

Mitochondrial Haplogroups Associated with Japanese Centenarians, Alzheimer's Patients, Parkinson's Patients, Type 2 Diabetes Patients, Healthy Non-Obese Young Males, and Obese Young Males

Shigeru Takasaki

Toyo University, Izumino 1-1-1, Ouragun Itakuracho, Gunma 374-0193, Japan

Abstract

There are strong connections among mitochondria, aging, and a wide variety of diseases. In this paper, the relations between eight classes of Japanese people—96 centenarians, 112 semi-supercentenarians (over 105 years old but less than 116 years old), 96 Alzheimer's disease (AD) patients, 96 Parkinson's disease (PD) patients, 96 type 2 diabetes (T2D) patients, 96 T2D patients with angiopathy, 96 healthy non-obese young males and 96 healthy obese young males—and their mitochondrial single nucleotide polymorphism (mtSNP) frequencies at individual mtDNA positions of the entire mitochondrial genome were examined using the radial basis function (RBF) network and a modified RBF method. New findings of mitochondrial haplogroups were obtained for individual classes. The centenarians were found to be associated with the haplogroups D4b2a, B5b, and M7b2; the semi-supercentenarians with B4c1a, F1, B4c1c1, B4c1b1, and M1; the AD patients with G2a and N9b1; the PD patients with N9a, G1a, B4e, and M7a1a; the T2D patients with D4b2b, M8a, and B5b; the T2D patients with angiopathy with N9a2, D4b1, and G2a; the healthy non-obese young males with D4g, N9a, D4b2b, and B4b/d/e; and the healthy obese young males with M7b2, D4b2b, B4c1, and M7a1a. These results are different from the previously reported haplogroup classifications. As the proposed analysis method can predict a person's mtSNP constitution and probabilities of becoming a centenarian, AD patient, PD patient, or T2D patient, it may be useful in the initial diagnosis of various diseases or longevity.

Keywords: mtSNPs; Haplogroup; Semi-supercentenarians; Centenarians; Alzheimer's disease; Parkinson's disease; Type 2 diabetes patients; Healthy non-obese young males; Obese young males; Radial basis function (RBF)

Introduction

Mitochondria are essential cytoplasmic organelles generating cellular energy in the form of adenosine triphosphate by oxidative phosphorylation. Because most cells contain hundreds of mitochondria, each having multiple copies of mitochondrial DNA (mtDNA), each cell contains several thousands of mtDNA copies. The mutation rate for mtDNA is a very high, and when mtDNA mutations occur the cells contain a mixture of wild-type and mutant mtDNAs. As the mutations accumulate, the percentage of mutant mtDNAs increases and the amount of energy produced within the cell can decline until it falls below the level necessary for the cell to function normally. When this bioenergetic threshold is crossed, disease symptoms appear and become progressively worse. Mitochondrial diseases encompass an extraordinary assemblage of clinical problems, usually involving tissues that require large amounts of energy, such as heart, muscle, kidney, and endocrine tissues [1-3].

Although mtDNA mutations have been reported to be related to aging and a wide variety of diseases—such as Parkinson's disease (PD), Alzheimer's disease (AD), type 2 diabetes, and various kinds of cancer [4-20]—those reports have been focused on the amino acid replacements caused by mtDNA mutations. Mitochondrial functions can of course be affected directly by amino acid replacements, but they can also be affected indirectly by mutations in mtDNA control regions. It is therefore important to examine the relations between all mtDNA mutations and centenarians or disease patients.

In the present study the relations between each of eight classes of Japanese people and their mitochondrial single nucleotide polymorphism (mtSNP) frequencies at mtDNA positions throughout

the mitochondrial genome were examined using a classification method based on the predictions of a radial basis function (RBF) network [21,22] and using a modified version of that classification method [23]. This examination revealed new mitochondrial haplogroups characteristic of these classes, and the relations between the haplogroups and classes differ from those reported previously [15,16,24,25].

Materials and Methods

mtSNPs for the eight classes of people

Tanaka et al. [26] sequenced the complete mitochondrial genomes of 672 Japanese individuals to construct an East Asia mitochondrial DNA (mtDNA) phylogeny [26]. Using these sequences and other published Asian sequences, they constructed the phylogenetic tree for macrohaplogroups M and N [26-28]. Kong et al. [29] recently corrected the above sequences by re-sequencing the dubious fragments and segments [29]. In the present study the mtSNPs in 112 Japanese semi-supercentenarians were obtained from the report by Bilal et al. [25] and the mtSNPs in the other classes—96 Japanese centenarians, 96 Japanese Alzheimer's disease (AD) patients, 96 Japanese Parkinson's disease (PD) patients, 96 Japanese type 2 diabetes (T2D) patients, 96 Japanese T2D patients with angiopathy, 96 Japanese healthy non-obese young males,

Corresponding author: Shigeru Takasaki, Toyo University, Izumino 1-1-1, Ouragun Itakuracho, Gunma 374-0193, Japan, Tel: +81-276-82-9024; Fax: +81-276-82-9033; E-mail: takasaki@khh.biglobe.ne.jp, s_takasaki@toyo.jp

Received May 05, 2011; Accepted May 25, 2011; Published May 27, 2011

Citation: Takasaki S (2011) Mitochondrial Haplogroups Associated with Japanese Centenarians, Alzheimer's Patients, Parkinson's Patients, Type 2 Diabetes Patients, Healthy Non-Obese Young Males, and Obese Young Males. J Proteomics Bioinform 4: 106-112. doi:10.4172/jpb.1000175

Copyright: © 2011 Takasaki S. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

and 96 Japanese healthy obese young males—were obtained from the GiiB Human Mitochondrial Genome Polymorphism Database (<http://mitsnp.tmig.or.jp/mitsnp>).

mtSNP classification using a RBF network

A RBF network is an artificial network used for supervised learning problems such as regression, classification, and time series prediction. In the supervised learning a function is inferred from the examples (training set) that a teacher supplies. The elements in the training set are paired values of the independent (input) variable and dependent (output) variable. The RBF network shown in Figure 1 is the supervised learning, and the mtSNP classification for the eight classes of people was carried out individually. In the mtSNP classification for the centenarians, the mtSNPs of the centenarians were regarded as correct and mtSNPs of the other seven classes of people (i.e., semi-supercentenarians, AD patients, PD patients, T2D patients, T2D patients with angiopathy, non-obese young males, and obese young males) were regarded as incorrect. The mtSNP classifications for the other seven classes were carried out in the same way as that for the centenarians (Figure 1).

The mitochondrial genome sequences of the eight classes of people were divided into two sets, one of training data and the other of validation data, and the processes of the classifications were carried out in two phases: training and validation. The steps are described in detail elsewhere [30].

Modified classification based on probabilities predicted by the RBF network

Since a RBF network can predict the probabilities that persons with certain mtSNPs belong to certain classes, these predicted probabilities were used to identify mtSNP features. By examining the relations between individual mtSNPs and the persons with high predicted probabilities of belonging to one of these classes, other mtSNPs useful

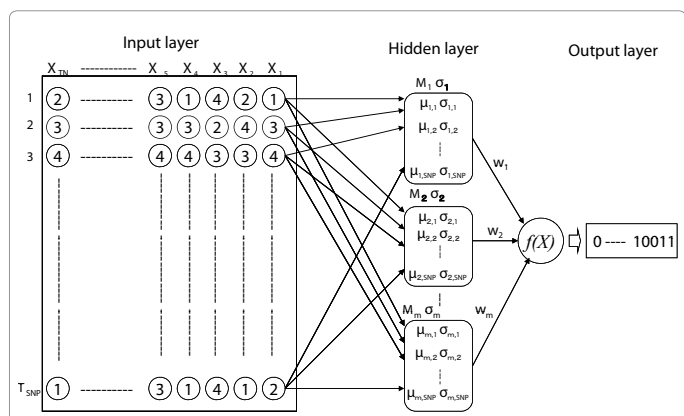


Figure 1: RBF network representation of the relations between individual mtSNPs and eight classes of people. The input layer is the set of mtSNP sequences represented numerically (A, G, C, and T are converted to 1, 2, 3, and 4). The hidden layer classifies the input vectors into several clusters depending on the similarities of individual input vectors. The output layer is determined depending on which analysis is carried out. In the case of centenarians, 1 corresponds to centenarians and 0 corresponds to other seven classes of people. In the case of AD patients, 1 corresponds to AD patients and 0 corresponds to other seven classes of people. The other classes of people (semi-supercentenarians, PD patients, T2D patients, healthy non-obese young males, and healthy obese young males) are carried out in a similar way. X_i : i -th input vector, T_N : maximum number of vectors (in this example, $T_N=448$ (64×7)), T_{SNP} : maximum number of mtSNPs (in this example, $T_{SNP}=562$), M_m : the location vector, m : the number of basis functions, μ : basis function, σ : standard deviation, w_i : i -th weighting variable, $f(X)$: weighted sum function.

Eight classes	Classification ID	Number of individuals	Predicted probability (%)
Centenarians	1	39	66.7
	2	9	55.6
	3	26	42.3
	4	17	41.1
	5	15	40
	6	22	13.6
	7	23	13
	8	28	7.1
	9	20	5
	10	7	0
	11	22	0
	12	113	0
	13	85	0
	14	97	0
Semi-supercentenarians	1	12	66.7
	2	29	48.3
	3	48	35.4
	4	26	30.8
	5	30	20
	6	55	16.4
	7	47	14.9
	8	30	13.3
	9	24	4.2
	10	99	1
	11	24	0
	12	99	0
AD patients	1	17	88.3
	2	33	45.5
	3	36	44.4
	4	27	40.7
	5	36	19.4
	6	39	0
	7	32	0
	8	303	0
PD patients	1	10	60
	2	16	56.3
	3	15	53.3
	4	17	41.2
	5	6	33.3
	6	13	30.8
	7	18	27.8
	8	21	19
	9	59	15.3
	10	17	11.8
	11	26	11.5
	12	10	10.7.7
	13	13	4.2
	14	24	3.4
	15	59	0
	16	9	0
	17	31	0
	18	9	0
	19	60	0
	20	6	0
	21	84	0
T2D patients	1	14	50
	2	48	35.4
	3	17	35.3
	4	31	29
	5	30	23.3
	6	40	17.5
	7	32	12.5
	8	36	8.3
	9	33	6.1
	10	47	4.3
	11	32	0
	12	89	0
	13	74	0
T2D patients with angiopathy	1	16	81.3
	2	19	57.9
	3	35	51.4
	4	39	43.6
	5	14	14.3
	6	32	9.4
	7	106	0
	8	262	0
Healthy non-obese young males	1	28	57.1
	2	11	54.5
	3	12	50
	4	5	40
	5	16	37.5
	6	6	33.3
	7	23	21.7
	8	10	20
	9	62	17.7
	10	10	10
	11	12	8.3
	12	24	8.3
	13	27	7.4
	14	19	5.3
	15	29	3.4
	16	27	0
	17	15	0
	18	72	0
	19	115	0
Obese young males	1	32	62.5
	2	38	52.6
	3	29	31
	4	24	29.2
	5	28	28.6
	6	97	0
	7	275	0

Table 1: mtSNP classifications based on the eight classes of people.

for distinguishing between the members in different classes were identified. The modified classification method based on the probabilities predicted by the RBF network was carried out in the following way [23].

- 1) Select the target class to be analyzed.
- 2) Rank individuals according to their predicted probabilities of belonging to the target class.
- 3) Either select individuals whose probabilities are greater than a certain

value or select the desired number of individuals and set them as a modified cluster.

Results and Discussion

Associations between Asian/Japanese haplogroups and mtSNPs for the eight classes of people

The mtSNP classifications obtained by the RBF network for the eight classes of people are listed in Table 1. The average predicted probabilities

		Centenarians	Semi-super centenarians	AD patients	PD patients	T2D patients	T2D patients with antiangiopathy	Non-obese young males	Obese young Males	
M	M1	O 7%								
	M7a1a				O 13%				O 13%	
	M7b2	O 13%							O 47%	
	M8a					O 27%				
	G1a				O 20%					
	G2a			O 53%			O 13%			
	D	D4b1						O 20%		
		D4b2a	O 27%							
		D4b2b					O 40%		O 13%	O 20%
		D4g							O 33%	
N	B4c1								O 13%	
	B4c1a		O 40%							
	B4c1b1		O 7%							
	B4c1c1		O 13%							
	B4b/d/e				O 13%					
	B4e					O 13%				
	B5b	O 27%			O 20%					
	F1		O 27%							
	N9a				O 20%			O 27%		
	N9a2						O 47%			
	N9b1			O 20%						

Table 2: Haplogroup-class relations determined using the 15 individuals with the highest predicted probabilities.

		Centenarians (69)	Semi-super centenarians (18)	AD patients (65)	PD patients (24)	T2D patients (7)	T2D patients with antiangiopathy (72)	Non-obese young males (16)	Obese young males (58)	
M	M1		O 6%						9%	
	M7a1								O 7%	
	M7a1a	3%			O 17%					
	M7b2	O 10%	6%						O 12%	
	M8a					O 14%				
	G1			6%						
	G1a				O 13%					
	G2a			O 12%			O 8%			
	D	D4a	12%					10%		
		D4b1						O 4%		
D4b1b									5%	
D4b2				5%						
D4b2a		O 7%							2%	
D4b2b		13%				O 29%		O 12%	O 5%	
D4d1a		6%								
N	D4e2			5%			10%			
	D4g							O 31%	7%	
	A1			6%						
	B4c	3%		8%					O 9%	
	B4c1									
	B4c1a		O 33%							
	B4c1b1		O 6%							
	B4c1c1		O 6%							
	B4b/d/e							O 12%		
	B4e				O 8%					
	B5b	O 6%				O 14%	7%		9%	
F1	3%	O 22%	11%			4%				
N9a				O 21%			O 25%			
N9a2			9%			O 12%				
N9b1			O 6%					9%		

O: haplogroups classified in the highest 15 individuals

Table 3: Haplogroup-class relations determined using the individuals whose predicted probabilities were greater than 50%.

	M1	M7a1	M7b2	G2a	G1	D4a	D4b1	D4e2	D4g	F1	B5b	B4c1a	B4c1b1	B4c1c1	B4b/d/e	A2	N9a
15 (53% - 74%)	7%									27%		40%	7%	6%		7%	
30 (51% - 74%)	3%		17%			3%				13%		20%	3%	3%	10%	17%	
45 (45% - 74%)	2%		11%	4%		29%				9%	2%	13%	4%	2%	7%	11%	
60 (37% - 74%)	5%	1%	10%	5%		25%	5%			7%	3%	10%	5%	5%	5%	8%	3%
75 (29% - 74%)	4%	1%	8%	5%		20%	4%	8%		7%	5%	8%	4%	4%	4%	8%	5%
90 (23% - 74%)	3%	2%	7%	4%	1%	18%	3%	7%		6%	4%	7%	3%	3%	3%	8%	4%
112 (4% - 74%)	3%	7%	5%	4%	3%	14%	3%	5%	2%	4%	4%	5%	3%	4%	4%	6%	4%

Table 4: Semi-supercentenarian haplogroups classifications obtained using various numbers of individuals with the highest predicted probabilities.

	Statistical technique	Proposed method
Technique	Relative relations between target and normal data	Supervised learning (RBF) by using correct and incorrect data
Analysis position	Each locus of mtDNA polymorphisms (independent position)	Entire loci of mtDNA polymorphisms (successive positions)
Input (required data)	Target (individual cases) and control (normal data)	Correct (individual cases) and incorrect (others except correct)
Output (results)	Odds ratio or relative risk	Clusters with predictions
Analysis	Check odds ratio or relative risk at each position	Check individuals in clusters based on prediction probabilities

Table 5: Differences between the statistical technique and the proposed (modified RBF) method.

of the highest cluster (classification ID 1) for the eight classes of people were 66.7% for both the centenarians, and the semi-supercentenarians, 88.3% for the AD patients, 60% for the PD patients, 50% for the T2D patients, 81.3% for the T2D patients with angiopathy, 57.1% for the healthy non-obese young males, and 62.5% for the healthy obese young males. As individual classes have different predicted probabilities, for each of the eight classes the 15 individuals with the highest predicted probabilities were selected to examine the relations between Asian/Japanese haplogroups and mtSNPs [26-28]. For the centenarians the association between the haplogroups and mtSNPs that was based on the highest 15 individuals is shown in detailed form in Figure 2A* (Provided in the Supplementary) and in simple form in Figure 2A (Provided in the Supplementary), and the corresponding associations for the other seven classes are shown in simple form in Figures 2B (Provided in the Supplementary) to 2H. The features of associations for the individual classes are described as follows: for the centenarians L3-N-B5b (27%), L3-M-D-D4b2a (27%), and L3-M-M7b2 (13%); for the semi-supercentenarians L3-N-B4c1a (40%), L3-N-F1 (27%), L3-N-B4c1c1 (13%), L3-N-B4c1b1 (7%), and L3-M-M1 (7%); for the AD patients L3-M-G2a (53%) and L3-N-N9b1 (20%); for the PD patients L3-N-N9a (20%), L3-M-G1a (20%), L3-M-M7a1a (13%), and L3-N-B4e (13%); for the T2D patients L3-M-D-D4b2b (40%), L3-M-M8a (27%), and L3-N-B5b (13%); for the T2D patients with angiopathy L3-N-N9a2 (47%), L3-M-D-D4b1 (20%), and L3-M-G2a (13%); for the healthy non-obese young males L3-M-D-D4g (33%), L3-N-N9a (27%), L3-M-D-D4b2b (13%), and L3-N-B4b/d/e (7%); and for the healthy obese young males L3-M-D-M7b2 (47%), L3-M-D-D4b2b (20%), L3-M-M7a1a (13%), and L3-N-B4c1 (13%).

The relations between the haplogroups for the eight classes of people are listed in Table 2. The haplogroup M7a1a was common in PD patients and obese young males; M7b2 was common in centenarians and obese young males; G2a was common in AD patients and T2D patients with angiopathy; D4b2b was common in T2D patients, non-obese young males, and obese young males; B5b was common in centenarians and T2D patients; and N9a was common in PD patients and non-obese young males.

Then the individuals whose probabilities predicted using the modified classification method were greater than 50% were selected and their nucleotide distributions at individual mtDNA positions were examined. The individuals selected were 69 centenarians, 18 semi-supercentenarians, 65 AD patients, 24 PD patients, 7 T2D patients, 72 T2D patients with angiopathy, 16 healthy non-obese young males, and 58 healthy obese young males. The associations between the

haplogroups and mtSNPs for the eight classes of people are shown in Figure 3A to H (Provided in the Supplementary). The relations among the haplogroups for the eight classes of people are listed in Table 3. In the mtSNP analysis for the individuals whose probabilities were greater than 50% there were 30 haplogroups for the eight classes, whereas in the analysis for the 15 individuals with the highest predicted probabilities there were only 21 haplogroups. As a result, the ratios of individual haplogroups of the individuals whose probabilities were greater than 50% tended to be lower than those of the 15 individuals with the highest predicted probabilities. In addition, the analysis based on the individuals with probabilities greater than 50% yielded 12 common haplogroups among the eight classes of people, whereas the analysis based on the 15 individuals with the highest predicted probabilities yielded only 6 common haplogroups among the eight classes.

The relations between haplogroups of individual people for two cases – case 1 is the 15 individuals with the highest predicted probabilities and case 2 is the individuals whose predicted probabilities are greater than 50% – are as follows: the centenarians were classified into more haplogroups in case 2 than in case 1; the ratios of the haplogroups D4b2a, B5b, and M7b2 were higher in case 1 than in case 2 (respectively 27% and 7%, 27% and 6%, and 13% and 10%); and new haplogroups D4a (12%), D4b2b (13%), D4d1a (6%), M7a1a (3%), F1 (3%), and B4c (3%) were classified in case 2.

Although the semi-supercentenarians were classified into similar haplogroups in the two cases, their ratios were lower in case 2. That is, in case 2 the ratios of B4c1a, F1, M1, B4c1b1, and B4c1c1 were respectively decreased from 40% to 33%, from 27% to 22%, from 7% to 6%, from 7% to 6%, and from 13% to 6%. In addition, a new haplogroup M7b2 (6%) was classified in case 2.

The AD patients were also classified into more haplogroups in case 2 than in case 1. The ratios of G2a and N9b1 were lower in case 2 than in case 1 (respectively 12% vs. 53%, and 6% vs. 20%), and new haplogroups F1 (11%), N9a2 (9%), B4c1 (8%), A1 (6%), G1 (6%), D4b2 (5%), and D4e2 (5%) were classified in case 2.

Although the PD patients were classified into the same haplogroups in the two cases, their ratios were different in the two cases. That is, the G1a ratio was 20% in case 1 and 13% in case 2, the N9a ratio was 20% in case 1 and 21% in case 2, the M7a1a ratio was 13% in case 1 and 17% in case 2, and the B4e ratio was 13% in case 1 and 8% in case 2.

Although the T2D patients were classified into the same haplogroups in both cases, the haplogroup ratios were different in the two cases. That

is, the D4b2b ratio was 40% in case 1 and 29% in case 2, the M8a ratio was 27% in case 1 and 14% in case 2, and the B5b was 13% in case 1 and 14% in case 2.

The T2D patients with angiopathy, though, were classified into more haplogroups in case 2 than in case 1, and the ratios of common haplogroups differed between the two cases. The ratios of N9a2, D4b1, and G2a were respectively 47% in case 1 and 12% in case 2, 20% in case 1 and 4% in case 2, and 13% in case 1 and 8% in case 2. Furthermore, the new haplogroups D4a (10%), D4e2 (10%), B5b (7%), and F1 (4%) were classified in case 2.

The healthy non-obese young males were classified into the same haplogroups in both cases and their ratios were also nearly the same in the two cases. That is, the haplogroup D4g, N9a, D4b2b, and B4b/d/e ratios in cases 1 and 2 were respectively 33% and 31%, 27% and 25%, 13% and 12%, and 7% and 12%. This similarity of haplogroup ratios is due to the numbers of selected individuals being nearly the same in both cases: 15 in case 1 and 16 in case 2.

The obese young males were also classified into more haplogroups in case 2 than in case 1. In cases 1 and 2 the ratios of common haplogroups M7b2, D4b2b, B4c1, and M7a1 were respectively 47% and 12%, 20% and 5%, 13% and 9%, and 13% and 7%, and the new haplogroups M1 (9%), B5b (9%), N9b1 (9%), D4g (7%), D4b1b (5%), and D4b2a (2%) were classified in case 2.

Then the relations between the haplogroups of pairs of related classes of people—centenarians and semi-supercentenarians, AD patients and PD patients, T2D patients and T2D patients with angiopathy, and healthy non-obese young males and healthy obese young males—were examined in cases 1 and 2.

Centenarians and semi-supercentenarians: Although in case 1 these two classes of people had no common haplogroups, in case 2 they had two common haplogroups M7b2 and F1.

AD patients and PD patients: Although AD and PD are both brain diseases, these patients had no common haplogroups in either case.

T2D patients and T2D patients with angiopathy: Although these two classes of people had no common haplogroups in case 1, they had a common haplogroup B5b in case 2.

Healthy non-obese young males and healthy obese young males: These two classes of people had a common haplogroup D4b2b in case 1 and had two common haplogroups D4g and D4b2b in case 2. Common haplogroups were found more often in case 2 because more haplogroups were classified in that case.

As there were 112 individuals in the class of semi-supercentenarians, changes in haplogroup classifications with changes in the number of highest-probability individuals selected were examined. As one sees in Table 4, the number of haplogroups classified increased from 6 for the 15 individuals with the highest predicted probabilities to 9 for the 30 with the highest predicted probabilities, to 11 for the 45 with the highest predicted probabilities, to 14 for the 60 with the highest predicted probabilities, to 15 for the 75 with the highest predicted probabilities, to 16 for the 90 with the highest predicted probabilities, and to 17 for all 112 semi-supercentenarians. The ratios of the haplogroups B4c1a and F1 were respectively 40% and 27% for the 15 individuals with the highest predicted probabilities, but they decreased as the numbers of

selected individuals increased and finally became respectively 5% and 4% when all 112 individuals were used. Although the haplogroup D4a was not classified when the 15 individuals with the highest predicted probabilities were used, its ratio was 3% when the 30 individuals with the highest predicted probabilities were used and was 29% when the 45 individuals with the highest predicted probabilities were used. This indicates that most of the semi-supercentenarians belonging to D4a were included in the range of the predicted probabilities 45% to 74%. From Table 4, it is implied that the feature of the semi-supercentenarians appears in the appropriate number of selected individuals used. In the case of the semi-supercentenarians, its number may be 45 individuals used.

Comparison with previous works

After analyzing the results of a large-scale study using hospital-based sampling data, Fuku et al. [16] reported that the mitochondrial haplogroup F in Japanese individuals is associated with a significantly increased risk of type 2 diabetes mellitus (T2DM) (odds ratio 1.53, $P=0.0032$) [16]. In the present study, on the other hand, the haplogroups (risks) of T2D patients and T2D patients with angiopathy were respectively D4b2b (40%), M8a (27%) and B5b (13%), and N9a (47%), D4b1 (20%) and G2a (13%) (Figures 2E, 2F (Provided in the Supplementary)). There were therefore big differences between the analysis of Fuku et al. [16] and results of this study.

In addition, Alexe et al. [15] used the same GiiB data used in the present study and reported the associations between Asian haplogroups and the longevity of Japanese people [15]. They showed the enrichment of longevity phenotype in mtDNA haplogroups D4b2b, D4a, and D5 in the Japanese population by using statistical techniques (*t*-test and *P*-value). In the present study, however, the haplogroups of the centenarians and semi-supercentenarians were respectively D4b2a (27%), B5b (27%), and M7b2 (13%), and B4c1a (40%), F1 (27%), B4c1 (13%), B4c1b1 (7%), and M1 (7%) (Figures 2A, 2B (Provided in the Supplementary)). There were therefore no common haplogroups in both analyses.

Bilal et al. [25] found that the haplogroup D4a was a marker for extreme longevity in Japan by analyzing the complete mtDNA sequences from 112 Japanese semi-supercentenarians (more than 105 but less than 116 years old) combined with previously published data [25]. In the present study, however, the haplogroups of the semi-supercentenarians in the 15 individuals with the highest predicted probabilities and the individuals whose predicted probabilities were more than 50% were respectively B4c1a (40%), F1 (27%), B4c1c1 (13%), B4c1b1 (7%), and M1 (7%), and B4c1a (33%), F1 (22%), B4c1c1 (6%), B4c1b1 (6%), M1 (6%), and M7b2 (6%) (Figure 2B, Figure 3B (Provided in the Supplementary)). In these analyses, the haplogroup D4a was not a marker of the semi-supercentenarians. The various haplogroup classifications were also examined using various numbers of individuals with the highest predicted probabilities (Table 4). In the case of the classifications using the 45 individuals with the highest predicted probabilities, D4a was characterized as the distinct haplogroup. However, there were other haplogroups indicating the characteristics of semi-supercentenarians. They were B4c1a (13%), M7b2 (11%), A2 (11%), F1 (9%), B4b/d/e (7%), B4c1b1 (4%), G2a (4%), B5b (2%), B4c1c1 (2%), and M1 (2%) listed in Table 4. This means that individuals with other haplogroups also have the possibility of becoming semi-supercentenarians.

Differences between statistical technique and the proposed method

Although the previously reported methods analyzed the relations between mtSNPs and Japanese T2D patients, centenarians, and semi-supercentenarians by using standard statistical techniques [15,25], the mutual relations among the other classes of people—AD patients, PD patients, healthy non-obese young males, and healthy obese young males—were not investigated. The differences between and mutual relations among the eight classes of people were described in this study. In addition, the predicted probabilities of associations between mtSNPs and the eight classes of people cannot be obtained by the statistical techniques used in the previous methods, whereas the proposed method is able to compute them from the results obtained when learning the mtSNPs of individual classes.

Although the previous methods used standard statistical techniques, a RBF network was used in the present study because the relations among individual mtSNPs for the eight classes of people should be analyzed as mutual mtSNP connections in the entire population of mtSNPs.

The differences between standard statistical technique and the proposed method are listed in Table 5. In the statistical technique, odds ratios or relative risks are analyzed on the basis of relative relations between target and control data at each polymorphic mtDNA locus. In the proposed method, on the other hand, clusters indicating predicted probabilities are examined on the basis of the RBF using correct and incorrect data for the entire set of polymorphic mtDNA loci. The statistical technique determines characteristics of haplogroups by using independent mtDNA polymorphisms that indicate high odds ratios, whereas the proposed method determines them by checking individuals with high predicted probabilities. This means that the statistical technique uses the results of independent mutation positions, whereas the proposed method uses the results of all mutation positions. As there are the differences between the two methods, which method is better will need to be determined in future research. Furthermore, the proposed method may have possibilities for use in the initial diagnosis of various diseases or longevity on the basis of the individual predicted probabilities.

Acknowledgements

I thank Dr. Y. Kawamura for his encouragement and technical support.

References

- Wallace DC (1999) Mitochondrial diseases in man and mouse. *Science* 283: 1482-1488.
- Vila M, Przedborski S (2003) Targeting programmed cell death neurodegenerative diseases. *Nat Rev Neurosci* 4: 365-375.
- Taylor RW, Turnbull DM (2005) Mitochondrial DNA mutations in human disease. *Nat Rev Genet* 6: 389-402.
- Lin F, Lin R, Wisniewski HM, Hwang Y, Grundke-Iqbal I, et al. (1992) Detection of point mutations in codon 331 of mitochondrial NADH dehydrogenase subunit 2 in Alzheimer's brains. *Biochem Biophys Res Commun* 182: 238-246.
- Schoffner JM, Brown MD, Torroni A, Lott MT, Cabell MF, et al. (1993) Mitochondrial DNA variants observed in Alzheimer disease and Parkinson disease patients. *Genomics* 17: 171-184.
- Kosel S, Egensperger R, Mehraein P, Graeber MB (1994) No association of mutations at nucleotide 5460 of mitochondrial NADH dehydrogenase with Alzheimer's disease. *Biochem Biophys Res Commun* 203: 745-749.
- Mayr-Wohlfart U, Paulus C, Rodel G (1996) Mitochondrial DNA mutations in multiple sclerosis patients with severe optic involvement. *Acta Neurol Scand* 94: 167-171.
- Schnopp NM, Kosel S, Egensperger R, Graeber MB (1996) Regional heterogeneity of mtDNA heteroplasmy in parkinsonian brain. *Clin Neuropathol* 15: 348-352.
- Simon DK, Mayeux R, Marder K, Kowall NW, Beal MF, et al. (2000) Mitochondrial DNA mutations in complex I and tRNA genes in Parkinson's disease. *Neurology* 54: 703-709.
- Tanaka M, Fuku N, Takeyasu T, Guo LJ, Hirose R, et al. (2002) Golden mean to longevity: rareness of mitochondrial cytochrome b variants in centenarians but not in patients with Parkinson's disease. *J Neurosci Res* 70: 347-355.
- Dawson TM, Dawson VL (2003) Molecular pathway of neurodegeneration in Parkinson's disease. *Science* 302: 819-822.
- Ross OA, McCormack R, Maxwell LD, Duguid RA, Quinn DJ, et al. (2003) mt4216C variant in linkage with the mtDNA TJ cluster may confer a susceptibility to mitochondrial dysfunction resulting in an increased risk of Parkinson's disease in the Irish Exp Gerontol 38: 397-405.
- Lustbader JW, Cirilli M, Lin C, Xu HW, Takuma K, et al. (2004) Aβ directly links Aβ to mitochondrial toxicity in Alzheimer's disease. *Science* 304: 448-452.
- Niemi AK, Moilanen JS, Tanaka M, Hervonen A, Hurme M, et al. (2005) A combination of three common inherited mitochondrial DNA polymorphisms promotes longevity in Finnish and Japanese subjects. *Eur J Hum Genet* 13: 166-170.
- Alexe G, Fuku N, Bilal E, Ueno H, Nishigaki Y, et al. (2007) Enrichment of longevity phenotype in mtDNA haplogroups D4b2b, D4a, and D4 in the Japanese population. *Hum Genet* 121: 347-356.
- Fuku N, Park KS, Yamada Y, Nishigaki Y, Cho YM, et al. (2007) Mitochondrial haplogroup N9a confers resistance against type 2 diabetes in Asians. *Am J Hum Genet* 80: 407-415.
- Chinnery PF, Mowbray C, Patel SK, Elson JL, Sampson M, et al. (2008) Mitochondrial DNA haplogroups and type 2 diabetes: a study of 897 cases and 1010 controls. *J Med Genet* 44: e80.
- Kim W, Yoo TK, Shin DJ, Rho HW, Jin HJ, et al. (2008) Mitochondrial haplogroup analysis reveals no association between the common genetic lineages and prostate cancer in the Korean population. *PLoS ONE* 3: e2211.
- Maruszak A, Canter JA, Styczynska M, Zekanowski C, Barcikowska M (2008) Mitochondrial haplogroup H and Alzheimer's disease—Is there a connection? *Neurobiol Aging* 30: 1749-55.
- Feder J, Blech I, Ovadia O, Amar S, Wainstein J, et al. (2008) Differences in mtDNA haplogroup distribution among 3 Jewish populations alter susceptibility to T2DM complications. *BMC Genomics* 9: 198.
- Poggio T, Girosi F (1990) Networks for approximation and learning. *Proc IEEE* 78: 1481-1497.
- Wu CH, McLarty JW (2000) Neural Networks and Genome Informatics.
- Takasaki S (2009) Mitochondrial haplogroups associated with Japanese centenarians, Alzheimer's patients, Parkinson's patients, type 2 diabetic patients and healthy non-obese young males. *J Genetics and Genomics* 36: 425-434.
- Saxena, R, de Bakker PI, Singer K, Mootha V, Burt N, et al. (2006) Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *Am J Hum Genet* 79: 54-61.
- Bilal E, Rabadan R, Alexe G, Fuku N, Ueno H, et al. (2008) Mitochondrial DNA haplogroup D4a is a marker for extreme longevity in Japan. *PLoS ONE* 3: e2421.
- Tanaka M, Cabrera VM, Gonzalez AM, Larruga JM, Takeyasu T, et al. (2004) Mitochondrial genome variation in Eastern Asia and the peopling of Japan. *Genome Res* 14: 1832-1850.
- Herrnstad C, Elson JL, Fahy E, Preston G, Turnbull DM, et al. (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian and European haplogroups. *Am J Hum Genet* 70: 1152-1171.
- Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, et al. (2003) Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet* 73: 671-676.

29. Kong QP, Salas A, Sun C, Fuku N, Tanaka M, et al. (2008). Distilling artificial recombinants from large sets of complete mtDNA genomes. PLoS ONE 3: e3016.
30. Takasaki S, Kawamura Y, Konagaya A (2006) Selecting effective siRNA sequences by using radial basis function network and decision tree learning. BMC Bioinformatics 7: S22.