Human Reproduction Update, Vol.24, No.5 pp. 519–534, 2018
Advanced Access publication on May 11, 2018 doi:10.1093/humupd/dmy018

human reproduction update

A systematic review and meta-analysis reveals pervasive effects of germline mitochondrial replacement on components of health

Ralph Dobler [1,*], Damian K. Dowling[2], Edward H. Morrow[3], and Klaus Reinhardt[1]

[1]Applied Zoology, Technische Universität Dresden, Zellescher Weg 20b, 01217 Dresden, Germany [2]School of Biological Sciences, Monash University, Clayton, Vic. 3800, Australia [3]Evolution, Behaviour and Environment Group, School of Life Sciences, University of Sussex, Brighton BN1 9QG, UK

*Correspondence address. E-mail: ralphdobler@gmx.ch orcid.org/0000-0003-4354-6508

Submitted on May 31, 2017; resubmitted on April 24, 2018; editorial decision on April 27, 2018; accepted on May 3, 2018

TABLE OF CONTENTS

BACKGROUND: Mitochondrial replacement, a form of nuclear transfer, has been proposed as a germline therapy to prevent the transmission of mitochondrial diseases. Mitochondrial replacement therapy has been licensed for clinical application in the UK, and already carried out in other countries, but little is known about negative or unintended effects on the health of offspring born using this technique.

OBJECTIVE AND RATIONALE: Studies in invertebrate models have used techniques that achieve mitochondrial replacement to create offspring with novel combinations of mitochondrial and nuclear genotype. These have demonstrated that the creation of novel mitochondrial–nuclear interactions can lead to alterations in offspring characteristics, such as development rates, fertility and longevity. However, it is currently unclear whether such interactions could similarly affect the outcomes of vertebrate biomedical studies, which have sought to assess the efficacy of the replacement therapy.

SEARCH METHODS: This systematic review addresses whether the effects of mitochondrial replacement on offspring characteristics differ in magnitude between biological (conducted on invertebrate models, with an ecological or evolutionary focus) and biomedical studies (conducted on vertebrate models, with a clinical focus). Studies were selected based on a key-word search in 'Web of Science', complemented by backward searches of reviews on the topic of mitochondrial–nuclear (mito-nuclear) interactions. In total, 43 of the resulting

116 publications identified in the search contained reliable data to estimate effect sizes of mitochondrial replacement. We found no evidence of publication bias when examining effect-size estimates across sample sizes.

OUTCOMES: Mitochondrial replacement consistently altered the phenotype, with significant effects at several levels of organismal performance and health, including gene expression, anatomy, metabolism and life-history. Biomedical and biological studies, while differing in the methods used to achieve mitochondrial replacement, showed only marginally significant differences in effect-size estimates (−0.233 [CI: −0.495 to −0.011]), with larger effect-size estimates in biomedical studies (0.697 [CI: 0.450–0.956]) than biological studies (0.462 [CI: 0.287–0.688]). Humans showed stronger effects than other species. Effects of mitochondrial replacement were also stronger in species with a higher basal metabolic rate. Based on our results, we conducted the first formal risk analysis of mitochondrial replacement, and conservatively estimate negative effects in at least one in every 130 resulting offspring born to the therapy.

WIDER IMPLICATIONS: Our findings suggest that mitochondrial replacement may routinely affect offspring characteristics across a wide array of animal species, and that such effects are likely to extend to humans. Studies in invertebrate models have confirmed mito-nuclear interactions as the underpinning cause of organismal effects following mitochondrial replacement. This therefore suggests that mito-nuclear interactions are also likely to be contributing to effects seen in biomedical studies, on vertebrate models, whose effect sizes exceeded those of biological studies. Our results advocate the use of safeguards that could offset any negative effects (defining any unintended effect as being negative) mediated by mito-nuclear interactions following mitochondrial replacement in humans, such as mitochondrial genetic matching between donor and recipient. Our results also suggest that further research into the molecular nature of mito-nuclear interactions would be beneficial in refining the clinical application of mitochondrial replacement, and in establishing what degree of variation between donor and patient mitochondrial DNA haplotypes is acceptable to ensure 'haplotype matching'.

Key words: biological / biomedical / epistasis / hybrid breakdown / maternal spindle transfer / mitochondrial disease / mito-nuclear mismatch / offspring / pronuclear transfer / three-parent baby

Introduction

Mitochondria provide most of the cell's energy, and are essential for a diverse range of other cellular and physiological processes, including (non-shivering) thermogenesis, amino acid metabolism, lipid metabolism, biosynthesis of haem and iron-sulphur clusters, calcium homoeostasis and apoptosis (Gorman *et al.*, 2016). Crucially, the integrity of most of these functions relies upon fine-tuned interactions between genes and gene products of the nuclear and mitochondrial genomes. The best-studied example is the precise co-ordination between mitochondrial- and nuclear-encoded subunits of the electron transport chain required to achieve uncompromised ATP conversion. Thus, mitochondrial–nuclear (hereafter, 'mito-nuclear') interactions are not only likely to have been of profound importance in facilitating the evolution of complex eukaryote life (Havird *et al.*, 2015), but also numerous studies have demonstrated that the genetic variation found within the mitochondrial genome of animals routinely affects the expression of key health-related traits, such as fertility, longevity and growth trajectories (Dobler *et al.*, 2014). The magnitude of the mitochondrial genetic effects on these traits is often moderated by the nuclear genetic background of the organism (e.g. Rand *et al.*, 2004; Burton and Barreto, 2012; Reinhardt *et al.*, 2013; Dobler *et al.*, 2014; Wolff *et al.*, 2014; Baris *et al.*, 2017; Kullar *et al.*, 2017; Marom *et al.*, 2017; Morrow and Camus, 2017; Loewen and Ganetzky, 2018). These findings also carry medical implications (Reinhardt *et al.*, 2013; Dowling, 2014; Picard *et al.*, 2016) because mito-nuclear interactions may play an important role in determining human health (Rand *et al.*, 2004; Burton and Barreto, 2012; Dobler *et al.*, 2014; Wolff *et al.*, 2014; Marom *et al.*, 2017; Milot *et al.*, 2017; Morrow and Camus, 2017). The occurrence of epistatic mito-nuclear interactions will generally mean that health effects attributable to particular mitochondrial variants (mitochondrial DNA [mtDNA] haplotypes or single-nucleotide polymorphisms in the mtDNA) may differ across different nuclear genetic backgrounds in which the variants are expressed, as has been confirmed empirically (James and Ballard, 2003; Dowling *et al.*, 2007c, 2010; Burgstaller *et al.*, 2014; Kenney *et al.*, 2014b; Latorre-Pellicer *et al.*, 2016; Mossman *et al.*, 2016a, b; Yamada *et al.*, 2016; Kullar *et al.*, 2017; Marom *et al.*, 2017; Morrow and Camus, 2017; Loewen and Ganetzky, 2018).

Health effects resulting from mito-nuclear interactions are particularly relevant for emergent mitochondrial replacement therapies. Here, the nuclear chromosomes (or fertilised pronuclei) from an oocyte of a female carrying a known pathogenic mtDNA mutation are translocated to an enucleated donor egg, which carries putatively healthy mtDNA, thus stopping the transmission of mitochondrial diseases (Tachibana *et al.*, 2009; Wolf *et al.*, 2015; Greenfield *et al.*, 2017). However, under simple models of mito-nuclear coevolution, mutations in mtDNA molecules will lead to selection of nuclear modifier alleles, within populations, which act to offset the negative effects associated with the mtDNA mutations (Rand *et al.*, 2004; Dowling *et al.*, 2008; Burton and Barreto, 2012; Sloan *et al.*, 2017; Connallon *et al.*, 2018). The existence of nuclear modifier alleles means that mtDNA variants that are healthy in some nuclear backgrounds (because they are counter balanced by modifier alleles) may be unhealthy in other backgrounds where there is no modifier (Reinhardt *et al.*, 2013; Kullar *et al.*, 2017; Morrow and Camus, 2017). The existence of modifier alleles would have an immediate implication for human mitochondrial replacement therapy. Ill-health of embryos might result from mitochondrial replacement if novel combinations of mito-nuclear genotype are poorly performing because a healthy donor's oocyte might not receive the appropriate nuclear modifiers required to offset one or more cryptic mutations in the donor's mtDNA genotype. The fact that phenotypic effects attributable to mito-nuclear interactions are commonly reported in the field of ecological and evolutionary genetics

(Rand *et al.*, 2004; Burton and Barreto, 2012; Reinhardt *et al.*, 2013; Dobler *et al.*, 2014; Wolff *et al.*, 2014; Hill, 2016; Baris *et al.*, 2017) suggests that the evolution of such nuclear modifier alleles might well have commonly occurred (see also Hao *et al.*, 1999 for an experimental *de novo* evolution of a modifier allele).

The effects of epistasis between genes spanning mitochondrial and nuclear genomes can be quantified empirically by examining data from organisms in which genetic strains have been created that allow mitochondrial genetic variation to be partitioned from confounding effects of the nuclear genome – via techniques that we will broadly refer to as achieving mitochondrial replacement. During mitochondrial replacement, mitochondrial variants are experimentally placed alongside novel nuclear genetic backgrounds. Comparing the resulting health and fitness outcomes (phenotypes) to those of the original genetic background provides a measure of the effect of mitochondrial replacement. In animals with short generation times, mitochondrial replacement can be achieved most simply by backcrossing females of a maternal lineage to males from a distinct paternally-derived lineage over sequential generations (resulting in the replacement of about 50% of the nuclear alleles that were found in the original maternal lineage, with each successive generation of backcrossing). This approach takes advantage of maternal transmission of the mtDNA but biparental transmission of the autosomal and X chromosomes. Using this method, more than 99.99% of the nuclear genome can be exchanged within 16 generations of backcrossing (technique described in Dowling *et al.*, 2008). Cytogenetic tools (e.g. balancer chromosomes in *Drosophila*, which suppress recombination) allow mitochondrial replacement to proceed via forced chromosome substitution over a shorter number of generations, but these tools are limited to research in a few model organisms (Herman *et al.*, 1976; Zheng *et al.*, 1999; Greenspan, 2004). Inter-oocyte nuclear transfer, a surgical technique where a nuclear karyoplast, or alternatively fertilised pronuclei, is implanted into an enucleated donor egg, results in near-complete mitochondrial replacement within a single generation. This technique is used in species where generation times are long, or where mitochondrial replacement can be more easily achieved because human medical technology can be applied. This method will be used to achieve mitochondrial replacement in the oocytes of female patients carrying mtDNA-mediated mitochondrial disease in humans (Hyslop *et al.*, 2016; Greenfield *et al.*, 2017).

The debate as to whether mito-nuclear interactions can affect human phenotypes has proven to be controversial (e.g. Reinhardt *et al.*, 2013; Chinnery *et al.*, 2014; Morrow *et al.*, 2015; Callaway, 2016; Milot *et al.*, 2017; Rishishwar and Jordan, 2017). Although no data exist, it has been argued that extensive population admixture and gene flow among human populations is likely to have impeded the capacity for mito-nuclear coadaptation in humans (Eyre-Walker, 2017; Rishishwar and Jordan, 2017). Accordingly, the potential for mito-nuclear interactions to modify the health and fitness outcomes of human offspring produced by mitochondrial replacement techniques has been largely dismissed (Human Fertilisation and Embryology Authority (HFEA) third review of the safety and efficacy report – update 2014, http://hfeaarchive.uksouth.cloudapp.azure.com/www.hfea.gov.uk/8807.html, Greenfield *et al.*, 2017), albeit without a systematic analysis of the literature into mito-nuclear interaction effects. After scientific debate, the regulatory authority in the UK has now stated that any clinic carrying out the procedure should 'consider matching the haplotypes of donors with recipients where possible' (HFEA Guidance on Mitochondrial Donation, section 33.23, https://www.hfea.gov.uk/code-of-practice/33). Recent cases outside the UK, whereby human offspring have been produced following nuclear transfer, have not taken haplotype-matching into account (Reardon, 2016; Zhang *et al.*, 2017). This discrepancy in guidelines, and in the application of mitochondrial replacement conducted so far, may be cause for concern if mito-nuclear interactions are relevant to humans.

During the review of evidence into the safety and efficacy of mitochondrial replacement therapies in the UK, a further discrepancy emerged between the interpretations of evidence from what we refer to here as 'biomedical' studies (studies that were conducted on vertebrates with direct biomedical applications) versus 'biological' studies (those conducted on invertebrates, usually coming from the field of evolutionary and ecological genetics). In the biomedical group of studies, most claimed that the phenotypic and health consequences of disrupting mito-nuclear interactions were negligible and likely to be irrelevant for human health (but see Nagao *et al.*, 1998; Moreno-Loshuertos *et al.*, 2006; Burgstaller *et al.*, 2014; Neupane *et al.*, 2015; Yamada *et al.*, 2016), and therefore concluded that mitochondrial replacement was safe for clinical use (Tachibana *et al.*, 2009, 2013; Craven *et al.*, 2010; Paull *et al.*, 2013; Wang *et al.*, 2014; Greenfield *et al.*, 2017). Biological studies, conversely, generally reported that organismal phenotypes following mitochondrial replacement are routinely modified, with relatively large effect sizes and a wide range of symptoms (reviewed by Reinhardt *et al.*, 2013; Dobler *et al.*, 2014). The results of the biological studies were suggested to have low inferential relevance to humans (Eyre-Walker, 2017, HFEA third review of the safety and efficacy report – update 2014, http://hfeaarchive.uksouth.cloudapp.azure.com/www.hfea.gov.uk/8807.html) because, first, studies of invertebrate models typically employed crosses between individuals exhibiting much larger levels of mitochondrial divergence than those found between human populations (an incorrect argument, see Morrow *et al.*, 2015) and, second, that inferences came from invertebrates that are only distantly related to humans (Chinnery *et al.*, 2014; Eyre-Walker, 2017, HFEA third review of the safety and efficacy report – update 2014, http://hfeaarchive.uksouth.cloudapp.azure.com/www.hfea.gov.uk/8807.html). Whether effect sizes between the two study types (biological and biomedical) are indeed different has, so far, not been subject to formal analysis. Neither has there been any attempt of the standard risk analysis, to estimate the frequency by which health risks might be expected after mitochondrial replacement. Both analyses would, however, be of great value because if the effect size across each set of studies converges on similar values, this would then suggest that biological studies can be used to make reliable predictions about effects on health and fitness of human offspring produced following mitochondrial replacement therapy.

Here, we conduct a systematic meta-analysis to formally estimate the strength of mito-nuclear effects across biomedical and biological studies. We test the hypothesis that differences exist between biomedical and biological studies, with the latter showing larger effect sizes of mitochondrial replacement than the former because of the larger levels of mitochondrial divergence. We also take the opportunity to explore three extra questions. First, given that mitochondrial diseases are often associated with the late onset in

adulthood, we explore the question of whether the magnitude of effects following mitochondrial replacement may possibly increase over ontogenetic development (such as from the two to the four-cell stage, or from 3-month to 6-month-old offspring and beyond), or whether there may be developmental stages that are particularly sensitive to replacement effects. Second, we ask whether certain trait types (effects on gene expression, metabolic traits or life-history traits such as adult fertility or longevity) might be particularly sensitive to the effects of mitochondrial replacement. Third, we address a novel hypothesis, which we call the 'Lane Hypothesis', which posits that species with high metabolic needs will experience stronger selection for tightly fine-tuned mito-nuclear compatibility than species with low metabolic needs (Lane, 2011). The hypothesis predicts that species with low metabolic needs will cope more easily with electron slippage in the electron transfer chain than species with high metabolic needs (Lane, 2011). If this hypothesis applies, then we expect smaller effects on health and fitness following mitochondrial replacement in species with low metabolic needs compared to species with high metabolic needs.

Methods

Systematic review – meta-analysis

We searched for literature containing relevant information about mito-nuclear health effects in two ways. First, we performed a key-word search in July 2016 for publication titles using the internet platform '*Web of Science*' for each of the following terms: mitochondria AND 'nuclear background', mitochondria AND 'compensatory epistasis', mitochondria AND compensatory, mitochondria AND penetrance, nucleo-mitochondria*, mito-nuclear, mitonuclear, mitonuclear AND epistasis, mitonuclear AND disease. Second, we did backwards searches of key reviews on the topic of mito-nuclear interactions. Overall, this search resulted in 116 publication titles. We employed several steps of quality check, resulting in 43 publications (Fig. 1). We conducted the meta-analysis according to the PRISMA statement (Moher *et al.*, 2009). We found no evidence for publication bias by plotting effect-size estimates on sample size (Supplementary Fig. S1). We assigned studies (study type) to be either 'biological' when the study organisms were invertebrates, and did not have an explicit medical focus, and instead addressed fundamental hypotheses in the field of ecological, evolutionary or population genetics; or 'biomedical' when the study organisms were vertebrates and had a direct biomedical focus. The distinction was rarely difficult to make. 'Biomedical' studies shared several characteristics – they were conducted on vertebrates, and they were published in biomedical journals examining clinically relevant parameters. In contrast, no invertebrate study was ever published in a biomedical journal, and none of these used surgical inter-oocyte techniques to achieve mitochondrial replacement. For the classification of the studies included, see Table I. The extracted literature allowed us to calculate 1004 (576 biological, 428 biomedical) effect-size estimates, Hedges' *g* (Hedges, 1981), based on mean values and test statistics (see below) collected from 43 studies (19 biological, 24 biomedical) (Table I, Supplementary Fig. S2). We only reported data for intra-specific (i.e. within-species) mito-nuclear genome combinations and did not consider the effects of any inter-specific combinations of mitochondrial and nuclear genomes between two species. We calculated the effect sizes by converting reported differences in means (and their associated SD), or the test statistics *F*-scores, *t*-scores, Chi-square-score, or log odds-ratios (Del Re, 2013). We used statistical values stated in the studies, extracted mean ± SD values from the figures provided in the articles using the programme Datathief (Tummers, 2006), and contacted the authors for missing statistical values when necessary.

An often overlooked aspect of studies on mito-nuclear interactions is that mito-nuclear genetic combinations that occur in natural populations do not represent a random sample of mito-nuclear combinations. Instead, these combinations are those that have presumably already been screened by, and survived, natural selection. Thus, to study the true effects of mito-nuclear interaction on health, experiments are required that create novel combinations of mito-nuclear genotypes. We were able to classify the literature accordingly and draw only effect sizes that compared naturally occurring (putatively 'matched') mito-nuclear combinations with combinations that had been experimentally engineered (via mitochondrial replacement techniques of backcrossing, forced chromosome substitution, or inter-oocyte nuclear transfer, and were thus putatively 'mismatched'). In these studies, the effect sizes directly compare the effects of mitochondrial replacement relative to the appropriate control (the putatively matched mito-nuclear combination). These effect-size estimates are thus meaningful for quantifying the relevance of mito-nuclear interaction effects of mitochondrial replacement, given that mitochondrial replacement results in an mtDNA genotype that is placed alongside a new nuclear genome.

We classified reported traits into one of four different levels (trait type) by responses of: individual genes or gene expression; metabolic traits; anatomical or morphological traits; or life-history traits such as lifespan or fecundity. The complete list of reported statistical values, additionally reported factors and calculated effect-size estimates can be found on Zenodo (http://doi.org/10.5281/zenodo.1212234). Our primary interest was in determining the strength (magnitude) of the

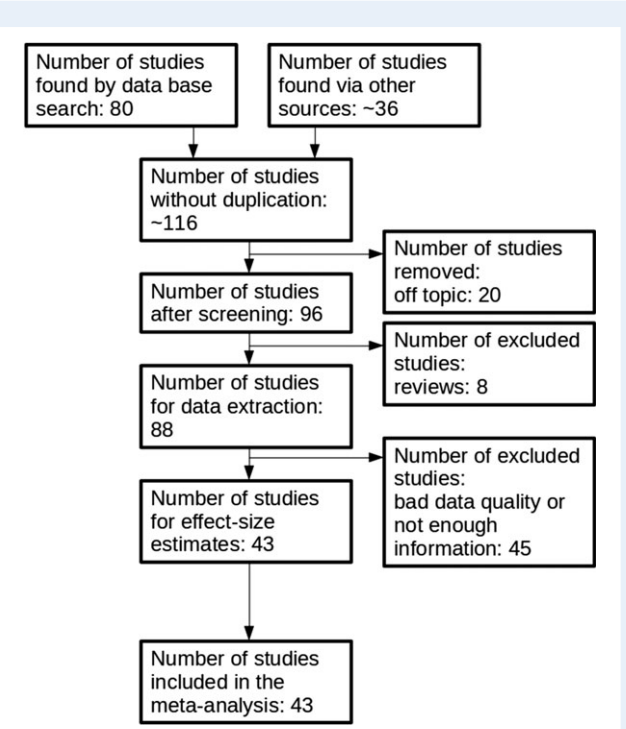

Figure 1 PRISMA flow chart for the selection process of relevant publications. 'Not enough data' as exclusion criterion means that either the sample size was not ascertainable or the error bars in figures were not characterised (i.e. SE, SD or CI). 'Bad data quality' refers to a bad (low) resolution in figures, which did not allow us to extract accurate means and/or error-bar length.

Table I List of extracted effect-size estimates by species.

Organism	Bio	Med	ES	Studies	References
Bovine	0	59	59	2	Yan *et al.* (2010, 2011)
Caenorhabditis briggsae	16	0	16	1	Chang *et al.* (2016)
Caenorhabditis elegans	4	0	4	1	Liau *et al.* (2007)
Callosobruchis maculatus	364	0	364	6	Dowling *et al.* (2007a,b, 2010) Messina and Jones (2011), Immonen *et al.* (2016a, b)
Drosophila melanogaster	16	0	16	3	Clancy (2008), Yee *et al.* (2013), Holmbeck *et al.* (2015)
Drosoophila simulans	24	0	24	2	James and Ballard (2003), Chatelain *et al.* (2011)
Drosophila subobscura	17	0	17	2	Kurbalija Novičić *et al.* (2015), Jelić *et al.* (2015)
Homo sapiens	0	169	169	10	Spees *et al.* (2006), Bellizzi *et al.* (2009), Ghelli *et al.* (2009), Lin *et al.* (2012), Paull *et al.* (2013), Strauss *et al.* (2013), Tachibana *et al.* (2013), Kenney *et al.* (2013, 2014a, b)
Macaca mulatta	0	33	33	2	Tachibana *et al.* (2009, 2013)
Mus musculus	0	114	114	8	Nagao *et al.* (1998), Johnson *et al.* (2001), Deng *et al.* (2006), Nakada *et al.* (2006), Gregorová *et al.* (2008), Neupane *et al.* (2014), Wang *et al.* (2014), Latorre-Pellicer *et al.* (2016)
Mus spp.	0	8	8	1	Kropáčková *et al.* (2015)
Rat	0	45	45	2	Pravenec *et al.* (2007), Kumarasamy *et al.* (2013)
Saccharomyces cerevisiae	4	0	4	1	Paliwal *et al.* (2014)
Tigriopus californicus	131	0	131	2	Ellison and Burton (2006, 2008), Barreto and Burton (2013)
Total	576	428	1004	44	

The list includes the organism used in the study, the number of effect-size estimates obtained by study type (Bio = biological: studies on invertebrate species, without explicit medical focus, Med = biomedical: studies on vertebrate species, with direct biomedical focus) and in total (ES) as well as the number of publications (Studies) used for effect-size estimation per organism. Some species were not clearly defined in the publication (i.e. Bovine, Rat, *Mus spp.*). Numbers of studies total 44 (not 43) because Tachibana *et al.* (2013) report data for two species (*H. sapiens* and *M. mulatta*).

reported effect-size estimates across biomedical and biological studies. However, the direction of these effects (i.e. whether a positive change in the effect size estimate has a positive effect on health or is favoured by natural selection) can be difficult to classify, except for a few traits (see below 'Sample sizes and incidences – risk analysis'). For example, in cases such as body weight or development time, it is rarely clear whether increased phenotypic expression has a positive or negative effect on overall organismal health and reproductive success ('Darwinian fitness'). Hence we consider any unintended phenotypic effects of mitochondrial replacement (regardless of direction) as negative because they were not the goal of the procedure. Accordingly, we employed a method previously used for similar cases where directionality of the effect size is less applicable than the magnitude of the effect (Kingsolver *et al.*, 2012; Dobler *et al.*, 2014), where Gaussian parameter estimates are used to calculate a 'folded-over' distribution (Hereford *et al.*, 2004; Morrissey and Hadfield, 2012) given by:

$$E(y) = \sigma * \sqrt{(2/\pi)} * \exp(-|\mu|^2/2\sigma^2) + |\mu| * (1 - 2\Phi * (-|\mu|/\sigma))$$

where $E(y)$ is the expectation on the folded-over distribution, σ the estimated SD, μ the mean and Φ denotes the cumulative distribution function of the standard normal distribution. For all statistical analyses we used the programme R, version 3.2.2. (R Development Core Team, 2017) and the specific packages are detailed at each section.

We used linear mixed models to estimate the means and variances needed to calculate the mean effect-sizes with their 95% credible intervals (CI – credible intervals are the Bayesian analogue of confidence intervals) on a folded-over distribution for each level of each factor. Such analyses depend on the estimates of specific means and variances for each factor level. Factors of interest were included as fixed effects (i.e. study type or trait type) into the model. Fitted variances (associated with the model-estimate random effect 'Publication ID' and 'residual') specific for each factor level were also included. Different data points within one specific study are unlikely to be statistically independent samples. We circumvented this issue by including 'Publication ID' as a random factor. Shared evolutionary ancestry may also produce inflated effects if closely related species are considered as independent samples (e.g. rat and mouse, or closely related *Drosophila* species) (Chamberlain *et al.*, 2012). We correct for evolutionary constraints and dependencies between different species by including the phylogenetic relationship among the investigated species into the statistical model. The estimation of group-specific variances is highly demanding of sample size, and we thus focussed on linear factor effects on effect-size estimates without any interaction terms (Dobler *et al.*, 2014). We can therefore not draw any conclusions about possible interaction effects between the factors study type and trait type. In order to specifically assess whether the effect-size estimates differ between studies on humans and other model organisms, we compared effect-size estimates from human studies (10 studies, 169 effect-size estimates) to effect size estimates of other studies (34 studies, 835 effect-size estimates).

We ran all models for the meta-analyses using the *MCMCglmm*-package (Hadfield, 2010) in R 3.3.2 (R Development Core Team, 2017), where the parameter estimates were transformed to a folded distribution, as described above. We used flat priors for the fixed effects and locally uninformative parameter-expanded priors for the random effects. Both represent little prior knowledge, and ensured that all autocorrelations were in the [–0.1, 0.1] interval. After a burn-in of 10 000 iterations,

we sampled the posterior chain of each one of the models with 510 000 iterations and a thinning interval of 500, yielding a total posterior sample of 1000. We used the 'pedigree' option of the *MCMCglmm*-package to include the phylogenetic structure of the collected data in our analyses. We created the phylogenetic tree of the species included in our meta-analysis with *timetree* (www.timetree.org) and transformed it with the R-packages *ape* (Popescu *et al.*, 2012) and *phytools* (Revell, 2012) to meet the *MCMCglmm* requirements. In addition to effect-size estimates and their 95% CI we calculated and report the I^2-values for meta-analyses. Instead of traditional *P*-value based statistics we used Bayesian statistics and whenever the 95% CI of an effect-size estimate does not overlap with zero, the effect-size estimate is significant. The I^2-values describe the amount of variation across studies in meta-analyses that is due to heterogeneity rather than chance. The R-code used for the meta-analysis is available on Zenodo (http://doi.org/10.5281/zenodo.1212234).

Finally, we used the dataset to assess how mitochondrial replacement affects survival rate over the course of ontogeny. We extracted survival data from all suitable studies ($n = 7$; all biomedical; $n = 0$ from biological), where survival data were reported for two or more developmental stages (11 development stages in total: oocyte, 1-cell stage, 2-cell stage, 8 cell stage, blastocyst, 20 days pregnancy, 60 days pregnancy, 90 days pregnancy, birth, 3 months old, 6 months old). We analysed the survival rate between ontogenetic stages as proportional change of surviving individuals between the control and the mitochondrial (mt) replacement treatment:

$$\frac{\text{change control} - \text{change mt replacement}}{\text{change control} + \text{change mt replacement}}$$

We calculated 92 proportional changes in survival from the reported data. Negative values indicate proportionally more individuals died in the mitochondrial replacement treatment group, positive values that proportionally more individuals died in the controls. We calculated survival rate in two ways. First, we calculated the survival rate between consecutively reported development stages (sequential survival rate – from stage A to B, from B to C, from C to D, etc.) in order to see whether survival would be particularly affected at some development stages by mitochondrial replacement. Second, we calculated the survival rate relative to the first reported development stage in the analysis (absolute survival rate – from stage A to D if A was the first reported stage or from C to E if C was the first reported stage). This procedure assesses whether small, non-significant effects in each ontogenetic stage nevertheless add up to an overall health effect later in life. We included all 92 calculated proportional changes in survival rate for the overall graphical inspection and took the mean estimate for each stage when illustrating the survival rate for each study separately. We performed a Spearman's rank correlation to assess whether there is a correlation between the development stage (ordinal factor, difference between stages is not linear and differs between species) and the survival rate (absolute values, continuous variable).

Coefficient of variation in biological and biomedical studies

One important difference between biomedical and biological studies lies in the way that outliers are interpreted. Biological studies focus on the mean or median in a population (the evolutionarily relevant unit). Outliers are treated with caution, as they indicate that factors other than those investigated may be the cause of the extreme value of the measurement, or that the value is the result of a measurement error, which may not be traceable or reproducible. In biomedical studies, outliers are not generally neglected because eventually they represent individual patient outcomes. Necessarily, biomedical studies sometimes focus on precisely such outliers, as they are of particular relevance for risk assessments of new therapies. Thus, in the case that mean effect sizes for mitochondrial replacement and control groups were different, it is important to identify whether the difference is driven by small, consistent differences in all individuals, or by few more extreme values. In the case of mitochondrial replacement therapy, if extreme values arose in only some women, risk mitigation factors might be specifically targeted for these women.

To assess whether the effects in biological and biomedical studies differed because of many small but consistent effects, or because of large effects in a few individuals, we separately analysed the coefficient of variation for all reported data in relation to their means. We calculated the coefficient of variation for each reported trait value by dividing its SD by the respective mean. This allowed us to compare studies with different mean values, as the coefficient of variation is no longer absolute but relative. A difference caused by many small effects will have a low coefficient of variation, and one caused by few extreme cases a large coefficient of variation. We analysed the coefficient of variation between biological and biomedical studies using a Kruskal–Wallis test. We further distinguished whether the trait was measured in animals with matched or mismatched mito-nuclear genomes. We also compared the sample sizes for each reported trait to assess the accuracy of the data. Higher sample sizes and lower coefficients of variation indicate a higher accuracy for reported trait data. We considered *P*-values <0.05 to be statistically significant.

Mitochondrial replacement effects and metabolic rate – Lane Hypothesis

Species with high energetic needs are expected to carry well-adapted (matched) mito-nuclear genome combinations and should hence be more prone to artificial mismatches of the two genomes (Lane, 2011). The reason this was predicted is that the strength of natural selection on key mito-nuclear combinations is likely to be greater in species with higher energetic demands, given that mismatches will have greater (potentially negative) effects on the energy production and consequently on health and fitness. In species with lower energetic demands, selection is expected to be relaxed as small deviations from the optimal energy production should be less likely to confer profound effects. As a consequence, species with high energetic needs (and therefore higher basal metabolic rates – BMR) should reveal larger effect-sizes in studies investigating mito-nuclear interactions. We were able to collect data on the BMR for 13 species included in the systematic review (no data on the BMR for *Caenorhabditis briggsae*) (Zenodo [http://doi.org/10.5281/zenodo.1212234]). We excluded yeast (*Saccharomyces cerevisiae*) from the analysis as it represented the only non-animal species. The BMR was either directly extracted as Watt per kilogram body weight [W/kg] or calculated via the oxygen consumption or carbon dioxide production per time unit. One litre of gas consumption or production was considered to be equivalent to 4.82 kcal (http://gasexchange.com/notes/metabolism/). The transformation from kcal/min to Watt was by a factor of 69.8 (http://www.convertunits.com/from/watt/to/kcal/min) and was adjusted to the proper time unit and one kilogram of body weight. Because in vertebrates BMR scales to body mass by the power of 3/4, we obtained mass-independent BMR estimates for the vertebrate species by dividing the reported BMR by $(\text{body mass})^{0.75}$ for each species (Kleiber, 1932; Savage *et al.*, 2004). We tested whether effect-size estimates increased with mass-independent BMRs (Lane, 2011). We *ln*-transformed the BMR values to apply a linear regression. We applied a linear regression because both the dependent variable (effect-size estimate) and the explaining variable (*ln*-transformed BMR-values) were continuous data. Because ectothermic invertebrates may be expected to have a different effect-size/BMR relationship than endothermic vertebrates, we analysed the data with

uncorrelated random slope and random intercept for biomedical and biological studies. We considered *P*-values <0.05 to be statistically significant. While the BMR of endothermic species should not be expected to rely heavily on the environmental temperature (at least not in the temperature range where BMRs are generally assessed), the environmental temperature can have a strong impact on the BMR measured for ectothermic species. We therefore also checked the temperature at which the reported BMR measurement was sampled. Studies in endothermic species measured the BMR at normal body temperature or did not provide the temperature. Ectotherm BMRs were assessed between 19°C and 25°C, representing the 'natural' conditions for each of the ectotherm species. While the variation in temperature across the various studies of ectothermic species may account for some of the variation in BMR measures across species in our analyses, we assume that our analysis of the BMR is nonetheless valid given the reported data points used in our analyses were always measured (and thus represent) values comparable to natural conditions in which each species would have evolved.

Sample sizes and incidences – risk analysis

A key consideration for introducing new technology to society is the risk analysis, an estimate of the likelihood of negative side effects of the technology. This standard procedure for all medical treatment can then be used to weigh the costs against the benefits of the technology. A formal risk analysis does not seem to exist for mitochondrial replacement therapy. Focusing explicitly on reproductive, health or performance indicators, where negative effects clearly indicate negative effects on overall health, we found sufficient data on performance outcomes in 19 studies. In 15 of these 19 studies (Table II), we were able to determine whether the mitochondrial replacement effect had an unambiguously negative or positive effect on health: for example, when lifespan was reduced or extended. For this subset of data, the direction of the effect was not arbitrary and we deemed directionality (i.e. an increase or a decrease in the measured trait) of biological (or biomedical) meaning. This means the direction of the mito-nuclear interaction effect can directly be attributed to either a negative or a positive fitness or health effect for the individual. The 15 studies resulted in a total of 532 trait-by-mitochondrial-by-nuclear-genome combinations (Zenodo [http://doi.org/10.5281/zenodo.1212234]) from which 391 matched-mismatched comparisons could be made. Because of the very low sample size per mito-nuclear combination in biomedical studies (resulting in low statistical power), a direct estimation of health risks based solely on biomedical studies might lead to unreliable inferences. We, therefore, used a different approach. We estimated the incidence of negative effects of mitochondrial replacement from biological studies (which possessed larger sample sizes and higher statistical power) and then used this figure to predict an expected risk of mitochondrial replacement in a series of biomedical studies for their given sample size. To produce the most conservative risk estimate, we then assumed the unlikely case that the negative effect in the mitochondrial replacement treatment group in a given study was entirely caused by a single offspring from a single mother. This assumption is unrealistic because an effect of one outlier would not result in a statistical difference between treatment and control in the first place, but the resulting value represents the lowest possible number of negatively affected offspring.

Results

Systematic review – meta-analysis

After the stringent quality check, 43 studies (19 biological and 24 biomedical) were retained (Table I, Fig. 1). Of the resulting 1004 (576 biological, 428 biomedical) effect-size estimates (Table I), Hedges' *g* (Hedges, 1981) yielded an overall effect-size estimate of 0.581 (95% CI: 0.379–0.880; consistency [amount of variation across studies due to heterogeneity rather than chance]: I^2 = 0.736, CI: 0.141–1.490).

Table II List of studies used for the risk analysis.

Source	Type	MM « M	MM < M	MM = M	MM > M	MM » M	Any sig
Dowling *et al.* (2007b)	Bio	4	0	0	1	0	Yes
Dowling *et al.* (2010)	Bio	7	7	1	10	10	Yes
Ellison and Burton (2008)	Bio	65	19	0	6	30	Yes
Gregorová *et al.* (2008)	Med	1	1	0	0	0	Yes
Johnson *et al.* (2001)	Med	2	2	4	0	0	Yes
Kumarasamy *et al.* (2013)	Med	12	10	17	15	12	Yes
Montooth *et al.* (2010)	Bio	1	0	7	2	4	Yes
Paull *et al.* (2013)	Med	0	3	0	1	0	No
Pravenec *et al.* (2007)	Med	5	3	2	6	5	Yes
Sackton *et al.* (2003)	Bio	0	2	0	4	2	Yes
Tachibana *et al.* (2013)	Med	1	3	0	0	0	Yes
Wang *et al.* (2014)	Med	1	5	11	4	0	Yes
Yan *et al.* (2010)	Med	22	22	5	1	0	Yes
Yan *et al.* (2011)	Med	9	10	7	1	0	Yes
Yee *et al.* (2013)	Bio	6	0	0	0	0	Yes

The list includes the study type (Type) and the direction of the reported effect (Bio = biological, Med = biomedical). MM « M indicates a significant negative effect of mitochondrial replacement, MM < M indicates a not significant negative effect of mitochondrial replacement, MM = M indicates no observable effects of mitochondrial replacement, MM > M indicates a not significant positive effect of mitochondrial replacement and MM » M indicates a significant positive effect of mitochondrial replacement. MM: mismatched mitochondrial–nuclear genome combination, M: matched mitochondrial–nuclear genome combination. 'Any sig' indicates, whether the study reveals any significant mitochondrial replacement consequences with a clear negative or positive effect.

The difference in effect size between biological studies (0.462, CI: 0.287–0.688) and biomedical studies (0.697, CI: 0.450–0.956) was just marginally significant (−0.233, CI: −0.495 to −0.011) (Supplementary Fig. S3), with effect sizes of biomedical studies being larger than those of biological studies. This indicates that previous claims about overestimated effects in biological studies (HFEA third review of the safety and efficacy report – update 2014, http://hfeaarchive.uksouth.cloudapp.azure.com/www.hfea.gov.uk/8807.html) are incorrect. Effect-size estimates were significantly larger for human (0.939, CI: 0.588–1.342) than non-human (0.463, CI: 0.267–0.679) data (difference: 0.476, CI: 0.127–0.900; I^2 = 0.351, CI: 0.152–0.573). The effect-size estimates were significant for all four trait types, with largest effect-size estimates for metabolic traits (Table III). There was a significant difference in effect-size estimates between metabolic traits and each of the three other trait types, and a marginally significant difference between anatomical and life-history traits, and traits at the gene level and life-history traits (Table III).

Over the course of ontogenetic development, there was no difference between sequential (i.e. between effect sizes of consecutively reported development stages within a study) and absolute (i.e. between a given development stage and the first reported development stage in a study) survival rates (Welch two sample *t*-test: t = 1.851, df = 37.564, P = 0.072). This suggests that either there is one critical developmental stage in which the effect on survival is decisive (rather than an accumulation of small effects), or alternatively, that few studies examined more than one or two developmental stages: with just seven out of 43 studies reporting data for several developmental stages (with 4.141 stages on average in these seven studies), the latter case appears to be a likely explanation, but the former cannot currently be ruled out. Plotting the survival rates on the rank-based development stage (Fig. 2) showed a significant decrease with developmental stage for the sequential (Spearman's rank correlation: ϱ = −0.521, P < 0.001, S = 203900, n = 7 studies) and the absolute (Spearman's rank correlation: ϱ = −0.631, P < 0.001, S = 155 450, n = 7 studies) approach. Within studies, survival rated decreased with progressing development in five cases and increased in one case (zero in two cases) (Supplementary Fig. S4).

Coefficient of variation in biological and biomedical studies

Coefficients of variation in biological and biomedical studies are based on extracted trait values from the studies. Any differences are therefore not comparing effect-size estimates, but the variance (or standard deviance) in the underlying trait measurements. The coefficient of variation for the reported traits was lower in biological studies than in biomedical studies (Kruskal–Wallis ranked sum test, χ^2 = 13.559, df = 1, P < 0.001, median [lower quantile – upper quantile] biological: 0.136 [0.045–0.261], biomedical: 0.157 [0.087–0.294]). It is noteworthy that the reported data include 33 trait measurements with a high variance (coefficient of variation >1) (Fig. 3). Although 22 of these measurements were from biological studies (and only 11 from biomedical studies), biological studies had, on average, a lower coefficient of variation.

The phenotypic effect of biological studies showed no significant difference in the coefficient of variation between matched and mismatched genomes (Kruskal–Wallis ranked sum test, χ^2 = 0.163, df = 1, P = 0.687; matched: 0.121 [0.053–0.245], mismatched: 0.144 [0.042–0.277]). The same was true for biomedical studies (Kruskal–Wallis ranked sum test, χ^2 = 0.002, df = 1, P = 0.969; matched: 0.168 [0.085–0.311], mismatched: 0.158 [0.089–0.286]). Average sample size for each trait measurement in biological studies with mismatched mito-nuclear combinations (i.e. the mitochondrial replacement group) was 37.2 [31.8–42.7], in biological studies with matched combinations (i.e. the control group) 37.6 [27.8–47.3], in biomedical studies with mismatched combinations 10.6 [8.1–13.1] and in biomedical studies with matched combinations 13.7 [10.6–16.8]. Although there was a significant difference in sample size for each trait measurement ($F_{3,964}$ = 32.650, P < 0.001), a post-hoc test revealed that the significant overall effect was caused by the consistent difference in sample sizes between the biological and the biomedical groups. These lower coefficients of variation and larger sample sizes indicate that biological groups achieve a more precise estimate of the trait measurement compared to the biomedical groups.

Table III Effect-size estimates for trait types for the subset directly comparing coevolved to non-coevolved mitochondrial–nuclear genome combinations.

Factor level	Effect-size estimate	95% CI	ES	Studies
Anatomy	0.531	0.317; 0.796	78	9
Gene level	0.660	0.293; 1.204	49	5
Life history	0.318	0.173; 0.492	618	28
Metabolic rate	1.260	0.843; 1.792	259	22
Δ anatom-gene	−0.148	−0.753; 0.309		
Δ anatom-life	0.195	−0.013; 0.406		
Δ anatom-meta	−0.747	−1.343; −0.245		
Δ gene-life	0.343	−0.015; 0.892		
Δ gene-meta	−0.599	−1.369; 0.148		
Δ life-meta	−0.941	−1.478; −0.471		

Consistency I^2 = 0.293, CI: 0.112–0.555. All trait types are significantly different from zero and show hence an effect of mitochondrial replacement. The differences between anatomy and metabolic rate and between metabolic rate and life-history are significant (the 95% CI does not cross zero). The differences between anatomy and life-history and between gene level and life-history are marginally significant (the 95% CI just touches zero). ES: the number of effect-size estimates per trait. Studies: the number of publication we used to extract the data.

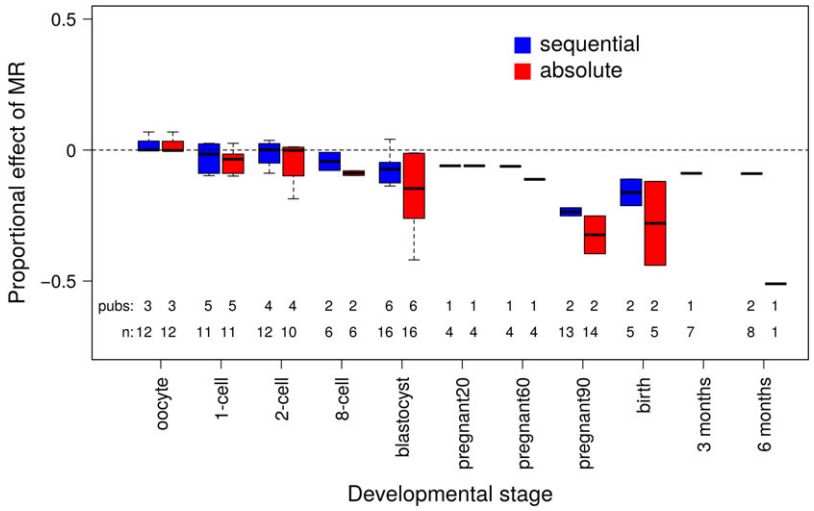

Figure 2 Mitochondrial replacement and average survival rates in the course of development. Median values of survival rates seem to decrease over the course of development stages. Data shown are based on the average survival rates in each study for a given developmental stage. The number of studies is indicated by 'pubs', the overall number of included data points was n. Sequential survival rates are defined as the proportional change in surviving individuals from the control treatment and the mitochondrial replacement (MR) treatment between two consecutively reported development stages in a study. Absolute survival rates are defined as the proportional change between a given development stage and the earliest reported development stage in a study. Bars represent median values, with 25% and 75% boxes, and whiskers are extreme values.

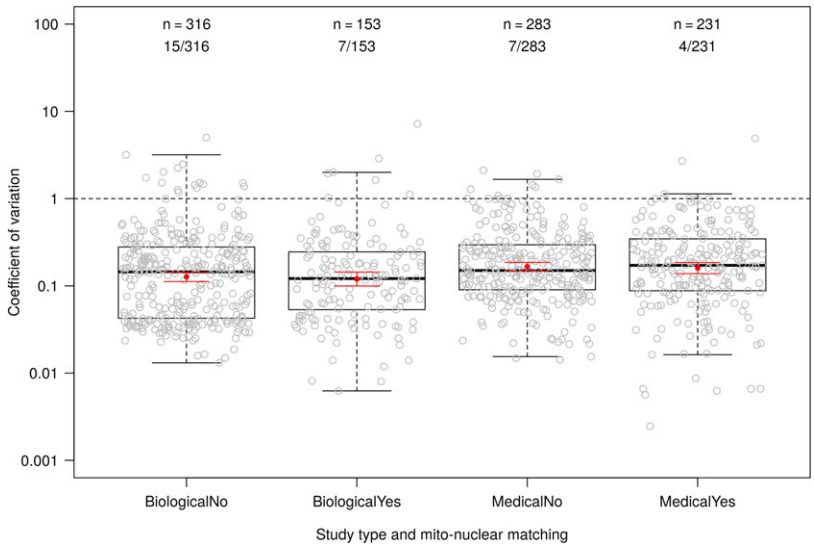

Figure 3 Variability in effect-sizes estimates of studies examining the effect of mitochondrial replacement. Data are grouped by study type (Biological: studies on invertebrate species, without explicit medical focus; Medical: studies on vertebrate species, with direct biomedical focus) and whether the mitochondrial and the nuclear genome were (Yes) or were not (No) matched Values below 1 represent data where the SD was smaller than the mean, indicating that measurements were clustered and the variance of measurements was low. Values above 1 represent data where the SD was larger than the mean, indicating non-clustered measurements and a high variance of measurements. Numbers above the boxes represent the number of reported data points (*n*) and the ratio indicates the amount of measurements with a coefficient of variation larger than 1. Box-plots show the median in a 25–75% quantile box. Whiskers indicate the maximal range with a 1.5-fold inter-quantile range. Points below the lower or above the upper whisker can be considered to be statistical outliers. Overlaid data in red are the mean (circle) and the SE (bars).

To conclude, mitochondrial replacement effects in biological studies appeared to be based more often on shifts of the mean caused by many small effects on measured traits (low coefficients of variation) than in biomedical studies, where large values dominated the effects (large coefficients of variation).

Mitochondrial replacement effects and metabolic rate – Lane Hypothesis

The mixed model with uncorrelated random slope and random intercept for vertebrates and invertebrates provided support for Lane's Hypothesis (Lane, 2011) that disrupting mito-nuclear combinations has a larger effect on organisms with higher BMR (Wald $\chi^2 = 3.865$, $P = 0.049$, Fig. 4). The BMR of humans and bovine are relatively low compared to the other vertebrate species, and human BMR clusters together with the invertebrate species (Fig. 4). Reducing sample size further, by analysing biological and biomedical studies in separate linear regressions (i.e. biological studies and biomedical studies only) revealed, however, no significant effects (biological: $F_{1,4} = 1.854$, $P = 0.245$; biomedical: $F_{1,4} = 1.255$, $P = 0.325$).

Sample sizes and incidences – risk analysis

Out of the 391 trait-by-mito-nuclear-combinations analysed (Table II), disruption of the matched combinations led to significant negative effects in performance for 136 combinations (~35%), and to significant positive effects in 63 combinations (~16%). We make the reader aware that we only used data for the risk analyses where we could unambiguously assign negative or positive effects of mitochondrial replacement on offspring fitness. In biological studies, 83 of 188 trait-by-mito-nuclear-interaction combinations showed significant negative effects of mitochondrial replacement (i.e. 1 in 2.3 combinations, 44.2%). Assuming the extreme case that all mitochondrial replacement effects observed in a particular biological study were caused by a single maternal subject in the mitochondrial replacement treatment group, the incidence of negative mitochondrial replacement effects would fall to 1 in 20.9 (4.8%) (given the average of 9.22 maternal subjects per trait-by-mito-nuclear-interaction combination across studies: 2.3 × 9.22 = 20.9). In the even more extreme case that observed mitochondrial replacement effects were always caused by a single offspring of a single mother, the incidence would be further reduced to 1 in 448.1 (0.2%) (given the average of 21.41 offspring per mother across studies: 20.9 × 21.41 = 448.1). Biological studies measured on average 2.4 traits per study and biomedical studies 8.4. Correcting for this 3.5 higher chance to detect differences in biomedical than biological studies, suggests that negative health consequences should appear in 1 in 129.9 (0.8%) (448.1 × 2.4 ÷ 8.4 = 129.9) mito-nuclear combinations. In other words, if mitochondrial replacement in vertebrates had the same effect as it does in invertebrates, and if all differences in trait means between treatment and control group were caused by only one offspring from one mother, then at least 1 in 130 unrelated offspring (i.e. unique mito-nuclear combinations) born are predicted to experience a negative risk of mitochondrial replacement. For comparison, it is assumed that 1 in 200 persons carry mtDNA mutations, but only 1 in 5000 to 10 000 adults express the disease phenotype, and this means 1 in 25–50 mutation carriers actually express the disease (calculated from Thorburn, 2004; Elliott *et al.*, 2008), that is, 2–4% of the mtDNA mutation carriers.

The figure that 1 out of 130 unrelated offspring is at risk to experience negative effects of mitochondrial replacement can now be used

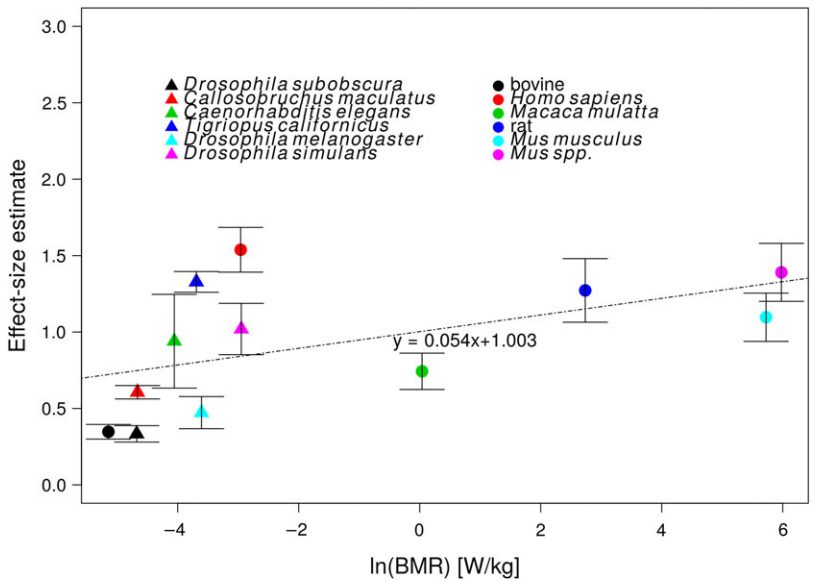

Figure 4 Relationship between effect-size estimates for mitochondrial–nuclear interactions and the basal metabolic rate (*ln*-transformed) of animals. Effect-size estimates for mitochondrial–nuclear interactions increase with increasing basal metabolic rate (BMRs). This supports Lane's Hypothesis that species with higher energetic needs are more susceptible to mitochondrial–nuclear mismatching. BMRs for vertebrate species are mass corrected. Symbols represent species means and error bars are SEs. Dotted line represents the regression on biological and biomedical samples with uncorrelated random slope and intercept. The formula indicates the slope and intercept for the linear regression.

to predict a very conservative incidence of expected negative health and fitness effects in five flagship studies (Tachibana *et al.*, 2009, 2013; Craven *et al.*, 2010; Paull *et al.*, 2013; Wang *et al.*, 2014), whose results provided key support for the approval of mitochondrial replacement therapy in the UK. These studies unfortunately often do not provide exact numbers, but we estimate they examined a total of 177 offspring from 14 mothers. Assuming that the mitochondrial replacement-procedure for each maternal subject resulted in a novel mito-nuclear combination (i.e. mitochondrial haplotype matching did not even occur accidentally), we would expect these five studies to collectively find negative effects of mitochondrial replacement in 0.11 combinations (14 ÷ 129.9 = 0.11, 11%). Hence, at most one of the five flagship studies should reveal negative effects of mitochondrial replacement. However, as Morrow *et al.* (2015) found, three of these studies reported data that would indicate differences between mitochondrial replacement and control treatment, that is, 1 in 1.67 studies, or 1 in 59 offspring (1.7%). This number is considerably higher than that predicted from biological studies (0.2%). The approximately eight-fold increase in observed negative effects may be used to adjust our conservative estimate of incidences being based on a single offspring from a single mother. The observed effects can be explained by a single mother with eight unhealthy offspring, two mothers with four affected offspring each, four mothers with two offspring each experiencing negative effects of mitochondrial replacement, or eight mothers with one offspring each displaying detrimental effects of mitochondrial replacement.

Discussion

We report significant effects of mitochondrial replacement at several levels of organismal performance and health, including gene expression, anatomy, metabolism and life-history. Furthermore, we show a significant difference exists in effect-size estimates between biological and biomedical studies resulting from mitochondrial replacement. Effect-size estimates were larger from biomedical studies than from biological studies. Therefore, our results support the view that biological studies, on invertebrates, may conservatively inform biomedical studies designed to test specific mitochondrial replacement treatment effects. This is true even though studies of vertebrates usually use surgical inter-oocyte nuclear transfer to achieve mitochondrial replacement, whereas studies on invertebrates generally used techniques based on controlled breeding schemes (backcrossing or forced chromosome replacement). One might argue that the larger effect-size estimates in biomedical studies are likely to result from the combined effects of the surgical method *per se* and the effects of creating novel mito-nuclear interactions following mitochondria replacement. This contention is currently difficult to assess because of an absence of biomedical studies that have utilised experimental designs with the power that would enable one to partition out the surgical effects from the mito-nuclear effects. Such a study requires the use of a procedural ('sham' treated) control, in which the surgical technique is used to enucleate an egg of its nuclear material, and then to reconstitute the same egg with the same nuclear chromosomes of the same mother. This was recently attempted by Hyslop *et al.* (2016) who used mitochondrial replacement technology to create two types of human zygotes. 'Autologous' zygotes were created by first removing and then replacing the same pronuclei back into the same zygote. Autologous zygotes were thus procedural controls, and carried precisely the same mito-nuclear genotype as they did prior to the procedure. The authors also created 'heterologous' zygotes, by transferring pronuclei from one zygote to another, either from the same female, or a different female. The authors then obtained single cell gene expression profiles but did not separately examine blastocysts created from the different females in their analysis nor statistically model their data, and therefore it is still not possible to home in on the true effect size caused by mito-nuclear interactions in this study (Morrow and Ingleby, 2017). However, Hyslop *et al.* (2016) did find lower rates of blastocyst formation in the heterologous zygotes than autologous zygotes, which suggests mito-nuclear interactions as prime candidates in exerting the negative effects.

To further assess the likely role of surgery *per se*, relative to mito-nuclear interactions, in shaping the effect size following mitochondrial replacement, we were able to partition 17 effect-size estimates obtained following mitochondrial replacement using non-surgical techniques, and 31 effect-size estimates obtained following mitochondrial replacement using surgical techniques, from six studies in mice. We found no evidence for a significant difference in life-history effect-size estimates between the two methods (0.072 [−1.004–0.833]) (Supplementary Fig. S5). Thus, the use of surgical versus non-surgical techniques does not seem to alter the effect sizes following mitochondrial replacement even when limiting the comparisons to studies of the same species for the same kind of trait. This strongly suggests, therefore, that the effects of surgery *per se* will make only a minor contribution to effect sizes generally following mitochondrial replacement, and that the large effect-size estimates observed in biomedical studies are likely to be largely attributable to mito-nuclear interactions. This result clearly challenges the view that any health effects attributable to mito-nuclear interactions in humans will be insignificant, of theoretical interest only, or lower than in model organisms (Tanaka *et al.*, 2013; Chinnery *et al.*, 2014; Greenfield *et al.*, 2017). To sum, none of the human studies in our meta-analysis used sham-treated control groups, and those non-human studies that included sham-treated control groups applied backcrossing, not surgical techniques. Future studies on humans need to include sham-treated control groups to separate unintended surgical from epistatic effects in offspring. In addition to this, it should be pointed out that we found both larger-than-predicted effect sizes in human studies, as well as larger-than-predicted incidences of negative effects in five of the studies that have been used as flagship evidence in support of progressing mitochondrial replacement therapy to the clinic (Tachibana *et al.*, 2009, 2013; Craven *et al.*, 2010; Paull *et al.*, 2013; Wang *et al.*, 2014).

The strong effects of mitochondrial replacement on metabolic traits were not unexpected given that mitochondria are the main producer of energy in cells. Although effect-size estimates differed substantially among the four reported and analysed trait types, our results clearly show that mitochondrial replacement affects traits at any level from gene expression to life-history traits. Why some trait types are affected more than others need to be further investigated. One possible explanation is that some trait types are under stronger selection against mito-nuclear mismatches than others. This means that some trait types can better cope with mismatches and show, therefore, larger effect-size estimates for mitochondrial replacement effects.

We also found lower coefficients of variation in biological than biomedical studies – indicating that the conclusions of biological studies are less often biased by outliers, or effects caused by single or few unusual individuals. This suggests that biological studies might therefore be better suited to draw generalised epidemiological conclusions about mito-nuclear interaction effects following mitochondrial replacement than biomedical studies. Biomedical studies may reveal essential interaction effects for individual cases (as the general focus of medical studies are individuals) but might be less suited for use in assessing the epidemiological risks of mitochondrial replacement (see also Morrow *et al.*, 2015). Taking together the considerations about the effects of mitochondrial-replacement techniques and differences in the coefficients of variation between biological and biomedical studies, we can be confident that models in biological studies are indeed suitable to estimate the magnitude of health effect after mitochondrial replacement. In concert with the type of health effects after mitochondrial replacement (Reinhardt *et al.*, 2013; Marom *et al.*, 2017; Morrow and Camus, 2017) and the frequency of effects (see 'Sample sizes and incidences – risk analysis'), our analysis provides a quantitative understanding of the effects of mitochondrial replacement.

A potential approach suggested to reduce risk of mitochondrial replacement is haplotype matching between the prospective mother and the donor. Here we show that the original nuclear genome paired with the original mitochondrial genome is often the best performing combination. Haplotype matching would then require high similarity between the coding regions of two mitochondrial genomes (with exception of the pathogenic mutation, of course), and perhaps also for the non-coding regions, given that previous studies have shown that even single point mutations can have severe effects on the offspring fitness when coexpressed in one but not another nuclear genome (Xu *et al.*, 2008; Clancy *et al.*, 2011; Camus *et al.*, 2015; Patel *et al.*, 2016; Loewen and Ganetzky, 2018). The alternative, to experimentally map the effects of mitochondrial haplotypes expressed in different nuclear backgrounds, would require that an impossibly large number of mito-nuclear genome combinations be tested. Whether 'the best possible' (or available) match between the mother's and the donor's haplotype could be sufficient to prevent unintended effects of mitochondrial replacement, therefore, needs further investigation – something that the fields of biology and biomedicine can both contribute to.

Our result of consistent effects of mito-nuclear interactions on the phenotype differs from a recent study by Eyre-Walker (2017), and a recent non-systematic review (Greenfield *et al.*, 2017). Eyre-Walker (2017) concluded that no evidence exists for negative effects associated with mito-nuclear mismatch effects in animals, and hence no evidence for the coevolution of mitochondrial and nuclear genomes. The difference between the results and conclusions of the Eyre-Walker (2017) study relative to ours is likely attributable to the larger sample size in our case, and by the fact that Eyre-Walker (2017) assigned the effects on traits a directionality, and disregarded data for traits in which directionality with fitness outcomes was difficult to assign. Assigning directionality has some important consequences. First, directionality of an effect (in terms of its effect on organismal fitness) is in many cases difficult to infer, unless fertility or mortality is measured (i.e. cases where the direction and form of natural selection is known). For example, developmental rate (or time) can be harmful if too fast as well as if too slow, body weight harmful if too large or too small, and the relationship of these traits with evolutionary fitness can change across species. Eyre-Walker (2017), however, assigned fast development and large body weight to be beneficial in the context of mitochondrial replacement, even though large body weight is associated with obesity and its related comorbidities, including shortened longevity in mammals (Kiilerich *et al.*, 2016) and fast development, would intuitively be maladaptive, for instance, in eutherians. Second, directionality becomes dubious if positive effects on one trait are outweighed by negative effects on another (e.g. Latorre-Pellicer *et al.*, 2016). Hence, our opinion is that the direction of the effect is not important when considering the likely health effects of mitochondrial replacement, but only the strength of the effect.

We also explored the evidence that effect sizes consistently differed in magnitude across developmental stages, or as a function of the BMR of the species under study. These analyses suffered from low sample sizes, but they do suggest that survival rates tend to decline as ontogenetic development proceeds. We encourage further empirical tests of the hypothesis that mito-nuclear interactions may exert age-specific effects throughout ontogeny. For example, Tachibana *et al.* (2009) presented treatment differences starting from spindle transfer to blastocyst. While not reporting any significant differences between any of the developmental stages (our 'sequential' survival rate), the overall effects accumulated and caused a significant treatment effect (Morrow *et al.*, 2015). Similarly, Yan *et al.* (2011) examined differences in treatments starting from the oocyte to 6-month post-natal survival. Although they did not find significant effects at each sequential step, the effects of the mitochondrial replacement treatment resulted in a significant decrease in survival rate over the course of ontogeny (our 'absolute' survival rate). The lack of significant differences between the different trait types in our meta-analysis (which in some respects also reflect a course over development, as different traits types emerge at different developmental stages) can be explained by the fact that most studies did not allow a direct comparison of survival rates over the course of development (only seven out of 43 studies). Our analyses of proportional changes in survival after mitochondrial replacement treatments in these seven studies revealed that effects of mitochondrial replacement can accumulate over development, and are only detectable when survival rates are compared between the initial and the last stage investigated in the study. That is, simply comparing survival between sequential steps can lead to the misleading conclusion that there is no significant effect of mitochondrial replacement.

The increase of effect sizes after mitochondrial replacement for species with higher metabolic rates is intriguing, because it suggests that species of high metabolic demand are associated with stronger selection for compatible mito-nuclear interactions, and that disruption of these compatible interactions, via mitochondrial replacement, led to greater effects on phenotypes. While this is consistent with the Lane Hypothesis, we note that information on BMR was only available for a relatively small number of species. More detailed studies with further species and with data from activity-based metabolic rates (e.g. during energetically demanding functions, such as during flight for insects) are desirable to get a complete picture of how an increase in metabolic needs across taxa may increase the magnitude of selection for precise mito-nuclear compatibility. Such studies could also reveal whether tasks requiring extremely high energetic demands

across short bursts of time might be alternatively powered by energy sources other than ATP production from oxidative phosphorylation (such as glycolysis). Such alternative energy sources hence might not affect the selection of mito-nuclear interactions since they would not cause oxidative stress and damage in the system. As a consequence, support for the Lane Hypothesis may help in identifying a meaningful donor in mitochondrial replacement studies. The presence of alternative energy sources does, however, not necessarily invalidate the Lane Hypothesis.

We note that effect-size estimates from a systematic meta-analysis might be used to help predict the incidence of potentially negative effects expected once mitochondrial replacement is introduced in the clinic. A risk analysis based on the likelihood of mito-nuclear incompatibilities has not previously been attempted. We attempted a risk analysis here, based on the most conservative assumptions possible (negative effects in any one study were attributed to negative effects manifesting in one offspring of one mother). When effect sizes estimated from biological studies were applied to studies that have previously been used to support legislation for clinical application of mitochondrial replacement therapy in the UK, we estimated that around one in every 59–130 children born to mitochondrial replacement therapy might be expected to suffer negative health effects brought about by mito-nuclear interactions as a result of the procedure. For comparison, it is assumed that 1 in 200 people carries a mitochondrial mutation, but only 1 in 5000 to 10 000 adults expresses a disease phenotype. This means 1 in 25–50 mutation carriers actually express the mitochondrial disease (calculated form Thorburn, 2004; Elliott *et al.*, 2008). Relaxing our assumption, such that more than one offspring per mother or more mothers would be causing the treatment effect, would lead to an increasing risk estimate of negative effects of mitochondrial replacement therapy. Indeed, relaxing our conservative assumptions as described in the 'Results' section (i.e. that more than a single mother and a single offspring were affected) would lead to a match of the predicted and observed occurrence of negative effects after mitochondrial replacement. For example, if we consider on average two mothers per study each giving birth to four offspring expressing the disease phenotype, the prediction would be that about 1.6% of the offspring following mitochondrial replacement would show unintended negative effects of the treatment. A previous estimate suggesting that several hundred women in the USA and UK could benefit from mitochondrial replacement therapy implicitly assumed that mitochondrial replacement therapy has zero risk (Gorman *et al.*, 2015) (because the majority of benefiting women and their offspring would be asymptomatic). Our study may provide useful data to refine this estimate.

Finally, we note that after the deadline of our literature search several papers were published that could not be considered in the formal meta-analysis but that show mito-nuclear epistatic effects on health such as impaired male fertility (Patel *et al.*, 2016) and Leigh syndrome effects in *D. melanogaster* (Loewen and Ganetzky, 2018), hearing loss in humans (Kullar *et al.*, 2016, 2017) and effects of mitochondrial replacement on blastocyst performance in humans (Hyslop *et al.*, 2016); the latter study was discussed above.

In conclusion, our results indicate that offspring trait expression is routinely affected by mito-nuclear interactions in animals, and that these effects extend to, and might even be stronger in, humans. Phenotypic modification as a result of mito-nuclear interactions is therefore a distinct risk of mitochondrial replacement therapy, and should be factored into future risk assessments of the technique. Our results further strengthen the call for haplotype matching as a strategy to mitigate the risks of mitochondrial replacement therapy.

Supplementary data

Supplementary data are available at *Human Reproduction Update* online.

Acknowledgements

We thank the two reviewers for comments on previous versions of the manuscript.

Authors' roles

R.D., D.K.D., E.H.M. and K.R. collected data, R.D. and K.R. analysed data, R.D., D.K.D., E.H.M. and K.R. interpreted the results and wrote the manuscript.

Funding

The study was financially supported by the DFG Excellence Initiative (via TU Dresden Zukunftskonzept to K.R. and R.D.), by a Royal Society University Research Fellowship (to E.H.M.) and the Australian Research Council (to D.K.D.).

References

Baris TZ, Wagner DN, Dayan DI, Du X, Blier PU, Pichaud N, Oleksiak MF, Crawford DL. Evolved genetic and phenotypic differences due to mitochondrial-nuclear interactions. *PLoS Genet* 2017;**13**:e1006517.

Barreto FS, Burton RS. Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. *Mol Biol Evol* 2013;**30**:310–314.

Bellizzi D, Taverna D, D'Aquila P, De Blasi S, De Benedictis G. Mitochondrial DNA variability modulates mRNA and intra-mitochondrial protein levels of HSP60 and HSP75: experimental evidence from cybrid lines. *Cell Stress Chaperones* 2009;**14**:265–271.

Burgstaller JP, Johnston IG, Jones NS, Albrechtova J, Kolbe T, Vogl C, Futschik A, Mayrhofer C, Klein D, Sabitzer S *et al*. mtDNA segregation in heteroplasmic tissues is common in vivo and modulated by haplotype differences and developmental stage. *Cell Rep* 2014;**7**:2031–2041.

Burton RS, Barreto FS. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol Ecol* 2012;**21**:4942–4957.

Callaway E. Three-person embryos may fail to vanquish mutant mitochondria. *Nature* 2016;**533**:445–446.

Camus MF, Wolf JB, Morrow EH, Dowling DK. Single nucleotides in the mtDNA sequence modify mitochondrial molecular function and are associated with sex-specific effects on fertility and aging. *Curr Biol* 2015;**25**:2717–2722.

Chamberlain SA, Hovick SM, Dibble CJ, Rasmussen NL, Van Allen BG, Maitner BS, Ahern JR, Bell-Dereske LP, Roy CL, Meza-Lopez M *et al*. Does phylogeny matter? Assessing the impact of phylogenetic information in ecological meta-analysis. *Ecol Lett* 2012;**15**:627–636.

Chang CC, Rodriguez J, Ross J. Mitochondrial-nuclear epistasis impacts fitness and mitochondrial physiology of interpopulation *Caenorhabditis briggsae* hybrids. *G3 (Bethesda)* 2016;**6**:209–219.

Chatelain EH, Pichaud N, Ballard JWO, Tanguay RM, Morrow G, Blier PU. Functional conservatism among *Drosophila simulans* flies experiencing different

thermal regimes and mitochondrial DNA introgression. *J Exp Zool B Mol Dev Evol* 2011;**316B**:188–198.
Chinnery PF, Craven L, Mitalipov S, Stewart JB, Herbert M, Turnbull DM. The challenges of mitochondrial replacement. *PLoS Genet* 2014;**10**:e1004315. 10.1371/journal.pgen.1004315.
Clancy DJ. Variation in mitochondrial genotype has substantial lifespan effects which may be modulated by nuclear background. *Aging Cell* 2008;**7**:795–804.
Clancy DJ, Hime GR, Shirras AD. Cytoplasmic male sterility in *Drosophila melanogaster* associated with a mitochondrial CYTB variant. *Heredity* 2011;**107**:374–376.
Connallon T, Camus MF, Morrow EH, Dowling DK. Coadaptation of mitochondrial and nuclear genes, and the cost of mother's curse. *Proc R Soc B* 2018;**285**:20172257.
Craven L, Tuppen HA, Greggains GD, Harbottle SJ, Murphy JL, Cree LM, Murdoch AP, Chinnery PF, Taylor RW, Lightowlers RN *et al*. Pronuclear transfer in human embryos to prevent transmission of mitochondrial DNA disease. *Nature* 2010;**465**:82–85.
Del Re A 2013. *compute.es: Compute Effect Sizes*. URL http://cran.r-project.org/web/packages/compute.es.
Deng JH, Li Y, Park JS, Wu J, Hu P, Lechleiter J, Bai Y. Nuclear suppression of mitochondrial defects in cells without the ND6 subunit. *Mol Cell Biol* 2006;**26**:1077–1086.
Dobler R, Rogell B, Budar F, Dowling DK. A meta-analysis of the strength and nature of cytoplasmic genetic effects. *J Evol Biol* 2014;**27**:2021–2034.
Dowling DK. Evolutionary perspectives on the links between mitochondrial genotype and disease phenotype. *Biochim Biophys Acta* 2014;**1840**:1393–1403.
Dowling DK, Abiega KC, Arnqvist G. Temperature-specific outcomes of cytoplasmic-nuclear interactions on egg-to-adult development time in seed beetles. *Evolution* 2007a;**61**:194–201.
Dowling DK, Friberg U, Arnqvist G. A comparison of nuclear and cytoplasmic genetic effects on sperm competitiveness and female remating in a seed beetle. *J Evol Biol* 2007b;**20**:2113–2125. 10.1111/j.1420-9101.2007.01433.x.
Dowling DK, Friberg U, Hailer F, Arnqvist G. Intergenomic epistasis for fitness: within-population interactions between cytoplasmic and nuclear genes in *Drosophila melanogaster*. *Genetics* 2007c;**175**:235–244.
Dowling DK, Friberg U, Lindell J. Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol Evol* 2008;**23**:546–554.
Dowling DK, Meerupati T, Arnqvist G. Cytonuclear interactions and the economics of mating in seed beetles. *Am Nat* 2010;**176**:131–140.
Elliott HR, Samuels DC, Eden JA, Relton CL, Chinnery PF. Pathogenic mitochondrial DNA mutations are common in the general population. *Am J Hum Genet* 2008;**83**:254–260.
Ellison CK, Burton RS. Disruption of mitochondrial function in interpopulation hybrids of *Tigriopus californicus*. *Evolution* 2006;**60**:1382–1391.
Ellison CK, Burton RS. Interpopulation hybrid breakdown maps to the mitochondrial genome. *Evolution* 2008;**62**:631–638.
Eyre-Walker A. Mitochondrial replacement therapy: are mito-nuclear interactions likely to be a problem? *Genetics* 2017;**205**:1365–1372.
Ghelli A, Porcelli AM, Zanna C, Vidoni S, Mattioli S, Barbieri A, Iommarini L, Pala M, Achilli A, Torroni A *et al*. The background of mitochondrial DNA haplogroup J increases the sensitivity of Leber's hereditary optic neuropathy cells to 2, 5-hexanedione toxicity. *PLoS ONE* 2009;**4**:e7922.
Gorman GS, Chinnery PF, DiMauro S, Hirano M, Koga Y, McFarland R, Suomalainen A, Thorburn DR, Zeviani M, Turnbull DM. Mitochondrial diseases. *Nat Rev Dis Primers* 2016;**2**:16080.
Gorman GS, Grady JP, Ng Y, Schaefer AM, McNally RJ, Chinnery PF, Yu-Wai-Man P, Herbert M, Taylor RW, McFarland R *et al*. Mitochondrial donation – how many women could benefit? *N Engl J Med* 2015;**372**:885–887.
Greenfield A, Braude P, Flinter F, Lovell-Badge R, Ogilvie C, Perry AC. Assisted reproductive technologies to prevent human mitochondrial disease transmission. *Nat Biotechnol* 2017;**35**:1059–1068.
Greenspan RJ *Fly Pushing: The Theory and Practice of Drosophila Genetics*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 2004.
Gregorová S, Divina P, Storchova R, Trachtulec Z, Fotopulosova V, Svenson KL, Donahue LR, Paigen B, Forejt J. Mouse consomic strains: exploiting genetic divergence between *Mus m. musculus* and *Mus m. domesticus* subspecies. *Genome Res* 2008;**18**:509–515.
Hadfield JD. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J Stat Softw* 2010;**33**:1–22.
Hao H, Morrison LE, Moraes CT. Suppression of a mitochondrial tRNA gene mutation phenotype associated with changes in the nuclear. *Hum Mol Genet* 1999;**8**:1117–1124.
Havird JC, Hall MD, Dowling DK. The evolution of sex: a new hypothesis based on mitochondrial mutational erosion. *Bioessays* 2015;**37**:951–958.
Hedges L. Distribution theory for Glass's estimator of effect size and related estimators. *J Educ Behav Stat* 1981;**6**:107–128.
Hereford J, Hansen TF, Houle D. Comparing strengths of directional selection: how strong is strong? *Evolution* 2004;**58**:2133–2143.
Herman RK, Albertson DG, Brenner S. Chromosome rearrangements in *Caenorhabditis elegans*. *Genetics* 1976;**83**:91–105.
Hill GE. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecol Evol* 2016;**6**:5831–5842.
Holmbeck MA, Donner JR, Villa-Cuesta E, Rand DM. A *Drosophila* model for mitonuclear diseases generated by an incompatible interaction between tRNA and tRNA synthetase. *Dis Model Mech* 2015;**8**:843–U530.
Hyslop LA, Blakeley P, Craven L, Richardson J, Fogarty NM, Fragouli E, Lamb M, Wamaitha SE, Prathalingam N, Zhang Q *et al*. Towards clinical application of pronuclear transfer to prevent mitochondrial DNA disease. *Nature* 2016;**534**:383.
Immonen E, Collet M, Goenaga J, Arnqvist G. Direct and indirect genetic effects of sex-specific mitonuclear epistasis on reproductive ageing. *Heredity* 2016a;**116**:329–338.
Immonen E, Ronn J, Watson C, Berger D, Arnqvist G. Complex mitonuclear interactions and metabolic costs of mating in male seed beetles. *J Evol Biol* 2016b;**29**:360–370.
James AC, Ballard JWO. Mitochondrial genotype affects fitness in *Drosophila simulans*. *Genetics* 2003;**164**:187–194.
Jelić M, Arnqvist G, Novičić ZK, Kenig B, Tanasković M, Anelković M, Stamenković-Radak M. Sex-specific effects of sympatric mitonuclear variation on fitness in *Drosophila subobscura*. *BMC Evol Biol* 2015;**15**:135.
Johnson KR, Zheng QY, Bykhovskaya Y, Spirina O, Fischel-Ghodsian N. A nuclear-mitochondrial DNA interaction affecting hearing impairment in mice. *Nat Genet* 2001;**27**:191–194.
Kenney MC, Chwa M, Atilano SR, Falatoonzadeh P, Ramirez C, Malik D, Tarek M, del Carpio JC, Nesburn AB, Boyer DS *et al*. Inherited mitochondrial DNA variants can affect complement, inflammation, and apoptosis pathways: insights into mitochondrial-nuclear interactions. *Hum Mol Genet* 2014a;**23**:3537–3551.
Kenney MC, Chwa M, Atilano SR, Falatoonzadeh P, Ramirez C, Malik D, Tarek M, del Carpio JC, Nesburn AB, Boyer DS *et al*. Molecular and bioenergetic differences between cells with African versus European inherited mitochondrial DNA haplogroups: implications for population susceptibility to diseases. *Biochim Biophys Acta* 2014b;**1842**:208–219.
Kenney MC, Chwa M, Atilano SR, Pavlis JM, Falatoonzadeh P, Ramirez C, Malik D, Hsu T, Woo G, Soe K *et al*. Mitochondrial DNA variants mediate energy production and expression levels for CFH, C3 and EFEMP1 genes: implications for age-related macular degeneration. *PLoS ONE* 2013;**8**:e54339.
Kiilerich P, Myrmel LS, Fjære E, Hao Q, Hugenholtz F, Sonne SB, Derrien M, Pedersen LM, Petersen RK, Mortensen A *et al*. Effect of a long-term high-protein diet on survival, obesity development, and gut microbiota in mice. *Am J Physiol Endocrinol Metab* 2016;**310**:E886–E899.
Kingsolver JG, Diamond SE, Siepielski AM, Carlson SM. Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evol Ecol* 2012;**26**:1101–1118.
Kleiber M. Body size and metabolism. *Hilgardia* 1932;**6**:315–353.
Kropáčková L, Piálek J, Gergelits V, Forejt J, Reifová R. Maternal-foetal genomic conflict and speciation: no evidence for hybrid placental dysplasia in crosses between two house mouse subspecies. *J Evol Biol* 2015;**28**:688–698.
Kullar PJ, Gomez-Duran A, Gammage PA, Garone C, Minczuk M, Golder Z, Wilson J, Montoya J, Häkli S, Kärppä M *et al*. Heterozygous SSBP1 start loss mutation co-segregates with hearing loss and the m.1555A G mtDNA variant in a large multigenerational family. *Brain* 2017;**141**:55–62.
Kullar PJ, Quail J, Lindsey P, Wilson JA, Horvath R, Yu-Wai-Man P, Gorman GS, Taylor RW, Ng Y, McFarland R *et al*. Both mitochondrial DNA and mitonuclear gene mutations cause hearing loss through cochlear dysfunction. *Brain* 2016;**139**:e33.

Kumarasamy S, Gopalakrishnan K, Abdul-Majeed S, Partow-Navid R, Farms P, Joe B. Construction of two novel reciprocal conplastic rat strains and characterization of cardiac mitochondria. *Am J Physiol Heart Circ Physiol* 2013;**304**:H22–H32.

Kurbalija Novičić Z, Immonen E, Jelić M, Anelković M, Stamenković-Radak M, Arnqvist G. Within-population genetic effects of mtDNA on metabolic rate in *Drosophila subobscura*. *J Evol Biol* 2015;**28**:338–346.

Lane N. The costs of breathing. *Science* 2011;**334**:184–185.

Latorre-Pellicer A, Moreno-Loshuertos R, Lechuga-Vieco AV, Sánchez-Cabo F, Torroja C, Acín-Pérez R, Calvo E, Aix E, González-Guerra A, Logan A *et al*. Mitochondrial and nuclear DNA matching shapes metabolism and healthy ageing. *Nature* 2016;**535**:561–565.

Liau WS, Gonzalez-Serricchio AS, Deshommes C, Chin K, LaMunyon CW. A persistent mitochondrial deletion reduces fitness and sperm performance in heteroplasmic populations of *C. elegans*. *BMC Genet* 2007;**8**:8.

Lin TK, Lin HY, Chen SD, Chuang YC, Chuang JH, Wang PW, Huang ST, Tiao MM, Chen JB, Liou CW. The creation of cybrids harboring mitochondrial haplogroups in the Taiwanese population of ethnic Chinese background: an extensive in vitro tool for the study of mitochondrial genomic variations. *Oxid Med Cell Longev* 2012;**2012**:824275.

Loewen CA, Ganetzky B. Mito-nuclear interactions affecting lifespan and neurodegeneration in a Drosophila model of Leigh syndrome. *Genetics* 2018;300818.

Marom S, Friger M, Mishmar D. MtDNA meta-analysis reveals both phenotype specificity and allele heterogeneity: a model for differential association. *Sci Rep* 2017;**7**:43449.

Messina FJ, Jones JC. Inheritance of traits mediating a major host shift by a seed beetle, *Callosobruchus maculatus* (Coleoptera: Chrysomelidae: Bruchinae). *Ann Entomol Soc Am* 2011;**104**:808–815.

Milot E, Moreau C, Gagnon A, Cohen AA, Brais B, Labuda D. Mother's curse neutralizes natural selection against a human genetic disease over three centuries. *Nat Ecol Evol* 2017;**1**:1400.

Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;**6**:e1000097.

Montooth KL, Meiklejohn CD, Abt DN, Rand DM. Mitochondrial-nuclear epistasis affects fitness within species but does not contribute to fixed incompatibilities between species of Drosophila. *Evolution* 2010;**64**:3364–3379.

Moreno-Loshuertos R, Acín-Pérez R, Fernández-Silva P, Movilla N, Pérez-Martos A, de Cordoba SR, Gallardo ME, Enríquez JA. Differences in reactive oxygen species production explain the phenotypes associated with common mouse mitochondrial DNA variants. *Nat Genet* 2006;**38**:1261–1268.

Morrissey MB, Hadfield JD. Directional selection in temporally replicated studies is remarkably consistent. *Evolution* 2012;**66**:435–442.

Morrow EH, Camus MF. Mitonuclear epistasis and mitochondrial disease. *Mitochondrion* 2017;**35**:119–122.

Morrow EH, Ingleby FC. Detecting differential gene expression in blastocysts following pronuclear transfer. *BMC Res Notes* 2017;**10**:97.

Morrow EH, Reinhardt K, Wolff JN, Dowling DK. Risks inherent to mitochondrial replacement. *EMBO Rep* 2015;**16**:541–544.

Mossman JA, Biancani LM, Zhu CT, Rand DM. Mitonuclear epistasis for development time and its modification by diet in *Drosophila*. *Genetics* 2016a;**203**:463–484.

Mossman JA, Tross JG, Li N, Wu Z, Rand DM. Mitochondrial-nuclear interactions mediate sex-specific transcriptional profiles in *Drosophila*. *Genetics* 2016b;**204**:613–630.

Nagao Y, Totsuka Y, Atomi Y, Kaneda H, Lindahl KF, Imai H, Yonekawa H. Decreased physical performance of congenic mice with mismatch between the nuclear and the mitochondrial genome. *Genes Genet Syst* 1998;**73**:21–27.

Nakada K, Sato A, Yoshida K, Morita T, Tanaka H, Inoue SI, Yonekawa H, Hayashi JI. Mitochondria-related male infertility. *Proc Natl Acad Sci* 2006;**103**:15148–15153.

Neupane J, Ghimire S, Vandewoestyne M, Lu YC, Gerris J, Van Coster R, Deroo T, Deforce D, Vansteelandt S, De Sutter P *et al*. Cellular heterogeneity in the level of mtDNA heteroplasmy in mouse embryonic stem cells. *Cell Rep* 2015;**13**:1304–1309.

Neupane J, Vandewoestyne M, Ghimire S, Lu Y, Qian C, Van Coster R, Gerris J, Deroo T, Deforce D, De Sutter P *et al*. Assessment of nuclear transfer techniques to prevent the transmission of heritable mitochondrial disorders without compromising embryonic development competence in mice. *Mitochondrion* 2014;**18**:27–33.

Paliwal S, Fiumera AC, Fiumera HL. Mitochondrial-nuclear epistasis contributes to phenotypic variation and coadaptation in natural isolates of *Saccharomyces cerevisiae*. *Genetics* 2014;**198**:1251–1265.

Patel MR, Miriyala GK, Littleton AJ, Yang H, Trinh K, Young JM, Kennedy SR, Yamashita YM, Pallanck LJ, Malik HS. A mitochondrial DNA hypomorph of cytochrome oxidase specifically impairs male fertility in *Drosophila melanogaster*. *eLife* 2016;**5**:e16923.

Paull D, Emmanuele V, Weiss KA, Treff N, Stewart L, Hua H, Zimmer M, Kahler DJ, Goland RS, Noggle SA *et al*. Nuclear genome transfer in human oocytes eliminates mitochondrial DNA variants. *Nature* 2013;**493**:632–637.

Picard M, Wallace DC, Burelle Y. The rise of mitochondria in medicine. *Mitochondrion* 2016;**30**:105–116.

Popescu AA, Huber KT, Paradis E. ape 3.0: new tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* 2012;**28**:1536–1537.

Pravenec M, Hyakukoku M, Houstek J, Zidek V, Landa V, Mlejnek P, Miksik I, Dudová-Mothejzikova K, Pecina P, Vrbacky M *et al*. Direct linkage of mitochondrial genome variation to risk factors for type 2 diabetes in conplastic strains. *Genome Res* 2007;**17**:1319–1326.

R Development Core Team 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Rand DM, Haney RA, Fry AJ. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol* 2004;**19**:645–653.

Reardon S. Mutated mitochondria could hold back stem-cell therapies. *Nature* 2016. doi:10.1038/nature.2016.20849.

Reinhardt K, Dowling DK, Morrow EH. Mitochondrial replacement, evolution, and the clinic. *Science* 2013;**341**:1345–1346.

Revell LJ. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* 2012;**3**:217–223.

Rishishwar L, Jordan IK. Implications of human evolution and admixture for mitochondrial replacement therapy. *BMC Genomics* 2017;**18**:140.

Sackton TB, Haney RA, Rand DM. Cytonuclear coadaptation in *Drosophila*: disruption of cytochrome c oxidase activity in backcross genotypes. *Evolution* 2003;**57**:2315–2325.

Savage VM, Gillooly J, Woodruff W, West G, Allen A, Enquist BJ, Brown J. The predominance of quarter-power scaling in biology. *Funct Ecol* 2004;**18**:257–282.

Sloan DB, Havird JC, Sharbrough J. The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol Ecol* 2017;**26**:2212–2236.

Spees JL, Olson SD, Whitney MJ, Prockop DJ. Mitochondrial transfer between cells can rescue aerobic respiration. *Proc Natl Acad Sci* 2006;**103**:1283–1288.

Strauss KA, DuBiner L, Simon M, Zaragoza M, Sengupta PP, Li P, Narula N, Dreike S, Platt J, Procaccio V *et al*. Severity of cardiomyopathy associated with adenine nucleotide translocator-1 deficiency correlates with mtDNA haplogroup. *Proc Natl Acad Sci* 2013;**110**:3453–3458.

Tachibana M, Amato P, Sparman M, Woodward J, Sanchis DM, Ma H, Gutierrez NM, Tippner-Hedges R, Kang E, Lee HS *et al*. Towards germline gene therapy of inherited mitochondrial diseases. *Nature* 2013;**493**:627–631.

Tachibana M, Sparman M, Sritanaudomchai H, Ma H, Clepper L, Woodward J, Li Y, Ramsey C, Kolotushkina O, Mitalipov S. Mitochondrial gene replacement in primate offspring and embryonic stem cells. *Nature* 2009;**461**:367–372.

Tanaka AJ, Sauer MV, Egli D, Kort DH. Harnessing the stem cell potential: the path to prevent mitochondrial disease. *Nat Med* 2013;**19**:1578–1579.

Thorburn D. Mitochondrial disorders: prevalence, myths and advances. *J Inherited Metab Dis* 2004;**27**:349–362.

Tummers B 2006. *DataThief III* http://datathief.org/ . URL http://datathief.org/.

Wang T, Sha H, Ji D, Zhang HL, Chen D, Cao Y, Zhu J. Polar body genome transfer for preventing the transmission of inherited mitochondrial diseases. *Cell* 2014;**157**:1591–1604.

Wolf DP, Mitalipov N, Mitalipov S. Mitochondrial replacement therapy in reproductive medicine. *Trends Mol Med* 2015;**21**:68–76.

Wolff JN, Ladoukakis ED, Enríquez JA, Dowling DK. Mitonuclear interactions: evolutionary consequences over multiple biological scales. *Phil Trans R Soc B* 2014;**369**:20130443.

Xu H, DeLuca SZ, O'farrell PH. Manipulating the metazoan mitochondrial genome with targeted restriction enzymes. *Science* 2008;**321**:575–577.

Yamada M, Emmanuele V, Sanchez-Quintero MJ, Sun B, Lallos G, Paull D, Zimmer M, Pagett S, Prosser RW, Sauer MV *et al*. Genetic drift can

compromise mitochondrial replacement by nuclear transfer in human oocytes. *Cell Stem Cell* 2016;**18**:749–754.

Yan H, Yan Z, Ma Q, Jiao F, Huang S, Zeng F, Zeng Y. Association between mitochondrial DNA haplotype compatibility and increased efficiency of bovine interspecies cloning. *J Genet Genomics* 2011;**38**:21–28.

Yan ZH, Zhou YY, Fu J, Jiao F, Zhao LW, Guan PF, Huang SZ, Zeng YT, Zeng F. Donor-host mitochondrial compatibility improves efficiency of bovine somatic cell nuclear transfer. *BMC Dev Biol* 2010;**10**:1.

Yee WKW, Sutton KL, Dowling DK. In vivo male fertility is affected by naturally occurring mitochondrial haplotypes. *Curr Biol* 2013;**23**:R55–R56.

Zhang J, Liu H, Luo S, Lu Z, Chávez-Badiola A, Liu Z, Yang M, Merhi Z, Silber SJ, Munné S *et al*. Live birth derived from oocyte spindle transfer to prevent mitochondrial disease. *Reprod Biomed Online* 2017;**34**:361–368.

Zheng B, Sage M, Cai WW, Thompson DM, Tavsanli BC, Cheah YC, Bradley A. Engineering a mouse balancer chromosome. *Nat Genet* 1999;**22**:375–378.