

MitImpact 3: modeling the residue interaction network of the Respiratory Chain subunits

Stefano Castellana¹, Tommaso Biagini¹, Francesco Petrizzelli^{1,2}, Luca Parca¹,
Noemi Panzironi², Viviana Caputo², Angelo Luigi Vescovi³, Massimo Carella⁴ and
Tommaso Mazza^{1,*}

¹Laboratory of Bioinformatics, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG), 71013, Italy, ²Department of Experimental Medicine, Sapienza University of Rome, Rome 00161, Italy, ³ISBRMIT Institute for Stem Cell Biology, Regenerative Medicine and Innovative Therapies, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG), 71013, Italy and ⁴Laboratory of Medical Genetics, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG) 71013, Italy

Received August 08, 2020; Revised October 14, 2020; Editorial Decision October 15, 2020; Accepted December 08, 2020

ABSTRACT

Numerous lines of evidence have shown that the interaction between the nuclear and mitochondrial genomes ensures the efficient functioning of the OXPHOS complexes, with substantial implications in bioenergetics, adaptation, and disease. Their interaction is a fascinating and complex trait of the eukaryotic cell that MitImpact explores with its third major release. MitImpact expands its collection of genomic, clinical, and functional annotations of all non-synonymous substitutions of the human mitochondrial genome with new information on putative Compensated Pathogenic Deviations and co-varying amino acid sites of the Respiratory Chain subunits. It further provides evidence of energetic and structural residue compensation by techniques of molecular dynamics simulation. MitImpact is freely accessible at <http://mitimpact.css-mendel.it>.

INTRODUCTION

The genetics of the mitochondria received a high impulse approximately thirty years ago when a missense variant in the mitochondrial MT-ND4 gene was discovered to cause Leber's Optic Neuropathy (LHON, MIM #535000) (1). Since then, several defects in mito-nuclear genes encoding structural proteins or proteins involved in mitochondrial functions were associated with mitochondrial disorders (2–4) and complex multifactorial diseases (5,6).

Around 1500 proteins appear to be necessary for the mitochondrial biogenesis and metabolism (2,7); the human mitochondrial genome (mtDNA) encodes only 13 of them. Then, despite their differences in terms of size, structure, complexity, and transmission mode, mitochondrial and nu-

clear genomes do necessarily interact, mainly involving the Respiratory Chain (RC) subunits (8–11), with substantial implications in bioenergetics, adaptation, and disease (12–14). These subunits are among the most complex components of the cellular proteome, which are continuously being investigated from structural (15–18), evolutionary (14), and genetic (2) perspectives.

Variations on mitochondrial proteins associate with a broad range of phenotypes that differ in terms of age at onset, single/multi-organ involvement and inheritance pattern (2). To date, hundreds of mutations in the nuclear and mitochondrial genomes were associated with mitochondrial disorders, although the diagnostic rate ranged only from 20% to 60% in several large cohorts (19–22). It could be assumed that a portion of the undiagnosed cases and cases with variable phenotypes might be caused by composites of underestimated interacting variants of uncertain significance.

Interaction is indeed a fundamental property of all biological systems, which is currently under active study (14,23–25). The evolution, structure, and function of a protein depend on the transient and stable interactions among residues of the same protein or interacting proteins. While more research was devoted to studying the epistasis of nuclear variants (26–29), it is important to notice here that at least 4–5% of all nuclear-encoded pathogenic amino acid substitutions are present as fixed residues within one or more mammalian species. The fact that such *Compensated Pathogenic Deviations* (CPDs) (30) were found in supposedly healthy, non-human species as reference amino acids implies that modulatory mechanisms must co-exist and that this should equally hold for nuclear and mitochondrial proteins.

Numerous lines of evidence showed that the nuclear and mitochondrial genomes need to co-evolve to ensure the efficient functioning of the OXPHOS complexes (14,31–33). The way this happens is fascinating and complicated since it

*To whom correspondence should be addressed. Tel: +39 06 44160526; Fax: +39 06 44160548; Email: t.mazza@css-mendel.it

depends on their structure, size, complexity, evolution and inheritance patterns (34,35). However, its comprehension is a critical goal since it would explain much of the mutation penetrance and disease phenotype variability.

The vast availability of genomics data, computational tools, and specialized databases contributed significantly to achieving these findings. Since its release in 2015, MitImpact (36) is one piece of this toolbox, when it first provided the most extensive set of pre-computed pathogenicity predictions for all possible non-synonymous variants of the mtDNA. It quickly became the reference resource to this regard, continuously expanding its content with the predictions of several third-party algorithms, and then including, two years later, the pathogenicity assessments of the best performing meta-predictor, APOGEE (37).

The third and current milestone of MitImpact paves the way to the study of a fascinating and complex trait of the eukaryotic cell: the interaction between amino acids of mitochondrial and nuclear proteins (38,39), by reviewing the interactions that may occur between the amino acids encoded by the mitochondrial genome. Although being complicated, studying these interactions is, in fact, possible using systematic protein-protein interaction screening (40), at the cost of expensive and time-consuming assays. Even targeted modification of the mitochondrial DNA has proven to be overly complicated *in-vitro* (41,42), although recent studies have introduced new gene-editing possibilities (43). Computational algorithms, as well, have faced numerous issues and limitations, including the reduced availability of genome sequences of species that are not economically relevant and of resolved protein structures.

In such a complex field of research, MitImpact expands its knowledge base with new information on protein co-variation, structural consequences of variants at co-varying sites, compensated pathogenic deviations, and with speculations on possible structural and evolutionary compensatory mechanisms.

MATERIALS AND METHODS

In this third major release (Figure 1), MitImpact updates its set of annotations with new content concerning dbSNP v153 (44), ClinVar v153 (45), COSMIC v90 (46) and MITOMAP (February 2020 release) (47), and adds the pathogenicity predictions of three new software packages: SNPdryad (48), DEOGEN2 (49), and MitoClass.1 (50).

CPDs (51,52) were inferred as described in Supplementary File S1B. In brief, the reference protein sequences of the 13 mitochondrial subunits were downloaded from NCBI Organelle (53), restricting the search to the Mammalia clade. The homologous sequences were aligned using ClustalOmega v1.2 (54). Sequences with >50% dissimilarity with the human sequences were discarded. The dissimilarity threshold was set to 30% for MT-ATP8 only. All human missense pathogenic variants were then retrieved from MITOMAP and ClinVar, and localized in the 13 ortholog protein alignments. We considered CPDs the variants that were reference residues in at least one non-human sequence, where at least seven out of the ten surroundings residues, five upstream and five downstream, were identical with the human sequence. This algorithm was inspired by the work

of Azevedo and colleagues (52). The final set of CPDs has been made available in MitImpact together with the specification of the species names and NCBI reference sequence identifiers.

We then used I-COMS to study the co-variation of mitochondrial residues (55). The analysis was conducted with two algorithms implemented in I-COMS, cMI (56) and EVFold-mfDCA (57), and was restricted to the orthologs belonging to the Mammalian clade (NCBI37 Taxon ID: 40674). Interactions were considered *inter-protein* whenever residues were located in different proteins of Complex I (twenty-one pairwise analyses), Complex IV (MT-CO1 + MT-CO2, MT-CO2 + MT-CO3, MT-CO1 + MT-CO3) or Complex V (MT-ATP6 + MT-ATP8), and *intra-protein* when both residues were located in the same protein (thirteen pairs), and then not in MT-ND1, which was concatenated by default to all proteins. From the TOP500 inter-protein pairs, in terms of cMI or mfDCA scores, we retained only the pairs whose residues did not fall in the same concatenated protein. From the TOP500 intra-protein pairs, we considered only the pairs where neither residue fell in the MT-ND1 concatenated sequence. Intra-protein interactions were assessed through EVMutation (58) also. We considered the TOP500 scoring pairs, where the higher a score, the stronger was the association between two residues (further details in Supplementary File S1A).

Inter- and intra-protein relationships between co-varying variants were also investigated from an energetic point of view. Atomic coordinates of all human mtDNA-encoded RC proteins, except MT-ATP8, were obtained by homology modeling using the SWISS-MODEL server (59). The MatchMaker extension of UCSF Chimera (60) was used to model the interactions between proteins by superimposing the relative *Bos taurus* structures, the evolutionary closest *Mammalia* whose proteins were available in the Protein Data Bank at the time of this analysis. We performed a minimization step conducted into the membrane environment to refine the interface regions and reduce local errors, such as clashes and Ramachandran outliers. Then, FoldX 4.0 (61) was employed to calculate the free-energy changes upon mutation of the residues lying at the interaction interface (average $\Delta\Delta G$ of 10 replicas) and, consequently, assess the protein stability and protein-protein interaction energy alteration. Alternative amino acids that caused a $\Delta\Delta G$ to exceed the suggested cutoffs (± 0.61 kcal/mol) for the single mutant were tagged as disruptive (62). Pairs of mutants with $\Delta\Delta G$ conservatively close to zero ($< \pm 0.1$ kcal/mol) were considered structurally compensative.

In parallel, we selected the putatively co-varying pairs of residues, where at least one of the pairs was reported as pathogenic in MITOMAP, and both were located in the interaction interface of proteins belonging to the same RC complex. We mapped these residues onto the 3D structure of the human RC I transmembrane arm and investigated the interacting properties of the wild-type complex and the single and double-mutated complexes for all the possible amino acids combinations of each co-varying pair through 50 ns-long molecular dynamics simulations performed using AMBER 18 (63), as described in (64,65). The root-mean-square deviation (RMSD), root mean square fluctuation (RMSF), binding energy components and per-residue

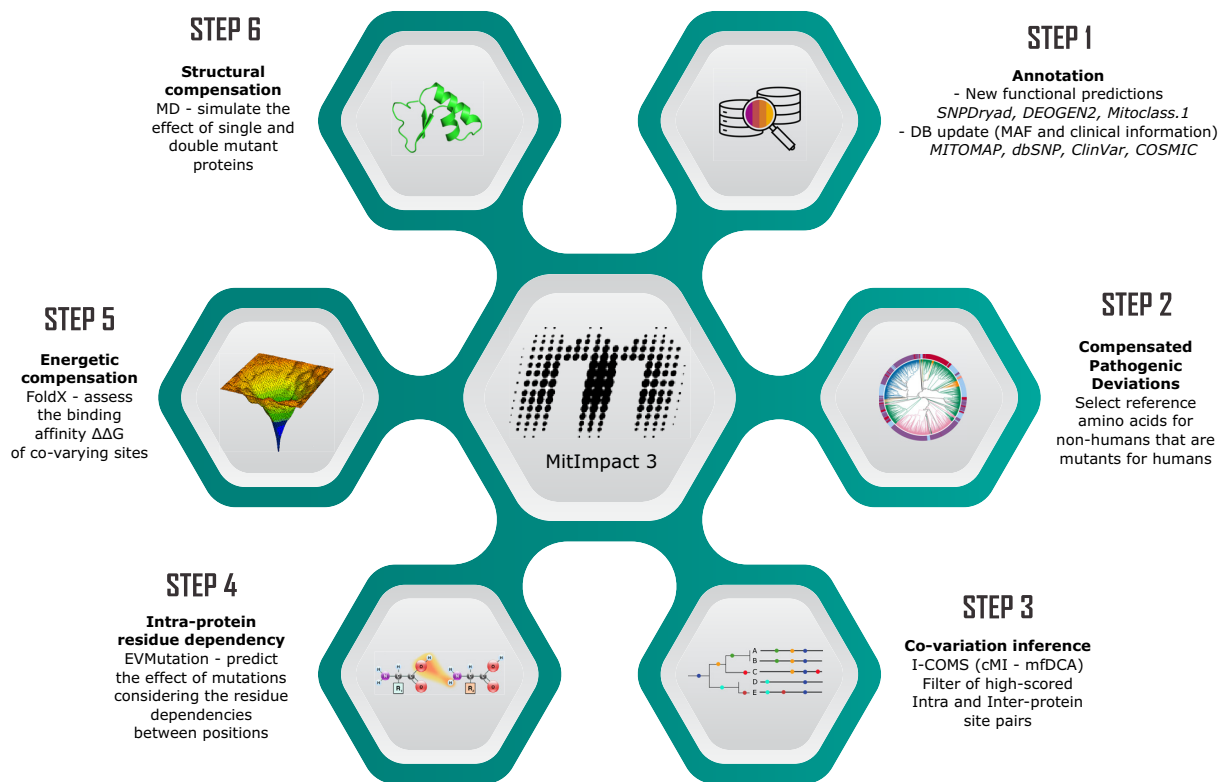


Figure 1. Content schema of MitImpact 3.

hydrogen bonds were measured on the simulated trajectories using GROMACS 4.5 (66) and the *g_mmpbsa* package (67). The essential protein motions' collective coordinates were inferred by principal component analysis of the atomic fluctuations. The 3D movies of the molecular dynamics simulations were obtained using the *g_anaig* GROMACS tool.

RESULTS

Annotations of individual variants

MitImpact 3 stores functional assessments of sixteen predictors, seven meta-predictors, and four cancer-specific predictors for 24,190 amino acid variations in the 13 protein-coding genes of the *Homo sapiens* mitochondrion (NC_012920.1).

CPD. Forty-one of these variants were classified as CPDs (Supplementary Data S2) following the workflow described in Supplementary File S1B, where six had already been described in two previous screenings (Table 1).

Only four out of the 41 CPDs were found in at least 10% of the screened mammalian orthologs (Supplementary Data S2), with the alternative residue (Isoleucine) in position 39 in MT-ND1 (MI.10998), which is reported in MITOMAP to be associated with maternally-inherited diabetes and deafness (MIM #520000), resulting in being the reference amino acid in 360/674 MT-ND1 orthologs (53.4%).

Residue interaction networks

Residue co-variation. The total number of inter-protein co-varying pairs of sites was 7,017 (Supplementary Data

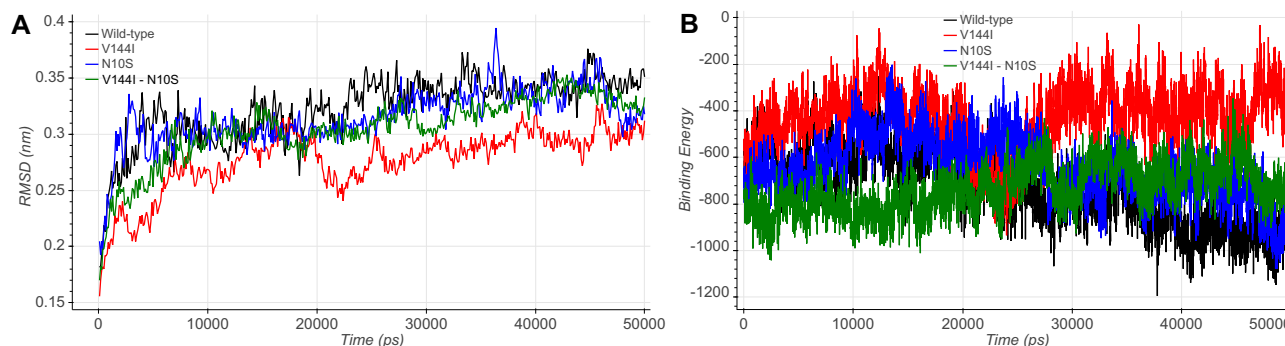
S3, worksheet 'cov inter'), 4,285 were found with the cMI method and 2,732 with mfDCA. Forty-seven pairs were in common between the two methods (first 47 rows of the 'cov common' worksheet). The intra-protein co-variation analysis generated 6,468 pairs ('cov intra'), with 3,038 cMI, 3,430 mfDCA and 339 shared pairs ('cov common'). Given the low congruence of the two analyses (details in Supplementary File S1A), we opted to consider the union of the results of the two methods (i.e. 4,285 and 2,732 inter- and 3,038 and 3,430 intra-protein pairs) for the following energetic analysis steps.

Energetically compensatory residues. The co-varying sites were then studied energetically. All the pairs encompassing residues that compensate for the energetic destabilization caused by the other members of the pairs were split into inter-protein ($N = 256$; Supplementary Data S3, 'comp inter'), intra-protein ($N = 3,039$; 'comp intra') and interface-protein ($N = 219$; 'comp interface'). This last subclass contains double mutants at the interaction interface between RC subunits, including nuclear proteins. An example case study where the energetic destabilization caused by a variant in a gene (i.e. S45P in ND3) can or cannot be restored by a variant in another gene (i.e. N126S and A64P in ND1, respectively) is described in Supplementary File S1C.

Structurally compensatory residues. The co-varying amino-acid site pairs containing at least one clinically relevant residue were further analyzed by molecular dynamics simulation. Comparing the wild-type protein complexes with the single or double mutants in terms of RMSD,

Table 1. Clinically significant amino acid variants in humans that have been identified in mammalian ortholog proteins by this and other published works. Details in Supplementary Data S2

MitImpact ID	Protein variant	Species	Ref.
MI.3124	MT-CO1:p.L196I	<i>Glis glis</i> (Rodentia)	(14)
MI.8247	MT-CO3:p.F251L	Many Primates (genus: <i>Cercocebus</i> , <i>Macaca</i> , <i>Cercopithecus</i>)	(14)
MI.9927	MT-CYB:p.G251S	Koala, many Primates (genus: e.g., <i>Pongo</i> , <i>Colobus</i> , <i>Trachypithecus</i>)	(14)
MI.11461	MT-ND1:p.V113A	<i>Galeopterus variegatus</i> , <i>Chlorocebus pygerythrus</i> , <i>Chlorocebus cynosures</i> (Primates)	(68)
MI.7210	MT-CO2:p.V91A	Several Primates and Cetaceans	(68)
MI.23991	MT-ND6:p.I33V	Many Monotremes and Primates	(68)

**Figure 2.** (A) The RMSD plot shows how the V144I mutant in MT-ND1 (in red) exhibited lower RMSD values during the simulation time with respect to the wild-type protein complex (in black). N10S in MT-ND3 (in blue) and the double mutant V144I:N10S (in green) showed comparable RMSD values with the wild-type. (B) The binding energy plot showed how the V144I variant (in red) causes an alteration of the binding energy values during the entire simulation time compared with the wild-type, while the simultaneous presence of both variants in the MT-ND1:MT-ND3 interface (green) resembles the wild-type condition (black).

RMSF, binding energy, essential dynamics and the number of hydrogen bonds established or abolished during the entire simulations, we found some cases where the effects of pathogenic alleles have chances to be compensated by co-occurring alleles. For example, the MT-ND3 S45P variant, which is associated with the Leigh syndrome (MIM #256000), is predicted by the previous energetic analysis to be stabilized by three co-occurring other residues located in position 126 of the interacting MT-ND3 protein. Similarly, the MT-ND1 V144I variant is associated with Leber hereditary optic neuropathy (LHON, MIM #535000). The analysis of the molecular dynamics trajectories of this mutant combined with the N10S variant on the interacting MT-ND3 protein showed how the co-occurrence of both mutations could partially restore the fundamental dynamical properties, such as stability and binding affinity (Figure 2). Further details on this were given in **Case study 2: rs201513497** available from the website. Together with 50 other pairs, the results of all MD analyses were stored and made interactively available in a dedicated section of MitImpact.

DISCUSSION

MitImpact 3 updates its knowledge base and becomes the largest collection of functional predictions of mitochondrial non-synonymous variants. Among all available sixteen pathogenicity predictors, seven meta-predictors, four cancer-specific predictors, our APOGEE (37) stands out for being the reference tool for assessing the pathogenicity of non-synonymous variants in molecular diagnosis (69). However, the most significant update regards the introduc-

tion of a new set of putative CPDs, the calculation of the amino acid sites that significantly co-vary within the RC subunits, and the assessment of the energetic and structural compensation caused by pairs of co-varying residues using techniques of molecular dynamics simulation.

These techniques have limitations, as standard procedures to benchmark their results do not exist. The immediate consequence is that different computational methods can yield incongruent results to an extent that is not currently accountable. Caution must also be used with the analyses based on molecular dynamics simulation. These yield results that are indeed more accurate and detailed than those of software predictors, but under the strong hypothesis that the causes of a variant's pathogenic character are ascribed to structural alterations (Supplementary File S1A).

Nevertheless, MitImpact 3 is one of the first specialized bioinformatics resources that adds information on binary epistatic events and their consequences to classical variant annotations and functional predictions, thereby disclosing sites that significantly co-vary through the mammalian phylogeny. Moreover, it stores the results of a great computational effort, which consisted of studying hundreds of co-varying amino acid pairs from energetic and molecular dynamics points of view. The aim of this study, which is still in progress, is that of assessing the structural consequences of known pathogenic variants and pinpointing any eventual compensatory effects. Within the frame of this task, we will create a submission form on the website where interested users can request to analyze energetically and structurally their sets of variants searching for epistatic interactions.

Future developments will deal with the low congruence of methods to assess the co-variation of RC subunits residues.

We will use other algorithms and will define a more reliable consensus set of results. We have also planned to release an updated version of APOGEE (37), which will be trained on an updated and finely curated set of known variants. In particular, APOGEE will aggregate other packages' predictions as in version 1 together with co-evolution information from EVMutation, cMI, mfDCA, and $\Delta\Delta G$ folding energy.

MitImpact has been supporting molecular and computational biologists, geneticists, and medical doctors for five years in conducting their work, and will keep doing it freely.

DATA AVAILABILITY

MitImpact is freely available at <http://mitimpact.css-mendel.it>. Raw I-COMS score matrices and protein alignments are available at https://mitimpact.css-mendel.it/static/supporting_data/ICOMS_mitimpact3.rar.

Atomic coordinates and structure factors of the human RC I transmembrane arm have been retrieved from the PDB with the following accession number 5xtc. Energetic compensation was calculated on smaller bovine models retrieved from PDB with the following accession numbers: 1be3, 1bgy, 1bmf, 1cow, 1e1q, 1e1r, 1e79, 1efr, 1h8e, 1h8h, 1l0l, 1l0n, 1mab, 1nbn, 1ntk, 1ntm, 1ntz, 1nu1, 1occ, 1oco, 1ocr, 1ocz, 1ohh, 1pp9, 1ppj, 1qcr, 1qo1, 1sqb, 1sqp, 1sqq, 1sqv, 1sqx, 1v54, 1v55, 1w0j, 1w0k, 2a06, 2bcc, 2ck3, 2cly, 2dyr, 2dys, 2eij, 2eik, 2eil, 2eim, 2ein, 2f43, 2fyu, 2jdi, 2jiz, 2jj1, 2jj2, 2occ, 2v7q, 2w6e, 2w6f, 2w6g, 2w6h, 2w6i, 2w6j, 2wss, 2xnd, 2y69, 2ybb, 2zxw, 3abk, 3abl, 3abm, 3ag1, 3ag2, 3ag3, 3ag4, 3asn, 3aso, 3bcc, 3wg7, 3x2q, 4asu, 4b2q, 4d6t, 4d6u, 4tsf, 4tt3, 4yxw, 4z1m, 5ara, 5are, 5arh, 5ari, 5b1a, 5b1b, 5b3s, 5fij, 5fik, 5fil, 5gpn, 5iy5, 5klv, 5lc5, 5ldw, 5ldx, 5lnk, 5luf, 5nmi, 5w97, 5wau, 5x19, 5x1b, 5x1f, 5xdq.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We gratefully acknowledge NVIDIA Corporation and the COSMIC and Amber teams for supporting this research.

FUNDING

Italian Ministry of Health (Ricerca Corrente 2018–2020); '5 × 1000' voluntary contribution. Funding for open access charge: Italian Ministry of Health.

Conflict of interest statement. None declared.

REFERENCES

- Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A., Schurr, T.G., Lezza, A.M., Elsas, L.J. 2nd and Nikoskelainen, E.K. (1988) Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science*, **242**, 1427–1430.
- Gorman, G.S., Chinnery, P.F., DiMauro, S., Hirano, M., Koga, Y., McFarland, R., Suomalainen, A., Thorburn, D.R., Zeviani, M. and Turnbull, D.M. (2016) Mitochondrial diseases. *Nat. Rev. Dis. Primers*, **2**, 16080.
- Ghezzi, D. and Zeviani, M. (2018) Human diseases associated with defects in assembly of OXPHOS complexes. *Essays Biochem.*, **62**, 271–286.
- El-Hattab, A.W. and Scaglia, F. (2016) Mitochondrial cardiomyopathies. *Front. Cardiovasc. Med.*, **3**, 25.
- Witte, M.E., Mahad, D.J., Lassmann, H. and van Horssen, J. (2014) Mitochondrial dysfunction contributes to neurodegeneration in multiple sclerosis. *Trends Mol. Med.*, **20**, 179–187.
- Wallace, D.C. (2012) Mitochondria and cancer. *Nat. Rev. Cancer*, **12**, 685–698.
- Palmfeldt, J. and Bross, P. (2017) Proteomics of human mitochondria. *Mitochondrion*, **33**, 2–14.
- Bar-Yaacov, D., Blumberg, A. and Mishmar, D. (2012) Mitochondrial-nuclear co-evolution and its effects on OXPHOS activity and regulation. *Biochim. Biophys. Acta*, **1819**, 1107–1111.
- Lloyd, R.E. and McGeehan, J.E. (2013) Structural analysis of mitochondrial mutations reveals a role for bigenomic protein interactions in human disease. *PLoS One*, **8**, e69003.
- Osada, N. and Akashi, H. (2012) Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome C oxidase complex. *Mol. Biol. Evol.*, **29**, 337–346.
- Aledo, J.C., Valverde, H., Ruiz-Camacho, M., Morilla, I. and Lopez, F.D. (2014) Protein-protein interfaces from cytochrome c oxidase I evolve faster than nonbinding surfaces, yet negative selection is the driving force. *Genome Biol. Evol.*, **6**, 3064–3076.
- Wallace, D.C. (2015) Mitochondrial DNA variation in human radiation and disease. *Cell*, **163**, 33–38.
- Gershoni, M., Levin, L., Ovadia, O., Toiw, Y., Shani, N., Dadon, S., Barzilai, N., Bergman, A., Atzmon, G., Wainstein, J. et al. (2014) Disrupting mitochondrial-nuclear coevolution affects OXPHOS complex I integrity and impacts human health. *Genome Biol. Evol.*, **6**, 2665–2680.
- Azevedo, L., Carneiro, J., van Asch, B., Moleirinho, A., Pereira, F. and Amorim, A. (2009) Epistatic interactions modulate the evolution of mammalian mitochondrial respiratory complex components. *BMC Genomics*, **10**, 266.
- Vinothkumar, K.R., Zhu, J. and Hirst, J. (2014) Architecture of mammalian respiratory complex I. *Nature*, **515**, 80–84.
- Chaban, Y., Boekema, E.J. and Dudkina, N.V. (2014) Structures of mitochondrial oxidative phosphorylation supercomplexes and mechanisms for their stabilisation. *Biochim. Biophys. Acta*, **1837**, 418–426.
- Signes, A. and Fernandez-Vizarra, E. (2018) Assembly of mammalian oxidative phosphorylation complexes I-V and supercomplexes. *Essays Biochem.*, **62**, 255–270.
- Stroud, D.A., Surgenor, E.E., Formosa, L.E., Reljic, B., Frazier, A.E., Dibley, M.G., Osellame, L.D., Stait, T., Beilharz, T.H., Thorburn, D.R. et al. (2016) Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature*, **538**, 123–126.
- Calvo, S.E., Compton, A.G., Hershman, S.G., Lim, S.C., Lieber, D.S., Tucker, E.J., Laskowski, A., Garone, C., Liu, S., Jaffe, D.B. et al. (2012) Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci. Transl. Med.*, **4**, 118ra110.
- Lieber, D.S., Calvo, S.E., Shanahan, K., Slate, N.G., Liu, S., Hershman, S.G., Gold, N.B., Chapman, B.A., Thorburn, D.R., Berry, G.T. et al. (2013) Targeted exome sequencing of suspected mitochondrial disorders. *Neurology*, **80**, 1762–1770.
- Haack, T.B., Haberberger, B., Frisch, E.M., Wieland, T., Iuso, A., Gorza, M., Strecker, V., Graf, E., Mayr, J.A., Herberg, U. et al. (2012) Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. *J. Med. Genet.*, **49**, 277–283.
- Taylor, R.W., Pyle, A., Griffin, H., Blakely, E.L., Duff, J., He, L., Smertenko, T., Alston, C.L., Neeve, V.C., Best, A. et al. (2014) Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies. *JAMA*, **312**, 68–77.
- Starr, T.N. and Thornton, J.W. (2016) Epistasis in protein evolution. *Protein Sci.*, **25**, 1204–1218.
- Jordan, D.M., Frangakis, S.G., Golzio, C., Cassa, C.A., Kurtzberg, J. and Task Force for Neonatal, G. Task Force for Neonatal, G., Davis, E.E., Sunyaev, S.R. and Katsanis, N. (2015) Identification of cis-suppression of human disease mutations by comparative genomics. *Nature*, **524**, 225–229.
- Xu, J. and Zhang, J. (2014) Why human disease-associated residues appear as the wild-type in other species: genome-scale structural evidence for the compensation hypothesis. *Mol. Biol. Evol.*, **31**, 1787–1792.

26. Mastroianno, S., Palumbo, P., Castellana, S., Leone, M.P., Massaro, R., Potenza, D.R., Mazza, T., Russo, A., Castori, M., Carella, M. *et al.* (2020) Double missense mutations in cardiac myosin-binding protein C and myopalladin genes: A case report with diffuse coronary disease, complete atrioventricular block, and progression to dilated cardiomyopathy. *Ann. Noninvasive Electrocardiol.*, **25**, e12687.
27. Gazzo, A.M., Daneels, D., Cilia, E., Bonduelle, M., Abramowicz, M., Van Dooren, S., Smits, G. and Lenaerts, T. (2016) DIDA: a curated and annotated digenic diseases database. *Nucleic Acids Res.*, **44**, D900–D907.
28. Deltas, C. (2018) Digenic inheritance and genetic modifiers. *Clin. Genet.*, **93**, 429–438.
29. König, E., Volpato, C.B., Motta, B.M., Blankenburg, H., Picard, A., Pramstaller, P., Casella, M., Rauhe, W., Pompilio, G., Meraviglia, V. *et al.* (2017) Exploring digenic inheritance in arrhythmogenic cardiomyopathy. *BMC Med. Genet.*, **18**, 145.
30. Kondrashov, A.S., Sunyaev, S. and Kondrashov, F.A. (2002) Dobzhansky-Muller incompatibilities in protein evolution. *PNAS*, **99**, 14878–14883.
31. Mishmar, D., Ruiz-Pesini, E., Mondragon-Palomino, M., Procaccio, V., Gaut, B. and Wallace, D.C. (2006) Adaptive selection of mitochondrial complex I subunits during primate radiation. *Gene*, **378**, 11–18.
32. Levin, L., Blumberg, A., Barshad, G. and Mishmar, D. (2014) Mito-nuclear co-evolution: the positive and negative sides of functional ancient mutations. *Front. Genet.*, **5**, 448.
33. Sunnucks, P., Morales, H.E., Lamb, A.M., Pavlova, A. and Greening, C. (2017) Integrative approaches for studying mitochondrial and nuclear genome co-evolution in oxidative phosphorylation. *Front. Genet.*, **8**, 25.
34. Castellana, S., Vicario, S. and Saccone, C. (2011) Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein coding genes. *Genome Biol. Evol.*, **3**, 1067–1079.
35. Havird, J.C. and Sloan, D.B. (2016) The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial versus nuclear genomes. *Mol. Biol. Evol.*, **33**, 3042–3053.
36. Castellana, S., Ronai, J. and Mazza, T. (2015) MitImpact: an exhaustive collection of pre-computed pathogenicity predictions of human mitochondrial non-synonymous variants. *Hum. Mutat.*, **36**, E2413–E2422.
37. Castellana, S., Fusilli, C., Mazzoccoli, G., Biagini, T., Capocéfalo, D., Carella, M., Vescovi, A.L. and Mazza, T. (2017) High-confidence assessment of functional impact of human mitochondrial non-synonymous genome variations by APOGEE. *PLoS Comput. Biol.*, **13**, e1005628.
38. Caporali, L., Iommarini, L., La Morgia, C., Olivieri, A., Achilli, A., Maresca, A., Valentino, M.L., Capristo, M., Tagliavini, F., Del Dotto, V. *et al.* (2018) Peculiar combinations of individually non-pathogenic missense mitochondrial DNA variants cause low penetrance Leber's hereditary optic neuropathy. *PLoS Genet.*, **14**, e1007210.
39. Morrow, E.H. and Camus, M.F. (2017) Mitonuclear epistasis and mitochondrial disease. *Mitochondrion*, **35**, 119–122.
40. Barshad, G., Zlotnikov-Poznianski, N., Gal, L., Schuldiner, M. and Mishmar, D. (2019) Disease-causing mutations in subunits of OXPHOS complex I affect certain physical interactions. *Sci. Rep.*, **9**, 9987.
41. Hamalainen, R.H. (2014) Induced pluripotent stem cell-derived models for mtDNA diseases. *Methods Enzymol.*, **547**, 399–415.
42. Fayzulin, R.Z., Perez, M., Kozhukhar, N., Spadafora, D., Wilson, G.L. and Alexeyev, M.F. (2015) A method for mutagenesis of mouse mtDNA and a resource of mouse mtDNA mutations for modeling human pathological conditions. *Nucleic Acids Res.*, **43**, e62.
43. Mok, B.Y., de Moraes, M.H., Zeng, J., Bosch, D.E., Kotrys, A.V., Raguram, A., Hsu, F., Radey, M.C., Peterson, S.B., Mootha, V.K. *et al.* (2020) A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature*, **583**, 631–637.
44. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
45. Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W. *et al.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*, **46**, D1062–D1067.
46. Forbes, S.A., Bindal, N., Bamford, S., Cole, C., Kok, C.Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A. *et al.* (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.*, **39**, D945–D950.
47. Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Sarmady, M., Procaccio, V. and Wallace, D.C. (2013) mtDNA variation and analysis using mitomap and mitomaster. *Curr. Protoc. Bioinformatics*, **44**, 1.23.1–1.23.26.
48. Wong, K.C. and Zhang, Z. (2014) SNPdryad: predicting deleterious non-synonymous human SNPs using only orthologous protein sequences. *Bioinformatics*, **30**, 1112–1119.
49. Raimondi, D., Tanyalcin, I., Ferte, J., Gazzo, A., Orlando, G., Lenaerts, T., Rooman, M. and Vranken, W. (2017) DEOGEN2: prediction and interactive visualization of single amino acid variant deleteriousness in human proteins. *Nucleic Acids Res.*, **45**, W201–W206.
50. Martin-Navarro, A., Gaudio-Simon, A., Alvarez-Jarreta, J., Montoya, J., Mayordomo, E. and Ruiz-Pesini, E. (2017) Machine learning classifier for identification of damaging missense mutations exclusive to human mitochondrial DNA-encoded polypeptides. *BMC Bioinformatics*, **18**, 158.
51. Baresic, A. and Martin, A.C. (2011) Compensated pathogenic deviations. *Biomol. Concepts*, **2**, 281–292.
52. Azevedo, L., Mort, M., Costa, A.C., Silva, R.M., Quelhas, D., Amorim, A. and Cooper, D.N. (2016) Improving the in silico assessment of pathogenicity for compensated variants. *Eur. J. Hum. Genet.: EJHG*, **25**, 2–7.
53. Coordinators, N.R. (2016) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **44**, D7–D19.
54. Sievers, F. and Higgins, D.G. (2014) Clustal omega. *Curr. Protoc. Bioinformatics*, **48**, 3.13.1–3.13.16.
55. Iserte, J., Simonetti, F.L., Zea, D.J., Teppa, E. and Marino-Buslje, C. (2015) I-COMS: Interprotein-CORrelated Mutations Server. *Nucleic Acids Res.*, **43**, W320–W325.
56. Buslje, C.M., Santos, J., Delfino, J.M. and Nielsen, M. (2009) Correction for phylogeny, small number of observations and data redundancy improves the identification of coevolving amino acid pairs using mutual information. *Bioinformatics*, **25**, 1125–1131.
57. Marks, D.S., Colwell, L.J., Sheridan, R., Hopf, T.A., Pagnani, A., Zecchina, R. and Sander, C. (2011) Protein 3D structure computed from evolutionary sequence variation. *PLoS One*, **6**, e28766.
58. Hopf, T.A., Ingraham, J.B., Poelwijk, F.J., Scharfe, C.P., Springer, M., Sander, C. and Marks, D.S. (2017) Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.*, **35**, 128–135.
59. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L. *et al.* (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.*, **46**, W296–W303.
60. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
61. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res.*, **33**, W382–W388.
62. Usmanova, D.R., Bogatyreva, N.S., Arino Bernad, J., Eremina, A.A., Gorshkova, A.A., Kanevskiy, G.M., Lonishin, L.R., Meister, A.V., Yakupova, A.G., Kondrashov, F.A. *et al.* (2018) Self-consistency test reveals systematic bias in programs for prediction change of stability upon mutation. *Bioinformatics*, **34**, 3653–3658.
63. Case, D.A., Cheatham, T.E. 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M. Jr, Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
64. Biagini, T., Chillemi, G., Mazzoccoli, G., Grottesi, A., Fusilli, C., Capocéfalo, D., Castellana, S., Vescovi, A.L. and Mazza, T. (2018) Molecular dynamics recipes for genome research. *Brief. Bioinform.*, **19**, 853–862.
65. Biagini, T., Pettrizzelli, F., Truglio, M., Cespa, R., Barbieri, A., Capocéfalo, D., Castellana, S., Tevy, M.F., Carella, M. and Mazza, T. (2019) Are gaming-enabled graphic processing unit cards convenient for molecular dynamics simulation? *Evol. Bioinformatics Online*, **15**, 1176934319850144.

66. Pronk,S., Pall,S., Schulz,R., Larsson,P., Bjelkmar,P., Apostolov,R., Shirts,M.R., Smith,J.C., Kasson,P.M., van der Spoel,D. *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **29**, 845–854.
67. Kumari,R., Kumar,R. and Open Source Drug Discovery, COpen Source Drug Discovery, C and Lynn,A. (2014) g_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.*, **54**, 1951–1962.
68. Klink,G.V., Golovin,A.V. and Bazykin,G.A. (2017) Substitutions into amino acids that are pathogenic in human mitochondrial proteins are more frequent in lineages closely related to human than in distant lineages. *PeerJ*, **5**, e4143.
69. McCormick,E.M., Lott,M.T., Dulik,M.C., Shen,L., Attimonelli,M., Vitale,O., Karaa,A., Bai,R., Pineda-Alvarez,D.E., Singh,L.N. *et al.* (2020) Specifications of the ACMG/AMP standards and guidelines for mitochondrial DNA variant interpretation. *Hum. Mutat.*, doi:10.1002/humu.24107.