 kit (Zymo Research). Template concentrations were calculated using the Quant-iT™ PicoGreen dsDNA Assay Kit (Invitrogen) following manufacturer's recommended protocol. Samples were then diluted to 0.2 ng μL^{-1} in 10 mM Tris, pH 8.0, 1 mM EDTA (TE). Fragmentation, adaptor ligation, and index ligation were accomplished using the Nextera XT DNA Sample Preparation Kit (Illumina) following the recommended protocol.

Because the common deletion breakpoint is within 100 bp of the 3' end of the amplicon, the normal tagmentation protocol could not be followed. Instead, adaptors were added directly via PCR using the following primers: 5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNN NCG TAT GGC CCA CCA TAA TTA CC (forward) and 5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GNN NNG AGG AAA GGT ATT CCT GCT AAT GCT-3' (reverse). Thermal cycling consisted of an initial denaturation at 95 °C for 10 min, followed by 8 cycles of 94 °C for 30 s, 58 °C for 30 s, and 63.5 °C for 4 min. Reactions were cleaned using the ZR-96 Clean and concentrator-5 kit (Zymo Research) at concentrations and dilutions performed above. 5 μL of 0.2 ng μL^{-1} DNA was mixed with 20 μL TD buffer prior to PCR amplification in the Nextera XT DNE Sample Prep workflow. The rest of the Nextera XT protocol was performed according to recommended procedures. Indexed ND1/ND2 and common deletion fragments were pooled for all patients and sequenced using the MiSeq Personal Sequencing System (Illumina) (see Methods S1). FASTQ files for each patient were deposited in the NCBI Sequence Read Archive (SRA) under project accession number SRP027401.

Reconstruction experiments

Genomic DNA was isolated from HCT 116 cells, chosen for its relatively low endogenous deletion frequency of 1.8×10^{-7} . Following *TaqI* digestion, a series of 10-fold serial dilutions of the genomic DNA were prepared, ranging over eight orders of magnitude. A 997-bp deletion was isolated, amplified, and cloned into a vector for use as a control molecule (Fig. 5B). Approximately 600 ng of the 3534 Δ 997 control plasmid was serially diluted 100 million fold and subjected to a preliminary 3D analysis to calculate the absolute concentration of molecules within the dilution. To each of the genomic dilutions, three copies of the 3534 Δ 997 control plasmid were added per microliter of reaction. The reaction mixtures were then partitioned, cycled, and the droplets analyzed to determine whether the small concentration of the control molecules could be accurately assessed even in the presence of high concentrations of background, HCT 116 DNA.

Heterogeneous population reconstruction experiments

Three control plasmids (3534 Δ 997, 3719 Δ 809, and 3871 Δ 492) were isolated from *POLG*^{D274A} HeLa cells as described above (see also Fig. 3). Each plasmid was serially diluted and subjected to preliminary 3D

analysis in order to calculate the concentration of molecules within each dilution. Based on these quantifications, 300 molecules μL^{-1} per template were subjected to another round of 3D analysis, either separately or combined into a single reaction.

Regression analysis

Linear regression analyses were performed in R using the built-in Stats package (R Core Team, 2013). Significance of linear models was calculated using the *F*-test against the null hypothesis of no correlation between the variables tested.

Acknowledgments

The authors thank J. Wanagat and K. Khrapko for insightful comments and critical reading of the manuscript, J. Bertout for advice and valuable discussion about development of the project, and M. Kulawiec and B. Hindson for excellent technical assistance. This work is funded by grants W81XWH-10-1-0563 from the CDMRP/U.S. Department of Defense, AG-NS-0577-09 from the Ellison Medical Foundation, and RO1 ES019319 from NIEHS. SDT is supported by the National Institute of Environmental Health Sciences of the National Institutes of Health under award number F32ES021703. JNB is supported by an NIH/NHGRI Genome Training Grant through UW Genome Sciences.

Conflict of Interest

The authors declare no conflict of interest.

8 Author contributions

JHB, NGE, and SDT involved in the conceptual and experimental design. SDT, NGE, and JNB performed data acquisition and data analysis. TAP and JRS provided the biological materials. SDT wrote the manuscript, with critical reading and intellectual contributions by NGE, JNB, JRS, JS, and JHB.

References

- Ameur A, Stewart JB, Freyer C, Hagström E, Ingman M, Larsson N-G, Gyllensten U (2011) Ultra-deep sequencing of mouse mitochondrial DNA: mutational patterns and their origins. *PLoS Genet.* **7**, ???-???.
- Berdanier CD, Everts HB (2001) Mitochondrial DNA in aging and degenerative disease. *Mutat. Res.* **475**, 169–183.
- Bielas JH, Loeb LA (2005) Quantification of random genomic mutations. *Nat. Methods* **2**, 285–290.
- Brierley EJ, Johnson MA, Lightowlers RN, James OF, Turnbull DM (1998) Role of mitochondrial DNA mutations in human aging: implications for the central nervous system and muscle. *Ann Neurol.* **43**, 217–223.
- Chabi B, Mousson de Camaret B, Duborjal H, Issartel J-P, Stepien G (2003) Quantification of mitochondrial DNA deletion, depletion, and over-replication: application to diagnosis. *Clin. Chem.* **49**, 1309–1317.
- Chinnery PF (1993) Mitochondrial disorders overview. In *GeneReviews*. (Pagon RA, Bird TD, Dolan CR, Stephens K, eds). Seattle, WA: ????, pp. ???-???.
- Coller HA, Khrapko K, Bodyak ND, Nekhaeva E, Herrero-Jimenez P, Thilly WG (2001) High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection. *Nat. Genet.* **28**, 147–150.
- Cortopassi GA, Arnheim N (1990) Detection of a specific mitochondrial DNA deletion in tissues of older humans. *Nucleic Acids Res.* **18**, 6927–6933.
- DiMauro S, Hirano M (1993). Mitochondrial DNA deletion syndromes. In *GeneReviews*. (RA Pagon, TD Bird, CR Dolan, K Stephens, eds). Seattle, WA: ????, pp. ???-???.
- Durham SE, Samuels DC, Chinnery PF (2006) Is selection required for the accumulation of somatic mitochondrial DNA mutations in post-mitotic cells? *Neuromuscul. Disord.* **16**, 381–386.
- Elson JL, Samuels DC, Turnbull DM, Chinnery PF (2001) Random intracellular drift explains the clonal expansion of mitochondrial DNA mutations with age. *Am. J. Hum. Genet.* **68**, 802–806.
- Foury F, Hu J, Vanderstraeten S (2004) Mitochondrial DNA mutators. *Cell. Mol. Life Sci.* **61**, 2799–2811.
- Freyer C, Cree LM, Mourier A, Stewart JB, Koolmeister C, Milenkovic D, Wai T, Floros VI, Hagstrom E, Chatzidaki EE, Wiesner RJ, Samuels DC, Larsson NG, Chinnery PF (2012) Variation in germline mtDNA heteroplasmy is determined prenatally but modified during subsequent transmission. *Nat. Genet.* **44**, 1282–1285.
- Fukui H, Moraes CT (2009) Mechanisms of formation and accumulation of mitochondrial DNA deletions in aging neurons. *Hum. Mol. Genet.* **18**, 1028–1036.
- Greaves LC, Reeve AK, Taylor RW, Turnbull DM (2012) Mitochondrial DNA and disease. *J. Pathol.* **226**, 274–286.
- de Grey AD (2009) How is mutant mitochondrial DNA clonally amplified? Much new evidence, still no answers. *Rejuvenation Res.* **12**, 217–219.
- Guo X, Kudryavtseva E, Bodyak N, Nicholas A, Dombrovsky I, Yang D, Kravtsov Y, Simon DK, Khrapko K (2010) Mitochondrial DNA deletions in mice in men: substantia nigra is much less affected in the mouse. *Biochim. Biophys. Acta* **1797**, 1159–1162.
- He L, Chinnery PF, Durham SE, Blakely EL, Wardell TM, Borthwick GM, Taylor RW, Turnbull DM (2002) Detection and quantification of mitochondrial DNA deletions in individual cells by real-time PCR. *Nucleic Acids Res.* **30**, e68.
- Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC, Kitano TK, Hodel MR, Petersen JF, Wyatt PW, Steenblock ER, Shah PH, Bousse LJ, Troup CB, Mellen JC, Wittmann DK, Erndt NG, Cauley TH, Koehler RT, So AP, Dube S, Rose KA, Montesclaros L, Wang S, Stumbo DP, Hodges SP, Romine S, Milanovich FP, White HE, Regan JF, Karlin-Neumann GA, Hindson CM, Saxonov S, Colston BW (2011) High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **83**, 8604–8610.
- Kato M, Nakamura M, Ichiba M, Tomiyasu A, Shimo H, Higuchi I, S-i U, Sano A (2011) Mitochondrial DNA deletion mutations in patients with neuropsychiatric symptoms. *Neurosci. Res.* **69**, 331–336.
- Khrapko K (2011) The timing of mitochondrial DNA mutations in aging. *Nat. Genet.* **43**, 726–727.
- Khrapko K, Vijg J (2009) Mitochondrial DNA mutations and aging: devils in the details? *Trends Genet.* **25**, 91–98.
- Khrapko K, Nekhaeva E, Kravtsov Y, Kunz W (2003) Clonal expansions of mitochondrial genomes: implications for *in vivo* mutational spectra. *Mutat. Res.* **522**, 13–19.
- Khrapko K, Ebralidze K, Kravtsov Y (2004) Where and when do somatic mtDNA mutations occur? *Ann. N. Y. Acad. Sci.* **1019**, 240–244.
- Kravtsov Y, Khrapko K (2005) Single-molecule PCR: an artifact-free PCR approach for the analysis of somatic mutations. *Expert. Rev. Mol. Diagn.* **5**, 809–815.
- Kravtsov Y, Nicholas A, Caro P, Khrapko K (2008) Single molecule PCR in mtDNA mutational analysis: genuine mutations vs. damage bypass-derived artifacts. *Methods* **46**, 269–273.
- Krishnan KJ, Reeve AK, Samuels DC, Chinnery PF, Blackwood JK, Taylor RW, Wanrooij S, Spelbrink JN, Lightowlers RN, Turnbull DM (2008) What causes mitochondrial DNA deletions in human cells? *Nat. Genet.* **40**, 275–279.
- Lee HC, Chang CM, Chi CW (2010) Somatic mutations of mitochondrial DNA in aging and cancer progression. *Ageing Res Rev.* **9**(Suppl 1), S47–S58.
- Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M (2012) Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, 11.
- Meissner C (2007) Mutations of mitochondrial DNA - cause or consequence of the ageing process? *Z. Gerontol. Geriatr.* **40**, 325–333.
- Meissner C, Bruse P, Mohamed SA, Schulz A, Warnk H, Storm T, Oehmichen M (2008) The 4977 bp deletion of mitochondrial DNA in human skeletal muscle, heart and different areas of the brain: a useful biomarker or more? *Exp. Gerontol.* **43**, 645–652.
- Nicholas A, Kravtsov Y, Guo X, Khrapko K (2009) On the timing and the extent of clonal expansion of mtDNA deletions: evidence from single-molecule PCR. *Exp. Neurol.* **218**, 316–319.
- Payne BA, Wilson IJ, Hateley CA, Horvath R, Santibanez-Koref M, Samuels DC, Price DA, Chinnery PF (2011) Mitochondrial aging is accelerated by anti-retroviral therapy through the clonal expansion of mtDNA mutations. *Nat. Genet.* **43**, 806–810.
- Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, Bhat S, Emslie KR (2012) Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal. Chem.* **84**, 1003–1011.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R

Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0,
<http://www.R-project.org/> [accessed on XX Xxxx XXXX].

- 12** Scheffler IE (2008) *Mitochondria*. Hoboken, NJ: Wiley-Liss.
- Sequeira A, Martin MV, Rollins B, Moon EA, Bunney WE, Macciardi F, Lupoli S, Smith EN, Kelsoe J, Magnan CN, van Oven M, Baldi P, Wallace DC, Vawter MP (2012) Mitochondrial mutations and polymorphisms in psychiatric disorders. *Front. genet.* **3**, 103.
- Song Z, Cao Y, Samuels DC (2011) Replication pauses of the wild-type and mutant mitochondrial DNA polymerase gamma: a simulation study. *PLoS Comput. Biol.* **7**, e1002287.
- Spelbrink JN, Toivonen JM, Hakkaart GA, Kurkela JM, Cooper HM, Lehtinen SK, Lecrenier N, Back JW, Speijer D, Foury F, Jacobs HT (2000) *In vivo* functional analysis of the human mitochondrial DNA polymerase POLG expressed in cultured human cells. *J. Biol. Chem.* **275**, 24818–24828.
- Vermulst M, Bielas JH, Loeb LA (2008a) Quantification of random mutations in the mitochondrial genome. *Methods* **46**, 263–268.
- Vermulst M, Wanagat J, Kujoth GC, Bielas JH, Rabinovitch PS, Prolla TA, Loeb LA (2008b) DNA deletions and clonal mutations drive premature aging in mitochondrial mutator mice. *Nat. Genet.* **40**, 392–394.
- Vermulst M, Khrapko K, Wanagat J (2012). Mitochondrial mutagenesis in aging and disease. *????, ????, ???–????*.
- 13** Wallace DC (1989) Mitochondrial DNA mutations and neuromuscular disease. *Trends Genet.* **5**, 9–13.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Fig. S1. Effect of primer concentration on amplified copy number.

Fig. S2. Relative recovery proportions at various stages of 3D analysis.

Fig. S3. 3D analysis of mtDNA isolated from PolgWT and PolgD257A knock-in mice.

Fig. S4. Analysis of mitochondrial DNA copy number from human brain tissue samples.

Fig. S5. Analysis of ddPCR droplet counts from sequenced patient samples.

Fig. S6. Deletion sites are not fully saturated.

Fig. S7. Distribution of positive droplets from ddPCR of human brain.

Fig. S8. High resolution analysis of deletion dynamics.

Fig. S9. XXXXXXXX. **14**

Data S1. Supplementary notes, methods and figures.

Data S1A. Optimization of amplification factors.

Data S1B. Residual amplification bias.

Data S1C. Biases inherent in downstream applications.

Data S1D. Validation in biological samples.

Data S1E. Validation of sampling and analysis.

Data S1F. Disruption of droplet emulsions

Data S2. Tables of all accepted deletions for each patient. Sequence data deposited in the NCBI Sequence Read Archive under project accession number SRP027401.

Jason H. Bielas, Ph.D.

Curriculum Vitae

Assistant Member, Translational Research Program
Division of Public Health Sciences
Fred Hutchinson Cancer Research Center
1100 Fairview Avenue North, M5-A885
Seattle, WA 98109-1024
Phone: (206) 667-3170
Fax: (206) 667-2537
Email: jbielas@fhcrc.org
<http://labs.fhcrc.org/bielas>

Department of Pathology
University of Washington
Box 358080
Seattle WA, 98195

EDUCATION

Ph.D., 2003 With Distinction, Governor General of Canada Gold Medal Recipient, Department of Biology, York University, Toronto ON
Thesis: The Role Of Proliferation in Mutagenesis
Advisor: John A. Heddle

B.Sc., 1998 Honours, First Class with Distinction, Department of Biology, York University, Toronto, ON

PROFESSIONAL POSITIONS AND APPOINTMENTS

2011 - present **Assistant Member**, Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle WA

2008 - present **Assistant Member**, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA

2009 - present **Affiliate Assistant Professor**, Department of Pathology, University of Washington, School of Medicine, Seattle WA

2009 - present **Graduate Faculty Member**, Molecular and Cellular Biology (MCB) Program, University of Washington/Fred Hutchinson Cancer Research Center

2009 - present **Member**, Genome Instability and Mutagenesis, Fred Hutchinson/University of Washington Cancer Consortium

2006 - 2008 **Acting Instructor**, Department of Pathology, University of Washington, School of Medicine, Seattle WA

2003 - 2008 **Postdoctoral Fellow**, Department of Pathology, University of Washington, School of Medicine, Seattle WA

HONORS AND AWARDS

Outstanding New Environmental Scientist (ONES) Award, National Institute of Environmental Health Sciences (NIEHS), 2010-2015

New Scholar Awardee in Aging, The Ellison Medical Foundation, 2009-2013

Postdoctoral Research Fellowship, Terry Fox Foundation, National Cancer Institute of Canada (NCIC), 2007-2010

Postdoctoral Fellowship, Canadian Institutes of Health Research (CIHR), 2006-2007

Postdoctoral Fellowship, Natural Sciences and Engineering Research Council of Canada (NSERC), 2004-2005

Gold Medal, Governor General's Academic Medal, Her Excellency the Right Honourable Adrienne Clarkson, Governor General of Canada, 2004

Haynes Scholarship for Academic Excellence in Doctoral Studies, 2003

Ontario Graduate Scholarship in Science and Technology, 2002-2003
Graduate Fellowship, International Association of Environmental Mutagen Societies, 2001
Graduate Scholarship, Natural Sciences and Engineering Research Council of Canada (NSERC), 2000-2002
Dean's Academic Excellence Scholarship, Faculty of Graduate Studies, York University, 2000
Health Science International Award, CANTOX, 2000
Ontario Graduate Scholarship (OGS), 1999-2000
Graduate Fellowship, Cancer Research Society Inc., 1998-1999
York University Graduate Entrance Scholarship, 1998

REFEREED TRAVEL AWARDS

International Conference on Environmental Mutagens and the 36th Annual Meeting of the EMS, San Francisco, CA, 2005
The Edward A. Smuckler Memorial Pathobiology of Cancer Workshop, Snowmass, CO, 2005
The 33rd Annual Society of Toxicology of Canada Symposium, Montréal, QC, 2000
Annual Environmental Mutagen Society Meeting, New Orleans, LA, 2000
Gordon Research Conferences on Genetic Toxicology, Oxford, UK, 1999
Annual Environmental Mutagen Society Meeting, Washington, DC, 1999

PROFESSIONAL ACTIVITIES AND NATIONAL AND INTERNATIONAL SERVICE

Editorial Responsibilities:

Member, Editorial Board, Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2011-
Member, Editorial Board, PLoS ONE, 2010 -
Member, Editorial Board, Mutation Research Reviews, 2008-

Manuscript Review for:

Biotechnology Advances,	2009-2010
Environmental and Molecular Mutagenesis, Genetics	1999-2005, 2008, 2010, 2012 2013
Human Mutation	2009, 2010
Nature Methods	2007-2011
Nucleic Acid Research	2011
PLoS ONE	2009-2013
Proceedings of the National Academy of Sciences (PNAS)	2007-2009, 2011-2013
Mutation Research	2002-2005, 2008, 2009, 2011, 2013
Science	2008

Elected Positions Held:

Councilor, Environmental Mutagen Society, USA, 2010-2012

Consulting, Membership on Grant Review, and Advisory:

Member, Special Emphasis Panel, Identification of Biomarkers for Early Detection of Environmentally Induced Mitochondrial Dysfunction (R01), ZES1 LWJ - J(MI), NIH/NIEHS, June 2011 (ad hoc)
Member, Interdisciplinary Research Consortium in Geroscience Advisory Panel, The Buck Institute for Age Research, October 2010 (external expert, ad hoc)
Grant Reviewer, Royalty Research Fund, Office of Research, University of Washington, March 2009 (ad hoc)
Grant Reviewer, Institute of Translational Health Sciences, Pilot Grants, NIH Clinical and Translational Science Award (CTSA), April 2008 (ad hoc)

National and International Meetings and Committees:

Chair, Mitochondria Mutagenesis and Disease Symposium, Environmental Mutagen Society's 40th Annual Meeting, Fort Worth, TX, 2010
Member, Task Force on Committees, Environmental Mutagen Society, USA, 2010-2011
Chair, In vivo Mutagenesis: Recent Advances and Future Prospects, 10th International Conference on Environmental Mutagens (ICEM), Florence, Italy, 2009

Program Committee, Environmental Mutagen Society Meeting, Complex Systems in Biology and Risk Assessment, 2009-2010
Chair, Mutagenic and Carcinogenic Mechanisms, Platform Session, Environmental Mutagen Society's 39th Annual Meeting, Puerto Rico, 2008
Public Relations and Communications Committee, Environmental Mutagen Society, USA, 2001-2004
Independent Consultant, Novartis Pharmaceuticals Corporation, East Hanover, NJ, 2006-2008
Independent Consultant, Koronis Pharmaceuticals, Redmond, WA, 2006-2008
Graduate, The American Association for Cancer Research Edward A. Smuckler Memorial Pathobiology of Cancer Workshop, 2005

University and FHRC Service:

Lecturer, Biological Basis of Neoplasia, Molecular Cellular Biology, 2012, 2013
Member, Faculty Search Committee, Translational Research Program, 2012-
Chair, Translational Research Program Seminar Series, 2012-
Member, Organizing Committee, Human Biology Division Retreat, 2012
Member, Senior Vice President and Public Health Sciences Division Director Search Committee, 2011-2012
Co-Chair, Retreat and Discussions on the Direction for Genomic Molecular Diagnosis of Solid Tumors at the Center, December 2010
Lead, Literature Review, Molecular and Cellular Biology Graduate Program, February 2009

PUBLICATIONS

Retinal ganglion cells accumulate a phosphorylated form of Mfn2 protein that correlates with glaucomatous disease progression. Nivison MP, Ericson NG, Bielas JH, Campbell JS, Horner PJ. **J. Neurochem.** 2013 *Under Review*

High-throughput sequencing of T cell receptors reveals a tumor-specific and homogeneous repertoire of tumor-infiltrating lymphocytes in ovarian cancer. Emerson RO, Sherwood AM, Rieder MJ, Guenthoer J, Williamson DW, Carlson CS, Drescher CW, Tewari M, Bielas JH, Robins HS. **J. Pathol.** 2013 *Under Review*

Digital quantification of tumor infiltrating lymphocytes. Robins RS, Ericson NG, Guenthoer J, O'Briant KC, Tewari M, Drescher CW, Bielas JH. **Sci Transl Med.** 2013 *In Revision*

Simultaneous digital quantification and fluorescence-based size characterization of massively parallel sequencing libraries. Laurie MT, Bertout JA, Taylor SD, Burton JN, Shendure JA, Bielas JH. **Biotechniques.** 2013 *In Press*

Targeted enrichment and high-resolution digital profiling of DNA deletions in mitochondria. Taylor SD, Ericson NG, Burton JN, Prolla TA, Silber JR, Shendure J, Bielas JH. **Aging Cell.** 2013 *In Press*

Tumor-induced changes in the expression of the human metabolic network. Hu J, Locasale JW, Bielas JH, O'Sullivan J, Sheahan K, Cantley LC, Vander Heiden MG, Vitkup D. **Nat. Biotechnol.** 2013 Jun 31; (6):522-9. PMID:23604282

Decreased mitochondrial DNA mutagenesis in human colorectal cancer. Ericson NG, Kulawiec M, Vermulst M, Sheahan K, O'Sullivan J, Salk JJ, Bielas JH. **PLoS Genet.** 2012 Jun; (6):e1002689
PMID:22685414

Comment in: Tipping the balance in the powerhouse of the cell to "protect" colorectal cancer. Kidane D, Sweasy JB. PLoS Genet. 2012 Jun;8(6):e1002758.

Aging neural progenitor cells have decreased mitochondrial content and lower oxidative metabolism. Stoll EA, Cheung W, Mikheev AM, Sweet IR, Bielas JH, Zhang J, Rostomily RC, Horner PJ. **J Biol Chem.** 2011 Nov;286(44):38592-601. PMID:21900249

Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication. Allen JM, Simcha DM, Ericson NG, Alexander DL, Marquette JT, Van Biber BP, Troll CJ, Karchin R, Bielas JH, Loeb LA, Camps M. **Nucleic Acids Res.** 2011 Sep;39(16):7020-33.

PMID:21622658

A random mutation capture assay to detect genomic point mutations in mouse tissue. Wright JH, Modjeski KL, Bielas JH, Preston BD, Fausto N, Loeb LA, Campbell JS. **Nucleic Acids Res.** 2011 Jun;39(11):e73.

PMID:21459851

Generation, function, and prognostic utility of somatic mitochondrial DNA mutations in cancer. Kulawiec M, Salk JJ, Ericson NG, Wanagat J, Bielas JH. **Environ Mol Mutagen.** 2010 Jun;51(5):427-39.

PMID:20544883

Molecularly evolved thymidylate synthase inhibits 5-fluorodeoxyuridine toxicity in human hematopoietic cells. Bielas JH, Schmitt MW, Icreverzi A, Ericson NG, Loeb LA. **Hum Gene Ther.** 2009 Dec;20(12):1703-7.

PMID:19694534

Quantification of random mutations in the mitochondrial genome. Vermulst M, Bielas JH, Loeb LA. **Methods.** 2008 Dec;46(4):263-8.

PMID:18948200

Genetic instability is not a requirement for tumor development. Bodmer W, Bielas JH, Beckman RA. **Cancer Res.** 2008 May 15;68(10):3558-60; discussion 3560-1.

PMID:18483234

Cancers exhibit a mutator phenotype: clinical implications. Loeb LA, Bielas JH, Beckman RA. **Cancer Res.** 2008 May 15;68(10):3551-7; discussion 3557.

PMID:18483233

DNA deletions and clonal mutations drive premature aging in mitochondrial mutator mice. Vermulst M, Wanagat J, Kujoth GC, Bielas JH, Rabinovitch PS, Prolla TA, Loeb LA. **Nat Genet.** 2008 Apr;40(4):392-4.

PMID:18311139

Mitochondrial point mutations do not limit the natural lifespan of mice. Vermulst M, Bielas JH, Kujoth GC, Ladiges WC, Rabinovitch PS, Prolla TA, Loeb LA. **Nat Genet.** 2007 Apr;39(4):540-3. Epub 2007 Mar 4.

PMID:17334366

Comment in: Mitochondrial DNA mutations and aging: a case closed? Khrapko K, Vijg J. Nat Genet. 2007 Apr;39(4):445-6.

LOH-proficient embryonic stem cells: a model of cancer progenitor cells? Bielas JH, Venkatesan RN, Loeb LA. **Trends Genet.** 2007 Apr;23(4):154-7.

PMID:17328987

Limits to the Human Cancer Genome Project? Loeb LA, Bielas JH. **Science.** 2007 Feb 9;315(5813):762; author reply 764-5.

PMID:17297724

Human cancers express a mutator phenotype. Bielas JH, Loeb KR, Rubin BP, True LD, Loeb LA. **Proc Natl Acad Sci U S A.** 2006 Nov 28;103(48):18238-42.

PMID:17108085

Comment in: Random mutations, selected mutations: A PIN opens the door to new genetic landscapes. Klein CA. Proc Natl Acad Sci U S A. 2006 Nov 28;103(48):18033-4.

Non-transcribed strand repair revealed in quiescent cells. Bielas JH. **Mutagenesis.** 2006 Jan;21(1):49-53.

PMID:16394029

Generation of mutator mutants during carcinogenesis. Venkatesan RN, Bielas JH, Loeb LA. **DNA Repair (Amst).** 2006 Mar 7;5(3):294-302.

PMID:16359931

Quantification of random genomic mutations. Bielas JH, Loeb LA. **Nat Methods.** 2005 Apr;2(4):285-90.

PMID:15782221

Unifying concept of DNA repair: the polymerase scanning hypothesis. Heddle JA, Bielas JH. *Environ Mol Mutagen*. 2005 Mar-Apr;45(2-3):143-9.

PMID:15672383

Mutator phenotype in cancer: timing and perspectives. Bielas JH, Loeb LA. *Environ Mol Mutagen*. 2005 Mar-Apr;45(2-3):206-13.

PMID:15672382

Quiescent murine cells lack global genomic repair but are proficient in transcription-coupled repair. Bielas JH, Heddle JA. *DNA Repair (Amst)*. 2004 Jul 2;3(7):711-7.

PMID:15177180

Elevated mutagenesis and decreased DNA repair at a transgene are associated with proliferation but not apoptosis in p53-deficient cells. Bielas JH, Heddle JA. *Proc Natl Acad Sci U S A*. 2003 Oct 28;100(22):12853-8.

PMID:14569010

A more efficient Big Blue protocol improves transgene rescue and accuracy in adduct and mutation measurement. Bielas JH. *Mutat Res*. 2002 Jul 25;518(2):107-12.

PMID:12113761

Proliferation is necessary for both repair and mutation in transgenic mouse cells. Bielas JH, Heddle JA. *Proc Natl Acad Sci U S A*. 2000 Oct 10;97(21):11391-6.

PMID:11005832

Comment in: DNA damage, DNA repair, cell proliferation, and DNA replication: How do gene mutations result? O'Neill JP. Proc Natl Acad Sci U S A. 2000 97 (21) 11137-11139

The cll locus in the MutaMouse system. Swiger RR, Cosentino L, Shima N, Bielas JH, Cruz-Munoz W, Heddle JA. *Environ Mol Mutagen*. 1999;34(2-3):201-7.

PMID:10529745

INTELLECTUAL PROPERTY

1. **Bielas, J** and Bertout JA. 2013. Compositions and Methods for Accurately Identifying Mutations. PCT/US2013/025,505, filed February 15, 2013, pending.
2. **Bielas, J** and Taylor SD. 2012. Compositions and Methods for Detecting Rare Nucleic Acid Molecule Mutations. US 61/654,236, Filed June 1, 2012, pending.
3. **Bielas, J**, Taylor SD, and Laurie MT. 2013. Compositions and Methods for Detecting Rare Nucleic Acid Molecule Mutations. US 61/783,815, Filed March 14, 2013, pending.
4. **Bielas, J**. 2012. Quantification of Adaptive Immune Cell Genomes in a Complex Mixture of Cells. PCT/US2012/061,193, Filed October 29, 2012, pending.
5. **Bielas, J**, Robins H, and Livingston JR. 2012. Quantification of Adaptive Immune Cell Genomes in a Complex Mixture of Cells. US 13/656,265, filed October 19, 2012, pending.
6. **Bielas, J**, 2012. Compositions and Methods for Sensitive Mutation Detection in Nucleic Acid Molecules. US61/659,837, filed June 14, 2012, pending.

INVITED LECTURES

11/2013 Mitochondrial DNA maintenance, 11th International Conference on Environmental Mutagens (11th ICEM), Foz do Iguassu, PR, Brazil

- 10/2013 Mechanism and Clinical Utility of Nuclear and Mitochondrial DNA Mutations in Cancer, 5th Meeting on Fundamental Aspects of DNA Repair and Mutagenesis, University of Sao Paulo, Sao Paulo, Brazil
- 09/2013 Ultra-sensitive Detection of Nuclear and Mitochondrial DNA Mutations, Mutation Detection Workshop, 15th International Conference on Chronic Myeloid Leukemia: Biology and Therapy, Estoril, Portugal
- 08/2013 Mitochondrial DNA Mutagenesis: Insight into Human Aging, Carcinogenesis, and Novel Anticancer Therapies, Ellison Medical Foundation Colloquium on the Biology of Aging, Woods Hole, MA
- 05/2013 Mechanism and Clinical Utility of Nuclear and Mitochondrial DNA Mutations in Cancer, Biochemistry and Molecular Biology, Faculty of Medicine, Dalhousie University, Halifax, NS, Canada
- 03/2013 Mechanism and Clinical Utility of Nuclear and Mitochondrial DNA Mutations in Cancer, Irish Association for Cancer Research Annual Meeting, Dublin, Ireland
- 12/2012 Digital Detection of Rare Mutations, First Annual Droplet Digital PCR User Meet, Boston, MA
- 10/2012 Digital Detection of Rare Mutations and Tumor Infiltrating T Cells: Clinical Application in Cancer and Disease, Digital PCR Applications and Advances, Cambridge Healthtech Institute, San Diego, CA
- 10/2012 Mechanism and Clinical Utility of Nuclear and Mitochondrial DNA Mutations in Cancer, External Scientific Advisory Board Meeting, Fred Hutchison and University of Washington Cancer Consortium, Seattle, WA
- 09/2012 Metabolism and Mitochondrial Mutagenesis in Human Colorectal Cancer, Metabolism and Metabolites Symposium, Fred Hutchinson Cancer Research Center, Seattle, WA
- 07/2012 Nuclear and Mitochondrial DNA Mutations: Mechanisms and Disease, Meeting of the Outstanding New Environmental Scientist (ONES) Grantee Forum, NIEHS, Research Triangle Park, NC
- 05/2012 Mutant Nuclear and Mitochondrial DNA-Based Biomarker Discovery: Validation for the Detection, Prognosis and Treatment of Cancer, Disease Biomarkers Conference, London, England
- 02/2012 Digital Detection of Ultra-Rare Nuclear and Mitochondrial Mutations: Fundamental and Clinical Implications in Cancer and Disease, Digital PCR Short Course, Molecular Med Tri-Con 2012, Cambridge Healthtech Institute, San Francisco, CA
- 02/2012 Finding the Needle in the Haystack, The Institute For Prostate Cancer Research Symposium, Fred Hutchinson Cancer Research Center, Seattle, WA
- 08/2011 Ultra-Sensitive Mutant Nuclear and Mitochondrial DNA-Based Biomarker Discovery: Validation for the Detection, Prognosis and Treatment of Cancer, QuantaLife Droplet Digital PCR Road show, Boston, MA, Potomac, MD, New York, NY, Chicago, IL, Seattle, WA, San Diego, CA, San Francisco, CA
- 06/2011 The Mechanism and Clinical Utility of Somatic Mitochondrial Mutagenesis in Cancer, 11th International Symposium on Mutations in the Genome, Santorini, Greece
- 07/2011 Mechanisms of Environmental Mitochondrial Mutagenesis, Meeting of the Outstanding New Environmental Scientist (ONES) Grantee Forum, NIEHS, Research Triangle Park, NC
- 10/2010 Mutagenesis in Mitochondrial Genome, Environmental Mutagen Society Meeting, Complex Systems in Biology and Risk Assessment, Fort Worth, TX
- 09/2010 Somatic Mitochondrial DNA Mutations in Aging and Disease: Cause and Consequences, Center for Exercise Science Seminar Series, University of Florida
- 08/2009 The mechanism and clinical utility of somatic mitochondrial mutagenesis in cancer, 10th *International Conference on Environmental Mutagens (ICEM)*, Florence, Italy
- 07/2009 Mitochondrial DNA mutations as biomarkers. *PNW Prostate Cancer SPORE Retreat*, Vancouver, BC
- 09/2008 Mutations in Cancer Evolution. 36th *Annual Association of Graduate Students in Biological Sciences (AGSBS) Symposium*. The Evolution of Biology at York: Past, Present, and Future. York University's 50th Anniversary
- 10/2008 Nuclear and Mitochondrial Mutations in Cancer. *Environmental Mutagen Society's 39th Annual Meeting*, Puerto Rico

LAB MEMBERS

Current

04/2013- Haley BrinJones, Lab Aide
07/2013- William Valente, B.S., Medical Scientist Train Program (MD/PhD Candidate)
09/2011- Jessica Bertout, V.M.D., Ph.D., Post-Doctoral Research Fellow
06/2010- Sean Taylor, Ph.D., Post-Doctoral Research Fellow
11/2008- Nolan Ericson B.S., Research Technician and Lab Manager

Rotation Students

Molecular Cellular Biology (MCB):

09/2012-12/2012 Ethan Ahler, B.S.
09/2011-12/2012 Andrew Mathewson, B.S.
03/2010-06/2010 Sam Lancaster, B.S.

Medical Scientist Train Program (MSTP):

06/2012-08/2012 William Valente, B.S.

Past

09/2009-07/2013 Mariola Kulawiec, Ph.D., Post-Doctoral Research Fellow
03/2010-07/2013 Mathew Laurie, B.S., Research Technician
06/2011-08/2011 Tyler Gable, Summer Undergraduate Research Program Intern, Eastern Washington University
01/2009-05/2010 Dorothy Park, B.S., Lab Aide
08/2009-10/2009 Katja Schmalbach, M.A. trainee, Host Ph.D. student, University of Würzburg
01/2009-07/2009 Danielle Harden, M.A., Development Consultant
09/2008-03/2009 Scott Paulson B.S., Project coordinator