

C. 本研究における対照群の選定

Barzilai ら(4)は人種的、文化的にも均一な Ashkenazy 民族を対象として、CETP 遺伝多型が長寿と関連のあることを報告した。彼らは長寿者として 95 歳以上の Ashkenazy 民族、長寿遺伝子を 50%の確率で受け継ぐものとして長寿者の子供 (offspring)、そして長寿者と環境を同じにするが遺伝的には関連の無い offspring の配偶者を対照群として設定した。この研究以来、長寿者、長寿者の子孫 (主に子)、そして子の配偶者の 3 群で遺伝多型を比較するという手法が Genetics of Healthy Aging study (GEHA) をはじめ多くの長寿関連遺伝子研究に取り入れられている。われわれもこれに習い、182 名の超百寿者、126 名の子孫、および 93 名の子孫の配偶者の血液サンプルを収集した。さらに現在、子孫およびその配偶者のサンプル収集を継続している (目標各 150 名)。しかし、長寿者の子孫やその配偶者は潜在的に 100 歳まで生きる可能性を持っている。先行研究ではこの確率を 1%程度と見積もっており、統計上は問題にならないとしている。理論的には百寿者と同年代に生まれ平均年齢で死亡した人が理想的な対照群ということになるが、今から数十年前に亡くなった人の遺伝子サンプルの入手は現実的ではない。そこでわれわれは東京都老人医療センター 剖検病理科 沢辺元司部長らと共同で 1990 年以降同施設で亡くなった連続剖検約 1400 例 (すでに遺伝子を抽出済み) のうち平均年齢付近で亡くなったもの約 400 例を対照群として選定することを思いついた。この対照群は 1920-30 年頃に生まれているが、この年代の本邦においては移民などによる人種構成の影響はほとんどなく、1900 年頃に生まれている百寿者、超百寿者の比較対照群として最適と考えられる。すでに死亡年齢、死因が確定されている対照群を設定している長寿遺伝子研究は国際的にも稀であり、われわれの長寿遺伝子研究の科学的価値を世界的レベルに高めるものと期待されている。

(倫理面への配慮)

本調査は慶應義塾大学医学部倫理委員会の承認を得た。

D. 結論

長寿関連遺伝子、老化制御遺伝子の研究において、長寿者の比較対照群として、長寿者の子孫とその配偶者、さらに平均年齢付近で死亡した剖検例を選定した。これにより本研究の科学的妥当性が世界的評価に耐えうるものとなることが期待される。

E. 健康危険情報

なし

F. 研究発表

1. 論文発表

新井康通, 広瀬信義. 寿命決定因子. 内分泌・糖尿病. 2006; 23: 422-430.

2. 学会発表

なし

G. 知的財産権の出願・登録状況

なし

H. 参考文献

- 1) Geesaman BJ, et al. Proc Natl Acad Sci U S A. 2003; 100, 14115-14120.
- 2) Beekman, et al. J Gerontol Bio Sci. 2006; 61: 355-362.
- 3) Nebel A, et al. Proc Natl Acad Sci U S A. 2005; 102: 7906-7909.
- 4) Barzilai N, et al. JAMA . 2003; 290, 2030-2040.

Proteomics は長寿科学で応用できるか？

共同研究者 佐藤雄治 東京都老人総合研究所ゲノム機能研究チーム
共同研究者 遠藤玉夫 東京都老人総合研究所ゲノム機能研究チーム
共同研究者 三浦ゆり 東京都老人総合研究所ゲノム機能研究チーム
主任研究者 広瀬信義 慶應義塾大学医学部内科（老年内科）講師

研究要旨

癌、動脈硬化研究で常用される Proteomics が人長寿科学でも応用できるかを検討した。超百寿者と若年対照群の血漿を元に 2 次元電気泳動を行った。PH4-7 の酸性領域において、明らかに濃度の異なるタンパク質が同定された。これらの中には機能が不明のタンパク質も存在した。Proteomics が人長寿科学に応用可能であることが示された。この方法を応用することにより網羅的なタンパク質の探索が可能となり、抗老化物質、老化指標の同定に有用と考えられた。

A. 研究目的

Proteomics は網羅的な蛋白分子の探索方法として、癌、動脈硬化などの研究で常用されている。従来はあるタンパク質にターゲットを絞って研究を行っていた。この方法により疾患や病態と特定のタンパク質の関係が明らかにされてきた。

網羅的な探索方法をとることにより思いがけないタンパク質が関与することが明らかになってきた。この成果により癌、動脈硬化の臨床、研究が飛躍的に進歩してきている。老化研究、特に人の老化研究で網羅的な蛋白分子の探索、即ち proteomics が可能であるかについての検討は我々の知る限りでは報告がない。

そこで我々は超百寿者（105 歳以上者）と若年対照群で血漿蛋白を用いた proteomics を行い、人長寿研究における proteomics の可能性について検討を試みた。

B. 研究方法

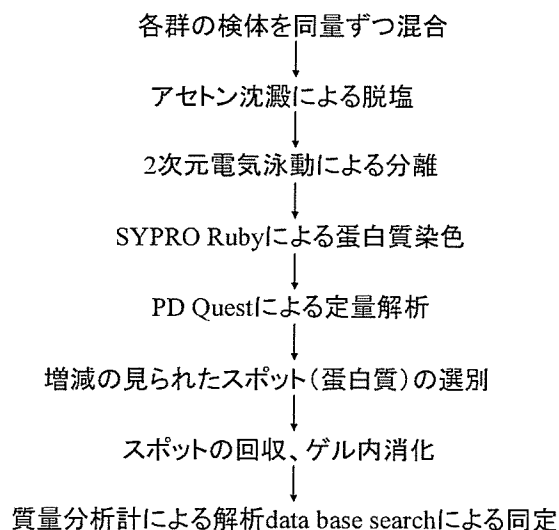
B-1, 対象

長寿形質の代表群とし超百寿者（年齢 106—109 歳）女性 10 名を設定した（SSC 群）。比較対照群として若年者（年齢 20—39 歳）女性 10 名を設定した（若年群）。

B-2, 方法

分析検体の調整、分離、同定（図 1）
SSC 群、若年群の検体を同量ずつ混合した。アセトン沈殿による脱塩後 2 次元電気泳動により分離した（PI 4—7）。ついで SYPRO Ruby によるタンパク質染色を行った。PD Quest により定量解析を行い増減の見られたスポットを選びスポットの回収を行った。ついでトリプシンで消化して質量分析計で解析した。得られたデータを基に Data base search を用いてタンパク質の同定を行った。

図 1. 解析の流れ



倫理面への配慮

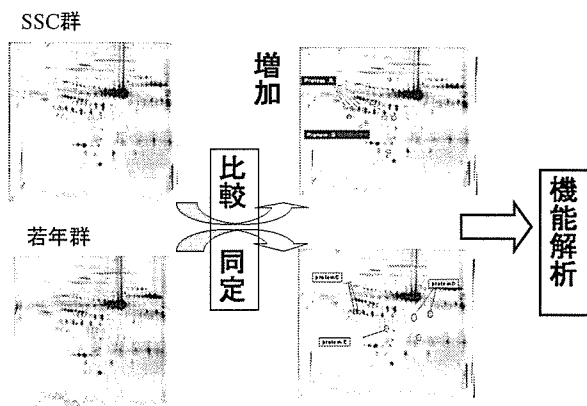
本研究は慶應大学医学部、東京都老人総合研究所の倫理委員会の承認を得た。調査については十分な説明を行い文書による同意を得た。

C. 研究結果 (図2)

今回の2次元電気泳動ではPI4-7までの解析を行った。

解析の結果SSC群では若年群に比較して増加するタンパク質と減少するタンパク質が数種類あることが判明した。増加するタンパク質としては4種(アイソフォームを除く)、減少するタンパク質としては4種類が存在した。このうち同定が出来なかったもの2種、機能が不明なものも2種であった。

図2. 解析結果



D. 考察

この結果は予備的なものであるが、癌、動脈硬化の研究で有力な解析方法である proteomics が人の老化研究でも応用可能であることを示した。今後この解析で判明したタンパク質の経年変化、機能解析を調べる予定である。従来のターゲット分子と同じタンパク質も同定されたがそれ以外のタンパク質も同定されており新しい物質の解明が予想される。同定されたタンパク質は、老化に伴って起こる炎症亢進に関連する可能性のあるもの、遺伝子多型性の違いによる可能性のあるものなどがあり、他の検査データとの相関、遺伝子多型との関連を調べる必要がある。この検討を通して判明したタンパク質の意義が明らかになると考えられる。すなわち老化指標や抗老化物質の同定に有用であることが期待できよう。

本研究では酸性領域での2次元電気泳動で血漿蛋白の分離を行った。アルカリ領域での

proteomics を行うことにより老化に伴い変化するタンパク質のさらなる同定が予想される。分析はタンパク質の分離を行ったが、同定の方法を検討することにより糖タンパク質の解析も可能である。我々はABO血液型の頻度が百寿者では若年群に比較して異なることを報告している。ABO血液型は糖鎖により決定されているため、長寿に糖鎖が関与していることが予想される。糖鎖は構造が不均一であり解析が困難であったが解析技術が急速に進歩しており包括的な解析として glycomics が提唱され様々な生物現象に応用されている。この結果は予備的なものであるが、老化研究に proteomics さらに glycomics が応用できることを示した報告であり有用な情報をもたらすことが期待される。

E. 結論

SSCと若年対照群の血漿蛋白を proteomics を用いて比較検討した。SSCでは若年群と濃度の異なるタンパク質があることが判明した。判明したタンパク質を検討することにより網羅的に老化指標、抗老化物質の探索が可能となろう。ABO血液型が百寿者と若年者で異なること、proteomics が可能であることから glycomics も人長寿研究で応用できる事が示唆された。今後新たな研究の展開が期待される。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表
なし
2. 学会発表
なし

H. 知的財産権の出願・登録状況

なし

加齢に伴う血中アディポネクチン濃度の変化ならびに
アディポネクチン遺伝子多型との相関

分担研究者 三木哲郎
愛媛大学大学院医学系研究科生命多様性医学講座加齢制御内科学・教授

研究要旨

百寿との関連が示唆されているアディポネクチン、およびその遺伝子多型の加齢性変化について、一般地域住民を対象に検討した。その結果、血中アディポネクチン濃度は、男女とも加齢にともなって有意に増加することが示された。一方、解析した-11377 C/G 多型、および 276 G/T 多型とアディポネクチン濃度とは、年齢階層を問わず、顕著な相関は認められなかった。高齢者においても、アディポネクチン高値はメタボリックシンドロームに対して予防的に作用することが示された。アディポネクチンが健康長寿の実現に対し、重要な役割を果たしている可能性が示唆された。

A. 研究目的

社会の高齢化に伴い、健康長寿の実現が個人的にも社会的にも切望されている。百寿あるいは超百寿を達成する背景には、医学・社会科学的要因とともに遺伝的要因も強く関与している。主任研究者らの従前の報告では、百寿の達成にはアディポネクチンが関与していることが示唆されている。そこで本研究では、愛媛県下における愛媛県下の一般地域住民を対象とし、加齢による血中アディポネクチン濃度の変化、ならびにアディポネクチン遺伝子多型との相関を検討することで、長寿との関連を探った。

B. 研究方法

対象は、愛媛県N町在住の一般地域住民とした。このうち、本研究の趣旨を十分に理解した上で、書面にて同意の得られた 1724 例を解析対象とした。一般臨床情報は、住民検診時の測定値を利用した。血中の高分子量アディポネクチンは、住民検診時に得た末梢血を用いて測定した。対象者のゲノム DNA は末梢血から定法に則って抽出した。

アディポネクチン遺伝子は、-11377 C/G 多型 (rs266729)、および 276 G/T 多型 (rs1501299) について分析した。分析は、TaqMan プローブ法により行った。用いたプライマー/プローブは、アプライドバイオシステムズ社より既成のものを購入して使用した (rs266729; C_2412786_10、rs1501299; C_7497299_10)。

(倫理面への配慮)

愛媛大学医学部ヒトゲノム・遺伝子解析研究倫理委員会において研究課題；生活習慣病、動脈硬

化性疾患、ならびに抗加齢に関する遺伝疫学研究として承認を得ている。対象者には、事前に十分な主旨説明の上、書面にて同意書を得た。個人情報、所定の手続きに則って匿名化し管理している。これらはヒトゲノム・遺伝子解析研究に関する倫理指針を遵守したものであり、本研究における倫理面への配慮は十分であると判断した。

C. 研究結果

対象者の臨床背景を表 1 に示した。年齢階層別のアディポネクチン濃度、および遺伝子多型頻度を検討するために、年齢を基準として対象者を 5 分割した。各年齢階層における平均アディポネクチン濃度を図 1 に示す。その結果、男女とも加齢に伴って血中アディポネクチン濃度が有意に上昇していた。年齢階層別の遺伝子多型頻度を表 2、3 に示した。いずれにおいても年齢と遺伝子多型との間に有意差は認められなかった。同様に年齢階級ごとに、遺伝子多型と血中アディポネクチン濃度との相関を表 4、5 に示した。いずれの多型においても、血中アディポネクチン濃度とは有意な相関を示さなかった。

表 1 対象者の臨床背景

年齢 (歳)	61±13
性別 (男/女)	734/990
Body mass index (kg/m ²)	23±3
収縮期血圧 (mmHg)	139±22
拡張期血圧 (mmHg)	82±12
総コレステロール (mg/dl)	204±34
HDL コレステロール (mg/dl)	62±16
中性脂肪 (mg/dl)	114±79
空腹時血糖 (mg/dl)	98±21

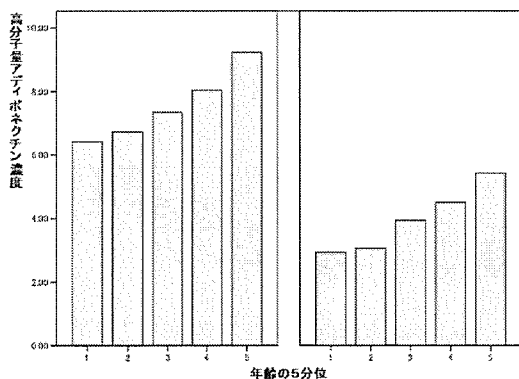


図1 年齢階級別アディポネクチン濃度 (左;女性、右;男性 いずれも p<0.001)

表2 -11377 C/G 多型 (p=0.248)

階級	GG	CG	CC	計
1	20	121	204	345
2	15	112	220	347
3	20	98	223	341
4	13	94	237	344
5	14	101	232	347

表3 276 G/T 多型 (p=0.074)

階級	TT	GT	GG	計
1	30	132	173	335
2	44	117	177	338
3	30	141	165	336
4	33	108	199	340
5	28	121	189	338

表4 -11377 C/G 多型

階級	GG	CG	CC	p
1	4.0±2.4	4.5±3.1	4.6±3.1	0.686
2	4.8±3.4	5.3±4.8	5.3±3.6	0.897
3	5.3±3.0	6.0±4.0	6.2±4.2	0.608
4	5.5±2.7	6.4±4.4	6.8±4.4	0.467
5	6.7±4.0	7.6±4.8	7.7±4.4	0.713

表5 276 G/T 多型

階級	TT	GT	GG	p
1	4.0±3.2	4.9±3.2	4.5±2.9	0.304
2	4.5±2.5	6.3±4.5	4.9±3.9	0.005
3	6.8±4.8	6.1±4.1	6.1±4.0	0.687
4	7.8±4.9	6.6±4.2	6.4±4.3	0.208
5	8.2±4.3	7.2±4.3	7.9±4.7	0.324

しかしながら、アディポネクチンは、高齢者においてもメタボリックシンドローム (MetS) に予防的に作用することが示された (図2)。これは、年齢階級ごとの検討でも同様であった。

D. 考察

血中アディポネクチン濃度は、男女とも加齢とともに有意に増加していた。高齢者において

も、アディポネクチン高値はメタボリックシンドロームに対して予防的に作用していたことから、アディポネクチンが健康長寿の実現に対して重要な役割を果たしている可能性が示唆された。一方、解析した-11377 C/G 多型、および 276 G/T 多型とアディポネクチン濃度とには、年齢階層を問わず、顕著な相関は認められなかった。今回は粗解析のみ実施したため、今後、当該遺伝子多型の影響力を検証するには、他の交絡因子を調整するなどした上で、更なる検討が必要といえる。

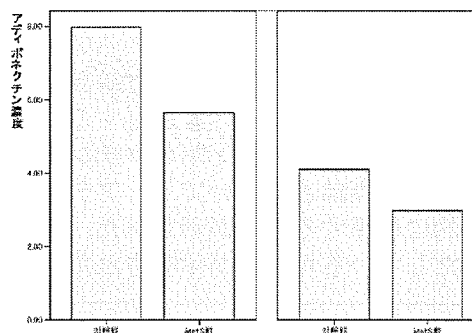


図2 MetS とアディポネクチン濃度 (女/男) (左;女性、右;男性 いずれも p<0.001)

E. 結論

一般地域住民において、アディポネクチンは健康長寿の実現に対し、重要な役割を果たす可能性が示唆された。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

Yamamoto M, Jin JJ, Wu Z, Abe M, Tabara Y, Nagai T, Yamasaki E, Igase M, Kohara K, Miki T, Nakura J. Interaction between serotonin 2A receptor and endothelin-1 variants in association with hypertension in Japanese. *Hypertens Res.* 2006;29:227-32.

2. 学会発表

第48回日本老年医学会学術集会・総会

「地域在住高齢者におけるアディポネクチンと代謝性疾患」 田原康玄・川本龍一・名倉潤・三木哲郎・小原克彦

H. 知的財産権の出願・登録状況

なし

健康長寿に関与する要因の研究—本年度のまとめと今後の研究展望—

主任研究者 広瀬信義 慶應義塾大学医学部内科（老年内科） 講師
分担研究者 三木哲郎 愛媛大学医学部老年医学講座 教授
分担研究者 小島俊男（独）理化学研究所比較システム解析研究チーム チームリーダー
分担研究者 稲垣宏樹 東京都老人総合研究所自立促進と介護予防研究チーム 研究員

研究要旨

はじめに長寿モデルの検討をおこない、百寿者ではなく超百寿者が長寿モデルの可能性を示した。ついで長寿モデルとしての超百寿者を対象に、50万 SNP を用いた whole genome scan を行った。その意義、予想される結果について考察した。Proteomics、glycomics が長寿科学でも応用できる可能性について示した。これらの技術を利用することにより網羅的なタンパク質、糖タンパク質の探索が可能となると事について考察を加えた。先行している癌、動脈硬化研究で常用される方法が可能であることより長寿研究が急速に進歩することを期待したい。

A. はじめに

少子高齢化に伴い健康長寿は社会的にも個人的にも強く望まれている。従来健康長寿のモデルと考えられる百寿者を対象に様々な研究が行われてきた。百寿者研究は対象が珍しいだけで老化研究の本筋からはずれているという見方がある。まず何故百寿者なのかという質問を受けることが多いが我々は明快な答えを出せなかった。さらに百寿者の急増は著しく本当に健康長寿のモデルかどうかについて研究者自身が疑問を抱く状況になっている。このような中で健康長寿の秘訣を知るために色々な研究方向が模索されている。この章では、今年の結果を基に人長寿科学の新しい研究方向について述べる。

B. 長寿モデルの検討—百寿者と超百寿者—

前節で述べたように百寿者の増加は急増している。h17年度の研究から自立した百寿者は約20%であった。残りの80%は寝たきりもしくは自立不可であり長寿ではあるが健康ではない人が圧倒的に多いことが分かった。さらに百寿者数は指数関数的に増加しており2006年には2万8千人を越えた。単純に人口比をみると4200名に一人の割合である。50年前では80万人に一人であったことを考えると百寿者が身近になっていることが分かる。ではどの年齢が長寿モデルとして最適である

うか？

2002年には105歳以上者は全国で849名生存していた。百歳以上者の死亡率を検討すると105歳から低下傾向を示した。また超百寿者になる確率は3200名に一人（100歳になる確率は380名に一人）であった。この人口動態的なデータから105歳が人の長寿モデルである可能性が示唆された。また100歳は正常老化の究極像であることも示唆された。これが何故百寿者を研究するかの回答になると考える。しかしいくら長寿モデルでも統計解析に十分な対象数が収集できるだろうか？東京百寿者調査では調査協力率は約15%であった。850名を母集団とすると130名の協力が予想された。実際に全国調査を行うと130名の方が調査に参加された。これに引き続き現在180名の方の調査を終了した。多くはないが統計解析を行うに足る数の対象者が収集できたことになる。以上、我々は死亡率の減少、対象数の確保の2点から、人長寿モデルは100歳ではなく105歳（超百寿者、semisupercentenarian:SSC）であると考えている。

C. 長寿モデルを対象に何を調べるのか

C-1, 長寿遺伝子

SSCを対象にして何を調べれば健康長寿の秘訣

が解明できるのか？長寿遺伝子の同定、抗老化物質、老化指標の同定が考えられる。長寿遺伝子があるかないかについては議論がある。生活習慣病や癌などにかからなければ100歳まで元気で生存可能であるという考えがある。この考えにたつと危険因子がない人が長寿であると考えため特別に長寿遺伝子を想定する必要がないということになる。医療関係者はこの立場をとる人が多い印象がある。もう一つの立場はただ単に危険因子がないだけでは100歳まで到達できず、別の防御因子または不死因子がなければ到達できないと言うものである。長寿科学者はこの立場に立つ人が多い。我々はこの第二の立場をとるものである。議論をしても結論が出るわけではない。実際にあるかないかを確認しなければ結論は出せない。Human genome project がほぼ終了に近づいている。この project の功績は genome 構造の解明と併せて配列決定などの技術を急速に押し進めたことにあると考えられる。このため50万 SNP を用いた whole genome scan も比較的価値がかつ短時間で出来るようになった。このような技術が利用できるようになったため長寿モデルを対象に解析を行えば長寿遺伝子があるのか無いのかについては結論が出せる様になってきた。すでに180名の超百寿者を収集し whole genome scan を終了しており、対照群と比較することによって相関解析による長寿遺伝子が2年以内に同定できると予想している。

C-2, DNA の構造変異の検討

50万 SNP を用いた whole genome scan を行うことにより DNA の構造変異を明らかにすることが可能となった。小島らはこの検討を報告した。構造変異が長寿、健康長寿にどのような影響を持つかを調べることは genome 研究の新しいかつ興味深い分野となろう。

C-3, 老化修飾因子、老化指標—proteomics, glycomics—

タンパク質の研究は長寿科学において重要な研究分野である。候補タンパクを選んで研究することが行われることが多い。proteomics が癌や動脈硬化の研究で頻用されている。この方法は血液タンパクなどを網羅的に分析するもので思いがけないタンパク質が抽出される可能性がある(遺伝子の candidate gene approach と whole genome scan に対応すると思われる)。Proteomics が人長寿科学で利用可能かはあまり検討されていなかった。我々は SSC 血液を元に proteomics が可能であることを示した。2D 電気泳動で分離したタンパクが SSC と若年群で異なるものがあることが判明した。今回は分

離したタンパクをタンパク染色したが、糖鎖染色を行えば糖タンパクの同定が可能である事も考えられる。前年度の報告書では ABO 血液型が百寿者と若年対照群で異なる事を示した。ABO 血液型は糖鎖により決定されているため糖鎖が SSC と若年で異なる事が示唆され、glycomics の応用が期待できる。Proteomics, glycomics など癌、動脈硬化の研究で頻用される技術が利用できることは、候補分子の研究から網羅的な分子研究が長寿科学でも可能であることを示し、長寿科学の新しい研究分野となろう。

Proteomics を用いて既にいくつかのタンパク質が同定されており、これらのタンパク質の経年変化、作用を明らかにすることにより抗老化物質、老化指標の同定に役立つと考えられる。

C-4, SSC の機能、医学所見一人の究極の老化像を知る—

SSC は長寿モデルである。そのためにより老化が進行していることが考えられる。医学所見では100歳と比較して検査所見が悪化していることが示されている。そこで SSC を調べることにより人の究極の老化像を知ることが出来るのではないかと予想している。

D. 百寿者を調べると何が分かるのか？

正常老化の究極である百寿者を調べると何が分かるのか？

D-1, 100歳以上の余命決定因子の検討

我々は東京百寿者調査に参加した百寿者300名の予後調査を行った。このシステムを使うことにより100歳以降の余命を決定する因子は何かを検討することが可能となった。アルブミン、ADL、性差などが予後を決めることが判明した。さらにアデイポカインの影響、ホルモンの影響など血中物質以外にも遺伝子多型性の影響も検討可能となった。また性格、幸福感が余命にどのような影響を示すかが明らかになりつつある。

100歳で得られた結果を超高齢者調査に応用することにより調査ターゲットの設定が容易になろう。また余命に影響する因子が判明することにより介入のターゲットを設定するのも容易になると考えられる。

D-2, 機能に影響する因子の解析

百寿者の機能分類の結果から高機能の百寿者は20%であった。SSC は retrospective にみると100歳の時点ではほとんどが高機能であったことが示唆された。百寿者250名と SSC 180名を解析対象とすることにより超高

齢期での機能に関連する因子が同定できるのではないかと予想している。すなわち百寿者 250名の20%+SSC180名(計230名)が高機能の百寿者、百寿者250名の80%(200名)が低機能である。50万SNPを用いた wholegenomescan を行い2群(高機能 vs 低機能)を比較して機能関連遺伝子多型性の同定が可能と予想している。

E. まとめ

既に SSC180名、百寿者250名の50万SNPを用いた wholegenomescan が終了した。2年以内に若年対照群と比較して相関解析による長寿関連遺伝子を明らかにする。Proteomics を用いた網羅的なタンパク質探索を行い抗老化物質、老化指標の同定を行う。Glycomics の可能性について検討を行う。糖鎖と長寿の関連について今後検討を行う。先行する癌、動脈硬化研究で常用されている様々な技術が応用することにより長寿科学の急速な進歩が期待される。

F. 謝辞

本調査にこころよく協力していただいた百寿者、超百寿者、その家族、介護提供者に深く感謝申し上げます。これらの方の惜しみない協力がなければこの研究は実行できませんでした。この方々の暖かいご協力を受けて、いただいたデータを今後の高齢者医療・福祉、抗老化医療の確立のために努力しなければならないと深く考えております。

III. 研究成果の刊行に関する一覧表

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Alexe G, Hirose N (9 番/11 人)	Enrichment of longevity phenotype in mtDNA haplogroups D4b2b, D4a, and D5 in the Japanese population	Hum Genet		Epub ahead	2007
Kojima T, Mukai W, Fuma D, Ueda Y, Okada M, Sakaki Y, Kaneko S.	Determination of genomic breakpoints in an epileptic patient using genotyping array	Biochem Biophys Res Commun	341	792-796	2006
Du Y, Hirose N, (2 番/12 人), Kojima T (5 番/12 人), Inagaki H (7 番/12 人), Gondo Y (8 番/12 人)	Analysis of growth hormone receptor polymorphism in Japanese semisupercentenarians	Geriatr Gerontol Int	6	82-86	2006
Yamamoto M, Miki T (10 番/11 人)	Interaction between Serotonin 2A Receptor and Endothelin-1 variants in Association with Hypertension in Japanese	Hypertens Res.	29	227-232	2006
Arai Y, Hirose N	Aging and high-density lipoprotein metabolism	Aging Health	2	611-621	2006
新井康通、広瀬信義	寿命決定因子	内分泌・糖尿病科	23	422-430	2006
高山美智代、広瀬信義	百寿者のライフスタイル	日本医師会雑誌	135	1258	2006
広瀬信義	百寿者に学ぶ	最新医学	61	118-131	2006
清水健一郎、飯沼利光、広瀬信義	高齢期	スタンダード栄養・食物. 応用栄養学	10	155-168	2005

Enrichment of longevity phenotype in mtDNA haplogroups D4b2b, D4a, and D5 in the Japanese population

Gabriela Alexe · Noriyuki Fuku · Erhan Bilal · Hitomi Ueno ·
Yutaka Nishigaki · Yasunori Fujita · Masafumi Ito · Yasumichi Arai ·
Nobuyoshi Hirose · Gyan Bhanot · Masashi Tanaka

Received: 27 October 2006 / Accepted: 9 January 2007
© Springer-Verlag 2007

Abstract We report new results from the re-analysis of 672 complete mitochondrial (mtDNA) genomes of unrelated Japanese individuals stratified into seven equal sized groups by the phenotypes: diabetic patients, diabetic patients with severe angiopathy, healthy non-obese young males, obese young males, patients with Alzheimer's disease, patients with Parkinson's disease and centenarians. Each phenotype had 96 samples over 27 known haplogroups: A, B4a, B4b, B4c, B*, B5, D*, F1, F2, M*, M7a, M7b, M8, M9, D4a, D4b1, D4b2, D4d, D4e, D4g, D4h, D5, G, Z, M*, N9a, and N9b. A *t*-test comparing the fraction of samples in a haplogroup to healthy young males showed a significant enrichment of haplogroups D4a, D5, and D4b2 in

centenarians. The D4b2 enrichment was limited to a subgroup of 40 of 61 samples which had the synonymous mutation 9296C > T. We identified this cluster as a distinct haplogroup and labeled it as D4b2b. Using an exhaustive procedure, we constructed the complete list of "mutation patterns" for centenarians and showed that the most significant patterns were in D4a, D5, and D4b2b. We argue that if a selection for longevity appeared only once, it was probably an autosomal event which could be dated to after the appearance of the D mega-group but before the coalescent time of D4a, D5, and D4b2b. Using a simple procedure, we estimated that this event occurred 24.4 ± 0.9 kYBP.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-007-0330-6) contains supplementary material, which is available to authorized users.

Gabriela Alexe and Noriyuki Fuku are joint first authors.

G. Alexe
The Broad Institute of MIT and Harvard, Cambridge,
MA 02142, USA

G. Alexe · G. Bhanot (✉)
The Simons Center for Systems Biology,
Institute for Advanced Study, Princeton, NJ 08540, USA
e-mail: gyanbhanot@gmail.com

N. Fuku · H. Ueno · Y. Nishigaki · M. Tanaka (✉)
Department of Genomics for Longevity and Health,
Tokyo Metropolitan Institute of Gerontology,
35-2 Sakae-cho, Itabashi-ku, Tokyo 173-0015, Japan
e-mail: mtanaka@tmig.or.jp

E. Bilal · G. Bhanot
BioMaPS Institute, Rutgers University, Piscataway,
NJ 08854, USA

Introduction

Life expectancy of humans continues to rise (Abrams 2004; de Grey 2002; Lewis 1999). While much of this

Y. Fujita · M. Ito
Department of Longevity and Aging Research,
Gifu International Institute of Biotechnology,
1-1 Naka-Fudogaoka, Kakamigahara,
Gifu 504-0838, Japan

Y. Arai · N. Hirose
Department of Geriatric Medicine,
Keio University School of Medicine,
35 Shinano-machi, Shinjuku-ku, Tokyo 160-8582, Japan

G. Bhanot
Department of Biomedical Engineering,
Rutgers University, Piscataway, NJ 08854, USA

G. Bhanot
Cancer Institute of New Jersey, 195 Little Albany Street,
New Brunswick, NJ 08903, USA

increase can be attributed to improved diet, sanitation and medical care, there may also be genetic reasons for longevity (Hawkes 2004; Hawkes et al. 1998). If so, they may arise from mutations, which might be identifiable by a suitable stratification of the population. Whereas there are several possible models for the correlation between mtDNA variants and longevity (e.g., see Elson et al. 2006) we will focus on one model based on coordination of mtDNA and nuclear mutations to explain the associations we observe in the data.

Mitochondria are non-recombinant and maternally inherited and humans readily stratify into mtDNA haplogroups (Brandon et al. 2005; Ingman and Gyllenstein 2006). The appearance of exclusively mitochondrial selective mutations would result in a distinct pattern of mtDNA mutations in a single haplogroup, which would persist over time without spread to other haplogroups. However, if the selective mutations required both mtDNA and nuclear chromosomes, a monophyletic signal would not exist (Elson et al. 2006). Mitochondria are subject to “hitch-hiking” (Bazin et al. 2006), a process by which a haplogroup is coherently amplified by selective mutations on other chromosomes. Hence, the signature of mtDNA and/or nuclear involvement in selective sweeps would be that a number of haplogroups appear enriched with enrichment spreading to other haplogroups over time. The time over which the enrichment was visible would depend on the degree of mixing between haplogroups. In a sufficiently isolated population, it might be visible in a number of haplogroups for some time. In this paper, we use an exhaustive search to correlate such patterns of mutation in haplogroups with several phenotypes including longevity.

An analysis of the mtDNA diversity in Japan using complete mtDNA sequences was reported recently (Tanaka et al. 2004). It was also shown that the D haplogroup in Japan is selected for longevity (Tanaka et al. 1998). We re-analyze the data of Tanaka et al. (2004) using novel techniques and identify a significant enhancement of the centenarian (longevity) phenotype compared to healthy normals in three haplogroups: D4a, D5, and D4b2b. The haplogroup D4b2b is a unique subhaplogroup identified in Kong et al. (2006) and Tanaka et al. (2004) and in the present study and is characterized by a synonymous mtSNP 9296C > T. Combining our results with previous studies, we conclude that the hierarchy of haplogroups: (D, (D5, D4, (D4a, D4b (D4b2, D4b2b)))) is enriched in centenarians. However, we find no single pattern of mutations that is unique to these haplogroups. This suggests that if the longevity phenotype was selected by a unique event, it cannot be exclusively from mtDNA mutations.

Since longevity is likely to have complex origins, if it has a genetic basis, it probably requires a coordination between mtDNA and nuclear mutations. If it appeared only once in the Japanese population, the phylogenetic hierarchy of enrichment in D and the observed correlation with haplogroups D4b2b, D4a, and D5 can be interpreted as relics of this ancestral event diffusing into mitochondrial haplogroups over time. By computing the coalescent time of these haplogroups, we estimate that selective mutations conferring longevity appeared in Japan at least (24.4 ± 0.9) kYBP. The other alternative is that this event has appeared multiple times for different reasons in the various haplogroups.

Our identification of haplogroups is not based on a presumed phylogenetic tree. Instead, we use unsupervised consensus ensemble clustering and agreement across multiple clustering techniques (Monti et al. 2003; Strehl et al. 2003) to determine strong groups of samples, which are similar within a group and dissimilar across groups. We find that, with few exceptions, the samples in the data fall into the haplogroups defined in the literature Tanaka et al. (2004). Robust patterns of mutations that distinguish the haplogroups are directly identified by a clustering algorithm (Alexe et al. 2004; Alexe and Hammer 2006) followed by a data averaging procedure that eliminates spurious mtSNPs and statistically meaningless haplogroup labels.

Materials and methods

A detailed description of the methods used in the collection, extraction, and sequencing of the mtDNA samples used here is given in Tanaka et al. (2004). The data were available as complete sequences of 96 samples in each of seven phenotypes: diabetic patients with severe angiopathy, diabetic patients without angiopathy, healthy non-obese young males, obese young males, patients with Alzheimer’s disease, patients with Parkinson’s disease and centenarians.

The subject groups are described in detail in the Technical Annex of the mtSNP database: http://www.giib.or.jp/mtsnp/annex_e.html. Briefly, we summarize this information below:

The Centenarian group consisted of 96 centenarians over 100 years of age who were registered by the Department of Gerontology, Keio University School of Medicine, in Tokyo. This comprised the Tokyo Centenarians (TC group) with 85 patients (60 females and 25 males) and the Gifu Centenarians (the GC group from the Gifu International Institute of Bio-

technology) with 11 samples (six females and five males). The Parkinson's disease group consisted of 96 patients with Parkinson's disease (mean age of 62.1 ± 8.9 years; range, 39–81 years) who were diagnosed at the Department of Neurology, Juntendo University School of Medicine, in Tokyo (53 were females and 43 males). The Alzheimer's disease group was comprised of 96 patients with Alzheimer's disease (mean age of 76.5 ± 9.7 years; range, 47–93) who were diagnosed at the Department of Neurology, Nihon Medical School, in Chiba ($n = 96$, 76 females, and 20 males). The Obesity group consisted of 96 young obese male subjects (mean age of 21 ± 2 years; range, 18–25 years) with a body mass index (BMI) of 30.1 ± 2.5 , who were registered at the Research Center of Health, Physical Fitness and Sports, Nagoya University, in Nagoya. The Healthy non-obese group consisted of 96 young healthy non-obese male subjects (mean age of 20 ± 3 years; range, 18–25) with a BMI of 20.2 ± 2.3 , who were registered at the same location as the obese group. The Diabetes group came from Nagoya and consisted of 96 patients (42 females and 54 males) with type-2 diabetes mellitus (mean age of 58 ± 5 years; range, 43–65), who were registered at the same location as the obese group. The diabetes patients with angiopathy came from a cohort of 96 patients with severe vascular disease. This group had 48 females and 48 males and a mean age of 65 ± 10 years ranging from 43 to 92. They were registered at the Department of Internal Medicine, Juntendo University School of Medicine, in Tokyo.

The complete mtDNA sequences from these patients were pair-wise aligned with the revised Cambridge Reference Sequence rCRS (Brandon et al. 2005) by applying the linear global alignment algorithm of Myers and Miller (1988) implemented in Emboss (<http://www.emboss.sourceforge.net>). We found 1,351 polymorphic loci out of which we retained 240 loci with a frequency greater than 1%.

The haplogroups used in our analysis are those defined in Tanaka et al. (2004). We used unsupervised consensus ensemble clustering techniques and bootstrap data perturbation techniques (Monti et al. 2003; Strehl et al. 2003) to divide the data into an optimum number of robust clusters and inferred their haplogroup labels using the specific haplogroup identifying mutations in Brandon et al. (2005) and Tanaka et al. (2004). The unsupervised consensus ensemble clustering method proceeds by first finding the optimum number of clusters (haplogroups) in the data using principal component analysis (see, e.g., Jolliffe 1986) and statistical measures such as the silhouette score (Rousseeuw 1987). Next, we create several datasets

with the 96 samples in each phenotype by a random sampling with replacement from each phenotype. A number of clustering techniques, representing hierarchical, agglomerative and probabilistic clustering methods, were then applied to each dataset. Each clustering technique divided the samples into clusters in slightly different ways. We averaged the sample assignments over the datasets to produce an agreement matrix for each clustering method used. The entries of the agreement matrix represent the fraction of times a pair of samples appears in the same haplogroup cluster. Averaging these agreement matrices over the clustering techniques and applying an appropriate threshold for the acceptable fraction of agreement (0.9 in our case) divided the samples into an optimum number of robust haplogroups. The markers defining each haplogroups were easily identified from the clusters by a frequency analysis of mtSNPs for all polymorphic loci. Our criterion for a good marker is one that is present in 90% or more samples in at least one haplogroups and in 10% or less in at least one other. Independent of the haplogroups clustering, we applied the method of Alexe et al. (2004) and Alexe and Hammer (2006) to exhaustively identify all subsets of samples that share a pattern of mutations. The algorithm used is based on an iterative consensus approach, runs in total polynomial time and is linear in the number of patterns found.

Results

The assignment of samples to clusters and clusters to haplogroups mostly agrees with the earlier analyses (Tanaka et al. 2004). However, our clustering revealed additional structure in the data not previously noticed. We found a new marker 9296C > T in 40 of 61 samples of D4b2, which defines a new haplogroup which we call D4b2b (Kong et al. 2006; Tanaka et al. 2004).

We stratified the samples by both phenotype and haplogroup. For each haplogroup, we compared the number of samples in each phenotype to the number of samples in healthy normals. As a preliminary analysis, we computed a *t*-statistic for each haplogroup using the sample frequencies in the non-centenarian phenotypes to compute the standard deviation. Table 1 shows our results. We see that the haplogroups enriched in centenarians are D4a, D5, and D4b2b. We also calculated an exact Fisher test odds ratio, which supported the *t*-test results, although with somewhat larger *P*-values. Other than the enrichment in centenarians, the only other significant signal was that haplogroup M7a is protected against the diabetes with angiopathy phenotype at $P \sim 0.005$. Table 1 also shows that there is an

Table 1 Number of samples in each phenotype for the haplogroups (in bold with *) where the centenarian (longevity) phenotype is enriched over healthy normals

Haplogrp	Diabetes with angio.	Diabetes without angio.	Healthy normal	Obese normal	Alzheimer disease patients	Parkinson disease patients	Centenarian	t_HN	P_val t-test	Odds ratio	P-value Fisher test
*D4b2b	3	7	2	4	9	4	11	3.41	0.007	6.03	0.009
D4b2a	3	1	6	1	2	2	6	0.00	0.500	1.00	0.617
D4b2	6	8	8	5	11	6	17	4.17	0.003	2.36	0.042
*D5	2	5	1	4	4	5	7	3.65	0.005	7.41	0.032
*D4a	7	6	4	6	5	5	8	3.81	0.004	2.08	0.186
D4b	9	9	12	6	14	10	19	2.54	0.022	1.72	0.120
D4	25	25	28	27	28	24	36	4.64	0.002	1.45	0.142
D	33	39	34	35	36	38	47	5.61	0.001	1.74	0.040
M	54	66	63	57	60	71	70	1.13	0.150	1.41	0.174

Results are sorted by phylogeny hierarchy in D and M, with the significantly enriched haplogroups in bold. The t_{HN} statistic is defined as t_{HN} = (number of centenarians – number of healthy normals)/sdev, where sdev is the standard deviation of the haplogroup samples over the six non-centenarians phenotypes. The P-value of t_{HN} was computed from a *t* distribution with 6 degrees of freedom. The Fisher P-value was computed by applying a Fisher exact test to check whether the frequency of centenarians versus healthy normals in each haplogroup was greater than the frequency of centenarians versus healthy normals outside the haplogroup. The estimate of the odds ratio was computed by using the conditional Maximum Likelihood Estimate (MLE)

enrichment in the parent group D4b2, but it is restricted to the subgroup D4b2b which is distinguished from the D4b2 haplogroup by the synonymous mutation 9296C > T.

Of the 61 samples in D4b2 in the data, 40 have mtSNP 9296C > T and 21 do not. Another marker enriched in centenarians is the synonymous mutation 14605A > G seen in a subgroup containing 29 of 40 samples in haplogroup D4b2b. This may define a further split of the D4b2b haplogroup (Tanaka et al. 2004), identified as D4b2b1 in Kong et al. (2006). The centenarian enrichment is not found in the subgroup D4b2a of D4b2 which does not have mtSNP 9296C > T. The enrichment is present in D4b2 but not visible at the level of D4b. It persists at the level of the D mega-group but disappears at the level of the M clade. We also do not find any single mtSNP significantly enriched in centenarians compared to the other phenotypes other than the defining mtSNPs for the D mega-group.

For all D samples taken as a single group, we observe the known (Tanaka et al. 1998) enrichment in centenarians correlated with two markers, a non-synonymous transversion 5178C > A (ND2: Leu237Met) and a synonymous transition 4883C > T. The subgroup D4b2a of D4b2 (which does not have the 9296C > T mutation), is not enriched. We find that the enrichment is not visible in D4b, but shows up in the parent and grandparent mega-groups D4 and D. However, the signal is again lost at the level of the M clade.

In summary, the hierarchy of haplogroups: (D, (D5, D4 (D4a, D4b (D4b2 (D4b2b)))))) is enriched in centenarians. However, there is no single pattern of

mutations enriched exclusively in the three subhaplogroups (D4a, D5, D4b2b), which is not seen in other D haplogroups. There are several mtSNPs in the control region: (16187C > T, 16189T > C, 16278C > T, 16311T > C, and 16362T > C) which are enriched in various combinations in these three haplogroups. However, these mtSNPs are also common in other haplogroups, which are not enriched in centenarians as well as outside the D clade (they also exist in N and L clade haplogroups). Thus we find no pattern of mtSNPs that is unique to the longevity enriched haplogroups.

A more detailed analysis was done using the methods of Alexe et al. (2004) and Alexe and Hammer (2006), which identified all maximal patterns of mutations carried by a minimum of five samples. For each pattern, we calculated the number of samples in each phenotype that displayed the pattern. The *t*-test scores for all observed patterns for centenarians compared to healthy normals are shown in Fig. 1. The 2,805 patterns with *t*-test scores >3.0 were selected for further analysis.

Figure 2 shows the haplogroups in which the significant patterns of Fig. 1 are found. Note again the appearance of the haplogroups D4a, D5, and D4b2b. Other haplogroups also show up in this figure but with lower frequency. Table 2 shows the mtSNPs in the enriched patterns. The mtSNPs are sorted by the D haplogroup hierarchy (Tanaka et al. 2004). The known enrichment in D shows up as the known mtSNPs 5178C > A and 4883C > T. The D4 mtSNP 8414C > T, the D4b and D4b2 mtSNPs 8020G > A, and (1382A > C, 8964C > T), respectively, are also identified as are several mutations characteristic of D4a, D5,

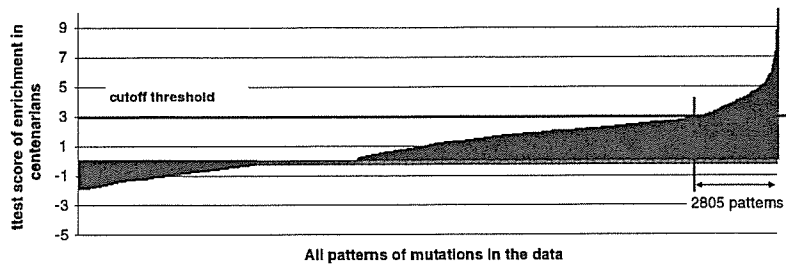
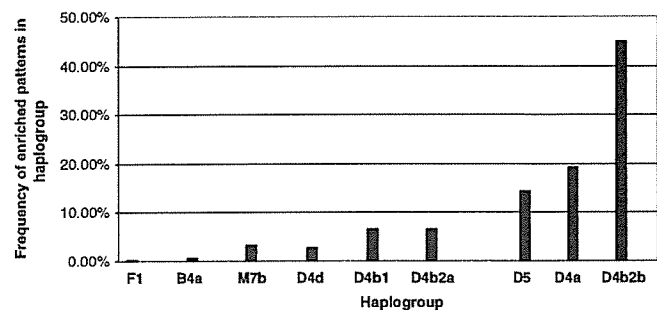


Fig. 1 The t -test scores for enrichment in centenarians for the ~20,000 patterns identified in the data. We define a pattern as a set of mutations exhibited by at least five samples. The patterns

were identified by applying an exhaustive search algorithm and are shown sorted with respect to the t -test scores. About 2,805 patterns with t -test score >3 were selected for further analysis

Fig. 2 Of enriched patterns in haplogroups. About 2,805 enriched patterns with a t -test score >3.0 were selected (see Fig. 1) and their frequency in haplogroups is shown. A pattern was considered to be associated to a haplogroup if it is present in at least 50% of the samples in the haplogroup



and D4b2b. In particular, the mtSNPs 9296C $>$ T and 14605A $>$ G for D4b2b are identified. There are also a few mutations characteristic of other haplogroups (D4d1, M7b, B4a, B5). Finally there are three mtSNPs (16362T $>$ C, 3010G $>$ A, and 10410T $>$ C), which are seen in many haplogroups outside the Japanese haplogroups in this study. These relate to older events in the evolution phylogeny (Brandon et al. 2005) and are not relevant in our analysis.

In Table 3, we show the best patterns enriched in centenarians. All the mtSNPs in Table 2 for D4b2b, D5, and D4a and in their ancestral groups appear in these patterns. Some ancient mtSNPs such as 150C $>$ T and 152T $>$ C reported previously (Niemi et al. 2005) are also seen in the patterns. There are patterns involving a number of mtSNPs in the control region: (16187C $>$ T, 16189T $>$ C, 16278C $>$ T, 16311T $>$ C, and 16362T $>$ C) which appear in the three haplogroups. However, these mutations are very common in the data and also appear in haplogroups not enriched in centenarians and in haplogroups not seen in Japan (Brandon et al. 2005). Overall, the patterns and mtSNPs in Tables 2 and 3 suggest that the only markers enriched in centenarians are the haplogroup markers for D4a, D5, and D4b2b. Since these three haplogroups are not in a subtree of the D phylogeny

(Tanaka et al. 2004), the longevity selection event (if it was unique) probably involved changes in the nuclear genes as well as on mitochondria. The selection event(s) suggested by the present analysis originated in D and spread to D4a, D5, and D4b2b. Since it is seen in D but not in M, it must have arisen after the appearance of the D clade but before the coalescent time of D4a, D5, and D4b2b.

To compute the second of these times, we use the following procedure: we compare the three haplogroups in pairs. For each haplogroup pair, we compute the distribution of the number of mutations between all the sample pairs, one from each haplogroup. Restricting to the mutations seen in more than 90% of the sample pairs, we compute the mean number of mutations n_M and its standard deviation σ_M . The time to most recent common ancestor (TMRCA) is then given by:

$$T_c = [n_M \pm \sigma_M] / (2 \mu),$$

where μ is the mutation rate in units of number of mutations per sample per year. Using the value 1.7×10^{-8} mutations per site per year (Ingman et al. 2000) we get $\mu = 2.82 \times 10^{-4}$ mutations/year/sample. We find the following results for each pair of haplogroups:

Table 2 mtSNPs in enriched patterns in centenarians (t -test value ≥ 3.00)

SNP	t -test statistic	t -test P -value	Haplogroups with SNP ^b	Haplogroup specificity	rCRS	Mutant	Type	Gene	Sense
9,296 ^a	3.41	0.007	1	D4b2b	C	T	TS	CO3	Syn
14,605 ^a	3.87	0.004	1	D4b2b	A	G	TS	ND6	Syn
11,944 ^a	3.95	0.004	1	D5	T	C	TS	ND4	Syn
12,026 ^a	3.95	0.004	1	D5	A	G	TS	ND4	Nsyn
1,107 ^a	4.26	0.003	3	D5	T	C	TS	12S rRNA	
5,301 ^a	4.26	0.003	3	D5	A	G	TS	ND2	Nsyn
10,397 ^a	4.26	0.003	3	D5	A	G	TS	ND3	Syn
752 ^a	3.95	0.004	1	D5	C	T	TS	12S rRNA	
14,979 ^a	3.30	0.008	1	D4a	T	C	TS	Cytb	Nsyn
3,206 ^a	3.81	0.004	1	D4a	C	T	TS	16S rRNA	
8,473 ^a	3.30	0.008	1	D4a	T	C	TS	ATP8	Syn
5,178 ^b	6.37	0.001	15	D	C	A	TV	ND2	Nsyn
4,883 ^b	5.81	0.001	15	D	C	T	TS	ND2	Syn
8,414 ^b	4.08	0.003	13	D4	C	T	TS	ATP8	Nsyn
8,020 ^b	3.30	0.008	5	D4b	G	A	TS	CO2	Syn
1,382 ^b	3.99	0.004	2	D4b2	A	C	TV	12S rRNA	
8,964 ^b	4.17	0.003	2	D4b2	C	T	TS	ATP6	Syn
194 ^c	4.58	0.002	2	D4b2b, D4d1a	C	T	TS	MajorNCR1	
199 ^c	6.10	0.001	3	M7b, B4a, D4d1a	T	C	TS	MajorNCR1	
5,460 ^c	3.02	0.012	2	M7b	G	A	TS	ND2	Nsyn
204 ^c	4.95	0.001	1	B5	T	C	TS	MajorNCR1	
4,071 ^c	4.06	0.003	1	M7b	C	T	TS	ND1	Syn
13,651 ^c	4.47	0.002	1	B4a	A	G	TS	ND5	Nsyn
16,362	4.95	0.001	19	None	C	T	TS		
3,010	4.00	0.004	13	None	G	A	TS	16S rRNA	
10,410	3.81	0.004	8	None	T	C	TS	tRNA-Arg	

Column 2 is the t -test score and column 3 is the P -value of the t -test score. The next two columns show the number of haplogroups enriched in the mtSNP (frequency above 70%) and the haplogroup or clade for which the mtSNP is a known marker. The next two columns present the nucleotide in rCRS and its mutated mtSNP form. In the column "Type," TV transversion and TS transition. The "Sense" column shows whether the mutation is synonymous or non-synonymous

^a mtSNPs that are specific for haplogroups D4a, D4b2b, and D5

^b mtSNPs that are specific for their ancestor megahaplogroup

^c mtSNPs in other haplogroups. The remaining mtDNA SNPs are highly polymorphic and not useful for classification

$$(D4a, D4b2b) : n_M = 10.9, \sigma_M = 0.3, \\ T_c = 19.3 \pm 0.5 \text{ kYBP},$$

$$(D4a, D5) : n_M = 11.9, \sigma_M = 0.4, \\ T_c = 21.1 \pm 0.7 \text{ kYBP},$$

$$(D4b2b, D5) : n_M = 13.8, \sigma_M = 0.5, \\ T_c = 24.4 \pm 0.9 \text{ kYBP}.$$

Since the coalescent time of all three haplogroups must be greater than the maximum of the T_c -values above, the longevity event, if it was unique, must have occurred at least 24.4 ± 0.9 kYBP. The coalescent time for each pairs of haplogroups within D can also be calculated using the same technique. We find that all

these times are smaller than 24.4 kY which corresponds to the oldest split of the D clade (the split of D5 and D4).

Summary and discussion

In this paper, we show that the data suggest that Japanese individuals are more likely to live longer if they are in haplogroups D4a, D5, and D4b2b. Our analysis raises the possibility that if longevity was selected for once, possibly by a gradual process of accumulated mutations, then it may not be an exclusively mitochondrial event but would have markers on nuclear chromosomes. The simplest way to study this possibility would be an association study correlating mtSNPs with nuclear SNPs in D4a, D5, and D4b2b haplogroups that show an enrichment in centenarians.

Table 3 Best patterns (t -test score >3.0) enriched in centenarians for haplogroups D4a, D4b2b, D5

No	Haplogroup	Coverage (%) of patterns across haplogroups			Pattern description
		D4a	D4b2b	D5	
1	D4a/ D4b2b/ D5	100.00	97.50	92.86	16187 16278 16362
1	D4a	100.00	0.00	0.00	489 3010 4883 5178 8414 8473 10400 14668 14783 14979 15043 15301 16187 16189 16278 16362
2	D4a	100.00	0.00	0.00	489 3010 3206 4883 5178 8414 8473 10400 14668 14783 14979 15043 15301 16187 16189 16278 1632
3	D4a	98.00	0.00	0.00	489 3010 3206 4883 5178 8414 8473 10400 14668 14783 14979 15043 15301 16129 16187 16189 16278 16362
4	D4a	98.00	0.00	0.00	489 3010 3206 4883 5178 8414 8473 10400 14668 14783 14979 15043 15301 16187 16189 16278 16311 16362
5	D4a	98.00	0.00	0.00	489 3010 4883 5178 8414 8473 10400 14668 14783 14979 15043 15301 16129 16187 16189 16278 16362
6	D4a	98.00	0.00	0.00	489 3010 4883 5178 8414 8473 10400 14668 14783 14979 15043 15301 16187 16189 16278 16311 16362
7	D4a	95.00	0.00	0.00	489 3010 4883 5178 8414 8473 10400 14668 14783 14979 15043 15301 16129 16187 16189 16278 16311 16362
8	D4a	95.00	0.00	0.00	489 3010 3206 4883 5178 8414 8473 10400 14668 14783 14979 15043 15301 16129 16187 16189 16278 16311 16362
9	D4a	80.00	0.00	0.00	489 3010 3206 4883 5178 8414 8473 10400 10410 14668 14783 14979 15043 15301 16187 16189 16278 16362
10	D4a	80.00	0.00	0.00	489 3010 4883 5178 8414 8473 10400 10410 14668 14783 14979 15043 15301 16187 16189 16278 16362
11	D4a	78.00	0.00	0.00	489 3010 4883 5178 8414 8473 10400 10410 14668 14783 14979 15043 15301 16129 16187 16189 16278 16362
12	D4a	78.00	0.00	0.00	489 3010 4883 5178 8414 8473 10400 10410 14668 14783 14979 15043 15301 16187 16189 16278 16311 16362
13	D4a	78.00	0.00	0.00	489 3010 3206 4883 5178 8414 8473 10400 10410 14668 14783 14979 15043 15301 16129 16187 16189 16278 16362
14	D4a	66.00	0.00	0.00	310 489 3010 4883 5178 8414 8473 10400 14668 14783 14979 15043 15301 16187 16189 16278 16311 16362
15	D4a	51.00	0.00	0.00	310 489 3010 4883 5178 8414 8473 10400 10410 14668 14783 14979 15043 15301 16129 16187 16189 16278 16311 16362
1	D4b2b	0.00	100.00	0.00	152 489 1382 3010 5178 8020 8414 8964 9296 9824 10400 14668 14783 15043 15301 16278 16311 16362
2	D4b2b	0.00	98.00	0.00	152 489 1382 3010 5178 8020 8414 8964 9296 9824 10400 14668 14783 15043 15301 16189 16278 16311 16362
3	D4b2b	0.00	98.00	0.00	152 489 1382 3010 5178 8020 8414 8964 9296 9824 10400 14668 14783 15043 15301 16278 16311 16362
4	D4b2b	0.00	98.00	0.00	152 489 1382 3010 5178 8020 8414 8964 9296 9824 10400 14668 14783 15043 15301 16187 16278 16311 16362
5	D4b2b	0.00	93.00	0.00	152 194 489 1382 3010 5178 8020 8414 8964 9296 9824 10400 14668 14783 15043 15301 16278 16311 16362
6	D4b2b	0.00	73.00	50.00	152 310 16311
7	D4b2b	0.00	70.00	0.00	152 489 1382 3010 5178 8020 8414 8964 9296 9824 10400 14605 14668 14783 15043 15301 16187 16189 16278 16311 16362
8	D4b2b	0.00	70.00	0.00	152 310 489 1382 3010 5178 8020 8414 8964 9296 9824 10400 14605 14668 14783 15043 15301 16187 16278 16311 16362
9	D4b2b	0.00	65.00	0.00	152 194 489 1382 3010 5178 8020 8414 8964 9296 9824 10400 14605 14668 14783 15043 15301 16187 16278 16311 16362
10	D4b2b	0.00	63.00	0.00	152 194 310 489 1382 3010 5178 8020 8414 8964 9296 9824 10400 14668 14783 15043 15301 16187 16189 16278 16311 16362
1	D5	0.00	8.00	100.00	150 489 16187 16278
2	D5	0.00	8.00	68.00	150 489 16278
3	D5	0.00	0.00	96.00	489 4883 5178 10397 10400 14783 15043 15301 16187 16278
4	D5	0.00	0.00	96.00	489 4883 5178 10397 10400 14783 15043 15301 16187 16278 16362
5	D5	0.00	0.00	96.00	489 1107 4883 5178 5301 10397 10400 14783 15043 15301 16187 16278

Table 3 continued

No	Haplogroup	Coverage (%) of patterns across haplogroups			Pattern description
		D4a	D4b2b	D5	
6	D5	0.00	0.00	93.00	310 489 1107 4883 5178 5301 10400 14783 15043 15301 16187 16278
7	D5	0.00	0.00	93.00	489 1107 4883 5178 5301 10400 14783 15043 15301 16187 16278
8	D5	0.00	0.00	93.00	489 1107 4883 5178 5301 10397 10400 14783 15043 15301 16187 16278 16311
9	D5	0.00	0.00	89.00	310 489 4883 5178 10397 10400 14783 15043 15301 16187 16278
10	D5	0.00	0.00	64.00	152 489 10400 14783 15043 15301 16183 16187 16278 16311 16362
11	D5	0.00	0.00	61.00	150 152 489 1107 4883 5178 5301 9180 10397 10400 14783 15043 15301 16182 16183 16187 16278 16362
12	D5	0.00	0.00	61.00	152 310 489 10400 14783 15043 15301 16183 16187 16311 16362
13	D5	0.00	0.00	57.00	489 752 1107 4883 5178 5301 9180 10397 10400 11944 12026 14783 15043 15301 16182 16183 16187 16278
14	D5	0.00	0.00	57.00	150 152 310 489 1107 4883 5178 5301 9180 10397 10400 14783 15043 15301 16182 16183 16187 16278 16362
15	D5	0.00	0.00	54.00	489 752 1107 4883 5178 5301 9180 10397 10400 11944 12026 14783 15043 15301 16182 16183 16187 16278 16311

Column 2 shows the haplogroup(s) characterized by the pattern. Columns 3–5 show the % of samples in each of the three haplogroups, which satisfy the pattern. Column 6 shows the list of mutation loci in the pattern

Parkinson's and Alzheimer's are late-onset diseases. Hence, one might expect that they would also show an enrichment in D4a, D5, and D4b2b relative to healthy normals. As can be seen from Table 1 by comparing columns 6/7 with column 4, there is a small signal for both these phenotypes. However, this signal is not statistically significant. The average age in the Parkinson's and Alzheimer's disease groups was 62.1 ± 8.9 and 76.5 ± 9.7 years, respectively, compared to the average age of centenarians, which was greater than 100 years. This difference in average age was sufficient to eliminate any signal for association of longevity haplogroups with the Parkinson's and Alzheimer's disease phenotypes because of the contamination from younger samples.

To check the consistency of our time estimates for coalescence, we computed the coalescence times of the other haplogroup clusters in our data. In the D clade, the largest coalescent time was between D4a and D4b2b (24.4 ± 0.9 kYBP) and the smallest was between D4e and D4h (3.5 ± 0.3 kYBP). We find that the D4b2b haplogroup (with mtSNP 9296C > T) split off from D4b2 approximately (6.8 ± 0.6) kYBP.

It would be interesting to see if longevity is also amplified in these haplogroups outside Japan. This would identify whether the enrichment is specific to some condition of diet or lifestyle in Japan in these haplogroups or whether it has a different origin. Another interesting study would be to see whether the N-clade haplogroups, which are genetically very distinct

and have undergone a different adaptation to climate, disease and diet from the D clade after emerging from Africa (<https://www3.nationalgeographic.com/genographic/>), have such an enrichment in haplogroups and whether it correlates with the same loci.

One may ask the question: "If the phenotype is from an autosomal event, why should the signal be visible in only a few haplogroups? Should it not rapidly diffuse to all haplogroups?" The relative isolation of the Japanese population and their social stratification may be one reason the longevity phenotype is still visible in this population. In addition, there may be other effects from lifestyle and diet that also play a role.

Alternatively, the longevity phenotype may have multiple causes (appeared more than once). In this case, each subhaplogroup identifying mtDNA mutation may have distinct effects on longevity. As shown in Table 2, subhaplogroup D4a is associated with mtSNP 3206C > T in the 16S rRNA gene and synonymous mtSNP 8473T > C in the ATP6 gene. In addition, subhaplogroup D4a is characterized by the mtSNP 14979T > C causing the Ile78Thr replacement in the cytochrome *b* subunit of respiratory chain complex III (cytochrome *bc*₁ complex, or ubiquinol-cytochrome *c* oxidoreductase, EC 1.10.2.2). We can speculate that this amino acid replacement might affect the functionality of the cytochrome *b* subunit, thereby conferring resistance against reactive oxygen species (ROS) or decreasing the production of ROS from mitochondria.

In a previous study it was shown that the combination of three mtSNPs 150C > T, 489T > C, and 10398A > G is associated with longevity in both Finland and Japan (Niemi et al. 2005). Although the mtSNP 150C > T was enriched in subhaplogroup D5, the present extensive search for combinations of mtSNPs did not detect the association of these three mtSNPs with the longevity phenotype. As shown in Table 2, subhaplogroup D5 is characterized by several mtSNPs including two non-synonymous mutations, the mtSNP 5301A > G causing the Ile278Val replacement in the ND2 subunit and the mtSNP 12026A > G causing the Ile423Val replacement in the ND4 subunit of respiratory chain complex I (NADH-ubiquinone oxidoreductase, EC 1.6.5.3). Since several pathogenic mutations in both ND2 and ND4 genes have been identified in patients with Leber's hereditary optic neuropathy (Yen et al. 2006), we cannot exclude the possibility that the mtSNPs in these genes are functionally associated with longevity.

Subhaplogroup D5a is specified by two mtSNPs 752C > T and 1107T > C in the 12S rRNA gene. The contribution of these mtSNPs to longevity also remains hypothetical. Subhaplogroup D4b2b is distinguished from its sister subhaplogroup D4b2a by two mtSNPs 9296C > T and 14605A > G. Because both of these mtSNPs are synonymous mutations, it is difficult to understand why these mtSNPs are enriched in Japanese centenarians. One possibility is that these mtSNPs confer resistance against ROS to the local mtDNA sequences surrounding them. An alternative is the "hitch-hiking" effect from nuclear SNPs that predispose individuals to longevity.

There are several limitations of the present study. The first is the rather small sample size. In each phenotype, including "longevity" and "control" or "Healthy Normals," we have only 96 samples. Larger sample sizes are necessary to validate our results. Second, we identified mtDNA SNP patterns solely on the basis of their dominance in phenotypes. Functional analysis of the identified mtSNPs for each haplogroup/phenotype using the trans-mitochondrial cybrid approach (Kazuno et al. 2006) is necessary to validate or disprove our results. Third, our analysis suggests but cannot prove or disprove a "hitch-hiking" effect correlating mtDNA SNPs with nuclear SNPs to create the longevity enrichment in D5, D4a, and D4b2b. Concurrent genome-wide analysis of centenarian DNA using GeneChip technology is necessary both to verify and to elucidate this relationship.

In this paper, we describe a robust and accurate method of identifying hidden enrichment of specific

haplogroups from a limited number of samples. Although the results need to be validated on larger sample sets, the method itself is reliable and can be used in any situation where a robust correlation between patterns of mtSNPs and phenotypes need to be identified. Combined with linkage analysis on nuclear chromosomal polymorphisms, our method should be useful in identifying correlated linked blocks under selection for complex phenotypes. In such a study, the correlation between the identified nuclear loci and mtDNA haplogroups would give key clues about possible population events that gave rise to the conditions under which the selection occurred.

Acknowledgments This work was supported in part by Support Project for Database Development from the Japan Science and Technology Corporation (to M. T.), Grants-in-Aid for Scientific Research (C2-10832009, A2-15200051) and for Priority Areas from the Ministry of Education, Science, Sports and Culture of Japan (to M. T.), and by grants BMC2001-3511 and COF2002-015 (to V. M. C.).

References

- Abrams PA (2004) Evolutionary biology: mortality and lifespan. *Nature* 431:1048
- Alexe G, Alexe S, Crama Y, Foldes S, Hammer PL, Simeone B (2004) Consensus algorithms for the generation of all maximal bicliques. *Discrete Appl Math* 145:11–21
- Alexe G, Hammer PL (2006) Spanned patterns for the logical analysis of data. *Discrete Appl Math* 154:1039–1049
- Bazin E, Glemin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* 312:570–572
- Brandon MC, Lott MT, Nguyen KC, Spolim S, Navathe SB, Baldi P, Wallace DC (2005) MITOMAP: a human mitochondrial genome database—2004 update. *Nucleic Acids Res* 33:D611–D613
- de Grey ADNJ (2002) The mitochondrial free radical theory of aging. Landes Bioscience, Austin
- Elson JL, Herrstadt C, Preston G, Thal L, Morris CM, Edwardson JA, Beal MF, Turnbull DM, Howell N (2006) Does the mitochondrial genome play a role in the etiology of Alzheimer's disease? *Hum genet* 119:241–254
- Hawkes K (2004) Human longevity: the grandmother effect. *Nature* 428:128–129
- Hawkes K, O'Connell JF, Jones NG, Alvarez H, Charnov EL (1998) Grandmothering, menopause, and the evolution of human life histories. *Proc Natl Acad Sci USA* 95:1336–1339
- Ingman M, Gyllensten U (2006) mtDB: human mitochondrial genome database, a resource for population genetics and medical sciences. *Nucleic Acids Res* 34:D749–D751
- Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713
- Jolliffe IT (1986) Principal component analysis. Springer, New York
- Kazuno AA, Munakata K, Nagai T, Shimozono S, Tanaka M, Yoneda M, Kato N, Miyawaki A, Kato T (2006) Identification of mitochondrial DNA polymorphisms that alter

- mitochondrial matrix pH and intracellular calcium dynamics. *PLoS Genet* 2:e128
- Kong QP, Bandelt HJ, Sun C, Yao YG, Salas A, Achilli A, Wang CY, Zhong L, Zhu CL, Wu SF, Torroni A, Zhang YP (2006) Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet* 15:2076–2086
- Lewis K (1999) Human longevity: an evolutionary approach. *Mech Ageing Dev* 109:43–51
- Monti S, Tamayo P, Mesirov J, Golub T (2003) Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data machine learning 52:91–118
- Myers E, Miller W (1988) Optimal alignments in linear space. *CABIOS* 4:11–17
- Niemi AK, Moilanen JS, Tanaka M, Hervonen A, Hurme M, Lehtimäki T, Arai Y, Hirose N, Majamaa K (2005) A combination of three common inherited mitochondrial DNA polymorphisms promotes longevity in Finnish and Japanese subjects. *Eur J Hum Genet* 13:166–170
- Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65
- Strehl A, Ghosh J, Cardie C (2003) Cluster ensembles: a knowledge reuse framework for combining multiple partitions. *J Mach Learn Res* 3:583–617
- Tanaka M, Cabrera VM, Gonzalez AM, Larruga JM, Takeyasu T, Fuku N, Guo LJ, et al (2004) Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res* 14:1832–1850
- Tanaka M, Gong JS, Zhang J, Yoneda M, Yagi K (1998) Mitochondrial genotype associated with longevity. *Lancet* 351:185–186
- Yen MY, Wang AG, Wei YH (2006) Leber's hereditary optic neuropathy: a multifactorial disease. *Prog Retin Eye Res* 25:381–396