Published in final edited form as:

Mitochondrion. 2014 July ; 0: 157–163. doi:10.1016/j.mito.2014.05.004.

High-Throughput Sequencing in Mitochondrial DNA Research

Fei Ye^a, David C. Samuels^b, Travis Clark^c, and Yan Guo^d

^aDepartment of Biostatistics, Vanderbilt University, Nashville, TN, 37232, USA

^bCenter for Human Genetics, Vanderbilt University, Nashville, TN, 37232, USA

^cVanderbilt Technology for Advanced Genomics, Vanderbilt University, Nashville, TN, 37232, USA

^dDepartment of Cancer Biology, Vanderbilt University, Nashville, TN, 37232, USA

Abstract

Next-generation sequencing, also known as high-throughput sequencing, has greatly enhanced researchers' ability to conduct biomedical research on all levels. Mitochondrial research has also benefitted greatly from high-throughput sequencing; sequencing technology now allows for screening of all 16569 base pairs of the mitochondrial genome simultaneously for SNPs and low level heteroplasmy and, in some cases, the estimation of mitochondrial DNA copy number. It is important to realize the full potential of high-throughput sequencing for the advancement of mitochondrial research. To this end, we review how high-throughput sequencing has impacted mitochondrial research in the categories of SNPs, low level heteroplasmy, copy number, and structural variants. We also discuss the different types of mitochondrial DNA sequencing and their pros and cons. Based on previous studies conducted by various groups, we provide strategies for processing mitochondrial DNA sequencing data, including assembly, variant calling, and quality control.

1. Introduction

Typically, there are approximately 100 mitochondria in each mammalian cell, and each mitochondrion harbors 2–10 copies of mitochondrial DNA (mtDNA) (Robin and Wong, 1988). Thus, mtDNA mutations are often heteroplasmic, with a mixture of normal and mutant mtDNA copies within a cell (Durbin et al., 2010; Ng et al., 2010). It has been found that heteroplasmy throughout the mitochondrial genome are common in normal individuals and moreover, that the frequency of heteroplasmic variants varies considerably between different tissues in the same individual (He et al., 2010). Mitochondria generate the majority

© 2014 Elsevier B.V. and Mitochondria Research Society. All rights reserved.

Corresponding Author: Yan Guo.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

of their cellular energy through oxidative phosphorylation, which produces ATP. Mitochondrial dysfunctions are important causes of many neurological diseases (Fernandez-Vizarra et al., 2007) and drug toxicities (Lemasters et al., 1999; Wallace and Starkov, 2000).

1.1 Older Methods to Sequence mtDNA

Previously, the two most popular complete mitochondrial genome sequencing methods were direct Sanger sequencing and mitochondrial DNA re-sequencing by Affymetrix's MitoChip v.2.0 (referred to as "MitoChip"). The MitoChip is based on microarray technology that contains 25-mer probes complementary to the revised Cambridge Reference Sequence (rCRS) (Andrews et al., 1999a). Several methods have been developed to quantify mtDNA heteroplasmy, such as real-time amplification refractory mutation system quantitative PCR (Bai and Wong, 2004), PCR-RFLP analysis (Holt et al., 1990), allele-specific oligonucleotide dot-blot analysis (Liang et al., 1998), and pyrosequencing (White et al., 2005). However, these methods are constrained by the limited number of targets they can scan. The maturity of high-throughput sequencing technology allows us to study the mitochondrial genome, including the level of mtDNA heteroplasmy at all sites in the mtDNA genome, in a reliable and cost-effective manner over large numbers of samples.

2. Direct sequencing of mtDNA

There have been three major sequencing platforms on the market: Illumina's HiSeq platform, Roche's 454 platform, and Applied Biosystems' SOLiD system. Mitochondrial DNA sequencing is possible with all three platforms (Craven et al., 2010; Payne et al., 2013), however the market has clearly been dominated by Illumina's sequencing platform for the last few years with no sign of diminishing. Thus, we focus on Illumina's sequencing technology in this review.

There are two typical ways to obtain information about the mitochondrial genome from high-throughput sequencing technology: direct and indirect. By "direct" we mean methods that sequence mtDNA directly through mtDNA enriched from total cellular DNA. There are several methods to enrich for mtDNA. Prior methods used ultra-centrifugation in CsCL density gradients to enrich mtDNA from nuclear DNA but this is a time-consuming and low-throughput procedure. Faster, high-throughput methods for mtDNA enrichment are microarray hybridization and PCR-based enrichment. For example, in the study of mitochondrial disorders by Vasta et al., a custom-designed Agilent microarray was used to capture the entire mitochondrial genome (Vasta et al., 2009). Similarly, in a radiation therapy study by Guo et al., the Affymetrix MitoChip v.2.0 was used to enrich mtDNA, though it was not used for the sequencing. Custom-designed primers can also be used to capture mtDNA (He et al., 2010) (Sosa et al., 2012). There is a major drawback for using overlapped primer capturing, however. For example, the MitoChip v.2.0 kit amplifies genomic DNA using PCR with two primer sets, mito3 and mito1-2. The two primer sets generate 7814bp and 9307bp long fragments respectively. Since, mtDNA are circular and only 16569 base pairs long, the two fragments will generate two overlap regions. The sequencing depth of the two overlapped regions is significantly higher than the non-overlapped regions, and the primer sequences need to be trimmed prior to variant calling. Common practice is to discard data obtained from the overlapped regions if overlapped

primers are used for enrichment (Guo et al., 2012a). Recently, a new PCR-based method using a single primer pair has been introduced to enrich the entire mitochondrial genome (Cui et al., 2013; Zhang et al., 2012). Using a single pair of primers readily avoids the pitfalls of using two or more sets of primers. Other advantages of this method include more uniform coverage, less interference from nuclear copies of the mitochondrial genome (nuMTs) (Hazkani-Covo et al., 2010; Li et al., 2012), and improved ability to estimate the breakpoints of large deletions. Additionally, several alternative commercial assays are available for mtDNA enrichment. For example, Qiagen SAbiosciences has a highly multiplexed PCR-based capture with an mtDNA GeneRead panel of 199 amplicons less than 300 bp covering 16146 bases (99.86%) and Integrated DNA Technologies (IDT) offers a solution phase capture of mtDNA with their xGen Lockdown probes.

A recent study comparing DNA isolation kits and mtDNA enrichment with and without PCR found that the Qiagen Miniprep kit had 22% of the reads aligned to mtDNA without a PCR enrichment step and 99% of the reads aligned to the mtDNA with a limited 10-cycle PCR step using a high fidelity enzyme (Quispe-Tintaya et al., 2013). The commercial mtDNA isolation kits from Miltenyi Biotec and BioVision both had ~10% of the reads aligned to mtDNA without PCR. With PCR enrichment, Miltenyi-prepped DNA increased to ~35% aligned to mtDNA, and BioVision increased to only ~15% indicating that the mtDNA isolation kits were inefficient in enrichment of mtDNA directly and the standard Qiagen Miniprep kit isolated a larger fraction of mtDNA, even though it was not optimized for mtDNA enrichment in the extraction procedure.

3. Indirect Sequencing of mtDNA

By “indirect sequencing”, we mean methods to obtain mitochondrial DNA sequences as byproducts of other types of high-throughput sequencing. Besides performing deep-sequencing specifically targeted at mtDNA, mtDNA sequences can also be extracted from other types of high-throughput sequencing data such as exome and whole genome sequencing data. In exome sequencing data, a significant amount of reads will align to the mitochondrial genome (around 1–5%), even when it is not the intended sequencing target (Samuels et al., 2013). Because the mitochondrial genome is not considered to be part of the exome, it is not included in the set of target DNA for exome sequencing methods in common use today. A recent study has shown that mtDNA content can be extracted from exome sequencing data (Larman et al., 2012) and that the fraction of captured mtDNA sequences is linked to the relative abundance of the corresponding mitochondrial genome in the original total DNA extract (Picardi and Pesole, 2012). The average coverage of the mitochondrial genome from exome sequencing is ~100, easily surpassing the average coverage of even the targeted genomic regions (Picardi and Pesole, 2012). The relatively high coverage is due to the high copy number of mtDNA per cell, on the order of hundreds to several hundred thousand copies per cell, depending on the tissue type (Bogenhagen and Clayton, 1974). The advance of high-throughput sequencing technologies and the typically high coverage of an mtDNA sequence provides a powerful tool for the study of mitochondrial DNA heteroplasmy in unprecedented detail (Durbin et al., 2010; Goto et al., 2011; Guo Y, 2012; Ng et al., 2010; Tang and Huang, 2010). However, this should be contrasted to techniques that specifically target the mtDNA sequence, which can produce an

average depth of tens of thousands of reads across the mitochondrial genome (Ameur et al., 2011; Guo et al., 2012a; He et al., 2010; Tang and Huang, 2010). Researchers have started to infer information about mtDNA mutation from exome sequencing data. The best examples are The Cancer Genome Atlas (TCGA) project, where all mitochondrial DNA somatic mutations were inferred from exome sequencing data. For example, the current somatic mutation results for breast cancer in TCGA(2012) contain exome sequencing data from 776 tumors and report 325 mtDNA somatic mutations derived from off-target reads from the exome sequencing data. Furthermore, by assessing mtDNA, exome sequencing mutation data has also been used to diagnose certain mitochondrial disorders (Dinwiddie et al., 2013).

Detecting mtDNA somatic mutation from exome sequencing data might contain false positive results caused by pseudogenes or homologous sequences (nuMTs). Because tumor tissue and the adjacent normal tissue often have different mtDNA content, the false results of heteroplasmic "mutation/variant" calling from the nuMTs could be different between the two samples, and thus at least some of the somatic mutations identified by using the exome data could be false. It is worth pointing out that false positive heteroplasmic variation due to nuMTs might also be improperly confirmed using a different method, unless the confirmation method carefully isolates the mtDNA from the nuclear DNA.

An important complication to consider in aligning DNA reads to the mitochondrial genome is the presence of nuMTs. The nuMTs can cause ambiguity about whether reads map to the nuclear or the mitochondrial genome. Aligning the raw reads against the mitochondrial reference genome directly and then filtering out the non-aligned reads, thus ignoring the nuMTs in the alignment is the simplest way to obtain the mitochondrial genome sequence. The disadvantage of this approach is that the reads from the nuMTs may map to the mtDNA, introducing false heteroplasmic variability in the reported mtDNA sequence. Another approach is to align the reads against both the nuclear and mitochondrial genomes simultaneously. When a read has multiple locations to which it may be mapped, such as the mtDNA and a nuMT, aligners such as BWA(Li et al., 2009) will randomly choose among the possible locations. This has the disadvantage of treating the nuMTs and the mitochondrial genome equally, ignoring the very large copy number difference between them. The effect of this choice will be that many of the reads coming from the mtDNA will be falsely aligned to the nuMTs, causing an artificially high coverage of the nuMTs and an artificially low coverage of the mtDNA. This artifact may limit the ability to detect mtDNA heteroplasmy, through lowering the mtDNA coverage. A third choice is to give precedence to the nuMTs by first aligning reads against the nuclear genome and then aligning only the remaining non-aligned reads to the mitochondrial genome. This approach will have the most extreme misalignment of true mtDNA reads to the nuclear DNA (potentially leading to false SNP calls in the nuclear DNA), which will also lower the coverage of the mitochondria genome and decrease the chance of detecting true variants or mtDNA heteroplasmy. The third approach is also the most conservative and time consuming approach, involving two alignment processes and leaving little chance of misaligning any nuMT reads to the mitochondrial genome. Of these three approaches, the second approach is the most balanced approach between time consumption and misalignment rate, and it has been implemented in

the recently developed software MitoSeek (Guo et al., 2013b), which can be used to extract mtDNA mutation and heteroplasmy information from exome sequencing data.

Indirect sequencing also has disadvantages compared to direct sequencing of mtDNA. It is important to note that even though mtDNA sequences extracted from exome sequencing data can be used to detect heteroplasmy, they lack the high depth of targeted mtDNA sequencing, which is needed to accurately detect low level heteroplasmy. Furthermore, exome sequencing capture based on hybridization technology cannot completely avoid capturing nuMTs.

4. SNP and Heteroplasmy Calling

Since mitochondrial DNA is haploid, a SNP in mtDNA is defined as a 100 percent deviation from the reference allele. Anything less would be considered heteroplasmy. Thus, the algorithm for SNP calling must be different for diploid and haploid genomes. In earlier studies of mtDNA (Guo et al., 2012a), researchers relied on SNP callers designed for haploid genomes such as the older version of GATK's Unified genotyper (DePristo et al., 2011) or custom designed algorithms to identify SNPs. The best example of using a diploid caller is the mtDNA SNP calling of the 1000 Genomes Project (Li et al., 2010) data, which was performed using a diploid genotype caller GLFTools v3 (Li et al., 2009). Using a caller designed for a diploid genome will result in calling high-level heteroplasmic sites as SNPs, possibly heterozygotic SNPs; therefore an additional filter will be needed to distinguish between SNPs and heteroplasmic sites. A custom calling algorithm usually involves counting the number of reads that support the non-reference allele (Fridjonsson et al., 2011). If all reads support the alternative allele, then this site is a SNP; otherwise it is heteroplasmy, and the heteroplasmy level can be estimated from the proportion of reads carrying the reference and non-reference alleles.

With the precise quantification achievable now with deep sequencing, researchers can determine whether the cancer-specific somatic mutations in mtDNA are heteroplasmic rather than homoplasmic. The level of heteroplasmy detectable in mtDNA is heavily dependent on the depth of coverage. In a study based on earlier sequencing technology, heteroplasmies as low as 5% were detectable (Li et al., 2010). Later studies have shown that with a read depth of tens of thousands, mtDNA heteroplasmies as low as 1% could be detected (Guo et al., 2012a). There are two requirements to detect heteroplasmies less than 1%: significantly increased depth and lowered sequencing error rates. The smallest heteroplasmy detectable is $1/D$, where D is the depth. However, as we lower the detectable threshold of heteroplasmy, it becomes increasingly difficult to distinguish between true heteroplasmies and sequencing errors (Guo et al., 2013a).

There are several sources that can contribute to errors. One of the most common is caused by PCR errors. Biased amplification from PCR can influence the minor allele frequency, and sequencing errors introduced during PCR amplification can be falsely detected as heteroplasmies (Calloway et al., 2000) (Grzybowski et al., 2003). PCR errors are very hard to prevent and identify, thus PCR errors must be understood as defining a lower limit to the detectable mtDNA heteroplasmy. One way to estimate the PCR error rate is by PCR and

sequence control libraries. For example, in He et al.'s mtDNA sequencing study, the authors calculated that the per-base mutation frequency is no greater than 0.82% from the control library (He et al., 2010). Based on this information, they made a conservative assumption that all heteroplasmy presented were at least twice this value (1.6%) (He et al., 2010). The downside is that not every study is designed with a control PCR library, and the mutation rate of the control library does often vary.

Another type of error is sequencing error from the sequencing platform. The majority of sequencing errors can be avoided by applying a stringent base quality filter. The Illumina platform outputs a Phred scale-based quality score (Ewing and Green, 1998; Ewing et al., 1998). The commonly used filter is $BQ < 20$ (more than 1% chance of a base being wrong). Alignment can also introduce errors. One way to minimize alignment error is by applying a mapping quality Phred score filter. However, not all aligners produce the same mapping quality scores. For example, in BWA (Li and Durbin, 2009), the mapping quality score is actually a Phred score to indicate the chances of incorrect mapping. However, the mapping quality score generated by Bowtie (Langmead and Salzberg, 2012) is used to indicate the uniqueness of the mapping. Thus, detailed attention needs to be paid to mapping quality score filtering based on the type of aligner used (Guo et al., 2013d). Another alignment related bias is the reference allele preferential bias, a phenomenon during alignment where there is preference toward the reference allele caused by alignment algorithms that penalize a mismatch from the reference. Degner et al. described such a bias in RNAseq data (Degner et al., 2009), and Guo et al. also described this in exome sequencing data (Guo et al., 2013c). This bias has an effect on minor allele frequency, usually in the range of 1–5% (Guo et al., 2013c). This seemingly small bias can potentially affect the detection of low level heteroplasmy. All of these types of sequencing errors will place a lower limit on the detectable heteroplasmy, no matter how deep the mtDNA read depth is.

One often ignored problem is caused by a peculiar feature in the standard reference, the rCRS: the artificially inserted “N” in the rCRS at position 3107 (Andrews et al., 1999b). When alignment is done directly using the rCRS, false heteroplasms, SNPs, and small indels will be detected around position 3107 (Guo et al., 2012a). To avoid this, the “N” should be deleted first from the reference before conducting the alignment. To maintain the standard site numbering system based on the rCRS, however, the false “N” will have to be inserted in the called sequence, shifting all position locations after 3107 by one.

Previously, sequencing accuracy was a concern in assessing the frequency of mutations in mitochondrial DNA, especially in tumor samples. Since the flanking sequence has a high effect on sequencing errors, these errors often occur in a biased manner between the reads on the opposite strands of DNA, an effect known as strand bias (Guo et al., 2012b). In the mitochondria DNA sequencing study by Guo et al., the authors computed a strand bias score and filtered out bases with extreme strand bias (Guo et al., 2012a). Also recently, another study (Schmitt et al., 2012) proposed a method called duplex sequencing. Duplex sequencing is performed by tagging and sequencing each of the two strands of a DNA duplex. Because the two strands are complementary, true mutations should be found on both strands. Thus, theoretically, this method can effectively detect and eliminate strand bias. In addition to the aforementioned sources of error, there are a few other factors that could

potentially generate false positives including bridging PCR, unintentionally sequenced nuMTS, and contamination in library preparation.

5. Copy Number and Structure Variation

Mitochondrial DNA copy number is highly variable and has been suggested to be associated with many diseases including cancer (Bai et al., 2011; Shen et al., 2010; Tseng et al., 2006; Yu et al., 2007). The traditional method for evaluating mtDNA copy number involves qPCR (Bhat and Epelboym, 2004). A more advanced method has been developed that relies upon a sequencing-based assay of mtDNA copy number that draws on the unbiased nature of next-generation sequencing and incorporates techniques developed for RNA expression profiling (Castle et al., 2010). The authors claimed that this assay reports absolute mtDNA copy number. However, the amount of library constructed may affect the reported copy number count, so we suggest that this method be interpreted as reporting a relative mtDNA copy number in arbitrary units, not an absolute count of mtDNA. For example, it has been shown that the fraction of captured mtDNA sequences in exome sequencing data is proportional to the relative abundance of the corresponding mitochondrial genome in the original total DNA extract (Picardi and Pesole, 2012). The mtDNA copy estimated from exome sequencing data can be useful when studying tumor samples for conducting association tests with phenotypes such as the tumor stage and metastasis stage. While researchers have already started to infer mtDNA copy number from exome sequencing data, (Guo et al., 2013b), it should be noted that this copy number estimation can be affected by many factors, including the method of DNA extraction (Guo et al., 2009).

Mitochondrial DNA deletions are a known disease-associated structural variation. The mtDNA deletions are often as long as a few thousand base pairs, such as the well-known 4977 base pair deletion referred to as the common deletion (Bogliolo et al., 1999; Maximo et al., 2001; Maximo et al., 1999). Such deletions have been linked to various diseases (Maximo et al., 2002); (Katada et al., 2013; Zhu et al., 2004). The methods used to detect large-scale mtDNA deletions usually involve PCR from two sets of primers overlapping at the region of a targeted deletion (Bogliolo et al., 1999; Maximo et al., 2001; Maximo et al., 1999). Such approaches are obviously limited by throughput and an inability to find the exact deletion break point. With high-throughput sequencing, researchers can detect novel deletions and can identify the exact break point. The detection method is usually based on detection of discordantly aligned paired-end reads. Paired-end read sequencing involves sequencing at both the 5' and 3' ends of a DNA fragment. After fragmentation by sonication, the DNA fragments are selected for a certain range, usually from 200 to 500 base pairs by electrophoresis. Thus, after alignment, the insert size (distance between two reads in a pair) should not exceed the DNA fragment range selected. A significantly larger insert size would indicate a large deletion. A more sophisticated approach involves splitting a read at the soft clip position of the read, and then aligning each half separately. Soft clips are proxies for split-reads that indicate that parts of the read map to different regions of the reference genome (Li et al., 2009). Two halves of a read aligned far apart on the reference genome would indicate a deletion, and the exact breaking point can be determined from the read.

Mitochondrial-nuclear genome integration is a known phenomenon where mtDNA fragments are integrated into nuclear chromosomes of eukaryotic cells during evolution (Zhang and Hewitt, 1996). Such integrations have been documented by multiple studies (Hazkani-Covo et al., 2010; Mourier et al., 2001; Timmis et al., 2004). Whether such integrations have any significant association with disease remains unknown. We can identify nuclear genome integration through high-throughput sequencing data by detecting discordantly aligned reads. Alignment of one read in the read pair to the mitochondrial DNA reference and the other to a nuclear chromosome would suggest an integration event. Like mtDNA deletions, the exact break point of the integration event can be identified by splitting reads around the soft clip point. The approaches described here have been implemented in programs such as MitoSeek, Pindel (Ye et al., 2009), and Dindel (Albers et al., 2011).

6. Quality Control on Mitochondrial DNA Sequencing Analysis

Quality control is an important component of sequencing analysis. Earlier, Guo et al. suggested that quality control on sequencing data should be performed at all stages of exome sequencing analysis, namely: raw data, alignment, and variant calling (Guo et al., 2013d). The same three-stage quality control strategies should be applied to mitochondrial DNA sequencing analysis. For raw data and alignment quality control, approaches similar to those for exome sequencing data can be used. One exception is that mtDNA capture efficiency must be substituted for exome capture efficiency. MtDNA capture efficiency is the number of reads mapped to the mtDNA reference genome divided by the total number of reads after quality control filtering is done. For variant SNPs and heteroplasmy, quality control is more challenging. In exome or whole genome sequencing, the transition/transversion (Ti/Tv) ratio can be used as an indicator of sequence quality (Durbin et al., 2010; Guo et al., 2012b; Guo et al., 2012c). However, this ratio is much different in the mitochondrial genome. For human nuclear genome data, the Ti/Tv ratio is around 3.0 for SNPs inside exons and around 2.0 elsewhere (Bainbridge et al., 2011); the ratio also differs between synonymous and non-synonymous SNPs (Yang and Nielsen, 1998). For mtDNA, previous studies have shown that the Ti/Tv ratio is much larger, between 21 (Pereira et al., 2009) and 38 (Guo et al., 2012a). These numbers can be used as guidance for future mitochondrial sequencing studies.

7. Mitochondrial DNA and Cancer

The investigation of somatic DNA changes in tumors is a major use of high-throughput sequencing. Mitochondrial DNA variations and somatic mutations may contribute to carcinogenesis and tumor progression (Chen, 2012; Modica-Napolitano and Singh, 2004), yet their role in tumorigenesis remains largely unknown. Early in the 1930s, Warburg discovered that cancer cells rely heavily on glycolysis to meet their metabolic demands (Kaiser-Wilhelm-Institut für Biologie and Warburg, 1930), and several hypothetical mechanisms have been proposed to explain this phenomenon (Kroemer, 2006). In 1998, somatic homoplasmic mutations in tumor mtDNA were found that were not present in matched control tissues (Polyak et al., 1998). Since then, various studies have shown that mutations in mtDNA can contribute to cancer etiology (Baysal et al., 2000; Vanharanta et al., 2004), and mtDNA mutations are associated with various types of cancer, including breast cancer (Canter et al., 2005) (Bai et al., 2007), prostate cancer (Herrmann et al., 2003;

Petrosillo et al., 2005), head and neck cancer (Sun et al., 2009), colon cancer (Ericson et al., 2012), and bladder cancer (Dasgupta et al., 2008; Fliss et al., 2000). It is likely that hereditary mitochondrial DNA variations may predispose individuals to cancer, while somatic mutations may possibly affect tumor progression. The functional consequences of mtDNA mutations largely result in metabolic alteration (Fischer et al., 1998; Lundholm et al., 1982; Mazurek et al., 1997; Ockner et al., 1993) and changes in the protein composition of the mitochondrial inner membrane. It also has been shown that an mtDNA mutation does not need to reach homoplasmy, i.e. all copies of mtDNA within a cell are mutated, to promote tumor growth (Lewis et al., 2000; Park et al., 2009). Researchers have found that heteroplasmy can promote tumor growth and that it is closely associated with aging (Kann et al., 1998; Smigrodzki and Khan, 2005; Sondheimer et al., 2011), although both points are somewhat controversial. In many earlier studies on mtDNA mutations, data were obtained by Sanger sequencing, which cannot detect low level heteroplasmies. For more than ten years, a growing number of articles have described somatic mutations of the mitochondrial genome in human tumors, identified by comparing tumor mtDNA and the mtDNA in adjacent normal tissue or blood.

In breast cancer, mutations in mitochondrial DNA have been linked to increased breast cancer risk through a deficiency in electron transport chain (ETC) function and altered reactive oxygen species (ROS) levels (Bai et al., 2007). Altered expression levels of the oxidative phosphorylation system (OXPHOS) subunits or mtDNA structural injury can impair ATP content and occur in breast-infiltrating ductal carcinoma (IDC) (Putignani et al., 2012). Variations in the mtDNA copy number may be the overall result of gene (hereditary) and environmental interactions (oxidative stress) caused by potential exogenous cancer risk factors such as age, hormones, diet and environmental oxidants/antioxidants, and reaction to oxidative damage (Lee et al., 1998; Renis et al., 1989; Verma et al., 2007). While it is difficult to prove the involvement of mtDNA mutations in triggering oncogenesis, there is increasing evidence that mutated mtDNA is a marker of poor survival in cancer prognosis. The first suggestions that mtDNA mutations may play a role in metastasis came from the comparison of frequency of somatic mtDNA mutations in non-small-cell lung cancer (NSCLC) at different stages of tumor formation, which show a significantly decreased survival among advanced NSCLC patients harboring mtDNA mutations. However, very few studies have investigated the role of mtDNA copy number in breast cancer patients and those studies have generated somewhat conflicting results. Researchers have found that both increased mtDNA copy number from whole blood DNA (Shen et al., 2010) and reduced mtDNA copy number in tissue may increase breast cancer risk (Bai et al., 2011; Tseng et al., 2006; Yu et al., 2007). This understudied area needs further exploration.

8. Conclusion

The introduction of high-throughput sequencing has greatly enhanced our abilities to conduct thorough and high-throughput mitochondrial genetics research. So far, high-throughput sequencing has been used to study mitochondrial disorders (Calvo et al., 2012; Vasta et al., 2009); (Dames et al., 2013); (van der Walt et al., 2012), mitochondrial DNA mutations due to radiation (Guo et al., 2012a), heteroplasmy inheritance (Payne et al., 2013), cancer (Yang Ai et al., 2013); (Lam et al., 2012) and many other related fields. Next-

generation methods targeting the mitochondrial genome result in very high depth mtDNA sequences, but the ability to detect low-level heteroplasmies is still limited by a number of quality control criteria that must be carefully handled. Even without targeted mtDNA sequencing, mtDNA sequences can be extracted from exome sequencing data, greatly increasing the range of mitochondrial genetics data available for research purposes.

Acknowledgments

We would like to thank Margot Bjoring for her editorial support. Yan Guo, and

References

- Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490:61–70. [PubMed: 23000897]
- Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, Durbin R. Dindel: accurate indel calls from short-read data. *Genome Res*. 2011; 21:961–973. [PubMed: 20980555]
- Ameur A, Stewart JB, Freyer C, Hagstrom E, Ingman M, Larsson NG, Gyllenstein U. Ultra-deep sequencing of mouse mitochondrial DNA: mutational patterns and their origins. *PLoS Genet*. 2011; 7:e1002028. [PubMed: 21455489]
- 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*. 1999a; 23:147. [PubMed: 10508508]
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature genetics*. 1999b; 23:147. [PubMed: 10508508]
- Bai RK, Chang J, Yeh KT, Lou MA, Lu JF, Tan DJ, Liu H, Wong LJ. Mitochondrial DNA content varies with pathological characteristics of breast cancer. *J Oncol*. 2011; 2011:496189. [PubMed: 22028711]
- Bai RK, Leal SM, Covarrubias D, Liu A, Wong LJ. Mitochondrial genetic background modifies breast cancer risk. *Cancer Res*. 2007; 67:4687–4694. [PubMed: 17510395]
- Bai RK, Wong LJ. Detection and quantification of heteroplasmic mutant mitochondrial DNA by real-time amplification refractory mutation system quantitative PCR analysis: a single-step approach. *Clin Chem*. 2004; 50:996–1001. [PubMed: 15073091]
- Bainbridge MN, Wang M, Wu Y, Newsham I, Muzny DM, Jefferies JL, Albert TJ, Burgess DL, Gibbs RA. Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome Biol*. 2011; 12:R68. [PubMed: 21787409]
- Baysal BE, Ferrell RE, Willett-Brozick JE, Lawrence EC, Myssiorek D, Bosch A, van der Mey A, Taschner PE, Rubinstein WS, Myers EN, Richard CW 3rd, Cornelisse CJ, Devilee P, Devlin B. Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. *Science*. 2000; 287:848–851. [PubMed: 10657297]
- Bhat HK, Epelboym I. Quantitative analysis of total mitochondrial DNA: competitive polymerase chain reaction versus real-time polymerase chain reaction. *J Biochem Mol Toxicol*. 2004; 18:180–186. [PubMed: 15452886]
- Bogenhagen D, Clayton DA. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid. *J Biol Chem*. 1974; 249:7991–7995. [PubMed: 4473454]
- Bogliolo M, Izzotti A, De Flora S, Carli C, Abbondandolo A, Degan P. Detection of the '4977 bp' mitochondrial DNA deletion in human atherosclerotic lesions. *Mutagenesis*. 1999; 14:77–82. [PubMed: 10474826]
- Calloway CD, Reynolds RL, Herrin GL, Anderson WW. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *American journal of human genetics*. 2000; 66:1384–1397. [PubMed: 10739761]

- Calvo SE, Compton AG, Hershman SG, Lim SC, Lieber DS, Tucker EJ, Laskowski A, Garone C, Liu S, Jaffe DB, Christodoulou J, Fletcher JM, Bruno DL, Goldblatt J, Dimauro S, Thorburn DR, Mootha VK. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci Transl Med*. 2012; 4:118ra110.
- Canter JA, Kallianpur AR, Parl FF, Millikan RC. Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Res*. 2005; 65:8028–8033. [PubMed: 16140977]
- Castle JC, Biery M, Bouzek H, Xie T, Chen R, Misura K, Jackson S, Armour CD, Johnson JM, Rohl CA, Raymond CK. DNA copy number, including telomeres and mitochondria, assayed using next-generation sequencing. *BMC genomics*. 2010; 11:244. [PubMed: 20398377]
- Chen EI. Mitochondrial dysfunction and cancer metastasis. *J Bioenerg Biomembr*. 2012; 44:619–622. [PubMed: 22892817]
- Craven L, Tuppen HA, Greggains GD, Harbottle SJ, Murphy JL, Cree LM, Murdoch AP, Chinnery PF, Taylor RW, Lightowlers RN, Herbert M, Turnbull DM. Pronuclear transfer in human embryos to prevent transmission of mitochondrial DNA disease. *Nature*. 2010; 465:82–85. [PubMed: 20393463]
- Cui H, Li F, Chen D, Wang G, Truong CK, Enns GM, Graham B, Milone M, Landsverk ML, Wang J, Zhang W, Wong LJ. Comprehensive next-generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders. *Genet Med*. 2013; 15:388–394. [PubMed: 23288206]
- Dames S, Chou LS, Xiao Y, Wayman T, Stocks J, Singleton M, Eilbeck K, Mao R. The development of next-generation sequencing assays for the mitochondrial genome and 108 nuclear genes associated with mitochondrial disorders. *J Mol Diagn*. 2013; 15:526–534. [PubMed: 23665194]
- Dasgupta S, Hoque MO, Upadhyay S, Sidransky D. Mitochondrial cytochrome B gene mutation promotes tumor growth in bladder cancer. *Cancer Res*. 2008; 68:700–706. [PubMed: 18245469]
- Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, Gilad Y, Pritchard JK. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics*. 2009; 25:3207–3212. [PubMed: 19808877]
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011; 43:491–498. [PubMed: 21478889]
- Dinwiddie DL, Smith LD, Miller NA, Atherton AM, Farrow EG, Strenk ME, Soden SE, Saunders CJ, Kingsmore SF. Diagnosis of mitochondrial disorders by concomitant next-generation sequencing of the exome and mitochondrial genome. *Genomics*. 2013
- Durbin RM, Altshuler DL, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Collins FS, De la Vega FM, Donnelly P, Egholm M, Flicek P, Gabriel SB, Gibbs RA, Knoppers BM, Lander ES, Leirach H, Mardis ER, McVean GA, Nickerson D, Peltonen L, Schafer AJ, Sherry ST, Wang J, Wilson RK, Gibbs RA, Deiros D, Metzker M, Muzny D, Reid J, Wheeler D, Wang J, Li JX, Jian M, Li G, Li RQ, Liang HQ, Tian G, Wang B, Wang J, Wang W, Yang HM, Zhang XQ, Zheng HS, Lander ES, Altshuler DL, Ambrogio L, Bloom T, Cibulskis K, Fennell TJ, Gabriel SB, Jaffe DB, Shefler E, Sougnez CL, Bentley DR, Gormley N, Humphray S, Kingsbury Z, Koko-Gonzales P, Stone J, McKernan KJ, Costa GL, Ichikawa JK, Lee CC, Sudbrak R, Leirach H, Borodina TA, Dahl A, Davydov AN, Marquardt P, Mertes F, Nietfeld W, Rosenstiel P, Schreiber S, Soldatov AV, Timmermann B, Tolzmann M, Egholm M, Affourtit J, Ashworth D, Attiya S, Bachorski M, Buglione E, Burke A, Caprio A, Celone C, Clark S, Connors D, Desany B, Gu L, Guccione L, Kao K, Kebbel A, Knowlton J, Labrecque M, McDade L, Mealmaker C, Minderman M, Nawrocki A, Niazi F, Pareja K, Ramenani R, Riches D, Song W, Turcotte C, Wang S, Mardis ER, Dooling D, Fulton L, Fulton R, Weinstock G, Durbin RM, Burton J, Carter DM, Churcher C, Coffey A, Cox A, Palotie A, Quail M, Skelly T, Stalker J, Swerdlow HP, Turner D, De Witte A, Giles S, Gibbs RA, Wheeler D, Bainbridge M, Challis D, Sabo A, Yu F, Yu J, Wang J, Fang XD, Guo XS, Li RQ, Li YR, Luo RB, Tai S, Wu HL, Zheng HC, Zheng XL, Zhou Y, Yang HM, Marth GT, Garrison EP, Huang W, Indap A, Kural D, Lee WP, Leong WF, Huang WC, Indap A, Kural D, Lee WP, Leong WF, Quinlan AR, Stewart C, Stromberg MP, Ward AN, Wu JT, Lee C, Mills RE, Shi XH, Daly MJ, DePristo MA, Altshuler DL, Ball AD, Banks E, Bloom T, Browning BL,

Cibulskis K, Fennell TJ, Garimella KV, Grossman SR, Handsaker RE, Hanna M, Hartl C, Jaffe DB, Kernysky AM, Korn JM, Li H, Maguire JR, McCarroll SA, McKenna A, Nemesh JC, Philippakis AA, Poplin RE, Price A, Rivas MA, Sabeti PC, Schaffner SF, Shefler E, Shlyakhter IA, Cooper DN, Ball EV, Mort M, Phillips AD, Stenson PD, Sebat J, Makarov V, Ye K, Yoon SC, Bustamante CD, Clark AG, Boyko A, Degenhardt J, Gravel S, Gutenkunst RN, Kaganovich M, Keinan A, Lacroute P, Ma X, Reynolds A, Clarke L, Flicek P, Cunningham F, Herrero J, Keenen S, Kulesha E, Leinonen R, McLaren W, Radhakrishnan R, Smith RE, Zalunin V, Zheng-Bradley XQ, Korbel JO, Stutz AM, Humphray S, Bauer M, Cheetham RK, Cox T, Eberle M, James T, Kahn S, Murray L, Ye K, De La Vega FM, Fu YT, Hyland FCL, Manning JM, McLaughlin SF, Peckham HE, Sakarya O, Sun YA, Tsung EF, Batzer MA, Konkel MK, Walker JA, Sudbrak R, Albrecht MW, Amstislavskiy VS, Herwig R, Parkhomchuk DV, Sherry ST, Agarwala R, Khouri H, Morgulis AO, Paschall JE, Phan LD, Rotmistrovsky KE, Sanders RD, Shumway MF, Xiao CL, McVean GA, Auton A, Iqbal Z, Lunter G, Marchini JL, Moutsianas L, Myers S, Tumian A, Desany B, Knight J, Winer R, Craig DW, Beckstrom-Sternberg SM, Christoforides A, Kurdoglu AA, Pearson J, Sinari SA, Tembe WD, Haussler D, Hinrichs AS, Katzman SJ, Kern A, Kuhn RM, Przeworski M, Hernandez RD, Howie B, Kelley JL, Melton SC, Abecasis GR, Li Y, Anderson P, Blackwell T, Chen W, Cookson WO, Ding J, Kang HM, Lathrop M, Liang LM, Moffatt MF, Scheet P, Sidore C, Snyder M, Zhan XW, Zollner S, Awadalla P, Casals F, Idaghdour Y, Keebler J, Stone EA, Zilversmit M, Jorde L, Xing JC, Eichler EE, Aksay G, Alkan C, Hajirasouliha I, Hormozdiari F, Kidd JM, Sahinalp SC, Sudmant PH, Mardis ER, Chen K, Chinwalla A, Ding L, Koboldt DC, McLellan MD, Dooling D, Weinstock G, Wallis JW, Wendl MC, Zhang QY, Durbin RM, Albers CA, Ayub Q, Balasubramaniam S, Barrett JC, Carter DM, Chen YA, Conrad DF, Danecek P, Dermitzakis ET, Hu M, Huang N, Hurles ME, Jin HJ, Jostins L, Keane TM, Keane TM, Le SQ, Lindsay S, Long QA, MacArthur DG, Montgomery SB, Parts L, Stalker J, Tyler-Smith C, Walter K, Zhang YJ, Gerstein MB, Snyder M, Abyzov A, Abyzov A, Balasubramanian S, Bjornson R, Du JA, Grubert F, Habegger L, Haraksingh R, Jee J, Khurana E, Lam HYK, Leng J, Mu XJ, Urban AE, Zhang ZD, Li YR, Luo RB, Marth GT, Garrison EP, Kural D, Quinlan AR, Stewart C, Stromberg MP, Ward AN, Wu JT, Lee C, Mills RE, Shi XH, McCarroll SA, Banks E, DePristo MA, Handsaker RE, Hartl C, Korn JM, Li H, Nemesh JC, Sebat J, Makarov V, Ye K, Yoon SC, Degenhardt J, Kaganovich M, Clarke L, Smith RE, Zheng-Bradley XQ, Korbel JO, Humphray S, Cheetham RK, Eberle M, Kahn S, Murray L, Ye K, De la Vega FM, Fu YT, Peckham HE, Sun YA, Batzer MA, Konkel MK, Xiao CL, Iqbal Z, Desany B, Blackwell T, Snyder M, Xing JC, Eichler EE, Aksay G, Alkan C, Hajirasouliha I, Hormozdiari F, Kidd JM, Chen K, Chinwalla A, Ding L, McLellan MD, Wallis JW, Hurles ME, Conrad DF, Walter K, Zhang YJ, Gerstein MB, Snyder M, Abyzov A, Du JA, Grubert F, Haraksingh R, Jee J, Khurana E, Lam HYK, Leng J, Mu XJ, Urban AE, Zhang ZD, Gibbs RA, Bainbridge M, Challis D, Coafra C, Dinh H, Kovar C, Lee S, Muzny D, Nazareth L, Reid J, Sabo A, Yu FL, Yu J, Marth GT, Garrison EP, Indap A, Leong WF, Quinlan AR, Stewart C, Ward AN, Wu JT, Cibulskis K, Fennell TJ, Gabriel SB, Garimella KV, Hartl C, Shefler E, Sougnez CL, Wilkinson J, Clark AG, Gravel S, Grubert F, Clarke L, Flicek P, Smith RE, Zheng-Bradley XQ, Sherry ST, Khouri HM, Paschall JE, Shumway MF, Xiao CL, McVean GA, Katzman SJ, Abecasis GR, Blackwell T, Mardis ER, Dooling D, Fulton L, Fulton R, Koboldt DC, Durbin RM, Balasubramaniam S, Coffey A, Keane TM, MacArthur DG, Palotie A, Scott C, Stalker J, Tyler-Smith C, Gerstein MB, Balasubramanian S, Chakravarti A, Knoppers BM, Peltonen L, Abecasis GR, Bustamante CD, Gharani N, Gibbs RA, Jorde L, Kaye JS, Kent A, Li T, McGuire AL, McVean GA, Ossorio PN, Rotimi CN, Su YY, Toji LH, Tyler-Smith C, Brooks LD, Felsenfeld AL, McEwen JE, Abdallah A, Christopher R, Clemm NC, Collins FS, Duncanson A, Green ED, Guyer MS, Peterson JL, Schafer AJ, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA, Consortium GP. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–1073. [PubMed: 20981092]

Ericson NG, Kulawiec M, Vermulst M, Sheahan K, O'Sullivan J, Salk JJ, Bielas JH. Decreased mitochondrial DNA mutagenesis in human colorectal cancer. *PLoS Genet*. 2012; 8:e1002689. [PubMed: 22685414]

Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*. 1998; 8:186–194. [PubMed: 9521922]

Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res*. 1998; 8:175–185. [PubMed: 9521921]

- Fernandez-Vizarra E, Bugiani M, Goffrini P, Carrara F, Farina L, Procopio E, Donati A, Uziel G, Ferrero I, Zeviani M. Impaired complex III assembly associated with BCS1L gene mutations in isolated mitochondrial encephalopathy. *Hum Mol Genet.* 2007; 16:1241–1252. [PubMed: 17403714]
- Fischer CP, Bode BP, Souba WW. Adaptive alterations in cellular metabolism with malignant transformation. *Ann Surg.* 1998; 227:627–634. discussion 634–626. [PubMed: 9605654]
- Fliss MS, Usadel H, Caballero OL, Wu L, Buta MR, Eleff SM, Jen J, Sidransky D. Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. *Science.* 2000; 287:2017–2019. [PubMed: 10720328]
- Fridjonsson O, Olafsson K, Tompsett S, Bjornsdottir S, Consuegra S, Knox D, de Leaniz CG, Magnusdottir S, Olafsdottir G, Verspoor E, Hjorleifsdottir S. Detection and mapping of mtDNA SNPs in Atlantic salmon using high throughput DNA sequencing. *BMC genomics.* 2011; 12:179. [PubMed: 21473771]
- Goto H, Dickins B, Afgan E, Paul IM, Taylor J, Makova KD, Nekrutenko A. Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. *Genome Biol.* 2011; 12:R59. [PubMed: 21699709]
- Grzybowski T, Malyarchuk BA, Czarny J, Miscicka-Sliwica D, Kotzbach R. High levels of mitochondrial DNA heteroplasmy in single hair roots: Reanalysis and revision. *Electrophoresis.* 2003; 24:1159–1165. [PubMed: 12707907]
- Guo W, Jiang L, Bhasin S, Khan SM, Swerdlow RH. DNA extraction procedures meaningfully influence qPCR-based mtDNA copy number determination. *Mitochondrion.* 2009; 9:261–265. [PubMed: 19324101]
- Guo Y, Cai Q, Samuels DC, Ye F, Long J, Li CI, Winther JF, Tawn EJ, Stovall M, Lahteenmaki P, Malila N, Levy S, Shaffer C, Shyr Y, Shu XO, Boice JD Jr. The use of next generation sequencing technology to study the effect of radiation therapy on mitochondrial DNA mutation. *Mutat Res.* 2012a; 744:154–160. [PubMed: 22387842]
- Guo Y, C Q, Samuels DC, Ye F, Long J, Li CI, Winther JF, Tawn EJ, Stovall M, Lähtenmäki P, Malila N, Levy S, Shaffer C, Shyr Y, Shu XO, Boice JD Jr. The use of next generation sequencing technology to study the effect of radiation therapy on mitochondrial DNA mutation. *Mutation Research - Genetic Toxicology and Environmental Mutagenesis.* 2012
- Guo Y, Li CI, Sheng Q, Winther JF, Cai Q, Boice JD, Shyr Y. Very low-level heteroplasmy mtDNA variations are inherited in humans. *J Genet Genomics.* 2013a; 40:607–615. [PubMed: 24377867]
- Guo Y, Li J, Li CI, Long J, Samuels DC, Shyr Y. The effect of strand bias in Illumina short-read sequencing data. *BMC genomics.* 2012b; 13:666. [PubMed: 23176052]
- Guo Y, Li J, Li CI, Shyr Y, Samuels DC. MitoSeek: Extracting Mitochondria Information and Performing High Throughput Mitochondria Sequencing Analysis. *Bioinformatics.* 2013b
- Guo Y, Long J, He J, Li CI, Cai Q, Shu XO, Zheng W, Li C. Exome sequencing generates high quality data in non-target regions. *BMC genomics.* 2012c; 13:194. [PubMed: 22607156]
- Guo Y, Samuels DC, Li J, Clark T, Li CI, Shyr Y. Evaluation of allele frequency estimation using pooled sequencing data simulation. *ScientificWorldJournal.* 2013c; 2013:895496. [PubMed: 23476151]
- Guo Y, Ye F, Sheng Q, Clark T, Samuels DC. Three-stage quality control strategies for DNA re-sequencing data. *Briefings in Bioinformatics.* 2013d
- Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.* 2010; 6:e1000834. [PubMed: 20168995]
- He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, Diaz LA Jr, Kinzler KW, Vogelstein B, Papadopoulos N. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature.* 2010; 464:610–614. [PubMed: 20200521]
- Herrmann PC, Gillespie JW, Charboneau L, Bichsel VE, Paweletz CP, Calvert VS, Kohn EC, Emmert-Buck MR, Liotta LA, Petricoin EF 3rd. Mitochondrial proteome: altered cytochrome c oxidase subunit levels in prostate cancer. *Proteomics.* 2003; 3:1801–1810. [PubMed: 12973739]
- Holt IJ, Harding AE, Petty RK, Morgan-Hughes JA. A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. *Am J Hum Genet.* 1990; 46:428–433. [PubMed: 2137962]

- Kaiser-Wilhelm-Institut für Biologie, B-D.; Warburg, OH. The metabolism of tumours : investigations from the Kaiser Wilhelm Institute for Biology. London: Berlin-Dahlem. Constable and Co.; 1930.
- Kann LM, Rosenblum EB, Rand DM. Aging, mating, and the evolution of mtDNA heteroplasmy in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95:2372–2377. [PubMed: 9482892]
- Katada S, Mito T, Ogasawara E, Hayashi J, Nakada K. Mitochondrial DNA with a large-scale deletion causes two distinct mitochondrial disease phenotypes in mice. *G3 (Bethesda)*. 2013; 3:1545–1552. [PubMed: 23853091]
- Kroemer G. Mitochondria in cancer. *Oncogene*. 2006; 25:4630–4632. [PubMed: 16892077]
- Lam ET, Bracci PM, Holly EA, Chu C, Poon A, Wan E, White K, Kwok PY, Pawlikowska L, Tranah GJ. Mitochondrial DNA sequence variation and risk of pancreatic cancer. *Cancer Res*. 2012; 72:686–695. [PubMed: 22174369]
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9:357–359. [PubMed: 22388286]
- Larman TC, Depalma SR, Hadjipanayis AG, Protopopov A, Zhang J, Gabriel SB, Chin L, Seidman CE, Kucherlapati R, Seidman JG. Spectrum of somatic mitochondrial mutations in five cancers. *Proc Natl Acad Sci U S A*. 2012; 109:14087–14091. [PubMed: 22891333]
- Lee HC, Lu CY, Fahn HJ, Wei YH. Aging- and smoking-associated alteration in the relative content of mitochondrial DNA in human lung. *FEBS Lett*. 1998; 441:292–296. [PubMed: 9883902]
- Lemasters JJ, Qian T, Bradham CA, Brenner DA, Cascio WE, Trost LC, Nishimura Y, Nieminen AL, Herman B. Mitochondrial dysfunction in the pathogenesis of necrotic and apoptotic cell death. *J Bioenerg Biomembr*. 1999; 31:305–319. [PubMed: 10665521]
- Lewis PD, Baxter P, Paul Griffiths A, Parry JM, Skibinski DO. Detection of damage to the mitochondrial genome in the oncocytic cells of Warthin's tumour. *J Pathol*. 2000; 191:274–281. [PubMed: 10878549]
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. [PubMed: 19451168]
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–2079. [PubMed: 19505943]
- Li M, Schonberg A, Schaefer M, Schroeder R, Nasidze I, Stoneking M. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *American journal of human genetics*. 2010; 87:237–249. [PubMed: 20696290]
- Li M, Schroeder R, Ko A, Stoneking M. Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic acids research*. 2012; 40:e137. [PubMed: 22649055]
- Liang MH, Johnson DR, Wong LJ. Preparation and validation of PCR-generated positive controls for diagnostic dot blotting. *Clin Chem*. 1998; 44:1578–1579. [PubMed: 9665449]
- Lundholm K, Edstrom S, Karlberg I, Ekman L, Schersten T. Glucose turnover, gluconeogenesis from glycerol, and estimation of net glucose cycling in cancer patients. *Cancer*. 1982; 50:1142–1150. [PubMed: 7104955]
- Maximo V, Soares P, Lima J, Cameselle-Teijeiro J, Sobrinho-Simoes M. Mitochondrial DNA somatic mutations (point mutations and large deletions) and mitochondrial DNA variants in human thyroid pathology: a study with emphasis on Hurthle cell tumors. *Am J Pathol*. 2002; 160:1857–1865. [PubMed: 12000737]
- Maximo V, Soares P, Seruca R, Rocha AS, Castro P, Sobrinho-Simoes M. Microsatellite instability, mitochondrial DNA large deletions, and mitochondrial DNA mutations in gastric carcinoma. *Genes Chromosomes Cancer*. 2001; 32:136–143. [PubMed: 11550281]
- Maximo V, Soares P, Seruca R, Sobrinho-Simoes M. Comments on: mutations in mitochondrial control region DNA in gastric tumours of Japanese patients, Tamura, et al. *Eur J Cancer*. 1999; 35:1407–1408. *Eur J Cancer* 1999, 35, 316–319. [PubMed: 10658535]
- Mazurek S, Boschek CB, Eigenbrodt E. The role of phosphometabolites in cell proliferation, energy metabolism, and tumor therapy. *J Bioenerg Biomembr*. 1997; 29:315–330. [PubMed: 9387092]

- Modica-Napolitano JS, Singh KK. Mitochondrial dysfunction in cancer. *Mitochondrion*. 2004; 4:755–762. [PubMed: 16120430]
- Mourier T, Hansen AJ, Willerslev E, Arctander P. The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Mol Biol Evol*. 2001; 18:1833–1837. [PubMed: 11504863]
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*. 2010; 42:30–35. [PubMed: 19915526]
- Ockner RK, Kaikaus RM, Bass NM. Fatty-acid metabolism and the pathogenesis of hepatocellular carcinoma: review and hypothesis. *Hepatology*. 1993; 18:669–676. [PubMed: 8395460]
- Park JS, Sharma LK, Li H, Xiang R, Holstein D, Wu J, Lechleiter J, Naylor SL, Deng JJ, Lu J, Bai Y. A heteroplasmic, not homoplasmic, mitochondrial DNA mutation promotes tumorigenesis via alteration in reactive oxygen species generation and apoptosis. *Hum Mol Genet*. 2009; 18:1578–1589. [PubMed: 19208652]
- Payne BA, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, Taylor RW, Samuels DC, Santibanez-Koref M, Chinnery PF. Universal heteroplasmy of human mitochondrial DNA. *Hum Mol Genet*. 2013; 22:384–390. [PubMed: 23077218]
- Pereira L, Freitas F, Fernandes V, Pereira JB, Costa MD, Costa S, Maximo V, Macaulay V, Rocha R, Samuels DC. The diversity present in 5140 human mitochondrial genomes. *Am J Hum Genet*. 2009; 84:628–640. [PubMed: 19426953]
- Petrosillo G, Di Venosa N, Ruggiero FM, Pistolese M, D'Agostino D, Tiravanti E, Fiore T, Paradies G. Mitochondrial dysfunction associated with cardiac ischemia/reperfusion can be attenuated by oxygen tension control. Role of oxygen-free radicals and cardiolipin. *Biochim Biophys Acta*. 2005; 1710:78–86. [PubMed: 16325647]
- Picardi E, Pesole G. Mitochondrial genomes gleaned from human whole-exome sequencing. *Nature methods*. 2012; 9:523–524. [PubMed: 22669646]
- Polyak K, Li Y, Zhu H, Lengauer C, Willson JK, Markowitz SD, Trush MA, Kinzler KW, Vogelstein B. Somatic mutations of the mitochondrial genome in human colorectal tumours. *Nat Genet*. 1998; 20:291–293. [PubMed: 9806551]
- Putignani L, Raffa S, Pescosolido R, Rizza T, Del Chierico F, Leone L, Aimati L, Signore F, Carrozzo R, Callea F, Torrisi MR, Grammatico P. Preliminary evidences on mitochondrial injury and impaired oxidative metabolism in breast cancer. *Mitochondrion*. 2012; 12:363–369. [PubMed: 22366096]
- Quispe-Tintaya W, White RR, Popov VN, Vijg J, Maslov AY. Fast mitochondrial DNA isolation from mammalian cells for next-generation sequencing. *Biotechniques*. 2013; 55:133–136. [PubMed: 24003945]
- Renis M, Cantatore P, Loguercio Polosa P, Fracasso F, Gadaleta MN. Content of mitochondrial DNA and of three mitochondrial RNAs in developing and adult rat cerebellum. *J Neurochem*. 1989; 52:750–754. [PubMed: 2465379]
- Robin ED, Wong R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *Journal of cellular physiology*. 1988; 136:507–513. [PubMed: 3170646]
- Samuels DC, Han L, Li J, Quangu S, Clark TA, Shyr Y, Guo Y. Finding the lost treasures in exome sequencing data. *Trends Genet*. 2013
- Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. *Proceedings of the National Academy of Sciences*. 2012; 109:14508–14513.
- Shen J, Platek M, Mahasneh A, Ambrosone CB, Zhao H. Mitochondrial copy number and risk of breast cancer: a pilot study. *Mitochondrion*. 2010; 10:62–68. [PubMed: 19788937]
- Smigrodzki RM, Khan SM. Mitochondrial microheteroplasmy and a theory of aging and age-related disease. *Rejuvenation research*. 2005; 8:172–198. [PubMed: 16144471]
- Sondheimer N, Glatz CE, Tirone JE, Deardorff MA, Krieger AM, Hakonarson H. Neutral mitochondrial heteroplasmy and the influence of aging. *Human Molecular Genetics*. 2011; 20:1653–1659. [PubMed: 21296868]

- Sosa MX, Sivakumar IK, Maragh S, Veeramachaneni V, Hariharan R, Parulekar M, Fredrikson KM, Harkins TT, Lin J, Feldman AB, Tata P, Ehret GB, Chakravarti A. Next-generation sequencing of human mitochondrial reference genomes uncovers high heteroplasmy frequency. *PLoS Comput Biol*. 2012; 8:e1002737. [PubMed: 23133345]
- Sun W, Zhou S, Chang SS, McFate T, Verma A, Califano JA. Mitochondrial mutations contribute to HIF1alpha accumulation via increased reactive oxygen species and up-regulated pyruvate dehydrogenase kinase 2 in head and neck squamous cell carcinoma. *Clin Cancer Res*. 2009; 15:476–484. [PubMed: 19147752]
- Tang S, Huang T. Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system. *Biotechniques*. 2010; 48:287–296. [PubMed: 20569205]
- Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet*. 2004; 5:123–135. [PubMed: 14735123]
- Tseng LM, Yin PH, Chi CW, Hsu CY, Wu CW, Lee LM, Wei YH, Lee HC. Mitochondrial DNA mutations and mitochondrial DNA depletion in breast cancer. *Genes Chromosomes Cancer*. 2006; 45:629–638. [PubMed: 16568452]
- van der Walt EM, Smuts I, Taylor RW, Elson JL, Turnbull DM, Louw R, van der Westhuizen FH. Characterization of mtDNA variation in a cohort of South African paediatric patients with mitochondrial disease. *Eur J Hum Genet*. 2012; 20:650–656. [PubMed: 22258525]
- Vanharanta S, Buchta M, McWhinney SR, Virta SK, Peczkowska M, Morrison CD, Lehtonen R, Januszewicz A, Jarvinen H, Juhola M, Mecklin JP, Pukkala E, Herva R, Kiuru M, Nupponen NN, Aaltonen LA, Neumann HP, Eng C. Early-onset renal cell carcinoma as a novel extraparaganglial component of SDHB-associated heritable paraganglioma. *Am J Hum Genet*. 2004; 74:153–159. [PubMed: 14685938]
- Vasta V, Ng SB, Turner EH, Shendure J, Hahn SH. Next generation sequence analysis for mitochondrial disorders. *Genome Medicine*. 2009; 1
- Verma M, Naviaux RK, Tanaka M, Kumar D, Franceschi C, Singh KK. Meeting report: mitochondrial DNA and cancer epidemiology. *Cancer Res*. 2007; 67:437–439. [PubMed: 17213255]
- Wallace KB, Starkov AA. Mitochondrial targets of drug toxicity. *Annu Rev Pharmacol Toxicol*. 2000; 40:353–388. [PubMed: 10836141]
- White HE, Durston VJ, Sellar A, Fratter C, Harvey JF, Cross NC. Accurate detection and quantitation of heteroplasmic mitochondrial point mutations by pyrosequencing. *Genet Test*. 2005; 9:190–199. [PubMed: 16225398]
- Yang Ai SS, Hsu K, Herbert C, Cheng Z, Hunt J, Lewis CR, Thomas PS. Mitochondrial DNA mutations in exhaled breath condensate of patients with lung cancer. *Respir Med*. 2013; 107:911–918. [PubMed: 23507584]
- Yang Z, Nielsen R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol*. 1998; 46:409–418. [PubMed: 9541535]
- Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009; 25:2865–2871. [PubMed: 19561018]
- Yu M, Zhou Y, Shi Y, Ning L, Yang Y, Wei X, Zhang N, Hao X, Niu R. Reduced mitochondrial DNA copy number is correlated with tumor progression and prognosis in Chinese breast cancer patients. *IUBMB Life*. 2007; 59:450–457. [PubMed: 17654121]
- Zhang DX, Hewitt GM. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends Ecol Evol*. 1996; 11:247–251. [PubMed: 21237827]
- Zhang W, Cui H, Wong LJ. Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing. *Clin Chem*. 2012; 58:1322–1331. [PubMed: 22777720]
- Zhu W, Qin W, Sauter ER. Large-scale mitochondrial DNA deletion mutations and nuclear genome instability in human breast cancer. *Cancer Detect Prev*. 2004; 28:119–126. [PubMed: 15068836]

Highlights

- High throughput sequencing technology allows researchers to conduct mitochondrial research at unprecedented details.
- Very low level heteroplasmy can be detected through high throughput sequencing
- mtDNA sequences can be extracted from exome sequencing data
- Quality control of high throughput sequencing mtDNA data is important to ensure study integrity