

RESEARCH

Open Access



Detection and quantification of mitochondrial DNA deletions from next-generation sequence data

Colleen M. Bosworth, Sneha Grandhi, Meetha P. Gould and Thomas LaFramboise*

From 12th International Symposium on Bioinformatics Research and Applications (ISBRA 2016)
Minsk, Belarus. 5-8 June 2016

Abstract

Background: Chromosomal deletions represent an important class of human genetic variation. Various methods have been developed to mine “next-generation” sequencing (NGS) data to detect deletions and quantify their clonal abundances. These methods have focused almost exclusively on the nuclear genome, ignoring the mitochondrial chromosome (mtDNA). Detecting mtDNA deletions requires special care. First, the chromosome’s relatively small size (16,569 bp) necessitates the ability to detect extremely focal events. Second, the chromosome can be present at thousands of copies in a single cell (in contrast to two copies of nuclear chromosomes), and mtDNA deletions may be present on only a very small percentage of chromosomes. Here we present a method, termed MitoDel, to detect mtDNA deletions from NGS data.

Results: We validate the method on simulated and real data, and show that MitoDel can detect novel and previously-reported mtDNA deletions. We establish that MitoDel can find deletions such as the “common deletion” at heteroplasmy levels well below 1%.

Conclusions: MitoDel is a tool for detecting large mitochondrial deletions at low heteroplasmy levels. The tool can be downloaded at <http://mendel.gene.cwru.edu/laframboiselab/>.

Keywords: Next-generation sequencing, Mitochondria DNA, Human genome, Chromosomal deletions

Background

Human genetic variation takes many forms, including single nucleotide variants, small insertions/deletions, larger chromosomal gains and losses, and inter-chromosomal translocations. A central pursuit in biomedical research is to determine those variants associated with human disease. Technological advances over the past several years have facilitated studies examining genetic variation at ever-increasing resolution, allowing better identification of variant-disease connections. Robust and accurate algorithms to detect all forms of human genetic variation from the ever-increasing number of large genomic data sets are necessary.

The vast majority of human DNA variant-detection algorithms focus exclusively on the 24 chromosomes (22 autosomes, X, and Y) comprising the nuclear genome. Usually ignored is the mitochondrial genome, despite the role of the mitochondrion in cellular bioenergetics and the known importance of mitochondrial mutations in a number of human diseases [1–7]. The mitochondrial genome (mtDNA) has features that distinguish it from its more commonly studied nuclear counterpart. First, the nuclear autosomal chromosomes are normally present in two copies per cell, while the number of copies of the mitochondrial chromosome varies widely from cell to cell, largely depending on tissue type. The mitochondrial chromosome may be present at hundreds, thousands, or tens of thousands of copies in a cell [8]. Second, the mutation rate of the mitochondrial genome is much higher than that of the nuclear genome

* Correspondence: thomas.laframboise@case.edu
Department of Genetics and Genome Sciences, Case Western Reserve
University School of Medicine, Cleveland, OH 44106, USA

and its repair mechanisms are far inferior to those in the nucleus [9]. The cell therefore carries considerably more mtDNA variants – both inherited and acquired – per base position than nuclear variants. Third, the mitochondrial genome is much smaller (16,569 bp) than the nuclear genome (~3.2 billion bp) and is circular rather than linear. Finally, mtDNA inheritance is strictly maternal. All of these differences present opportunities and challenges from an analytic perspective.

Since established computational tools used to identify biologically important nuclear DNA variants are often not adaptable to the mitochondrial genome, it is vitally important to develop new approaches to assess and quantify mtDNA genomic variation. Robust assessment of this variation in humans will allow identification of those variants that drive phenotypes, both benign and pathogenic. Owing to the limitations of established methods, this will necessitate the formulation of novel approaches particularly suited to the unique data types and biological scenarios inherent to mitochondrial genomics.

This study focuses on detecting deletions within the mitochondrial chromosome (Fig. 1). With the advent of genome-wide technologies, a great deal of research has been devoted to developing methodology to identify sub-chromosomal gains and losses from “next-generation” sequencing (NGS) data [10–12]. Few of these approaches have been applied to the mitochondrial genome. One of the reasons for this is the fluidity of mtDNA abundance and content. For instance, although generally

only two haplotypes per nuclear chromosome exist in an individual human (the exception being tumor cells), many distinct mitochondrial haplotypes may exist within a single individual, even in the same cell [13]. More than one distinct mtDNA haplotype being present in a single cell, tissue, or individual is known as heteroplasmy. MtDNA chromosomes harboring deletions are often present at very low heteroplasmy levels, making them difficult to detect.

In this paper, we describe MitoDel, the computational procedure we have developed to infer mtDNA deletions and their abundances from NGS data. We assess the theoretical sensitivity of our approach, and test its sensitivity and specificity using simulated data. We apply MitoDel to previously published data from sequencing experiments involving aging human brain tissue, and to the large public 1000 Genomes dataset [14]. We conclude the manuscript with discussion of the results and future directions.

Software implementing MitoDel is available at the LaFramboise laboratory website (<http://mendel.gene.cwru.edu/laframboiselab/>).

Methods

Acquisition of previously-published data

Courtesy of Dr. Sion Williams of the University of Miami, we obtained raw sequence data from the Williams et al. study [15]. Whole-genome .bam files of aligned and unaligned reads were downloaded from the 1000 Genomes website [14].

Simulated data

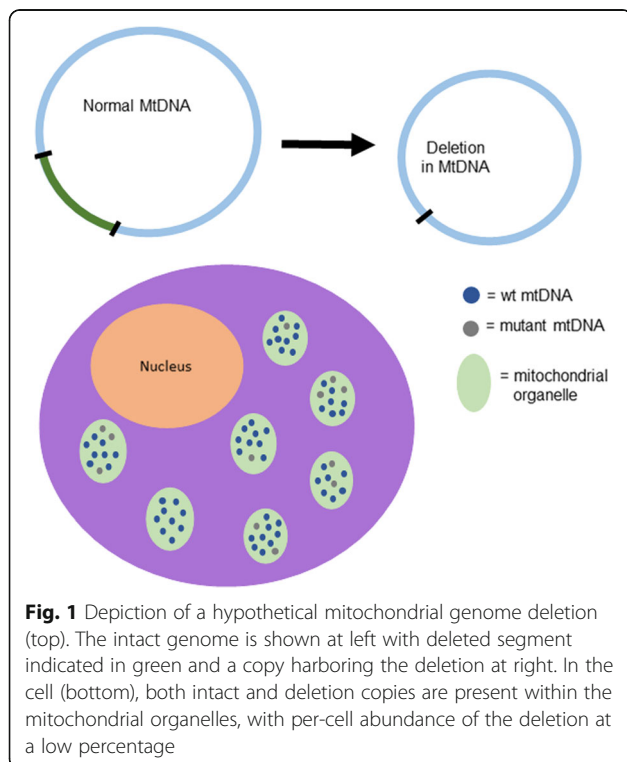
Read data from samples harboring deletions of various sizes and heteroplasmy levels were simulated using the ART simulator (version ART-ChocolateCherryCake-03-19-2015). To simulate an experiment generating R paired-end Illumina reads from a sample with a given deletion present in proportion p of mtDNA copies, we first modified the .fasta file containing the revised Cambridge Reference Sequence (rCRS; NC_012920.1) [16], removing a string of bases corresponding to the desired deletion. We then used ART to simulate $(1 - p) \times R$ reads from the rCRS reference, and $p \times R$ reads from the deleted version.

Raw read preprocessing

All .fastq files were first aligned to a modified human genome build hg19 using BWA [17]. Hg19 was modified by removing the original chrM and replacing it with the rCRS. Reads were not realigned if a .bam file was available.

MitoDel’s bioinformatic pipeline to detect mitochondrial DNA deletions

The mitochondrial genome is described as circular chromosome 16,569 bases in length. In the reference



genome, the base positions are numbered in a clock-like manner, from 5' to 3' on the “light” strand, from base position 1 to base position 16,569 (Fig. 2). When a deletion occurs, it has the effect of moving two base positions that are distant in the intact genome to being adjacent. It follows that reads harboring the resulting fusion point will either: i) not be deemed by the standard NGS aligner as having come from the mitochondrial genome, and will therefore be unaligned (Fig. 2); or ii) only be aligned after clipping or other modifications to the read. These modifications will be recorded in the CIGAR string field of the resulting .sam/.bam file [18], and the modified reads may thus be identified. Recovering these sequences and mining them for recurrent fusion points is the procedure that underpins our approach, as briefly described in a published abstract [19]. Furthermore, the relative abundance of mtDNA haplotypes harboring the deletion may be inferred by comparing the number of reads harboring the fusion point with the average read depth across the mitochondrial chromosome.

Formally (notation also shown in Fig. 2), suppose that the region from mitochondrial base position $s + 1$ to base position $e - 1$ is deleted in proportion p of mtDNA copies, and suppose that the NGS experiment generates reads of length l bases. Suppose further that n reads harbor the deletion fusion point. For the i^{th} of these reads, let x_i ($i = 1, \dots, n$) denote the position in the read (oriented from lower mtDNA base position to higher) harboring base position s in the mitochondrial genome ($1 \leq x_i \leq l$). Many of these reads will not align to anywhere on the reference genome, and will be therefore be marked as “unaligned” in the resulting .bam file output by BWA. We extract these unaligned reads, plus all reads with CIGAR strings indicating potential structural variants. This set of reads is then aligned to rCRS using BLAT [20].

Unlike BWA and other aligners designed for NGS data, BLAT is able to find splits of reads into multiple segments that each align to separate sites in a reference genome.

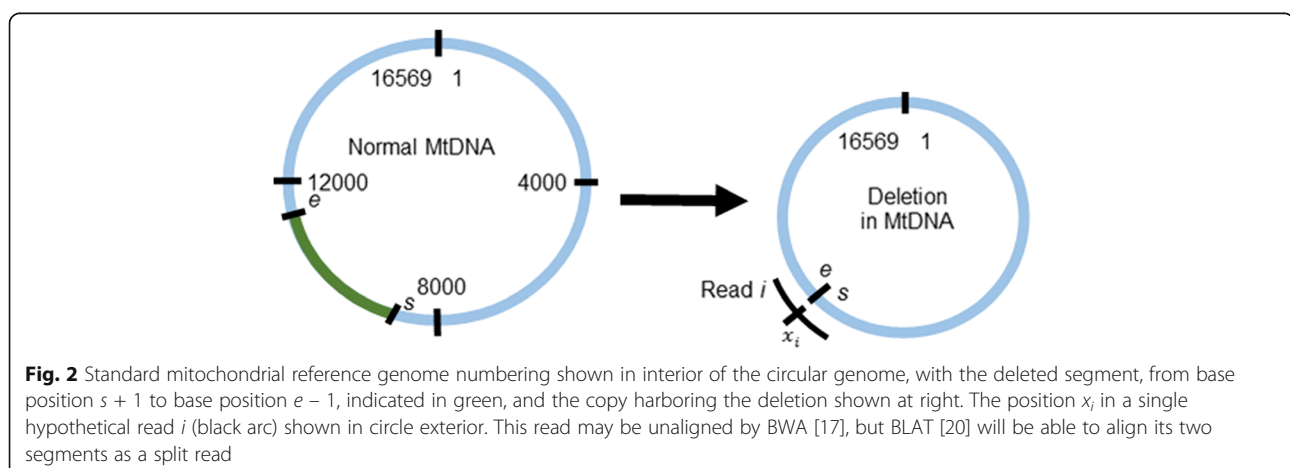
This capability comes at an extremely high computational cost, which is among the reasons that NGS aligners do not include it. However, since the mitochondrial genome is relatively extremely small, and since we filter out reads that map perfectly a priori, we are able to take advantage of BLAT without excessive computational burden (see “Compute time considerations” subsection below).

BLAT’s output for split reads includes the start and end read positions of each aligned segment of the read. In the above notation, this would correspond to two segments with (start, end) positions $(1, x_i)$ and $(x_i + 1, l)$ for read i . BLAT’s output also includes the beginning and ending genomic coordinates (mtDNA base position) to which each segment aligns. In the above notation, this would correspond to mtDNA positions $(s - x_i + 1, s)$ and $(e, e + l - x_i - 1)$ for the two read segments. It follows that we may interrogate the BLAT output for a set of n split reads that:

1. each split into two segments,
2. each have both segments map to the same strand of the mitochondrial genome,
3. all suggest the same deleted segment, and
4. collectively have the fusion point appear in at least five different locations in the read, i.e. the set $\{x_1, \dots, x_n\}$ contains at least five unique elements.

This last requirement helps avoid false positive deletions such as those that are the result of well-known sequencing artifacts such as PCR errors or the aligner splitting a read due to a single nucleotide substitution difference from the reference genome.

If the number of reads suggesting precisely the same breakpoint is sufficiently large, enough evidence is deemed to have been produced to report the breakpoint as biologically real. This number n (where a deletion is called if at least n split reads support it) is a tuning parameter. Clearly, higher values of n will increase specificity and decrease sensitivity. We use $n = 10$ as a default value in



MitoDel (see “Results from simulated reads” subsection below for justification), though this may be adjusted in the corresponding software.

An overview of the MitoDel procedure is shown in Fig. 3.

Estimating heteroplasmy level and confidence interval

The number N of reads harboring a deletion given the total reads in the experiment would be expected to approximately follow a binomial distribution $\text{Bin}(r, q)$, where r is the number of reads from the mitochondrial genome, and q is the proportion of reads that harbor the deletion. Since there are 16,569 possible starting positions for mtDNA reads, reads from the deleted copy of the genome will harbor the deletion with probability $l/16569$, where l is the length of the reads. Therefore, if we estimate q as $\hat{q} = n/r$ we may estimate the heteroplasmy level as $\hat{q} \times \frac{16569}{l}$. The $1 - \alpha$ confidence intervals on q may be computed analytically [21] as

$$\left(\frac{1}{1 + \frac{r-n+1}{n} F_{2(r-n+1), 2n, \frac{\alpha}{2}}}, \frac{\frac{n+1}{r-n} F_{2(n+1), 2(r-n), \frac{\alpha}{2}}}{1 + \frac{n+1}{r-n} F_{2(n+1), 2(r-n), \frac{\alpha}{2}}} \right),$$

where $F_{a,b,c}$ denotes the $1 - c$ quantile of the F distribution with a and b degrees of freedom. We can then transform this confidence interval on q to a confidence interval on the heteroplasmy level.

Results

Theoretical sensitivity

As mentioned above, mtDNA deletions have typically been observed at extremely low abundances, frequently a fraction of 1 %. Therefore, sequencing at high read depths is necessary to detect deletions. When designing NGS experiments for this purpose, researchers also must take into account the high numbers of nuclear genome reads present in the sequencing data, which will decrease the average number of reads per mitochondrial base position. Indeed, unless an mtDNA enrichment procedure is applied in the laboratory prior to DNA sequencing, only approximately 0.2% of DNA is expected to be mitochondrial [22]. Even with enrichment, the sensitivity of our computational procedure clearly depends on the number of mtDNA reads, which is a function of the enrichment protocol’s efficiency. We and others have performed studies developing and comparing various mtDNA enrichment protocols [22, 23], with varying results depending on the tissue type and other factors. Theoretical sensitivity for a sequencing experiment therefore here takes into account various enrichment levels.

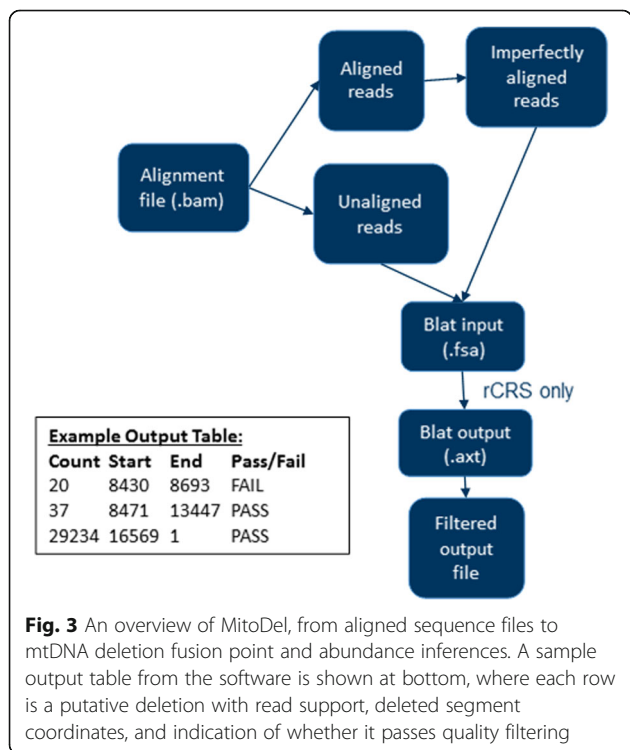
Given an NGS experiment with M total reads and mtDNA enrichment level E (enrichment here is defined as the proportion of DNA in the sample that is mitochondrial as opposed to nuclear), the number of reads harboring a given mitochondrial base position is expected to be approximately

$$N \approx (M \times E \times l) / 16569.$$

Computations using the binomial distribution show, for example, that a typical run on a standard Illumina MiSeq of ~50 million reads from a sample subjected to a protocol yielding 60% mtDNA enrichment would allow for detection of deletions as low as 0.006% with 95% probability, using our default threshold of 10 reads supporting the deletion. Sequencing experiments with higher numbers of reads and/or better enrichment protocols could find even lower-level deletions.

Results from simulated reads

We simulated sequencing experiments for three different mitochondrial deletions (small, medium, and large), generating paired-end Illumina reads of 150 bp each, with mean 300 and standard deviation 100 for the distance between reads. The simulated deletions were 15 bp (m.700_715del), 200 bp (m.5000_5200del), and one comparable in size to the well-known “common deletion” [24] at 4900 bp (m.6930_11830del). For each deletion, 100 replicates of the corresponding simulated sequencing experiment were run. Each iteration sampled the deleted genome at 70x coverage and the intact mitochondrial genome at 69930x coverage, thereby simulating a heteroplasmy level of 0.1%,



yielding paired-end .fastq files. These files were aligned and then run through MitoDel. We also performed 100 iterations of a simulation sampling the genome with the 200 bp deletion at 700× coverage and the intact mitochondrial genome at 69300× coverage, thereby simulating a heteroplasmy level of 1%.

We used the simulations to assess the sensitivity and the false positive rate of MitoDel, varying the number of reads harboring a deletion necessary for it to be called (the n parameter from above) from 1 to 50. For each value of n , we computed the proportion of the 100 replicates calling the deletion at least that threshold. This value was used as an estimate of MitoDel’s sensitivity. We also tallied the average number of deletions called (all but one of which are false positives since we only “spiked in” one deletion at a time) across the iterations for each threshold n . This was used to estimate the average number of false positives. These averages, (Fig. 4) show that for a variety of deletions at very low heteroplasmy levels, MitoDel remains highly specific. The average number of false positive calls falls steeply with increasing threshold n until about 10 reads, a threshold at which all called deletions are true positives, and the true deletion is always called. These simulation experiments led us to choose 10 as the default value for n in the MitoDel software.

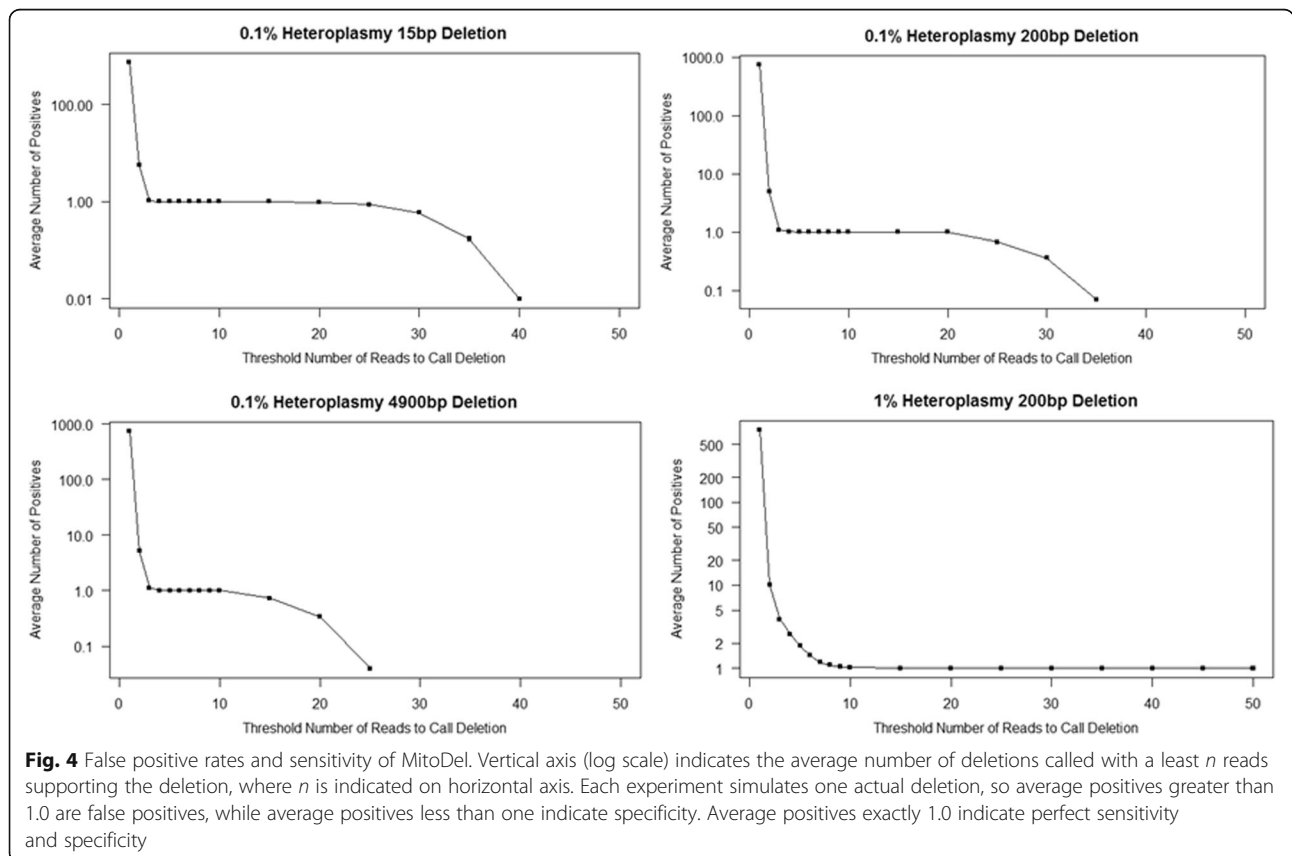
Detection of low-level deletions in brain tissue

We applied MitoDel to NGS data generated from human brain tissue for a previously-published study [15]. Using a method with no software and few computational details provided, the authors analyzed tissue from young (< 35 years old) and aged (> 66 years old) individuals. The study reported a ~ 5000 bp deletion (m.8483_13459del4977, the well-known common deletion [24]) present in the majority of the aged individuals but a minority of young individuals.

We acquired the raw sequencing data for 10 of these individuals directly from the authors and applied MitoDel agnostically, without targeting the common deletion specifically. Our presence/absence largely agreed with the authors’ assessments, except that we found evidence for a low-abundance deletion (0.58%) in an individual that the Williams et al. study deemed absent of deletions (Table 1). Manual inspection of the reads gives evidence that the deletion is indeed present, and our simulation results suggest that a false positive is unlikely. Generally, the heteroplasmy levels reported in the Williams et al. study were lower than, but were correlated with, our inferences.

MtDNA deletions in 1000 genomes data

Applying MitoDel to 10 .bam files from phase 3 of the 1000 Genomes Project, we found a 27-bp deletion (m.16306_16333del27) in the D-loop of individual



greatly reducing mapping burden for BLAT. The fact that the mtDNA reference genome is so small compared to the nuclear genome also reduces the burden. As demonstrated by our simulation results, the high precision of BLAT split read mapping confers extremely favorable precision and recall levels to MitoDel, even without more sophisticated statistical modeling of sequencing errors, mapping errors, and other sources of noise.

To our knowledge, there have been three methods published that could conceivably be used to detect low-level mtDNA deletions specifically. Mitoseek [26] is designed to detect all types of mitochondrial DNA-level variation. However, its deletion tool only reports read pairs whose mapped distance apart exceeds a user-specified threshold. It does not actually call the deletions or specify their coordinates, and therefore cannot be directly compared with MitoDel for accuracy. Delly [10] is designed to detect structural variants in a cancer context. While it allows for polyploidy when making calls, it does not handle heteroplasmy levels below 1% as MitoDel does. We were unable to successfully install and run the third method, MToolBox [27].

Conclusions

Here we have presented a computational method, MitoDel, to detect and quantify mtDNA deletions from next-generation sequencing experiments. Our method meets a need for software to identify aberrations present at extremely low levels. Our results demonstrate the ability to call deletions present at well below 1% heteroplasmy levels, with a very low false positive rate. Similar methods for detection of chromosomal aberrations have been developed for the nuclear genome, but these are tuned for much higher abundances. Indeed, a deletion in the nuclear genome will be present at at least 50% abundance in a cell. In heterogeneous tumor samples, the level may be lower in the overall sample, but available methods are not suitable for the extremely low abundances (< 1%) that MitoDel targets.

We can see a number of extensions to the work presented here. The most obvious would be to detect other types of mitochondrial chromosomal aberrations such as tandem duplications and inversions. Although these classes of aberration have rarely been described in mtDNA, only newer technology can detect them when present at low abundances, which could explain the lack of prior studies reporting them. Even mitochondrial-nuclear translocations have recently been described in cancer samples [28]. Theoretically, the approach described here could be easily modified to detect all of these lesion types.

Abbreviations

mtDNA: Mitochondrial DNA; NGS: Next-generation sequencing; rCRS: Revised Cambridge reference sequence

Acknowledgements

This work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University, as well as the Ohio Supercomputing Center. The authors would like to gratefully acknowledge Dr. Sion Williams of University of Miami for making raw sequence data available to us. The study was supported by American Cancer Society Research Scholar Grant 123436-RSG-12-159-01-DMC to T.L. This work was firstly presented at the 10th International Symposium on Bioinformatics Research and Applications (ISBRA 2016), June 5-8, 2016, Minsk, Belarus. A two-page abstract of this work was included in Lecture notes in computer science: Bioinformatics research and applications. Springer; 2016. p. 339-341.

Funding

Publication charges for this article have been funded by Case Western Reserve University institutional funding to T.L. The funding body played no role in study design or conclusions.

Availability of data and materials

The MitoDel tool can be downloaded at <http://mendel.gene.cwru.edu/laframboiselab/>.

About this supplement

This article has been published as part of BMC Bioinformatics Volume 18 Supplement 12, 2017: Selected articles from the 12th International Symposium on Bioinformatics Research and Applications (ISBRA-16): bioinformatics. The full contents of the supplement are available online at <<https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-18-supplement-12>>.

Authors' contributions

CMB and TL conceived the study, designed the software/experiments, and wrote the manuscript. SG designed the upstream NGS pipeline for mitochondria. MPG performed the sequencing validation experiments. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Publisher's Note

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

Published: 16 October 2017

References

- Shoffner JM, Fernhoff PM, Krawiecki NS, Caplan DB, Holt PJ, Koontz DA, Takei Y, Newman NJ, Ortiz RG, Polak M, et al. Subacute necrotizing encephalopathy: oxidative phosphorylation defects and the ATPase 6 point mutation. *Neurology*. 1992;42(11):2168–74.
- Johns DR, Neufeld MJ, Park RD. An ND-6 mitochondrial DNA mutation associated with Leber hereditary optic neuropathy. *Biochem Biophys Res Commun*. 1992;187(3):1551–7.
- Holt IJ, Harding AE, Petty RK, Morgan-Hughes JA. A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. *Am J Hum Genet*. 1990;46(3):428–33.
- Shoffner JM, Lott MT, Lezza AM, Seibel P, Ballinger SW, Wallace DC. Myoclonic epilepsy and ragged-red fiber disease (MERRF) is associated with a mitochondrial DNA tRNA(Lys) mutation. *Cell*. 1990;61(6):931–7.
- Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AM, Elsas LJ 2nd, Nikoskelainen EK. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science*. 1988;242(4884):1427–30.
- Prezant TR, Agopian JV, Bohlman MC, Bu X, Oztas S, Qiu WQ, Arnos KS, Cortopassi GA, Jaber L, Rotter JJ, et al. Mitochondrial ribosomal RNA

- mutation associated with both antibiotic-induced and non-syndromic deafness. *Nat Genet.* 1993;4(3):289–94.
7. Wallace DC. Mitochondrial DNA mutations in disease and aging. *Environ Mol Mutagen.* 2010;51(5):440–50.
 8. Robin ED, Wong R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol.* 1988;136(3):507–13.
 9. Schneider S, Excoffier L. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics.* 1999; 152(3):1079–89.
 10. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics.* 2012;28(18):i333–9.
 11. Leung WY, Marschall T, Paudel Y, Falquet L, Mei H, Schonhuth A, Mao Moss TY. SV-AUTOPILLOT: optimized, automated construction of structural variation discovery and benchmarking pipelines. *BMC Genomics.* 2015;16:238.
 12. Bartenhagen C, Dugas M. Robust and exact structural variation detection with paired-end and soft-clipped alignments: SoftSV compared with eight algorithms. *Brief Bioinform.* 2016;17(1):51–62.
 13. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet.* 2015;16(9):530–42.
 14. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, GA MV, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68–74.
 15. Williams SL, Mash DC, Zuchner S, Moraes CT. Somatic mtDNA mutation spectra in the aging human putamen. *PLoS Genet.* 2013;9(12):e1003990.
 16. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 1999;23(2):147.
 17. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
 18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
 19. Bosworth CM, Grandhi S, Gould MP, LaFramboise T. MitoDel: a method to detect and quantify mitochondrial DNA deletions from next-generation sequence data. *Lect N Bioinform.* 2016;9683:339–41.
 20. Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res.* 2002;12(4):656–64.
 21. Clopper CP, E. S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika.* 1934;26:404–13.
 22. Gould MP, Bosworth CM, McMahon S, Grandhi S, Grimerg BT, LaFramboise T. PCR-free enrichment of mitochondrial DNA from human blood and cell lines for high quality next-generation DNA sequencing. *PLoS One.* 2015; 10(10):e0139253.
 23. Jayaprakash AD, Benson EK, Gone S, Liang R, Shim J, Lambertini L, Toloue MM, Wigler M, Aaronson SA, Sachidanandam R. Stable heteroplasmy at the single-cell level is facilitated by intercellular exchange of mtDNA. *Nucleic Acids Res.* 2015;43(4):2177–87.
 24. Chen T, He J, Shen L, Fang H, Nie H, Jin T, Wei X, Xin Y, Jiang Y, Li H, et al. The mitochondrial DNA 4,977-bp deletion and its implication in copy number alteration in colorectal cancer. *BMC Med Genet.* 2011;12:8.
 25. Damas J, Carneiro J, Amorim A, Pereira F. MitoBreak: the mitochondrial DNA breakpoints database. *Nucleic Acids Res.* 2014;42(Database issue):D1261–8.
 26. Guo Y, Li J, Li C, Shyr Y, Samuels DC. MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis. *Bioinformatics.* 2013;29(9):1210–1.
 27. Calabrese C, Simone D, Diroma MA, Santorsola M, Gutta C, Gasparre G, Picardi E, Pesole G, Attimonelli M. MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. *Bioinformatics.* 2014; 30(21):3115–7.
 28. Ju YS, Tubio JM, Mifsud W, Fu B, Davies HR, Ramakrishna M, Li Y, Yates L, Gundem G, Tarpey PS, et al. Frequent somatic transfer of mitochondrial DNA into the nuclear genome of human cancer cells. *Genome Res.* 2015; 25(6):814–24.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

