Replication Errors Made During Oogenesis Lead to Detectable De Novo mtDNA Mutations in Zebrafish Oocytes with a Low…

Article *in* Genetics · October 2016

DOI: 10.1534/genetics.116.194035

12 authors, including:

Mike Gerards
Maastricht University
45 PUBLICATIONS **473** CITATIONS
SEE PROFILE

Sabina Vanherle
Maastricht University
6 PUBLICATIONS **227** CITATIONS
SEE PROFILE

Jo Vanoevelen
Maastricht Universitair Medisch Centrum
38 PUBLICATIONS **1,132** CITATIONS
SEE PROFILE

Marc Muller
University of Liège
173 PUBLICATIONS **2,804** CITATIONS
SEE PROFILE

Genetics: Early Online, published on October 21, 2016 as 10.1534/genetics.116.194035

Replication errors made during oogenesis lead to detectable *de novo* mtDNA mutations in zebrafish oocytes with a low mtDNA copy number

Auke B.C. Otten*, Alphons P.M. Stassen*, Michiel Adriaens†, Mike Gerards†, Richard G.J. Dohmen*, Adriana J. Timmer*, Sabina J.V. Vanherle*, Rick Kamps*, Iris B.W. Boesten*, Jo M. Vanoevelen*, Marc Muller‡, Hubert J.M. Smeets*,†

*Department of Genetics & Cell Biology, Unit Clinical Genomics, School for Oncology and Developmental Biology (GROW), Maastricht University, 6200MD, Maastricht, the Netherlands

†Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, 6200MD, the Netherlands

‡Laboratory of Organogenesis and Regeneration, GIGA-Resarch, Univérsité de Liège, 4000 Liège, Belgium

Running title: Prevalence of mitochondrial DNA mutations in zebrafish oocytes

Keywords: Mitochondrial DNA, de novo mutations, Next generation sequencing, zebrafish, oogenesis

Corresponding Author:

Hubert J.M. Smeets (HS)

Postbus 616 (box 16)

6200 MD Maastricht

Abstract

Of all pathogenic mitochondrial DNA (mtDNA) mutations in humans, ~25% is *de novo*, although the occurrence in oocytes has never been directly assessed. We used next generation sequencing to detect point mutations directly in the mtDNA of 3-15 individual mature oocytes and three somatic tissues from eight zebrafish females. Various statistical and biological filters allowed reliable detection of *de novo* variants with heteroplasmy ≥1.5%. In total, we detected 38 *de novo* base substitutions, but no insertions or deletions. These 38 *de novo* mutations were present in 19 of 103 mature oocytes, indicating that ~20% of the mature oocytes carry at least one *de novo* mutation with heteroplasmy ≥1.5%. This frequency of *de novo* mutations is close to that deducted from the reported error rate of polymerase gamma, the mitochondrial replication enzyme, implying that mtDNA replication errors made during oogenesis are a likely explanation. Substantial variation in the mutation prevalence among mature oocytes can be explained by the highly variable mtDNA copy number, since we previously reported that ~20% of the primordial germ cells have a mtDNA copy number of ≤73 and would lead to detectable mutation loads. In conclusion, replication errors made during oogenesis are an important source of *de novo* mtDNA base substitutions and their location and heteroplasmy level determine their significance.

Article Summary

Approximately 25% of the mitochondrial DNA (mtDNA) mutations are thought to occur *de novo*. To detect *de novo* mutations, we screened the mitochondrial DNA of 103 zebrafish oocytes. We detected 38 *de novo* base substitutions in ~20% of the zebrafish oocytes tested. The mutation frequency is similar to the reported error rate

of the mitochondrial replication enzyme polymerase gamma, suggesting replication errors made during oogenesis are an important source of *de novo* mtDNA mutations. Variation in the mtDNA copy number among oocytes explains the high variation in the prevalence of *de novo* mutations.

Introduction

Comparative sequence analysis of the mitochondrial DNA (mtDNA) has revealed a high degree of variability, much higher than its nuclear counterpart (LYNCH *et al.* 2006). This is generally explained by limited recombination and recombination-mediated mtDNA repair events to counteract errors made during mtDNA replication (BARR *et al.* 2005) and the close proximity of the unprotected mtDNA to the oxidative phosphorylation (OXPHOS) machinery, which produces (potentially) mutagenic reactive oxygen species (ROS) (BRAND 2010). As a result, mtDNA mutations are an important cause of a group of devastating inherited diseases (TAYLOR and TURNBULL 2005). To date, over 150 pathogenic mtDNA mutations have been identified, as well as many more polymorphisms of unknown significance (HELLEBREKERS *et al.* 2012). As human and animal cells have a high mtDNA copy number, wild-type and variant mtDNA genotypes can co-exist, a state referred to as heteroplasmy. The high mtDNA copy number compensates low-level pathogenic mtDNA mutations and avoids disease manifestation. A pathogenic mtDNA mutation will only manifest if its heteroplasmy value exceeds a certain threshold (HELLEBREKERS *et al.* 2012). The mtDNA inherits maternally, and a female carrying an mtDNA mutation can transmit this mutation to her offspring through the mtDNA of her oocytes.

Maternal inheritance of pre-existing mtDNA mutation does not explain all patients suffering from mtDNA mutations. In 25% of these patients, the disease-

causing mutation cannot be detected in the maternal mtDNA (SALLEVELT *et al.* 2016). Although we cannot exclude some of these mutation were present at undetectable low-level heteroplasmy levels in the maternal mtDNA, this suggests that most of these mutations have occurred *de novo* during germ line development. The inheritance of the mtDNA occurs through a segregational bottleneck: only a limited number of the mtDNA molecules, the so-called bottleneck size, from the oocyte are transmitted to the (primordial) germline cells (PGCs) of the next generation (CREE *et al.* 2008). We hypothesize that, in case of a constant mutation rate, a low mtDNA copy number at the bottom of the bottleneck could lead to *de novo* mutations reaching detectable heteroplasmy levels, which, dependent on the nature of the mutations, can be of functional and/or pathogenic significance. Experimental evidence for this hypothesis has long been difficult to obtain, because of the low heteroplasmy level at which these mutations occur, which were generally below the detection level of conventional sequencing techniques. Over the past decade, next-generation sequencing (NGS) technologies have been developed, allowing an in depth determination of the mutations and heteroplasmy levels directly in the mtDNA of individual oocytes.

An estimation of the prevalence of *de novo* mtDNA mutation requires a significant number of oocytes to be sequenced. For both biological and ethical reasons this is difficult to achieve in humans and most animals. In contrast, oocyte collection from zebrafish is relatively easy and efficient and therefore we used zebrafish mature oocytes to assess the *de novo* mtDNA mutation risk. We characterized *de novo* mutations, their location and their heteroplasmy level in 103 oocytes and three somatic (maternal) tissues from eight different female zebrafish using NGS with a minimal coverage of 1,700. Furthermore, for all oocytes with one or

more *de novo* mutation(s), we estimated the mtDNA copy number at which these mutations arose, based on their heteroplasmy levels. Given the high mtDNA sequence similarity between humans and zebrafish (BROUGHTON *et al.* 2001) and the conservation of the mtDNA bottleneck within the animal kingdom (HOWELL *et al.* 1992; CREE *et al.* 2008; WOLFF *et al.* 2011; LEE *et al.* 2012), our findings have insinuations for the occurrence of *de novo* mtDNA disease in humans.

Materials and Methods

Zebrafish maintenance and sample collection

Wild-type female zebrafish from the *AB* strain were used. Raising and housing was according to standard procedures at 28˚C (KIMMEL *et al.* 1995) in the zebrafish facility of Liège University, where local ethical approval by the committee of Animal Research was obtained. Mature unfertilized oocytes were collected by squeezing the abdomen of anaesthetized females. The oocytes used were normal in morphological appearance. After oocyte collection, the female fish were sacrificed in ice-cold water, after which biopsies of brain, liver and muscle were obtained. An overview of the 127 samples is given in Table 1.

Table 1. Overview of samples collected

Zebrafish #	1	2	3	4	5	6	7	8
Biopsies	3	3	3	3	3	3	3	3
Number of oocytes	3	13	15	14	15	14	15	14

Isolation procedure of mtDNA

Oocytes were collected in sterile tubes and lysed for 4 hours at 50˚C in 500 µl DNA lysis buffer containing 75 mM NaCl, 50 mM EDTA, 20 mM HEPES, 0.4% SDS and

200 µg proteinase K (Sigma). Subsequently, isopropanol was added and samples were precipitated overnight at -20˚C. After thorough centrifugation, the DNA pellet was washed with 70% ethanol and dissolved in TE buffer. The biopsies of brain, liver and muscle from the adult female fish were collected in sterile tubes containing Nuclei Lysis solution from the Wizard Genomic Purification Kit (Promega). Subsequently, mtDNA was extracted according to manufacturer's instructions and dissolved in FG3-buffer.

mtDNA amplification and sequencing

The mtDNA (reference NCBI: NC002333.2) was amplified in three ~5.6kb amplicons (A-C). Fragment A (Forward: CACACCCCTGACTCCCAAAG, Reverse: GGTCGTTTGTACCCGTCAGT) amplified a target spanning nucleotide 16,594 (gene: *trna-pro*) to 5,952 (*nd2*), fragment B (Forward: AAATTAACACCCTAACAACGACCTG, Reverse: GGGGATCAGTACTTTTAGCATTGTAGT) an amplicon from nucleotide 5,669 (*nd2*) to 11,319 (*nd4*) and fragment C amplified the mtDNA from nucleotide 11,170 (*nd4l*) to 295 (*D-loop*). Primers (designed with Primer3) were specific for the mtDNA to avoid the amplification of nuclear-encoded mitochondrial pseudogenes. PCR amplification was performed using Phusion Hot Start II DNA polymerase in GC-buffer (ThermoScientific): 30s at 98˚C, followed by 40 cycles of 10s at 98˚C (denaturation), 20s at 58˚C (annealing) and 8 min at 72˚C (extension), with a final step for 10 min at 72˚C. The PCR product was checked using electrophoresis on a 1% agarose gel containing ethidium bromide, allowing also the detection of large deletions. Amplicons were purified using the Agencourt AMPure XP system (Beckman-Coulter), according to the protocol of the manufacturer. Subsequently, the three purified amplicons were (equimolar) mixed and processed using the customized Nextera XT protocol (McElhoe *et al.* 2014). The library of a random

subset of the samples was analyzed using a Bioanalyzer 2100 High Sensitivity DNA chip (Agilent Technologies) to confirm quantity and size of the library. Libraries were indexed and 18 libraries were pooled per lane and analyzed on the HiSeq 2000 system (Illumina), using a read length of 1,000 base pairs. PhiX (1%) was spiked in every lane as an internal control.

Pre-processing of Next-Generation sequencing data

As the mtDNA is a multicopy genome (resulting in many biological duplicates), duplicate reads were included in the analysis. Demultiplexing of the data was performed using Illumina CASAVA software (v.1.8.2) and reads were aligned against the mitochondrial reference sequence for the zebrafish (NCBI: NC002333.2) using Burrows-Wheeler Alligner (BWA) software (v.0.5.9) (LI and DURBIN 2010). For variant calling, we used Phyton 2.6.6., Phyton Package pysam 0.7.8 and SAMTools 0.1.19. In-house built Perl tools were used to process the variants. As the prevalence of any of the four nucleotides per position was counted to call a variant, the heteroplasmy value was calculated as the ratio of one of the nucleotides over the coverage, which was defined as the total count of any nucleotide at a certain position.

Identification of heteroplasmic *de novo* point mutations

A statistical algorithm was developed to distinguish variant calls from the noise signal, as well as to determine whether point mutations, either a single base substitution or a small insertion of deletion (indel), reported in an oocyte was absent in the corresponding female fish, and thus arose *de novo*. A call from a sample is included in the analysis if its coverage is above the threshold, which is determined by calculating the median coverage for every position of the mtDNA genome, based on the coverage data of all 127 samples (Figure 1). Assuming sequencing quality is independent of the nucleotide position, the median value is the most robust estimate

of the coverage across the entire mtDNA genome. This implies that the lowest median coverage value is a robust estimate of the minimal reliable coverage and of the maximum background (noise) signal. As the lowest median coverage value was ~1,700 (Figure 1A), this was chosen as a cut-off value for a position to be included in the analysis, preventing variants with lower coverage from influencing the statistical calculations we applied to exclude false positives.

For base substitutions, we discriminated false positives in the higher coverage group from true variants by comparing the percentage heteroplasmy of the variant nucleotide of a particular sample with the average percentage heteroplasmy of all other samples (female tissues and oocytes, but without littermate oocytes and tissues from the mother). To this end, for each substitution a probability distribution of the heteroplasmy values was generated (using all samples) and transformed to a Gaussian distribution using 'a rank-transformation'. We assumed this distribution to be a representation of the noise signal for this variant, which is inherent to the NGS procedure (GUO *et al.* 2013). Hence, we calculated a *z*-score and *P*-value (one tailed) and considered substitutions with a *P*-value ≤0.01 as true variants, being statistically different from the noise signal. The check for *de novo* substitutions was only performed for variants for which the coverage of the oocyte and of ≥2 tissues of its mother was ≥1,700. Finally, a substitution was assumed *de novo* if the *P*-value of this variant in the oocyte was ≤0.01, while in the tissues from the corresponding female the *P*-value was >0.01 (= absent or in noise signal) and if the heteroplasmy percentage was ≥1.0%.

As the sensitivity to identify small indels is at least 2-fold lower as for single nucleotide substitutions (KRAWITZ *et al.* 2010; NEUMAN *et al.* 2013; SENECA *et al.* 2015), we only considered *de novo* indels with a coverage ≥3,400x. All suspected *de*

novo indels were inspected manually, as the software tools used are known be less reliable than for the substitutions, using the Integrative Genomics Viewer (IGV) (THORVALDSDOTTIR *et al.* 2013).

Estimation of mtDNA copy number and mutation rate

From the average heteroplasmy value (het-%) of all *de novo* variants within one mature oocyte, we estimated the mtDNA copy number at the time the mutation occurred. Under the assumption that heteroplasmy levels remain stable during oogenesis, as observed in mice (JENUTH *et al.* 1996) and during stem cell culturing (YAMADA *et al.* 2016) and only one copy is mutated, the mtDNA copy number (z) at the time the mutation occurred can be calculated using: $z = 100/\text{het-\%}$.Subsequently, we aimed to estimate the mutation rate for mutations ≥1.5%. To do this, we first calculated for every oocyte the total number of detectable nucleotides (y) that were present at the time the mutation arose. Therefore, we multiplied the estimated mtDNA copy number (z) with the number of nucleotides that had coverage >1,700: $y = z$ * [#nucleotides >1,700 coverage]. The mutation rate (x) was then calculated by dividing the number of *de novo* mutations in an oocyte by the total number of analyzable nucleotides: x = [#mutations detected] / y.

Data Availability

The authors state that all detected *de novo* mutations resulting from our statistical analysis are presented within the article.

Results

We studied the occurrence of detectable *de novo* mutations in 103 mature oocytes derived from eight different female zebrafish. No large deletions were observed, but we did detect significant numbers of point mutations. Of the 1,592,714 different

nucleotide positions that had a sequence-coverage >1,700, we identified 2,624 different base substitutions, which were statistically different from the sequencing noise signal and absent in all three maternal tissues and therefore considered as potential *de novo* mutations. To increase the reliability of our *de novo* base substitution detection, we applied three additional biological filter steps (Figure 1). 1) All oocyte variants for which the heteroplasmy level in one of the corresponding maternal tissues was ≥1.0% were considered to be pre-existing and were therefore excluded. 2) The heteroplasmy level of the variants was corrected for the noise signal by subtracting the maternal (M) heteroplasmy level (=the average of the heteroplasmy level in the three maternal tissues) from the oocyte (O) heteroplasmy level (O-M) and variants with small O-M values (<1.5%) were excluded (205 variants had an O-M value of 1.0-1.5% and 929 variants of <1.0%). This filter restricted our analysis to *de novo* point mutations with a mutation load ≥1.5%. 3) A *de novo* variant was rejected if, within the same oocyte, another variant was detected in the two nucleotides adjacent (upstream or downstream) to this variant, as variants located in close proximity to each other are most likely the result of alignment artefacts (LI and DURBIN 2009). By applying these filters, 38 single base substitutions were considered to be *de novo* (Table 2), of which 36 were unique, as two mutations (m.7247T>G and m.10578C>A) were found in two oocytes isolated from the same zebrafish.

We also checked for the presence of small *de novo* indels. We did identify three indels with a coverage >3,400: one in oocyte 2.1 (m.7352insC, O-M het: 1.63%, coverage 5,868), one in oocyte 5.10 (m.12323del, O-M het: 1.58%, coverage 3,848) and in oocyte 7.15 (m.11305del, O-M het: 2.41%, coverage 4,689). Using IGV we found that the insertion was only detected in ambiguous reads, while both deletions

were only detected in duplicate reads. Therefore, these indels were most likely alignment artifacts and excluded from further analysis.

The 38 *de novo* mutations were detected in only 19 different oocytes (Table 2), indicating that the majority (82%) did not have a detectable *de novo* mutation. In three oocytes (5.2, 5.10 and 7.9) more than three *de novo* mutations were found. No *de novo* mutations were detected in any oocyte from zebrafish 6. The heteroplasmy level of all the *de novo* point mutations ranged from 1.5% to 9.0% with an average of 2.7%.

Table 2. Heteroplasmic mtDNA base substitution mutations in all oocytes

Oocyte ID	Corr. Het-% (O-M)	Gene location	Nucleotide change	Am. Acid change
1.1	2.5%	*ND1*	m.4164C>A	Ser>STOP
2.3	1.8%	*COI*	m.6489T>C	Phe>Ser
2.6	5.4%	*D-loop*	m.250C>T	-
2.6	5.9%	*COII*	m.8700G>A	Val>Lys
2.10	1.6%	*COI*	m.6510T>A	Val>Glu
3.2	3.3%	*D-loop*	m.283G>T	-
3.2	1.9%	*ND5*	m14296T>A	Ile>Asn
3.9	4.1%	*tRNA-Trp*	m.6089G>A	-
3.15	1.9%	*ND4L/ ND4*	m.11303T>A	Stop>Lys Leu>Gln
4.4	3.1%	*D-loop*	m.1818A>T	-
4.10	1.9%	*D-loop*	m.532G>A	-
4.10	2.1%	*ND4*	m.11708G>A	Trp>STOP
5.2	2.7%	*ND1*	m.4077G>A	Trp>STOP
5.2	2.2%	*COI*	m.7112C>T	Leu>Phe
5.2	1.6%	*COI*	m.7247T>G	Trp>Gly
5.2	2.2%	*COI*	m.7574G>T	Gly>Trp
5.2	2.2%	*COI*	m.7580G>A	Val>Met
5.3	1.8%	*ND6*	m.14894T>A	Leu>Phe
5.10	1.6%	*12s rRNA*	m.1220A>G	-
5.10	2.4%	*COI*	m.7112C>T	Leu>Phe
5.10	2.1%	*COI*	m.7247T>G	Trp>Gly
5.10	1.9%	*COI*	m.7253A>G	Met>Val
5.10	1.8%	*COIII*	m.9909C>A	Arg>STOP
5.10	1.6%	*ND4*	m.11500G>C	Val>Leu
5.10	1.5%	*tRNA-Leu*	m.12838T>G	-
5.10	1.6%	*ND5*	m.13472C>T	Tyr>Tyr
5.13	8.2%	*ND6*	m.14761G>A	Leu>Leu

7.3	1.5%	*16s rRNA*	m.2632A>G	-
7.8	1.7%	*16s rRNA*	m.2537A>G	-
7.9	2.5%	*12s rRNA*	m.1550G>C	-
7.9	1.8%	*ND4*	m.12263T>G	Leu>Arg
7.9	1.6%	*ND5*	m.13205T>G	Phe>Leu
7.9	1.6%	*CytB*	m.16232A>C	Thr>Pro
7.9	1.5%	*CytB*	m.16324A>T	Gly>Gly
8.5	2.8%	*ND5*	m.14400C>T	Leu>Leu
8.6	9.0%	*tRNA-Gly*	m.10578C>A	-
8.6	1.5%	*ND4*	m.12464G>T	Trp>Leu
8.10	4.6%	*tRNA-Gly*	m.10578C>A	-

Corr. Het-%, Oocyte heteroplasmy value corrected from noise signal by subtracting the average heteroplasmy of the corresponding female tissues from the heteroplasmy value detected in the oocyte. Am. Acid change, change in amino acid due to change in the codon sequence.

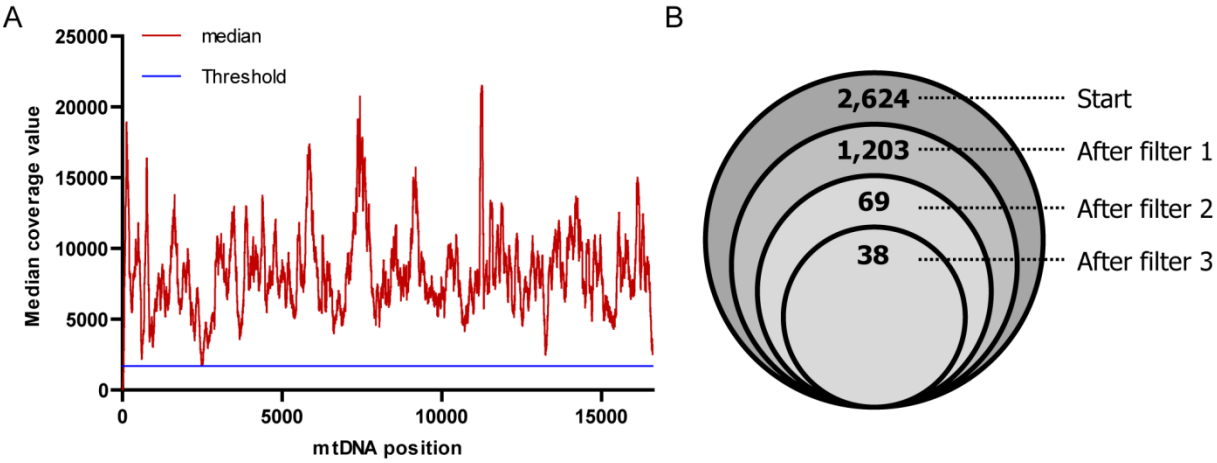

Figure 1. Median reads per position and filtering steps used to determine *de novo* mutations. A) Overview of the median reads per position of the mtDNA genome (red line) across all samples. The blue line indicates the threshold value for the coverage value to include samples. B) Results of filtering steps performed to determine *de novo* mutations. All reads with coverage above 1,700x (1,592,714 nucleotides) were statistically analyzed, resulting in 2,624 variants with a *P*-value ≤ 0.01. Subsequent filter steps performed: 1) prevalence tissues adult female zebrafish <1.0%, 2) heteroplasmy difference between tissues from adult zebrafish and her oocytes (O-M) ≥ 1.5% and 3) mutations not detected in the two nucleotides adjacent (upstream or downstream) of another variant. After filtering, 38 *de novo* mutations remained.

For every mature oocyte with at least one *de novo* mutation, we estimated the mtDNA copy number present at the time the mutation occurred. This number ranged

290 from 18 to 67 (Table 3). Based on this, we calculated the mutation rate for these
291 oocytes (Table 3). For most oocytes this mutation rate was in the range of 10^{-6}
292 mutations per nucleotide, while oocytes 4.4, 5.10 and 7.3 had a higher mutation rate.
293 On average, the mutation rate in these 19 oocytes was $4.3x10^{-6}$ mutations per
294 nucleotide. Strikingly, for oocyte 8.6 the heteroplasmy values of the two reported *de*
295 *novo* mutations greatly differed from each other, resulting in two different estimations
296 of the mutation rate.
297
298 **Table 3. mtDNA copy number and *de novo* mutation rate ≥1.5% at a given**
299 **nucleotide for all oocytes in which at least one *de novo* mutation was detected**

Oocyte ID	# *de novo* mutations	Average Het-% of *de novo* mutation	Copy number from het-%	Positions analyzed (coverage >1700)	Analyzable nucleotides when mutation occurred	Mutation rate (>1.5%) per nucleotide
1.1	1	2.5	40	05,643	225,720	$4.4x10^{-6}$
2.3	1	1.8	56	16,532	925,792	$1.1x10^{-6}$
2.6	2	5.7	18	16,580	298,440	$6.7x10^{-6}$
2.10	1	1.6	63	14,543	916,209	$1.1x10^{-6}$
3.2	2	1.9	53	16,532	876,196	$2.3x10^{-6}$
3.9	1	4.1	24	16,555	397,320	$2.5x10^{-6}$
3.15	1	1.9	53	14,902	789,806	$1.3x10^{-6}$
4.4	1	3.1	32	02,816	90,112	$1.1x10^{-5}$
4.10	2	2.0	50	16,583	829,150	$2.4x10^{-6}$
5.2	5	2.2	45	06,991	314,595	$1.6x10^{-6}$
5.3	1	1.8	56	12,505	700,280	$1.4x10^{-6}$
5.10	8	1.7	59	10,740	633,660	$1.3x10^{-5}$
5.13	1	8.2	12	16,570	198,840	$5.0x10^{-6}$
7.3	1	1.5	67	16,528	1,107,376	$9.0x10^{-7}$
7.8	1	1.7	59	16,569	977,571	$1.0x10^{-6}$
7.9	5	1.8	56	16,426	919,856	$5.4x10^{-6}$
8.5	1	2.8	36	16,569	596,484	$1.7x10^{-6}$
8.6*	1	9.0	11	16,239	178,629	$5.6x10^{-6}$
8.6*	1	1.5	67	16,239	1,088,013	$9.2x10^{-7}$
8.10	1	4.6	22	16,533	363,726	$2.7x10^{-6}$
Average	2	3.2	56	-	-	$4.3x10^{-6}$

de novo mutations, number of *de novo* mutations detected in the oocyte. Average Het-% of *de novo* mut., the average heteroplasmy value for all *de novo* mutations found in the oocyte. Copy number from het-%, the copy number the oocytes had when the mutation occurred. *the heteroplasmy values of the variants in oocyte 8.6 differed markedly from each other and therefore the calculations were performed for both heteroplasmy values separately.

The number of unique *de novo* mutations per base was assessed for every gene (Figure 2). In case the *de novo* mutations were classified per gene function (tRNA genes, rRNA genes, protein coding or D-loop separately), the prevalence of *de novo* mutations appeared to be slightly higher in the D-loop (4 variants per 1000 bases). 26 variants were in protein-coding genes (*ND1, ND4, ND5, ND6, COI, COIII and cyt-b*), with little difference in the prevalence among the different protein-coding genes. Four of these mutations were synonymous and 22 non-synonymous, including four mutations leading to a premature stop codon (Table 2).

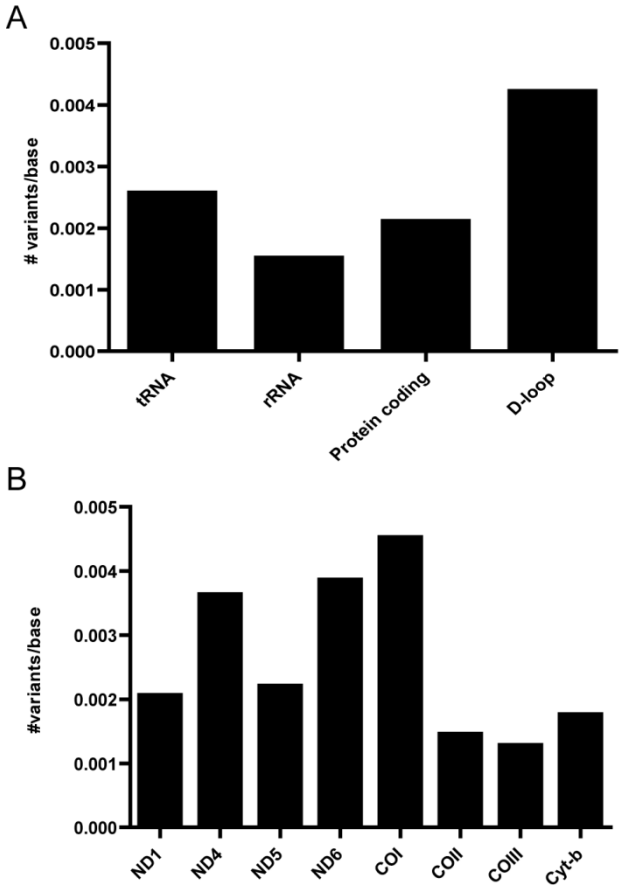

Figure 2. Prevalence of *de novo* variants. Prevalence A) per gene type (protein coding genes, tRNA, rRNA genes and D-loop) and B) per protein-coding gene. The observed number of variants in all oocytes is expressed as the number of variants per nucleotide. If a *de novo* mutation occurred multiple times, it is only counted once in the construction of these distributions.

Discussion

Robustness of identification *de novo* mutations with heteroplasmy levels ≥1.5%

We used the Nextera XT protocol to prepare libraries for sequencing on the HiSeq2000 platform, a system which has been successfully applied before for sequencing of the mitochondrial genome (McElhoe *et al.* 2014; Rebolledo-Jaramillo *et al.* 2014). The identification of specific mutations with low heteroplasmy is limited by the noise level of the sequencing procedure (Guo *et al.* 2013), which is around 1% (Guo *et al.* 2013; Rebolledo-Jaramillo *et al.* 2014; Ma *et al.* 2015). As we cannot be sure noise levels will be equal among different runs and detection thresholds differ accordingly, we estimated the noise signal of our sequence run using a non-parametric, data-driven approach. Our approach is twofold. First, we estimated the median minimal coverage (Figure 1), which is a robust estimate of the coverage of the entire mitochondrial genome, rendering a threshold coverage above which the calculated heteroplasmy value is reliable. Second, the noise level of the procedure is estimated for every position in the mtDNA based on the heteroplasmy levels reported for this position in all samples. By applying statistics, this allows identification of those variants for which the heteroplasmy level was significantly different from the background (or noise) signal (*P*-value ≤0.01), thereby excluding variants with a high occurrence in all samples, something which is not expected. In a previous report, power calculations were used to estimate the reliability of the reads

(REBOLLEDO-JARAMILLO *et al.* 2014). However, power calculations are theoretical and the use of such post-hoc calculations for the interpretation of available experimental results is debatable (HOENIG and HEISEY 2001). Another step in our analysis involves correction for the noise signal by subtracting the maternal heteroplasmy value from the detected heteroplasmy value. This correction is essential to know at which heteroplasmy value the *de novo* variant arose. A last step involved exclusion of variants that most likely occurred as a result of alignment artefacts. Based on our and others (MCELHOE *et al.* 2014; REBOLLEDO-JARAMILLO *et al.* 2014) experience with next generation sequencing, we excluded variants if they arose in close proximity to each other. Altogether, our data-driven statistical approach allows detection of *de novo* variants (heteroplasmy levels ≥1.5%) with high reliability, which is corrected for potential differences in quality between sequence runs.

De novo mutations are detected in oocytes with a low mtDNA copy number

After applying statistical and biological filters, we characterized 38 *de novo* base substitutions with an average heteroplasmy level of 2.7% in 18% of the oocytes. No large or small indels were detected in our analysis. Our analysis pipeline allowed the detection of variants with a heteroplasmy value ≥1.5%, which equals detection of a single mutated mtDNA molecule in a population of 65 or less. The estimated mtDNA copy number at which a *de novo* mutation occurred, ranged, based on the detected heteroplasmy levels of the mutations in the oocytes, from 11-67 (Table 2). In our analysis we only detected an mtDNA mutation in 18.9% of the mature oocytes. The inheritance of the mtDNA through a bottleneck leads to low mtDNA levels in primordial germ cells (PGCs) at the bottom of the bottleneck (CREE *et al.* 2008) and a mutation originating at this point may lead to higher heteroplasmy levels. In a previous study (OTTEN *et al.* 2016), we have determined the mtDNA copy number in

zebrafish PGCs isolated from several embryonic stages and found, on average, 171 mtDNA molecules at the bottom of the bottleneck, but with high variation in this number (StDev: 111). Based on these parameters we constructed a Gaussian distribution with mean 171 and standard deviation 111. As 18.9% of the mature oocytes harbored a *de novo* mutation, this distribution allowed us to estimate that lower 18.9% (left tail of the distribution, *z*-score -0.88) of the PGCs possess ≤72.8 mtDNA molecules. An mtDNA copy number of 72.8 correspond to a heteroplasmy level at mutation manifestation of 1.4%, which is close to our detection limit of 1.5%. This means our pipeline allows detection of *de novo* mutations in oocytes that were generated from germ cells with the lowest mtDNA content, which includes those germ cells most prone for acquiring a *de novo* mutation reaching a detectable heteroplasmy of 1.5% or higher, after mtDNA replication. Together, this implies ~20% of the oocytes had a bottleneck size at which *de novo* mutations could reach detectable heteroplasmy levels (≥1.5%). This also indicates that the oocytes (the other 82%) with a higher mtDNA copy number will equally carry *de novo* point mutations with a heteroplasmy level ≤1.5%, but the sensitivity at this level was too low for accurate mtDNA heteroplasmy analysis, which was in line with a recent study (HAMMOND *et al.* 2016). As the current noise level of sequencing is around 1% (GUO *et al.* 2013; REBOLLEDO-JARAMILLO *et al.* 2014; MA *et al.* 2015), further improvements of sequencing technologies are needed to detect mtDNA point mutations in all oocytes, including those with a higher mtDNA copy number.

During oogenesis in zebrafish, the mtDNA of the PGC is replicated extensively. mtDNA replication is expected to occur almost exclusively by polymerase gamma (POLG) (COPELAND and LONGLEY 2003; KAGUNI 2004), which is a two-subunit holoenzyme with high fidelity in nucleotide selection and incorporation, alongside with

389 3'-5' exonuclease proofreading functionality (KAGUNI 2004). For all mature oocytes in
390 which we detected a *de novo* mutation, the calculated average mutation rate was
391 $4.3x10^{-6}$ per nucleotide. This is in the same order of the reported POLG error
392 frequencies, which range from $2x10^{-6}$ to $10x10^{-6}$ per nucleotide in different animals
393 (KUNKEL and MOSBAUGH 1989; LONGLEY *et al.* 2001). Although POLG has a high
394 accuracy, this suggests that errors made by POLG during the extensive replication
395 during oogenesis are the main cause of *de novo* mutations detected in mature
396 oocytes. The absence of *de novo* indels is in line with a study in human germline
397 mtDNA (REBOLLEDO-JARAMILLO *et al.* 2014) and suggests that error rate of POLG for
398 insertions and deletions is lower than for substitutions, which has been reported
399 before (LONGLEY *et al.* 2001).

400 Although replication errors are made continuously in all oocytes, only
401 replication errors made when mtDNA copy numbers are low (e.g. 65 or lower) lead to
402 detectable (heteroplasmy levels of ≥1.5%) *de novo* mutations in mature oocytes,
403 which was likely the case in 19 of the oocytes. Four of these oocytes had a higher
404 mutation rate (range of 10^{-5}). Although this can be a chance event, this could also be
405 a reflection of individual differences in error rates of the mtDNA replicative machinery,
406 or due to another mutagenic source being active in these oocytes. During OXPHOS,
407 which is highly active during oogenesis (VAN BLERKOM *et al.* 1995; DUMOLLARD *et al.*
408 2007), mutagenic ROS are being produced and this could be a factor contributing to
409 differences in the mutation rate between oocytes. However, based on the mutation
410 rates in most oocytes, errors made by POLG are most likely the dominant source of
411 *de novo* mtDNA mutations during oogenesis.

412 Our calculations on the mtDNA copy number and mutation rate are only
413 applicable if random processes prevail and every mtDNA molecule is equally

amplified. Studies in mice have suggested that heteroplasmy remains stable during oogenesis (JENUTH *et al.* 1996). However, due to genetic drift, leading to loss or fixation of mutations, especially in small sample sizes, and preferential selection, mutation loads can shift. This could explain the high mtDNA mutation load for one of the two *de novo* mutations detected in oocyte 8.6, although these mutations could also have manifested during separate replication cycles, as an extremely low mtDNA bottleneck size creates multiple cycles at which the mutation can manifest at heteroplasmy levels ≥1.5%. Negative and positive selection has been demonstrated for specific some mutations (STEFFANN *et al.* 2015), further corroborating the possibility that non-random processes also influence the heteroplasmy level of *de novo* mtDNA mutations. In the case of selection, the physical and effective bottleneck sizes are different. The mtDNA molecules that actively replicate determine the effective bottleneck size, which can be lower than the physical bottleneck size when selective events result in only a subpopulation of mtDNA molecules being more actively replicated. The mtDNA copy number we have estimated here based on the heteroplasmy levels (11-67) correspond to effective mtDNA copy numbers, and might therefore be an underestimation of the physical bottleneck size.

De novo mutations in oocytes are potentially pathogenic

The 38 *de novo* mutations with a frequency of >1.5% were randomly distributed over the mtDNA genome. After correction for the size of the gene, the *tRNA-Leu*, *tRNA-Trp* and *tRNA-Gly* genes had a high number of mutations per nucleotide. However, the numbers were too low to estimate this correctly [only one mutation in the *tRNA-Leu* and *tRNA-Trp* genes were observed and two mutations in the *tRNA-Gly* gene] and mutations occurred in two oocytes from the same zebrafish. Furthermore, the tRNA genes as a group do not support a higher prevalence of mutations in the tRNA

genes. The observed higher prevalence in the three tRNA genes is most likely due to the relatively small group of *de novo* mutations. On the contrary, a higher prevalence for mutations in the D-loop exists (Figure 2A). This is corroborated by the many variants observed in this mitochondrial control region (CHINNERY *et al.* 1999). Although preferences for the D-loop might exist from an evolutionary perspective, mutations in the mtDNA can arise anywhere in the mtDNA genome. In total, eight protein-coding genes were affected with little differences in the prevalence. 22 mutations were non-synonymous, including four mutations leading to a premature stop codon. No differences in prevalence in one of the three codon positions were found. This indicates the effect of a *de novo* mutation can be of any kind. The non-synonymous mutations, especially those causing a premature stop codon, are likely pathogenic, implying that also these severe mutations, which are rarely found in human patients, can occur *de novo*. Most likely in humans, these pathogenic mutations are filtered out by mitophagy (SONG *et al.* 2014) or are at high levels not compatible with embryonic survival and remain at low levels unnoticed.

Given the high sequence homology (72%; NCBI blast performed) between the mtDNA genome of zebrafish and humans and the high evolutionary conservation of the mtDNA bottleneck in animal species (WOLFF *et al.* 2011; GUO *et al.* 2013; OTTEN and SMEETS 2015; OTTEN *et al.* 2016) our results indicate that the *de novo* risk might be similar among zebrafish and humans. Indeed, a study in 26 human oocytes (JACOBS *et al.* 2007), 7 oocytes (26.9%) were found to harbor *de novo* variants. This is close to the frequency of 18.9% we report her for zebrafish. This is further corroborated by a similar degree of variation in mtDNA copy number in human oocytes (OTTEN and SMEETS 2015), which suggest also variation in the mtDNA bottleneck size and subsequent differences in the *de novo* risk. In humans, it has

been estimated that about 5% of the mutations in the mtDNA alter a conserved nucleotide and are thus potentially pathogenic (JACOBS *et al.* 2007). As we found *de novo* mutations in ~20% of the oocytes, this implies ~1% of oocytes will carry a pathogenic *de novo* mutation [with a heteroplasmy level ≥1.5%]. The presence of low-level mtDNA mutations in the oocyte could, after fertilization, lead to mtDNA disease later in life due to genetic drift, which could lead to fixation of the mutation (GREAVES *et al.* 2014; YIN *et al.* 2015) or in the offspring of the following generation, as inheritance through the mtDNA bottleneck can cause shifts in the heteroplasmy level between mother and child, also leading to fixation of the mutant mtDNA (BLOK *et al.* 1997).

Despite the described similarities between zebrafish and humans, important reproductive and mtDNA differences should be taken into account. Zebrafish oocytes (OTTEN *et al.* 2016) possess a much higher absolute mtDNA content as human oocytes (DURAN *et al.* 2011; MURAKOSHI *et al.* 2013) (factor 100), mostly due to different implantation patterns. In humans, implantation occurs rapidly (WIMSATT 1975), allowing a fast shift to the uterus for energy supply, while in zebrafish implantation is absent and energy must be supplied by the embryo itself. This lower mtDNA copy number in human oocytes might result in lower mtDNA numbers at the bottom of the bottleneck and the mtDNA genome might be even at higher risk for a *de novo* mutation to reach detectable heteroplasmy levels. This is supported by the high mutation frequency reported for the mammalian mtDNA compared to other animals, including fish (LYNCH 2006). In conclusion, our study in zebrafish has revealed that replication errors made during oogenesis are an important source of *de novo* mtDNA mutations and their location and heteroplasmy determine the eventual significance.

Acknowledgements

We thank Marie Winandy and Hélène Pendeville from the Zebrafish facility of Liège University for their assistance in collecting the zebrafish material.

Funding

This work was supported by the Interreg IV program of the European council [the alma-in-silico project to M.M. and H.J.M.S.] and the E-RARE 2 project GENOMIT [R 50.02.12F to M.M.]. Part of this work has been made possible with the support of the Dutch Province of Limburg [M.A., M.G. and H.J.M.S.]

References

Barr, C. M., M. Neiman and D. R. Taylor, 2005 Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytol 168**:** 39-50.

Blok, R. B., D. A. Gook, D. R. Thorburn and H. H. Dahl, 1997 Skewed segregation of the mtDNA nt 8993 (T-->G) mutation in human oocytes. Am J Hum Genet 60**:** 1495-1501.

Brand, M. D., 2010 The sites and topology of mitochondrial superoxide production. Exp Gerontol 45**:** 466-472.

Broughton, R. E., J. E. Milam and B. A. Roe, 2001 The complete sequence of the zebrafish (Danio rerio) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. Genome Res 11**:** 1958-1967.

Chinnery, P. F., N. Howell, R. M. Andrews and D. M. Turnbull, 1999 Mitochondrial DNA analysis: polymorphisms and pathogenicity. J Med Genet 36**:** 505-510.

Copeland, W. C., and M. J. Longley, 2003 DNA polymerase gamma in mitochondrial DNA replication and repair. ScientificWorldJournal 3**:** 34-44.

Cree, L. M., D. C. Samuels, S. C. de Sousa Lopes, H. K. Rajasimha, P. Wonnapinij *et al.*, 2008 A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. Nat Genet 40**:** 249-254.

Dumollard, R., M. Duchen and J. Carroll, 2007 The role of mitochondrial function in the oocyte and embryo. Curr Top Dev Biol 77**:** 21-49.

Duran, H. E., F. Simsek-Duran, S. C. Oehninger, H. W. Jones, Jr. and F. J. Castora, 2011 The association of reproductive senescence with mitochondrial quantity,

function, and DNA integrity in human oocytes at different stages of maturation. Fertil Steril 96**:** 384-388.
Greaves, L. C., M. Nooteboom, J. L. Elson, H. A. Tuppen, G. A. Taylor *et al.*, 2014 Clonal Expansion of Early to Mid-Life Mitochondrial DNA Point Mutations Drives Mitochondrial Dysfunction during Human Ageing. PLoS Genet 10**:** e1004620.
Guo, Y., C. I. Li, Q. Sheng, J. F. Winther, Q. Cai *et al.*, 2013 Very low-level heteroplasmy mtDNA variations are inherited in humans. J Genet Genomics 40**:** 607-615.
Hammond, E. R., M. P. Green, A. N. Shelling, M. C. Berg, J. C. Peek *et al.*, 2016 Oocyte mitochondrial deletions and heteroplasmy in a bovine model of ageing and ovarian stimulation. Mol Hum Reprod 22**:** 261-271.
Hellebrekers, D. M., R. Wolfe, A. T. Hendrickx, I. F. de Coo, C. E. de Die *et al.*, 2012 PGD and heteroplasmic mitochondrial DNA point mutations: a systematic review estimating the chance of healthy offspring. Hum Reprod Update 18**:** 341-349.
Hoenig, J. M., and D. M. Heisey, 2001 The Abuse of Power The Pervasive Fallacy of Power Calculations for Data Analysis Am. Stat. 55**:** 19-24.
Howell, N., S. Halvorson, I. Kubacka, D. A. McCullough, L. A. Bindoff *et al.*, 1992 Mitochondrial gene segregation in mammals: is the bottleneck always narrow? Hum Genet 90**:** 117-120.
Jacobs, L., M. Gerards, P. Chinnery, J. Dumoulin, I. de Coo *et al.*, 2007 mtDNA point mutations are present at various levels of heteroplasmy in human oocytes. Mol Hum Reprod 13**:** 149-154.
Jenuth, J. P., A. C. Peterson, K. Fu and E. A. Shoubridge, 1996 Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nat Genet 14**:** 146-151.
Kaguni, L. S., 2004 DNA polymerase gamma, the mitochondrial replicase. Annu Rev Biochem 73**:** 293-320.
Kimmel, C. B., W. W. Ballard, S. R. Kimmel, B. Ullmann and T. F. Schilling, 1995 Stages of embryonic development of the zebrafish. Dev Dyn 203**:** 253-310.
Krawitz, P., C. Rodelsperger, M. Jager, L. Jostins, S. Bauer *et al.*, 2010 Microindel detection in short-read sequence data. Bioinformatics 26**:** 722-729.
Kunkel, T. A., and D. W. Mosbaugh, 1989 Exonucleolytic proofreading by a mammalian DNA polymerase. Biochemistry 28**:** 988-995.
Lee, H. S., H. Ma, R. C. Juanes, M. Tachibana, M. Sparman *et al.*, 2012 Rapid mitochondrial DNA segregation in primate preimplantation embryos precedes somatic and germline bottleneck. Cell Rep 1**:** 506-515.
Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25**:** 1754-1760.
Li, H., and R. Durbin, 2010 Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26**:** 589-595.
Longley, M. J., D. Nguyen, T. A. Kunkel and W. C. Copeland, 2001 The fidelity of human DNA polymerase gamma with and without exonucleolytic proofreading and the p55 accessory subunit. J Biol Chem 276**:** 38555-38562.
Lynch, M., 2006 The origins of eukaryotic gene structure. Mol Biol Evol 23**:** 450-468.
Lynch, M., B. Koskella and S. Schaack, 2006 Mutation pressure and the evolution of organelle genomic architecture. Science 311**:** 1727-1730.

570 Ma, J., H. Purcell, L. Showalter and K. M. Aagaard, 2015 Mitochondrial DNA
571 sequence variation is largely conserved at birth with rare de novo mutations in
572 neonates. Am J Obstet Gynecol 212**:** 530 e531-538.
573 McElhoe, J. A., M. M. Holland, K. D. Makova, M. S. Su, I. M. Paul *et al.*, 2014
574 Development and assessment of an optimized next-generation DNA
575 sequencing approach for the mtgenome using the Illumina MiSeq. Forensic
576 Sci Int Genet 13**:** 20-29.
577 Murakoshi, Y., K. Sueoka, K. Takahashi, S. Sato, T. Sakurai *et al.*, 2013 Embryo
578 developmental capability and pregnancy outcome are related to the
579 mitochondrial DNA copy number and ooplasmic volume. J Assist Reprod
580 Genet 30**:** 1367-1375.
581 Neuman, J. A., O. Isakov and N. Shomron, 2013 Analysis of insertion-deletion from
582 deep-sequencing data: software evaluation for optimal detection. Briefings in
583 bioinformatics 14**:** 46-55.
584 Otten, A. B., and H. J. Smeets, 2015 Evolutionary defined role of the mitochondrial
585 DNA in fertility, disease and ageing. Hum Reprod Update.
586 Otten, A. B., T. E. Theunissen, J. G. Derhaag, E. H. Lambrichs, I. B. Boesten *et al.*,
587 2016 Differences in Strength and Timing of the mtDNA Bottleneck between
588 Zebrafish Germline and Non-germline Cells. Cell Rep 16**:** 622-630.
589 Rebolledo-Jaramillo, B., M. S. Su, N. Stoler, J. A. McElhoe, B. Dickins *et al.*, 2014
590 Maternal age effect and severe germ-line bottleneck in the inheritance of
591 human mitochondrial DNA. Proc Natl Acad Sci U S A 111**:** 15474-15479.
592 Sallevelt, S. C., C. E. de Die-Smulders, A. T. Hendrickx, D. M. Hellebrekers, I. F. de
593 Coo *et al.*, 2016 De novo mtDNA point mutations are common and have a low
594 recurrence risk. J Med Genet.
595 Seneca, S., K. Vancampenhout, R. Van Coster, J. Smet, W. Lissens *et al.*, 2015
596 Analysis of the whole mitochondrial genome: translation of the Ion Torrent
597 Personal Genome Machine system to the diagnostic bench? European journal
598 of human genetics : EJHG 23**:** 41-48.
599 Song, W. H., J. W. Ballard, Y. J. Yi and P. Sutovsky, 2014 Regulation of
600 mitochondrial genome inheritance by autophagy and ubiquitin-proteasome
601 system: implications for health, fitness, and fertility. Biomed Res Int 2014**:**
602 981867.
603 Steffann, J., S. Monnot and J. P. Bonnefont, 2015 mtDNA mutations variously impact
604 mtDNA maintenance throughout the human embryofetal development. Clinical
605 genetics 88**:** 416-424.
606 Taylor, R. W., and D. M. Turnbull, 2005 Mitochondrial DNA mutations in human
607 disease. Nat Rev Genet 6**:** 389-402.
608 Thorvaldsdottir, H., J. T. Robinson and J. P. Mesirov, 2013 Integrative Genomics
609 Viewer (IGV): high-performance genomics data visualization and exploration.
610 Briefings in bioinformatics 14**:** 178-192.
611 Van Blerkom, J., P. W. Davis and J. Lee, 1995 ATP content of human oocytes and
612 developmental potential and outcome after in-vitro fertilization and embryo
613 transfer. Hum Reprod 10**:** 415-424.
614 Wimsatt, W. A., 1975 Some comparative aspects of implantation. Biol Reprod 12**:** 1-
615 40.
616 Wolff, J. N., D. J. White, M. Woodhams, H. E. White and N. J. Gemmell, 2011 The
617 strength and timing of the mitochondrial bottleneck in salmon suggests a
618 conserved mechanism in vertebrates. PLoS One 6**:** e20522.

Yamada, M., V. Emmanuele, M. J. Sanchez-Quintero, B. Sun, G. Lallos *et al.*, 2016 Genetic Drift Can Compromise Mitochondrial Replacement by Nuclear Transfer in Human Oocytes. Cell stem cell.

Yin, A. H., C. F. Peng, X. Zhao, B. A. Caughey, J. X. Yang *et al.*, 2015 Noninvasive detection of fetal subchromosomal abnormalities by semiconductor sequencing of maternal plasma DNA. Proc Natl Acad Sci U S A.