Research Paper

A novel diagnostic tool reveals mitochondrial pathology in human diseases and aging

Morten Scheibye-Knudsen[1], Karsten Scheibye-Alsing[2], Chandrika Canugovi[1], Deborah L. Croteau[1], and Vilhelm A. Bohr[1]

[1] *Laboratory of Molecular Gerontology, National Institute on Aging, National Institutes of Health, Baltimore, MD 20892, USA;*
[2] *Section of Genetics and Bioinformatics, IBHV, Faculty of Health and Medical sciences, University of Copenhagen, Denmark*

Key words: *aging, mitochondria, progeria, bioenergetics, diagnostics, bioinformatics*
Received: *2/24/13; Accepted: 3/22/13; Published: 3/24/13*
Correspondence to: *Vilhelm A. Bohr, PhD;* ***E-mail:*** *vbohr@nih.gov*

Abstract: The inherent complex and pleiotropic phenotype of mitochondrial diseases poses a significant diagnostic challenge for clinicians as well as an analytical barrier for scientists. To overcome these obstacles we compiled a novel database, www.mitodb.com, containing the clinical features of primary mitochondrial diseases. Based on this we developed a number of qualitative and quantitative measures, enabling us to determine whether a disorder can be characterized as mitochondrial. These included a clustering algorithm, a disease network, a mitochondrial barcode and two scoring algorithms. Using these tools we detected mitochondrial involvement in a number of diseases not previously recorded as mitochondrial. As a proof of principle Cockayne syndrome, ataxia with oculomotor apraxia 1 (AOA1), spinocerebellar ataxia with axonal neuropathy 1 (SCAN1) and ataxia-telangiectasia have recently been shown to have mitochondrial dysfunction and those diseases showed strong association with mitochondrial disorders. We next evaluated mitochondrial involvement in aging and detected two distinct categories of accelerated aging disorders, one of them being associated with mitochondrial dysfunction. Normal aging seemed to associate stronger with the mitochondrial diseases than the non-mitochondrial partially supporting a mitochondrial theory of aging.

INTRODUCTION

Bona fide mitochondrial diseases represent a heterogeneous group of genetic syndromes with a combined incidence of around 1:5000 [1]. The clinical diversity represents a considerable diagnostic challenge for pediatricians often leading to a delay in diagnosis [1]. Because of the early onset and rapid progression of many of these diseases an early diagnosis will become increasingly important with the accelerating development of mitochondrial therapeutic strategies. Although the complete pathogenesis remains unknown, energy deficiency in affected tissues is believed to be the causative agent in most of these disorders [2].

Besides their cardinal role in ATP metabolism mitochondria are the main producers of endogenous oxidative radicals. These highly volatile species react with lipids, proteins and nucleic acids in their vicinity. The mitochondrial theory of aging states that an accumulation of damage to these macromolecules throughout the lifetime of an organism leads to cellular decay, loss of tissue homeostasis, and finally to death [3]. Multiple lines of evidence have corroborated this theory and suggested that mitochondrial maintenance may be important in promoting longevity and healthy aging [4-9]. Indeed, mitochondria have been implicated in most age related diseases such as neurodegeneration [10], cardiovascular disease [11] and diabetes [12]. If

mitochondrial dysfunction is causative in aging, we would expect the accelerated aging disorders [13] to exhibit features of mitochondrial disease.

To investigate this, we compiled a database of the clinical parameters seen in mitochondrial diseases, www.mitodb.com. Based on this database we developed extensive bioinformatics tools to dissect whether a disease could be characterized as mitochondrial or not. These tools include three qualitative and two quantitative measures. Using these tools we identified a number of diseases as mitochondrially associated that had not previously been considered as mitochondrial. Recently a number of diseases, such as ataxia-telangiectasia, Cockayne syndrome, ataxia with oculomotor apraxia 1 (AOA1) and spinocerebellar ataxia with axonal neuropathy 1 (SCAN1) have been suggested to have mitochondrial dysfunction and these disorders were also identified by our tools [14-17]. With the validation of the tools we went on to investigate the mitochondrial involvement in a number of monogenic diseases. Interestingly, Parkinson's disease, Huntington's disease and amyotrophic lateral sclerosis all showed a substantial mitochondrial involvement. Further, when adding the accelerated aging disorders to the database two groups of progeria appeared; one group associated with chromosomal instability and another group clustered with mitochondrial diseases. Normal aging seemed to associate closer with the mitochondrial group in the clustering algorithm but showed mixed mitochondrial and non-mitochondrial values in the support vector machine and mitoscore. Taken together these findings indicate at least two separate causes of aging, one of them possibly being mitochondrial.

RESULTS

www.mitodb.com

Using various sources such as the United Mitochondrial Disease Foundation webpage (http://www.umdf.org), Pubmed (http://www.ncbi.nlm.nih.gov/pubmed) and the Online Mendelian Inheritance in Man database (http://www.omim.org) we identified 31 monogenic diseases that all have been characterized as mitochondrial describing a total of 1,265 patients (Supplementary table 1). To get as broad a spectrum of mitochondrial phenotypes as possible we added respiratory chain defects, fatty acid oxidation defects and other primary mitochondrial diseases. 25 non-mitochondrial monogenic diseases, used as controls, describing a total of 21,032 patients, were selected based on well established pathogenesis with no or very minimal involvement of mitochondria. A database, www.mitodb.com, was constructed by writing ~8,000 lines of code in the languages html and php and using the aforementioned data sources, referenced signs, symptoms, laboratory and paraclinical findings (collectively referred to as clinical parameters or parameters) were recorded for each disease by clinically experienced physicians. The prevalence of each parameter was listed based on what was reported in the most authoritative literature we could identify. The publication date was as current as possible (Supplementary figure 1). For any particular disease we used the largest cohorts we could identify. Case reports were excluded. If we used data from two or more studies the prevalence was the average of the combined number of patients. All the references including their abstracts, publication date, periodical and links to Pubmed can be found online at www.mitodb.com. Each disease was thus characterized by a number of clinical parameters with a prevalence from 0-100 percent. See supplementary figure 2 or www.mitodb.com for visual examples of how the diseases are represented in the database. We will present a number of figures that can be generated online at www.mitodb.com. We also recommend reading our tutorial on the following URL: www.mitodb.com/tutorial.html.

Mitochondrial diseases form a distinct clinical group

Based on the data we could now analyze the clinical parameters in the mitochondrial group by averaging the occurrence of a parameter across all the diseases in each group. This allowed us to investigate whether any particular parameter would be more prevalent among the mitochondrial diseases than among the non-mitochondrial diseases. Figure 1A shows the top 20 most prevalent parameters and supplementary figure 3 and 4 shows all the parameters in the mitochondrial group and the non-mitochondrial group, respectively. Our analysis shows that, out of 117 clinical parameters seen in mitochondrial diseases, lactate accumulation, hypotonia, muscle weakness, developmental delay and seizures are among the most common. Since it has been suggested that the clinical parameters of mitochondrial diseases depend upon the age of onset of the disease we investigated whether this was the case for our dataset. We identified the mean age of onset and standard deviation for each disease in the database (See supplementary table 1). Using this data we could then model the age of onset of each clinical parameter letting us identify the most common symptoms for diseases with onset before age 20 and most common symptoms for diseases with onset after age 20 (Figure 1B and C) and their evolution over time (Supplementary figure 5). Lactate accumulation, hypotonia, developmental delay, muscle weakness and seizures were the most prevalent

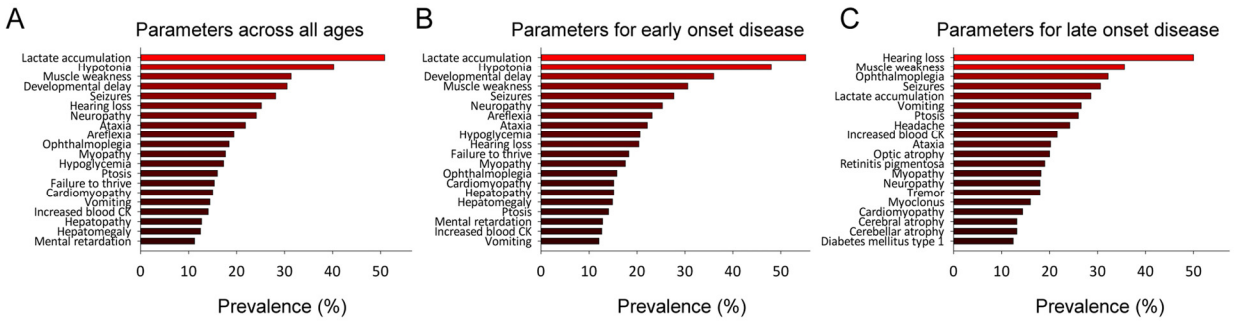


Figure 1. Mitochondrial diseases have a defined clinical spectrum. (**A**) The top-20 clinical parameters seen in all mitochondrial diseases. (**B**) The top-20 clinical parameters seen in mitochondrial diseases with a mean age of onset before 20. (**C**) The top-20 clinical parameters seen in mitochondrial diseases with a mean age of onset after 20.

parameters in the early onset diseases while hearing loss, muscle weakness, ophthalmoplegia, seizures and lactate accumulation were the most prevalent in the late onset diseases. Some differences in clinical parameters were thus found when comparing early onset to late onset disease. However, quantifying these differences as a Pearson coefficient resulted in a value of 0.5 indicating an overall similar phenotype in early and late onset diseases. In addition, there was no published data for the individual parameter onsets for every disease preventing us from using age of onset for further analysis. We therefore decided to continue our analysis without including age of onset.

By overlaying the prevalences of all the mitochondrial and non-mitochondrial parameters it became evident that some mitochondrial clinical parameters occurred at a high frequency indicating a high degree of similarity between the diseases in the mitochondrial group while there was much more heterogeneity amongst the clinical features in the non-mitochondrial disease group (Supplementary figure 6A). Based on this resemblance among the mitochondrial diseases, we investigated how these diseases would associate with each other when clustering them using the software cluster 3.0 [18]. Supplementary figure 6B shows an example of how some mitochondrial and some non-mitochondrial diseases cluster with each other based on clinical parameters. The clustering of the diseases is shown in the top, while the clinical parameters are shown on the right with the clustering of the parameters on the left. The tint of each square in the center of the cluster map reflects the prevalence of a parameter in the disease based on published literature. Interestingly, there seemed to be excellent separation between the mitochondrial and non-mitochondrial diseases. Supplementary figure 6C shows an overview of the cluster while a detailed look can be seen online at www.mitodb.com. In summary, we were able to use cluster analysis to compare the diseases in the database.

The qualitative determination of a mitochondrial phenotype using mitodb.com

With this established database we proceeded to develop tools to test diseases for potential mitochondrial involvement. We started by implementing hierarchical clustering algorithms using the coding languages html, javascript and php. The phenotype of a disease is entered under the "Test-disease" tab on mitodb.com by adding the prevalence of all the observed parameters. After the analysis has been completed a circular graphic illustrating (dendrogram) how the diseases cluster with each other will be generated. The connecting lines (branches of the tree) illustrate how similar two diseases are. As can be seen in figure 2A there is substantial separation of mitochondrial (red) and non-mitochondrial (green) diseases using this approach. To further visualize how the clinical parameters are interconnected we generated a disease network tool (Figure 2B) showing the association between diseases. Each dot represents a disease and each line represents a shared parameter, with the distance between the dots and the thickness of the line representing the degree of similarity (See supplementary equations). Interestingly, a central tightly associated hub of mitochondrial diseases formed with the non-mitochondrial diseases loosely associated in the periphery (Figure 2B). This is caused by the lack of overlapping parameters seen in the non-mitochondrial group compared with the highly overlapping parameters seen in the mitochondrial. Due to the substantial complexity we recommend investigat-ing this network online on www.mitodb.com, where different thresholds can be applied to selectively visualize the most common connections. To better depict the correlation between diseases we developed a mitochondrial barcode-

algorithm that shows which parameters are found in other diseases in the database (Figure 2C). Each vertical colored bar represents a disease that shares a clinical parameter with the disease being tested. The tint is given by the prevalence of the parameter in the disease tested, multiplied by the prevalence in the disease in the database that shares the parameter. Thus, the barcode appears more red with increasing mitochondrial features and more green or blue with fewer or no mitochondrial features. The list of shared clinical parameters for a tested disease can be found online underneath the barcode.

The quantitative determination of a mitochondrial phenotype using mitodb.com

To get more quantitative results we developed a tool assigning a mito-score to a disease from 0-100 with a score above 50 indicating a larger overlap with mitochondrial diseases than non-mitochondrial diseases (See supplementary equations). This method gave consistent results with mitochondrial diseases scoring high and non-mitochondrial scoring low (Figure 2D).
We next developed a machine learning algorithm to separate diseases based on clinical parameters and to get a quantitative score. Briefly, we chose a model of supervised learning, a support vector machine (SVM), that is ideal for investigating complicated datasets. This type of machine learning has been used to analyze various datasets including domain mapping of amino acid sequences in proteins and facial recognition in pictures on the web. In this method, the clinical pattern of the mitochondrial diseases and the pattern of the non-mitochondrial diseases are fed to a computer that will learn from these training examples and subsequently predict whether a tested disease may show mitochondrial involvement. A negative score reflects a likely non-mitochondrial disease while a positive score indicates a mitochondrial disease (See methods). We implemented several variations of the SVM, one unfiltered taking all parameters in the database into account and various filtered ones using only parameters displaying a cumulative prevalence of 25%, 50%, 75%, 100% or 200%. A cumulative prevalence of 25% is defined as one disease with a 25% prevalence of a parameter or five diseases with a 5% prevalence each. Thus, with increasing percentages progressively more common clinical parameters will be the only ones considered by the SVM. Overall, the performance of the various classifiers was excellent with 100% separation and a correlation coefficient (measured by cross-validation) around 0.85, with the 75% filtered version yielding a marginally better correlation coefficient of 0.87 (See supplementary equations). The great performance of all the SVM's show that the method is robust, and that the prevalence of the clinical parameters can be used to indicate whether a disease show mitochondrial involvement. To keep the data as unbiased as possible we chose the unfiltered version as our standard SVM algorithm (Figure 2E, and supplementary equations).

The bioinformatics tools reveal mitochondrial dysfunction in ataxia-telangiectasia, Cockayne syndrome and others

Recent studies that have shown mitochondrial dysfunction in ataxia-telangiectasia [17] , AOA1 [16], SCAN1 [15] and Cockayne syndrome [14,20]. Interestingly, these diseases clustered tightly with the mitochondrial diseases, grouped with mitochondrial disease in the disease network and displayed a relatively high SVM and mito-score (Figure 2A-E). This indicated that our tools could be used to investigate mitochondrial dysfunction in diseases of unknown pathogenesis simply by comparing the clinical parameters with the parameters in our database.

Aging and some progerias display mitochondrial dysfunction

Many diseases with unknown pathogenesis were strongly associated with mitochondrial pathology (Figure 3A and supplementary table 2). Because a close correlation between aging and mitochondrial dysfunction has been proposed, we next tested whether diseases characterized by accelerated aging shared clinical parameters with the mitochondrial diseases in the database. Of all the segmental progerias tested, Cockayne syndrome and ataxia-telangiectasia correlated tightly with mitochondrial pathology in the dendrogram, while Werner syndrome, Bloom syndrome, Rothmund-Thomson syndrome, trichothiodystrophy, dyskeratosis congenita, Fanconi anemia, Nijmegen breakage syndrome and Hutchinson-Gilford progeria formed an independent cluster (Figure 3A). This was corroborated by the finding in the disease network that Cockayne syndrome and ataxia-telangiectasia appeared to lie deeper in the mitochondrial hub compared to the other progerias (Figure 3B). When looking at the barcode for Cockayne syndrome and ataxia-telangiectasia, these diseases also showed a predominantly red impression while the other progerias where less red (Figure 2C and Figure 3C). The SVM and mito-score supported all these results (Figure 2D and E and figure 3D and E). The association between the mitochondrially associated progerias, Cockayne syndrome and ataxia-telangiectasia, was related to the high degree of neurological involvement in these diseases, since removing the neuronal parameters led these diseases to cluster with the other progerias (Supplementary figure 7).

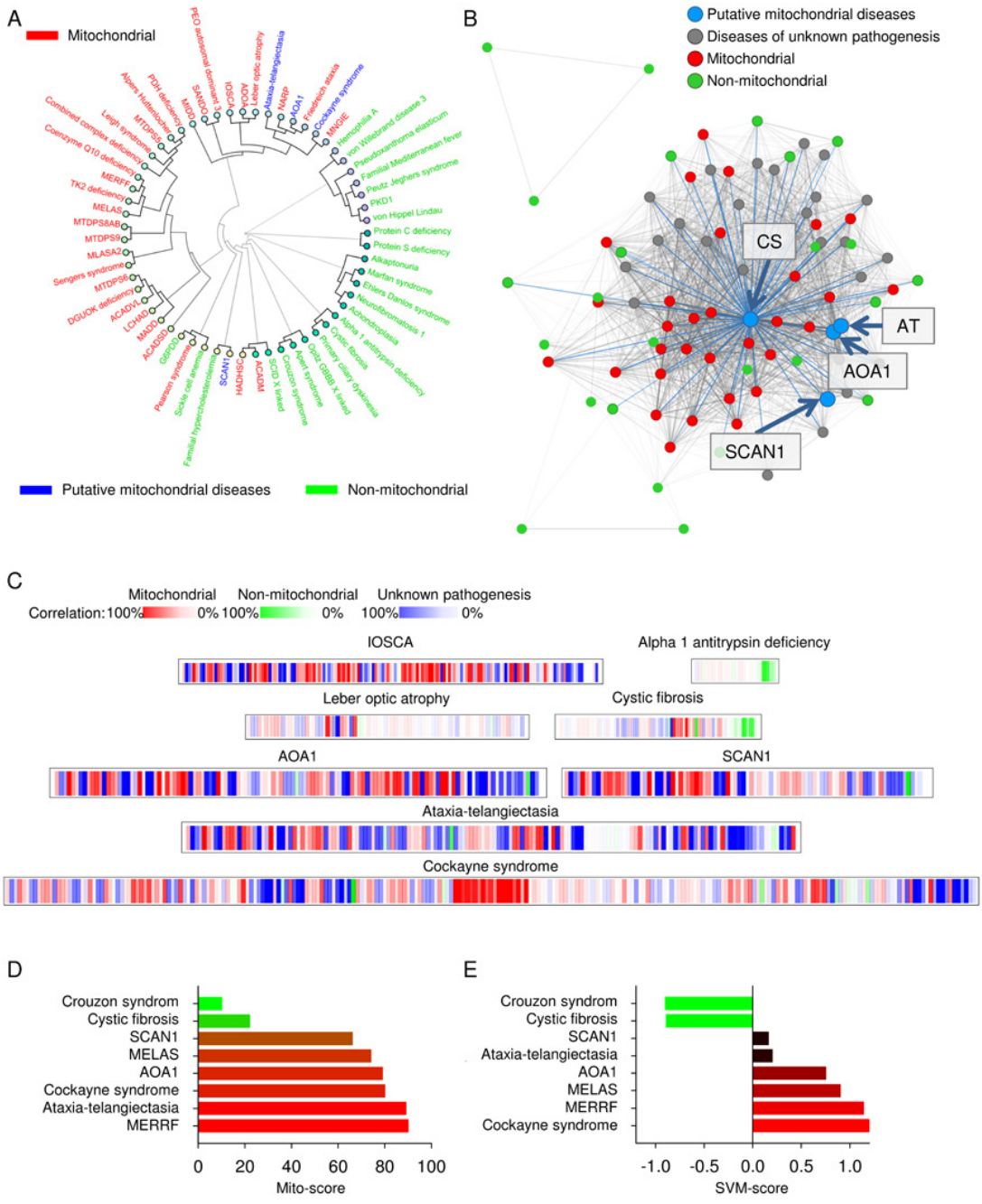


Figure 2. Putative mitochondrial diseases are identified using www.mitodb.com. (**A**) The clustering of several diseases of unknown pathogenesis with recently recognized mitochondrial dysfunction (blue) and mitochondrial (red) and non-mitochondrial diseases (green). AOA1: Ataxia with oculomotor apraxia 1; SCAN1: Spinocerebellar ataxia with axonal neuropathy 1. (**B**) A representation of how the putative mitochondrial diseases (blue) associate within the disease network. Each dot represents a disease and the closer two diseases are connected the shorter the distance between them. Mitochondrial diseases: red; non-mitochondrial diseases: green. (**C**) The mitochondrial barcode of a number of diseases. Each bar represents a clinical parameter that is shared with another disease in the database. Red is mitochondrial diseases, green is non-mitochondrial and blue is diseases of unknown pathogenesis. The tint is given by the percentage of patients that are affected in the disease tested multiplied by the percentage of patients that are affected in the disease in the database that shares the parameter. IOSCA: infantile onset spinocerebellar ataxia (**D**) The mito score of the putative mitochondrial diseases and two bona fide mitochondrial (MERFF and MELAS) and two non-mitochondrial (cystic fibrosis and Crouzon syndrome) diseases. (**E**) The SVM score of the tested diseases and two mitochondrial (MERFF and MELAS) and two non-mitochondrial (cystic fibrosis and Crouzon syndrome) diseases.

We finally investigated if normal human aging was associated with mitochondrial pathology. A number of important papers were found to describe the phenotype of normal human aging as accurately as possible (Figure 4) [21-30]. By entering the parameters of aging an interesting picture emerged. Normal aging seemed to cluster with the mitochondrial diseases where the mitochondrial segmental progerias also clustered indicating a potential mitochondrial phenotype (Figure 3A-C). Reflecting this, the barcode showed a mitochondrial imprint that was corroborated by a high mito-score of 70. However, the SVM was inconclusive reflected by a score of around 0. Interestingly, using the 200% filtered SVM, aging received a score of 1.46, indicating a probable mitochondrial impact when considering only very common clinical parameters. It therefore seems likely that some features of aging, eg. cancers, may be independent of mitochondria while others, eg. neurological features, are closely correlated with mitochondrial health.

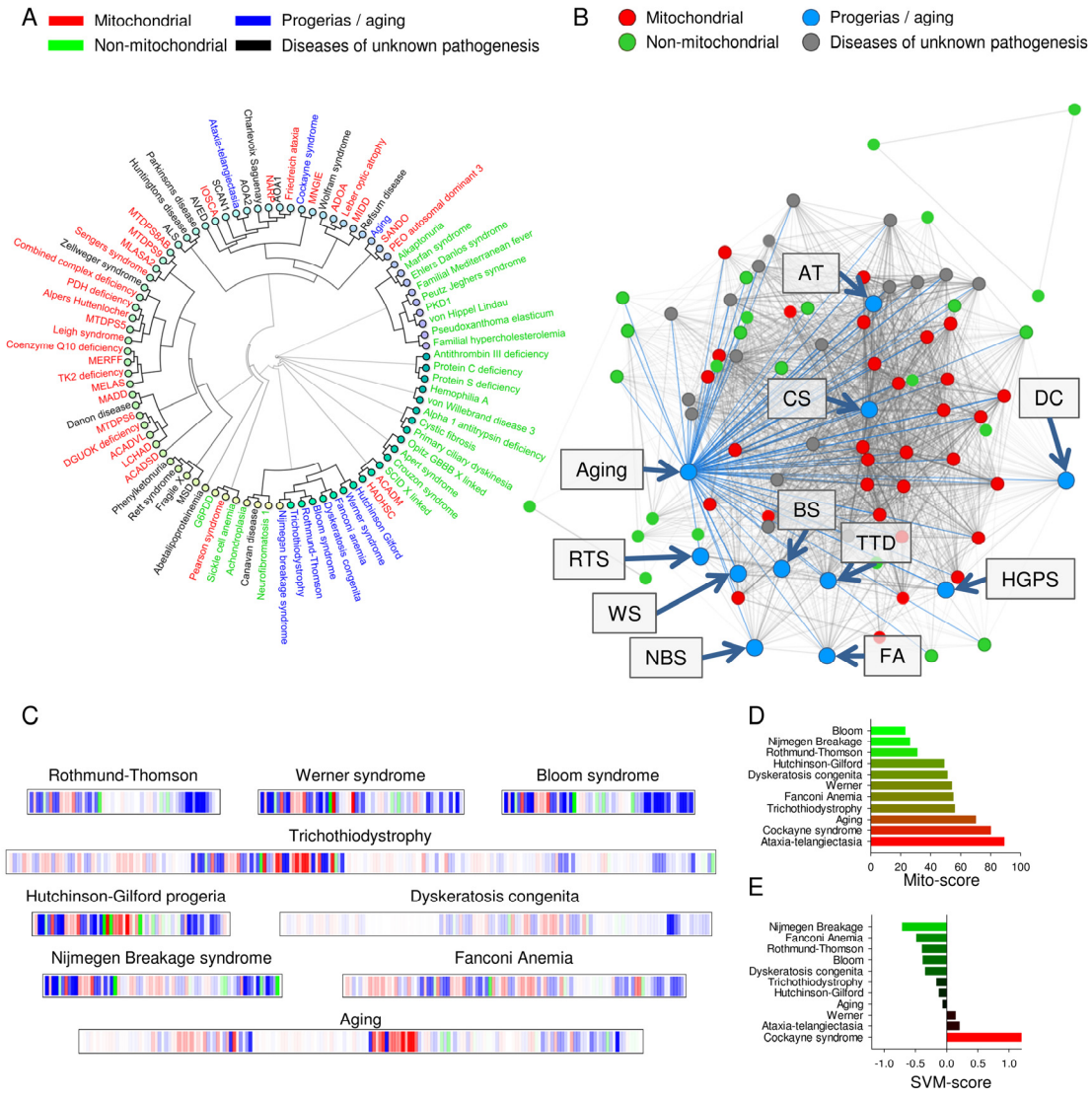


Figure 3. Normal aging and some accelerated aging disorders display phenotypical similarities to mitochondrial diseases. (**A**) Clustermap using uncentered similarity and average linkage of all the diseases in the database. Blue represents aging and the accelerated aging disorders. (**B**) A representation of how aging and the progerias (blue dots) associate within the disease network. (**C**) Mitochondrial barcode of some accelerated aging disorders. (**D**) The mito score of the tested diseases. (**E**) The SVM score of the tested diseases.

.

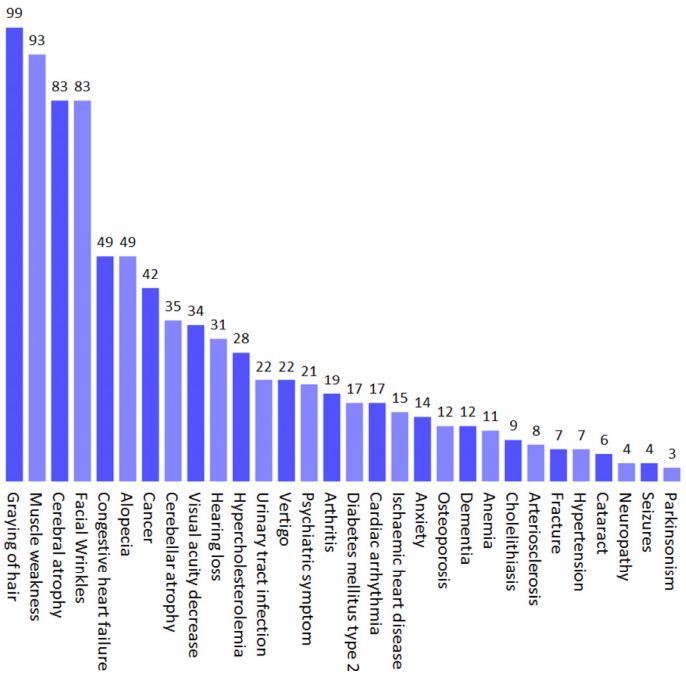


Figure 4. The phenotype of normal human aging based on published studies. Values represent the prevalence (%) of a given parameter.

.

DISCUSSION

The use of bioinformatics in medical pattern recognition is a powerful tool ideally suited to support clinicians and scientists when investigating large datasets. Mitochondrial diseases represent a bioinformatically ideal group of pathologies because they are etiologically well defined while presenting a complex clinical picture. Here we introduce a novel mitochondrial database and a unique array of useful tools to investigate potential mitochondrial involvement in diseases of unknown pathogenesis. We have placed the database online to ensure the highest usability and a continuous evolution based on the addition of new diseases and revisions of clinical parameters. Mitochondrial diseases are highly complex involving multiple organ systems with a large variability in age of onset and clinical severity. It is therefore not surprising that there is a very substantial lag phase in disease onset and time of diagnosis in many of these diseases. We believe that the tools we have presented here can be helpful in the diagnosis of these diseases potentially allowing earlier treatments for this group of severe disorders.

The neuropathology observed in some of the accelerated aging disorders seems to drive the association with mitochondrial diseases in the database. This is not unexpected since mitochondrial dysfunction has been implicated in many age related neurological disorders such as Parkinson's and Huntington's disease. In recent years increasing evidence of mitochondrial involvement has emerged in these diseases. For Parkinson's disease mutations in PINK1 and parkin have implicated the degradation of damaged mitochondria through autophagy as part of the pathogenesis [31]. Interestingly, this pathway has also been suggested to be defective in Huntington's disease perhaps indicating a common pathogenesis [32]. In support of a mitochondrial dysfunction in these neurodegenerative disorders the basal ganglia are often affected in mitochondrial diseases such as Leigh syndrome and Neuropathy, Ataxia and Retinitis Pigementosum (NARP syndrome) [33,34]. Indeed, Parkinson's disease receives a mito-score of 100 while Huntington's disease receives a mito-score of 99 and both have a positive SVM score.

In addition to the clinical uses the tools presented herein could be valuable for scientists investigating the cause of diseases of unknown pathogenesis. Indeed, using the tools, we corroborate the findings recently published for ataxia-telangiectasia and Cockayne syndrome [14,17]. We further show tight associations between mitochondrial pathology and AOA1 and SCAN1, two diseases recently shown to have mitochondrial defects [15,16]. In this context, it is important to point out that these tools do not specify whether a disease is primarily mitochondrial or if the exposed dysfunction stems from a secondary unknown mechanism.

According to our hypothesis premature aging disorders as well as normal aging should share features with mitochondrial disease. This seems to be partially true. Based on the hierarchical clustering it appears that two distinct types of accelerated aging disorders exist. One type consisting of Werner syndrome, Bloom syndrome, Rothmund-Thomson syndrome, trichothiodystrophy, dyskeratosis congenita, Fanconi anemia, Nijmegen breakage syndrome and Hutchinson-Gilford progeria form a separate cluster. For these disorders the common molecular denominator involves chromosome instability in relation to DNA replication, repair of double stranded DNA breaks and telomere maintenance [35-37]. The second group of accelerated aging disorders consists of Cockayne syndrome and ataxia-telangiectasia. These two disorders are characterized by a substantial neurological phenotype that drives the clustering with the mitochondrial diseases. Interestingly, normal aging itself, seems to associate closer with the mitochondrial diseases that the non-mitochondrial diseases perhaps corroborating the mitochondrial theory of aging.

Noticeably, during the development of the database we encountered a number of syndromes of unknown pathology displaying strong correlation with mitochondrial diseases (Supplementary table 2). Several of these diseases exhibit severe neuropathology supporting the idea that the central nervous system is particularly prone to mitochondrial dysfunction [10,38]. Disorders of this nature would thus be prime targets for further investigation and potentially for treatment with drugs known to augment mitochondrial function.

METHODS

Hierarchical clustering. Hierarchical clustering was implemented using common similarity measures and linkage methods (See supplementary equations) [18,39]. For each clustering the cophenetic correlation is calculated as a measure of how well the clustering was achieved revealing uncentered similarity and average linkage producing the best result (see supplementary figure 9 and supplementary equations). Other similarity metrics and linkage methods can be chosen online by clicking "advanced" under the "test-disease" tab on www.mitodb.com.

Support vector machine (SVM). The support vector classifier was created by exporting the symptom-vectors of the known mitochondrial and non-mitochondrial diseases. The symptom-vectors were processed with the statistical program "R" using the "e1071" package (http://cran.r-project.org/web/packages/e1071/index.html), which is based on the library "libSVM" to train the SVMs. The SVMs were trained using a "linear kernel" and "eps-regression", and displays (using 11-fold cross-correlation) a total mean squared error of 0.32, a 0.84 correlation coefficient, and 100% separation. The parameters of the SVMs were then exported from R and the classifier functions for the web-page was created (with php) using these parameters.

Further methods can be found in the supporting materials.

ACKNOWLEDGMENTS

We would like to thank Dr. Peter Sykora for critically reading the manuscript. This work was supported by funds from the intramural program of the National Institutes of Health.

Conflict of Interest Statement

The authors declare no conflicts of interest.

REFERENCES

1. Haas RH, Parikh S, Falk MJ, Saneto RP, Wolf NI, Darin N, and Cohen BH Mitochondrial disease: a practical approach for primary care physicians. Pediatrics 2007; 120: 1326-1333.

2. Schapira AH Mitochondrial disease. Lancet 2006; 368: 70-82.

3. Balaban RS, Nemoto S, and Finkel T Mitochondria, oxidants, and aging. Cell 2005; 120: 483-495.

4. Corral-Debrinski M, Horton T, Lott MT, Shoffner JM, Beal MF, and Wallace DC Mitochondrial DNA deletions in human brain: regional variability and increase with advanced age. Nat. Genet. 1992; 2: 324-329.

5. Hudson EK, Hogue BA, Souza-Pinto NC, Croteau DL, Anson RM, Bohr VA, and Hansford RG Age-associated change in mitochondrial DNA damage. Free Radical Research 1998; 29: 573-579.

6. Schriner SE, Linford NJ, Martin GM, Treuting P, Ogburn CE, Emond M, Coskun PE, Ladiges W, Wolf N, Van RH, Wallace DC, and Rabinovitch PS Extension of murine life span by overexpression of catalase targeted to mitochondria. Science 2005; 308: 1909-1911.

7. Trifunovic A, Wredenberg A, Falkenberg M, Spelbrink JN, Rovio AT, Bruder CE, Bohlooly Y, Gidlof S, Oldfors A, Wibom R,

Tornell J, Jacobs HT, and Larsson NG Premature ageing in mice expressing defective mitochondrial DNA polymerase. Nature 2004; 429: 417-423.

8. Petersen KF, Befroy D, Dufour S, Dziura J, Ariyan C, Rothman DL, DiPietro L, Cline GW, and Shulman GI Mitochondrial dysfunction in the elderly: possible role in insulin resistance. Science 2003; 300: 1140-1142.

9. Someya S, Yu W, Hallows WC, Xu J, Vann JM, Leeuwenburgh C, Tanokura M, Denu JM, and Prolla TA Sirt3 mediates reduction of oxidative damage and prevention of age-related hearing loss under caloric restriction. Cell 2010; 143: 802-812.

10. Lin MT and Beal MF Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. Nature 2006; 443: 787-795.

11. Mercer JR, Cheng KK, Figg N, Gorenne I, Mahmoudi M, Griffin J, Vidal-Puig A, Logan A, Murphy MP, and Bennett M DNA damage links mitochondrial dysfunction to atherosclerosis and the metabolic syndrome. Circ Res 2010; 107: 1021-1031.

12. Lowell BB and Shulman GI Mitochondrial dysfunction and type 2 diabetes. Science 2005; 307: 384-387.

13. Martin GM and Oshima J Lessons from human progeroid syndromes. Nature 2000; 408: 263-266.

14. Scheibye-Knudsen M, Ramamoorthy M, Sykora P, Maynard S, Lin P, K.Minor R, M.Wilson III D, Cooper M, Spencer R, de Cabo R, L.Croteau D, A.Bohr, and V Cockayne syndrome group B protein prevents the accumulation of damaged mitochondria by promoting mitochondrial autophagy. J. Exp. Med. 2012

15. Das BB, Dexheimer TS, Maddali K, and Pommier Y Role of tyrosyl-DNA phosphodiesterase (TDP1) in mitochondria. Proc. Natl. Acad. Sci. U. S. A 2010; 107: 19790-19795.

16. Sykora P, Croteau DL, Bohr VA, and Wilson DM, III Aprataxin localizes to mitochondria and preserves mitochondrial function. Proc. Natl. Acad. Sci. U. S A 2011; 108: 7437-7442.

17. Valentin-Vega YA, Maclean KH, Tait-Mulder J, Milasta S, Steeves M, Dorsey FC, Cleveland JL, Green DR, and Kastan MB Mitochondrial dysfunction in ataxia-telangiectasia. Blood 2012; 119: 1490-1500.

18. de Hoon MJ, Imoto S, Nolan J, and Miyano S Open source clustering software. Bioinformatics. 2004; 20: 1453-1454.

19. Holten D Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. IEEE Trans. Vis. Comput. Graph. 2006; 12: 741-748.

20. Aamann MD, Sorensen MM, Hvitby C, Berquist BR, Muftuoglu M, Tian J, de Souza-Pinto NC, Scheibye-Knudsen M, Wilson Iii DM, Stevnsner T, and Bohr VA Cockayne syndrome group B protein promotes mitochondrial DNA stability by supporting the DNA repair association with the mitochondrial membrane. FASEB J. 2010

21. Beghi E and Monticelli ML Chronic symmetric symptomatic polyneuropathy in the elderly: a field screening investigation of risk factors for polyneuropathy in two Italian communities. Italian General Practitioner Study Group (IGPST). J. Clin. Epidemiol. 1998; 51: 697-702.

22. Guralnik JM, Eisenstaedt RS, Ferrucci L, Klein HG, and Woodman RC Prevalence of anemia in persons 65 years and older in the United States: evidence for a high rate of unexplained anemia. Blood 2004; 104: 2263-2268.

23. Haavisto M, Geiger U, Mattila K, and Rajala S A health survey of the very aged in Tampere, Finland. Age Ageing 1984; 13: 266-272.

24. Hayat MJ, Howlader N, Reichman ME, and Edwards BK Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER) Program. Oncologist. 2007; 12: 20-37.

25. Ito M, Hatazawa J, Yamaura H, and Matsuzawa T Age-related brain atrophy and mental deterioration--a study with computed tomography. Br. J. Radiol. 1981; 54: 384-390.

26. Koller WC, Glatt SL, Fox JH, Kaszniak AW, Wilson RS, and Huckman MS Cerebellar atrophy: relationship to aging and cerebral atrophy. Neurology 1981; 31: 1486-1488.

27. Melov S, Tarnopolsky MA, Beckman K, Felkey K, and Hubbard A Resistance exercise reverses aging in human skeletal muscle. PLoS. One. 2007; 2: e465.

28. Naughton C, Bennett K, and Feely J Prevalence of chronic disease in the elderly based on a national pharmacy claims database. Age Ageing 2006; 35: 633-636.

29. Schnohr P, Lange P, Nyboe J, Appleyard M, and Jensen G Gray hair, baldness, and wrinkles in relation to myocardial infarction: the Copenhagen City Heart Study. Am. Heart J. 1995; 130: 1003-1010.

30. Tenenhouse A, Joseph L, Kreiger N, Poliquin S, Murray TM, Blondeau L, Berger C, Hanley DA, and Prior JC Estimation of the prevalence of low bone density in Canadian women and men using a population-specific DXA reference standard: the Canadian Multicentre Osteoporosis Study (CaMos). Osteoporos. Int. 2000; 11: 897-904.

31. Matsuda N, Sato S, Shiba K, Okatsu K, Saisho K, Gautier CA, Sou YS, Saiki S, Kawajiri S, Sato F, Kimura M, Komatsu M, Hattori N, and Tanaka K PINK1 stabilized by mitochondrial depolarization recruits Parkin to damaged mitochondria and activates latent Parkin for mitophagy. J. Cell Biol. 2010; 189: 211-221.

32. Martinez-Vicente M, Talloczy Z, Wong E, Tang G, Koga H, Kaushik S, de VR, Arias E, Harris S, Sulzer D, and Cuervo AM Cargo recognition failure is responsible for inefficient autophagy in Huntington's disease. Nat. Neurosci. 2010; 13: 567-576.

33. Gelfand JM, Duncan JL, Racine CA, Gillum LA, Chin CT, Zhang Y, Zhang Q, Wong LJ, Roorda A, and Green AJ Heterogeneous patterns of tissue injury in NARP syndrome. J. Neurol. 2011; 258: 440-448.

34. Lee HF, Tsai CR, Chi CS, Lee HJ, and Chen CC Leigh syndrome: clinical and neuroimaging follow-up. Pediatr. Neurol. 2009; 40: 88-93.

35. Bernstein KA and Rothstein R At loose ends: resecting a double-strand break. Cell 2009; 137: 807-810.

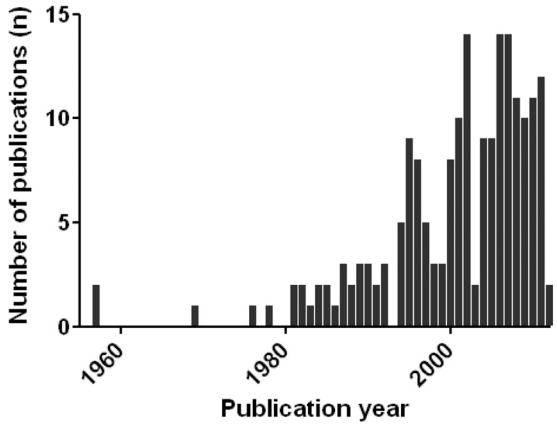
36. Chu WK and Hickson ID RecQ helicases: multifunctional genome caretakers. Nat. Rev. Cancer 2009; 9: 644-654.

37. Collado M, Blasco MA, and Serrano M Cellular senescence in cancer and aging. Cell 2007; 130: 223-233.

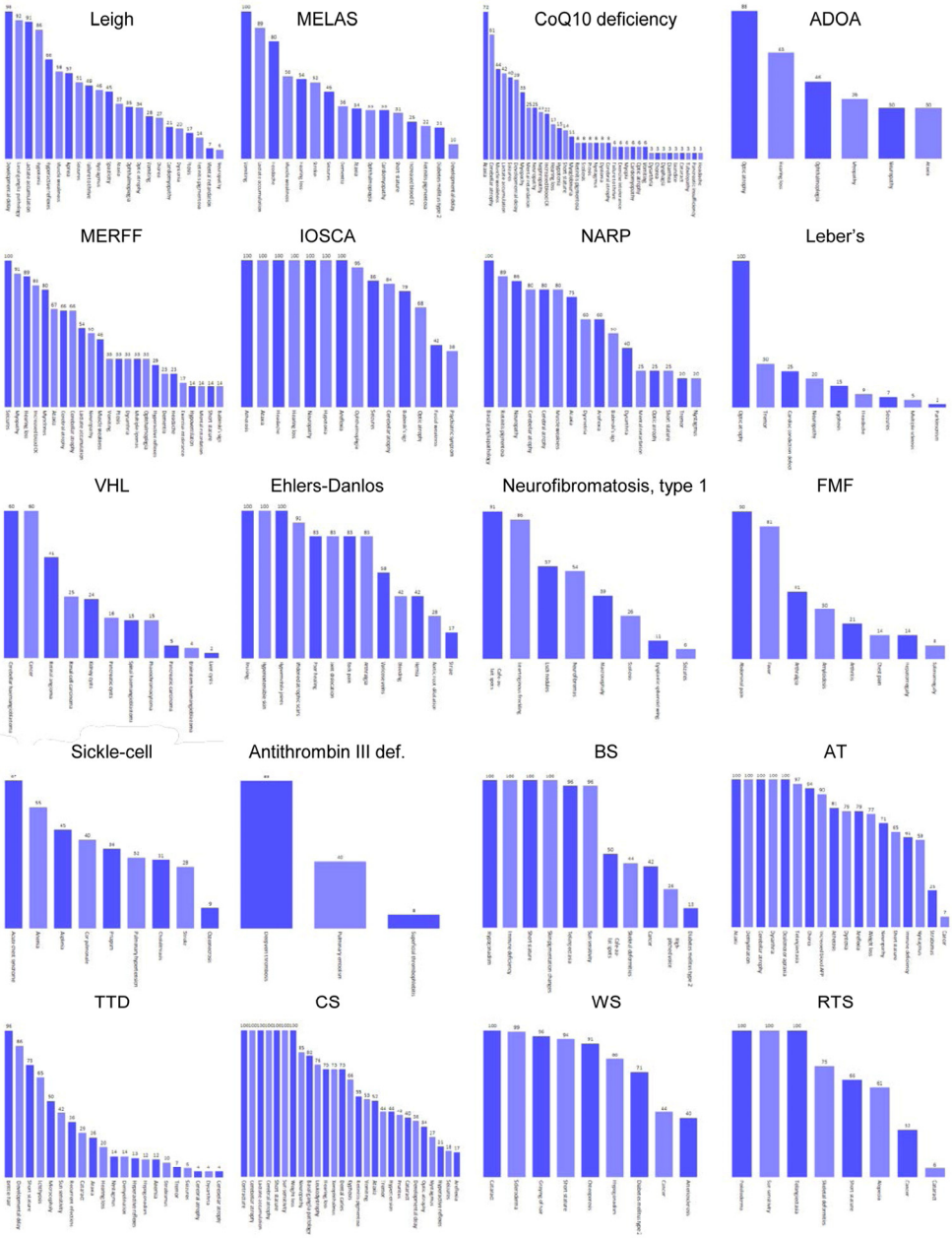
38. McFarland R, Taylor RW, and Turnbull DM The neurology of mitochondrial DNA disease. Lancet Neurol. 2002; 1: 343-351.

39. Hastie,T., Tibshirani,R., & Friedman,J. The Elements of Statistical Learning (Springer,2009).

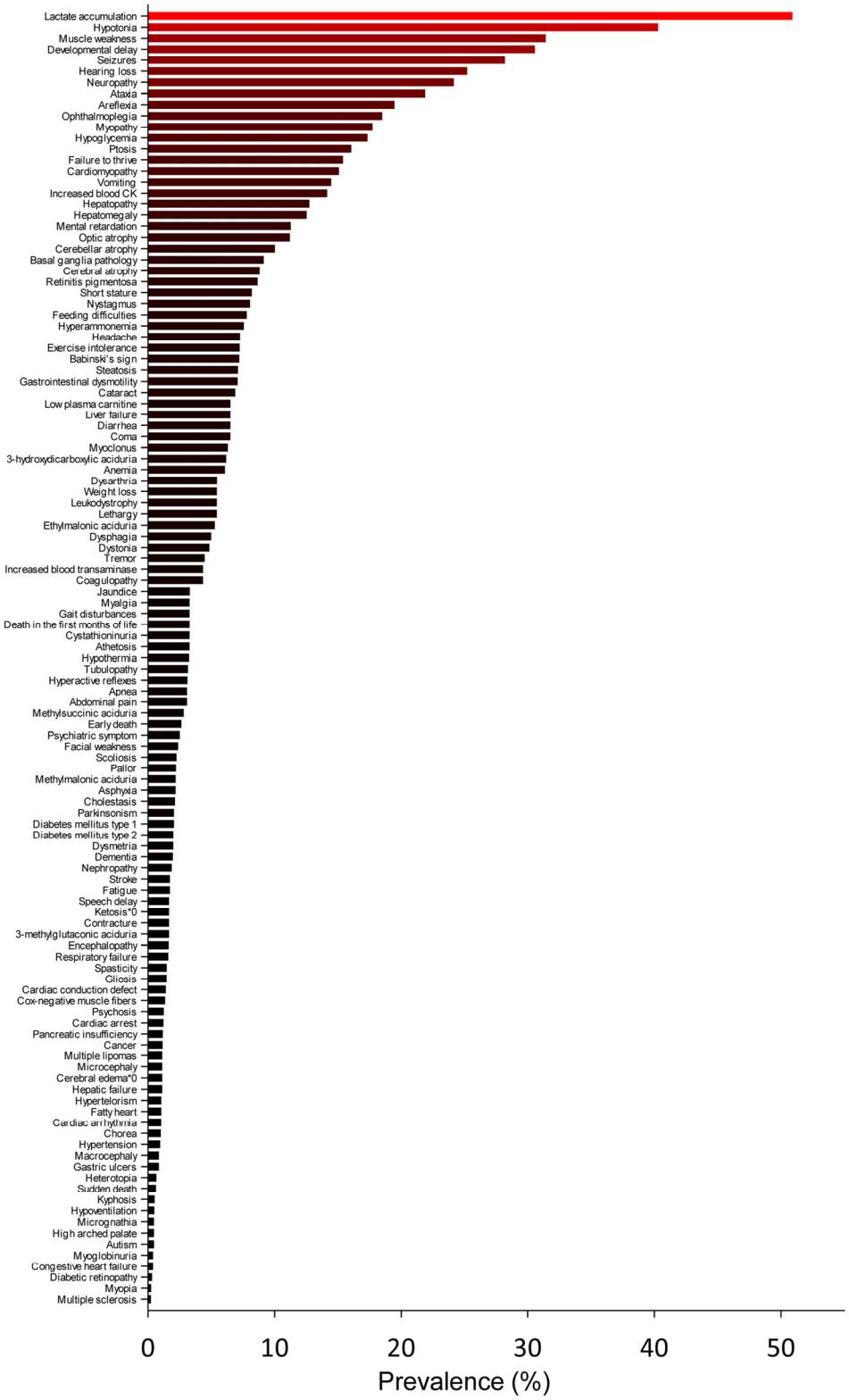
SUPPORTING INFORMATION



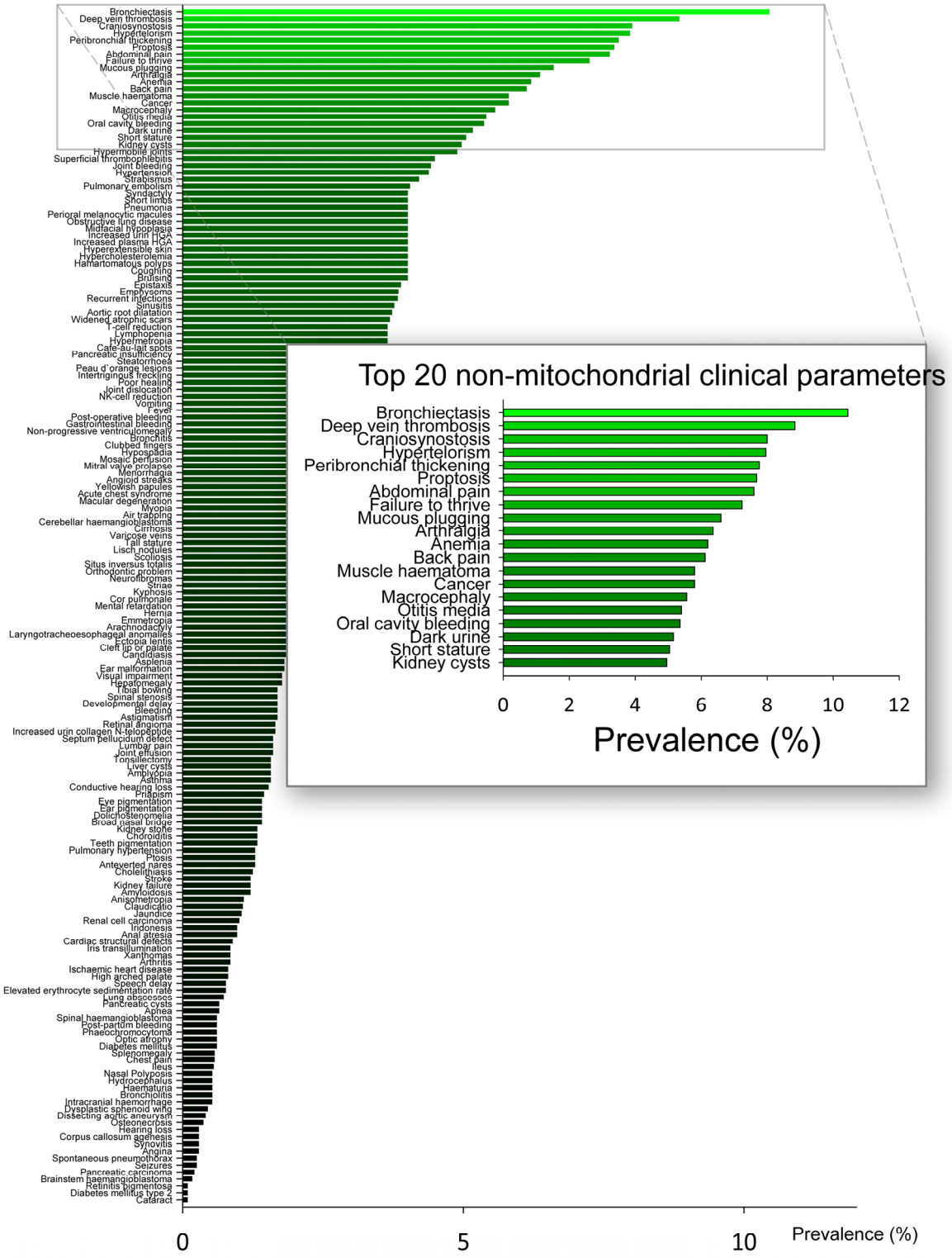
Supplementary Figure 1. The publication dates of the references used in the database.



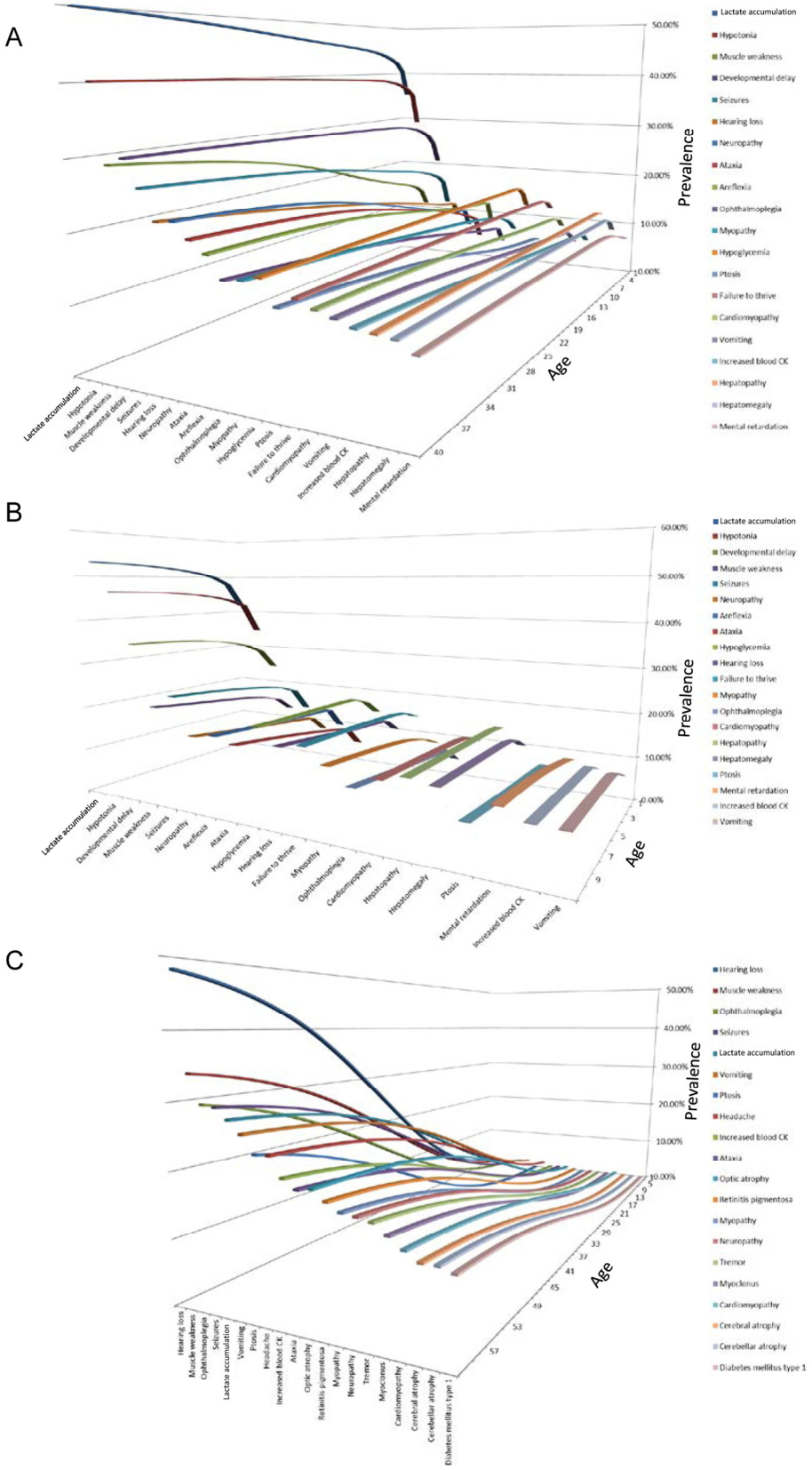
Supplementary Figure 2. Visual representation of some diseases and their clinical parameters as seen in the database. See supplementary table 1 for abbreviations.



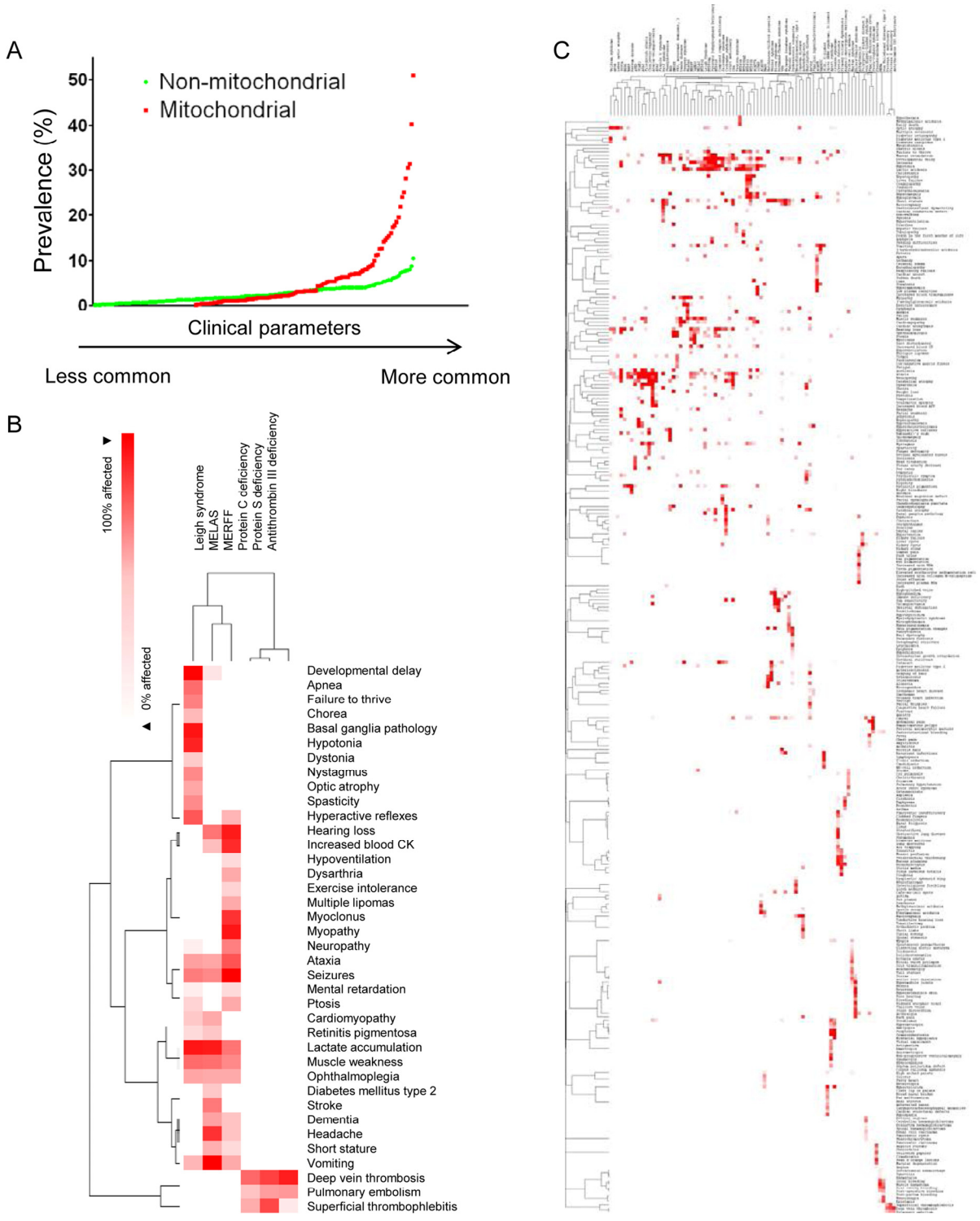
Supplementary Figure 3 Mitochondrial diseases in the database have a defined clinical spectrum. The average prevalence of clinical parameters across all the non-mitochondrial diseases in the database. Inset: Close-up of the top-20 clinical parameters seen in non-mitochondrial diseases.



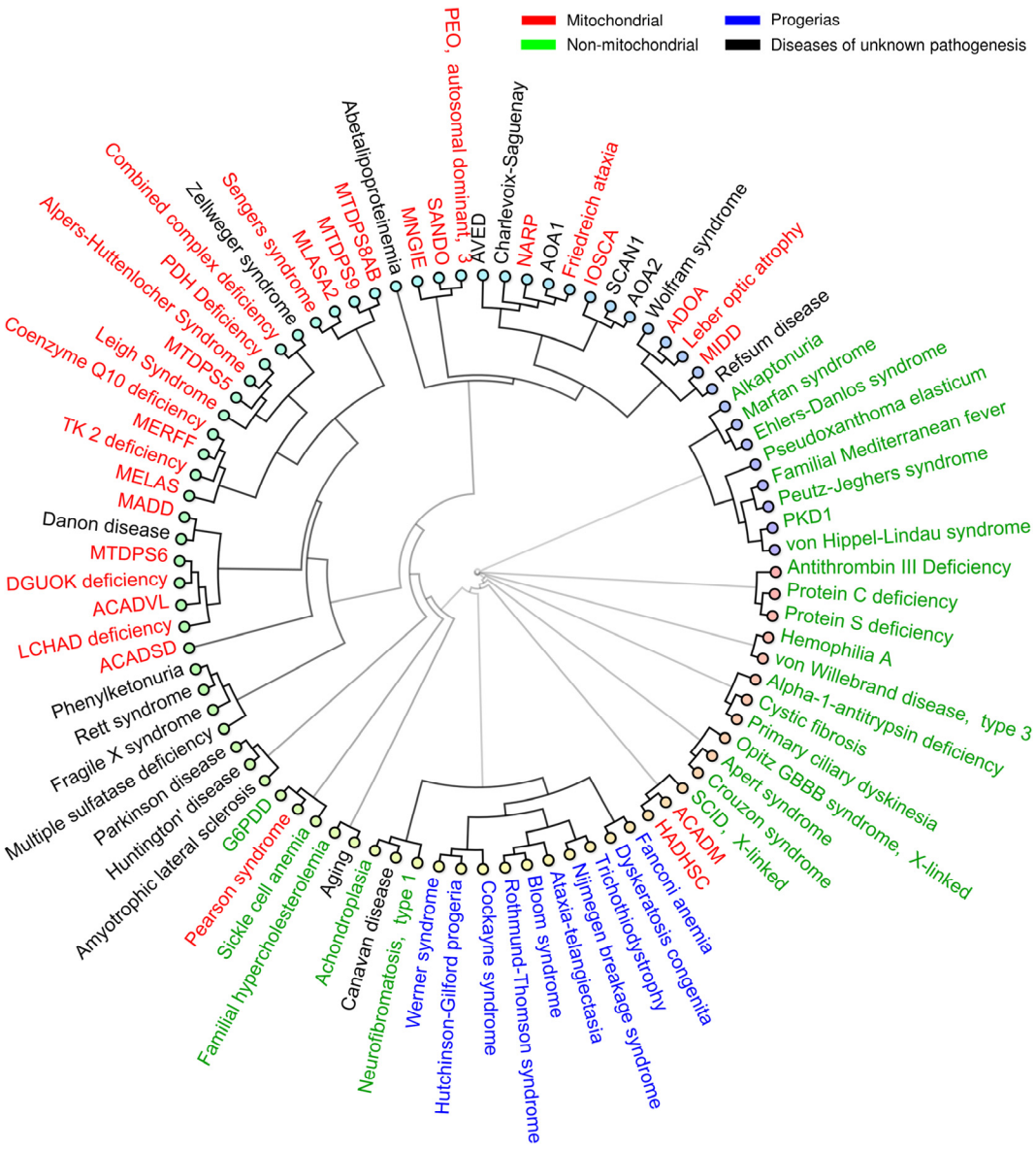
Supplementary Figure 4. Non-mitochondrial diseases in the database do not have a defined clinical spectrum. The average prevalence of clinical parameters across all the non-mitochondrial diseases in the database. Inset: Close-up of the top-20 clinical parameters seen in non-mitochondrial diseases.



Supplementary Figure 5. The approximated age of onset of parameters in mitochondrial diseases. (**A**) The average prevalence of clinical parameters across all the mitochondrial diseases as a function of age. (**B**) The average prevalence of clinical parameters across the mitochondrial diseases with an onset before age 20 as a function of age. (**C**) The average prevalence of clinical parameters across the mitochondrial diseases with an onset after age 20 as a function of age.



Supplementary Figure 6. **Mitochondrial diseases cluster well with each other.** (**A**) An overlay of the prevalence of clinical parameters seen in mitochondrial and non-mitochondrial diseases in the database. Note: The clinical parameters in the mitochondrial and non-mitochondrial diseases do not correspond to each other. For the non-mitochondrial clinical parameters see supplementary figures. (**B**) An example of the clustering of diseases in the database using uncentered similarity metrics and average linkage. The diseases are listed across the top, the dendrogram below, with the clinical parameters denoted on the right. The tint of the square represents the prevalence of the parameter in the given disease. (**C**) The complete clustermap of all the diseases in the database. Each square represents a referenced parameter that can be found online on www.mitodb.com.

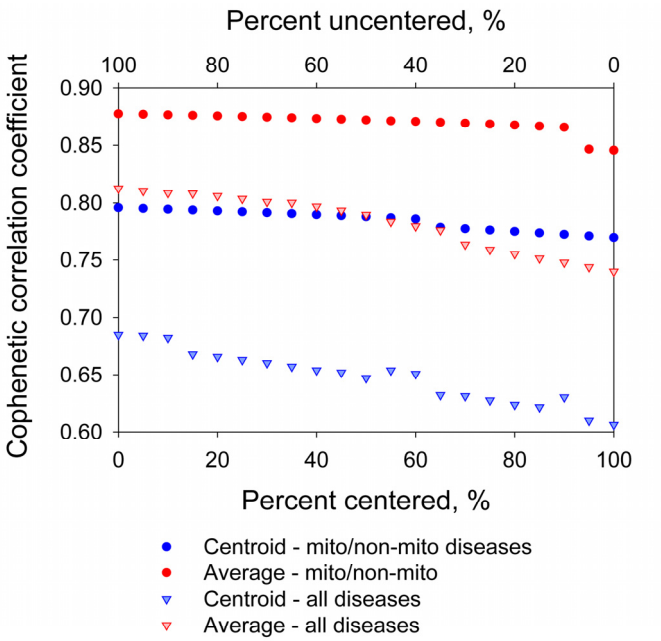


Supplementary Figure 7. Removing neuronal parameters allow the accelerated aging disorders to cluster together. Clustermap using uncentered similarity and average linkage of all the diseases in the database. Blue represents the accelerated aging disorders where the neuronal parameters have been removed.

Supplementary table 1 **A list of the diseases currently in the database and their mean age of onset.**

Mitochondrial	Abbreviation	OMIM		Age of onset	SD	PMID
Autosomal dominant optic atrophy	ADOA	165500		7.21	3.57	20417570
Medium-Chain Acyl-CoA Dehydrongenase Deficiency	ACADM	201450		1.83	0.72	6646897
Short Chain Deficiency of ACYL-CoA Dehydrogenase	ACADSD	201470		1.55	2.23	18054510
Very Long Chain Deficiency of ACYL-CoA Dehydrogenase	ACADVL	201475		0.14	0.15	7769092
Alpers-Huttenlocher Syndrome	MTDPS4A	203700		0.11	0.23	20142534
Sengers syndrome		212350		0.28	0.35	16736096
Friedreich ataxia		229300		15.50	8.00	8815938
3-Hydroxyacyl-CoA Dehydrogenase Deficiency	HADHSC	231530		0.22	0.29	10347277
Multiple Acyl-CoA Dehydrogenase Deficiency	MADD	231680		15.94	13.52	17977044, 20370797
Encephalomyopathy with methylmalonic aciduria	MTDPS9	245400		0.00	0.00	17668387
Deoxyguanosine kinase deficiency	DGUOK deficiency	251880		0.11	0.12	16908739
Leigh Syndrome		256000		0.94	1.04	8602753
Mitochondrial DNA depletion syndrome 6	MTDPS6	256810		0.57	0.36	16909392
Infantile onset spinocerebellar ataxia ,MTDPS7	IOSCA	271245		1.20	0.27	8133312
Maternally transmitted diabetes and deafness syndrome	MIDD	520000		34.60	13.90	11329229
Leber optic atrophy		535000		24.34	13.98	7735876
Mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes	MELAS	540000		20.33	14.53	11708999
Myoclonic Epilepsy associated with Ragged Red Fibers	MERFF	545000		34.67	15.06	21303704
Neuropathy Ataxia and Retinitis Pigmentosa	NARP	551500		14.25	14.01	20953793
Pearson syndrome		557000		0.55	0.71	7581370
Mitochondrial neurogastrointestinal encephalopathy	MNGIE	603041		17.58	11.03	21933806
Coenzyme Q10 deficiency		607426		2.99	3.96	17442627
Sensory ataxic neuropathy, dysarthria and ophthalmoparesis	SANDO	607459		14.33	6.19	15824347
Pyruvate Dehydrogenase Deficiency	PDH deficiency	608782		2.47	3.06	1327585
Long-chain 3-hydroxyacyl–coenzyme A dehydrogenase deficiency	LCHAD	609016		0.67	0.73	9003853
Progressive external ophthalmoplegia, autosomal dominant, 3	PEO, autosomal dominant, 3	609286		40.94	15.14	20479361
Thymidine kinase 2 deficiency	TK2 deficiency	609560		0.61	0.55	18819985
Encephalomyopathy with methylmalonic aciduria	MTDPS5	612073		0.12	0.15	17301081
Encephalomyopathy with renal tubulopathy	MTDPS8AB	612075		0.04	0.05	17486094
Myopathy, mitochondrial progressive, with congenital cataract, hearing loss, and developmental delay	Combined complex deficiency	613076		0.00	0.00	19409522
Myopathy, Lactic acidosis, and Sideroblastic Anemia 2	MLASA2	613561		1.10	1.65	20598274
Non mitochondrial						
Achondroplasia		100800		0.00	0.00	690757
Apert syndrome		101200		0.00	0.00	7645606
Crouzon syndrome		123500		0.00	0.00	15885794
Ehlers-Danlos syndrome		130000		12.67	2.08	17762038
Familial hypercholesterolemia		143890		29.00	18.00	227426
Marfan syndrome		154700		11.40	3.95	9059160
Neurofibromatosis, type 1		162200		14.90	12.90	18277076
Polycystic kidney disease 1	PKD1	173900		54.00	19.00	22367170
Peutz-Jeghers syndrome		175200		19.00	9.45	20126809
Congenital protein C deficiency		176860		24.10	11.90	20126809
von Hippel-Lindau syndrome	VHL	193300		26.25	12.53	8929948
Alkaptonuria		203500		55.38	60.52	21927854
Cystic fibrosis, CF		219700		3.70	7.20	12001283
Primary ciliary dyskinesia		244400		21.04	13.59	9387968
Familial Mediterranean fever		249100		5.50	3.40	9266193
Pseudoxanthoma elasticum		264800		30.50	16.50	17693525
von Willebrand disease, type 3		277480		3.00	6.71	19473418
Opitz GBBB syndrome, X-linked		300000		0.00	0.00	15558842
Severe combined immunodeficiency, X-linked	SCID, X-linked	300400		2.38	1.70	8185357
Glucose-6-phosphate dehydrogenase deficiency	G6PDD	305900		1.79	2.78	7110809
Hemophilia A	HEMA	306700		0.50	0.50	10650861
Sickle cell anemia		603903		2.18	1.19	22224796
Protein S deficiency		612336		23.88	5.84	2959350
Antithrombin III Deficiency		613118		31.67	14.03	1489375
Alpha-1-antitrypsin deficiency		613490		42.73	16.79	309708
Unknown pathogenesis						
Aging		0				
Amyotrophic lateral sclerosis	ALS	105400		54.62	10.90	17296839
Dyskeratosis congenita	DC	127550		29.10	20.23	21436073
Huntington disease		143100		42.30	14.90	11574110
Parkinson disease		168600		62.20	10.60	22362919
Hutchinson-Gilford progeria	HGPS	176670	est:	2.00	2.00	18256394
Abetalipoproteinemia		200100		14.67	5.65	10679949
Ataxia-telangiectasia	AT	208900		2.32	1.20	1377828
Ataxia with oculomotor apraxia type 1	AOA1	208920		5.53	1.30	21465257
Bloom syndrome	BS	210900	est:	0.00	0.00	16763388
Zellweger syndrome		214100		0.00	0.00	9818927
Cockayne syndrome	CS	216400		3.29	2.54	17092472
Wolfram syndrome		222300		4.14	1.95	21968327
Fanconi anemia	FA	227650		6.34	3.97	12393516
Nijmegen breakage syndrome	NBS	251260		0.69	0.05	15033202
Phenylketonuria		261600		3.76	7.32	13452670
Refsum disease		266500		19.73	9.65	2433405
Rothmund-Thomson syndrome	RTS	268400		1.64	1.99	13393794
Charlevoix-Saguenay	ARSACS	270550		5.50	1.60	16961075
Canavan disease		271900		0.81	2.18	9568915
Multiple sulfatase deficiency	MSD	272200		1.00	0.82	18509892
Ataxia with selective vitamin E deficiency	AVED	277460		13.25	8.25	15300460
Werner syndrome	WS	277700		34.71	8.00	16673358
Danon disease		300257		12.43	3.82	19318653
Fragile X syndrome		300624		3.10	2.18	22134579
Rett syndrome		312750		4.70	3.70	20728410
Trichothiodystrophy	TTD	601675		2.75	0.35	11709541
Ataxia with oculomotor apraxia type 2	AOA2	606002		14.60	3.40	19696032
Spinocerebellar Ataxia with Axonal Neuropathy	SCAN1	607250		13.67	1.15	12244316

*Some diseases are represented with the abbreviated titles in the main figures.



Supplementary Figure 8. Uncentered similarity metrics and average linkage yields the best clustering. The cophenetic correlation coefficient (see supplementary equations) plotted using different weighting of uncentered and centered similarity and average and centroid linkage and clustering of all the diseases or only mitochondrial and non-mitochondrial diseases.

Supplementary table 2 **Some diseases of unknown pathogenesis display a mitochondrial phenotype.**

Disease	OMIM ID	Mito clustering?	Mito score	SVM	Mitochondrial?
ALS	105400	No	100	0.23	++
Huntington's disease	143100	No	99	0.17	++
Parkinson's disease	168600	No	100	0.26	++
Abetalipoproteinemia	200100	No	53	-0.18	-
Zellweger syndrome	214100	Yes	75	0.37	+ + +
Wolfram syndrome	222300	Yes	94	1.09	+ + +
Phenylketonuria	261600	Yes	59	-0.26	+
Refsum disease	266500	Yes	96	0.33	+ + +
Charlevoix-Saguenay	270550	Yes	100	0.4	+ + +
Canavan disease	271900	No	67	0.74	+ +
Multiple sulfatase deficiency	272200	Yes	73	-1.05	+
AVED	277460	Yes	92	0.38	+ + +
Danon disease	300257	Yes	83	1.65	+ + +
Fragile X syndrome	300624	Yes	36	0.07	+
Rett syndrome	312750	Yes	68	0.61	+ + +
AOA2	606002	Yes	97	0.02	+ + +

The last column represents our interpretation of how strong the mitochondrial features are based on the tests we have done. Each disease receives a +/- sign for clustering with mitochondrial diseases, scoring more than 50 in the mito-score or receiving a positive SVM-score.