

デルと仮定した。この仮説の検討を中心に調査を行うことにした。残念ながら研究が採択されなかったため、東京百寿者調査参加者の予後調査、データの整理、解析を行ってきた。

## C-2, 平成14年以降の研究成果

### I, 百寿者調査の成果

#### I-1, 研究方法

平成12年より平成14年に東京都23区には推定1735名の100歳以上者が在住していた。このうち住所の判明した1194名に調査勧誘の手紙を送った。調査協力を承諾した百寿者にアンケートを送り、生活歴、嗜好品、現病歴、既往歴、内服薬、ADL、性格などを調べた。さらに訪問調査に同意した百寿者については医師（老年医学専門医）、看護師、心理学士による訪問調査を行い、1) 身体測定、2) 理学所見、3) 採血、4) 病歴、治療歴の確認、5) 認知機能（MMSEおよびCDR）、6) ADL（Barthel Index）の確認などを行った。

#### <倫理面への配慮>

本調査に当たって、本人及び家族に文書により十分な説明をおこなった。その後本人（可能な場合は）または家族より同意文書への署名を得た。この調査は慶應義塾大学医学部倫理委員会、東京都老人総合研究所、理化学研究所、愛媛大学医学部の承認を得た。

#### I-2, 研究結果

##### i) 健康長寿モデル探索のための百寿者の機能分類

長寿遺伝子探索のターゲット集団の絞り込みを行うために、東京百寿者研究の参加者を視聴覚機能、認知機能、身体機能により分類した。304名の訪問調査参加者を分類した結果、「非常に優秀」、「正常」、「虚弱」、「非常に虚弱」は、それぞれ、2%、18%、55%、25%となった。この分類は、血清アルブミン濃度、調査後1年の死亡率、といった外的な基準や、過去の習慣的飲酒や喫煙と関連していた。死亡率や機能状態との関係から本分類によって分類された、「優秀」および「正常」群は、長寿遺伝子探索における調査対象として超百寿者に比する可能性が示された。

ii), 超高齢者用認知機能評価尺度の開発  
超高齢者の認知機能を第三者の観察によって評定する尺度を作成した。対象者は東京23区在住の100歳以上高齢者233名であった。対象者の日常生活の様子をよく知る家族や施設職員によって尺度の評定が行われた。項目反応理

論を用いて、19項目の完全版と10項目版の短縮版を構成した。尺度の信頼性は高かった（ $\alpha=0.94, 0.87$ ）。MMSEとの得点との相関も高く（ $r=0.85, 0.83$ ）、妥当性も高いことが示された。また、MMSEでは測定不可能な低いレベルの認知機能を持つ者にも適用できること、視聴覚障害の影響も少ないことが示された。本尺度は応答困難や視聴覚障害で認知テストの実施が困難な超高齢者にも簡便に実施できるため、面接調査のみならず郵送調査や電話調査での活用が期待される。

iii) 百寿者の病歴 profile に関する研究  
百寿者は健康長寿モデルと考えられている。近年、百寿者の増加は著しく本当に健康長寿モデルかどうかについて議論されるようになった。我々は平成12年より東京地区の百寿者を対象に調査を行い、1) どの様な疾患を持っているか、2) どの様な治療が行われているか、3) 疾患 profile と機能について検討した。対象は東京都在住の百寿者302名（男性65名、女性237名）である。97%の百寿者が何らかの病気を持っていた。病歴で多いものは順に高血圧、骨折、白内障、心疾患、消化器疾患、脳卒中であった。糖尿病の罹患率は6%と低かった。日本人の3大死因である脳卒中、心疾患、癌に100歳までならなかった割合は61%であった。百寿者の機能に影響を与える疾患は、脳卒中と骨折でいずれも負の影響を示した。興味深いことに高血圧は良好な機能と関連していた。長寿達成には生活習慣病にならないことが重要であるが、100歳時の機能を保つためには脳卒中と骨折の予防が重要であることが示唆された。

##### iv) コレステロール、HDL-コレステロールと血圧の機能への影響

コレステロール、HDL、血圧は生活習慣病の重要な危険因子である。これらの危険因子を是正することにより動脈硬化性疾患の発症が予防されることが知られている。どのレベルまで治療するかについて治療ガイドラインが専門学会より提唱されている。動脈硬化性疾患を免れてきた多くの百寿者において危険因子が機能にどのような影響を持つかを調べた。総コレステロール、HDLは平均値より高い群の方が機能が有意に高いことが判明した。驚いたことに血圧では140/90mmHgを越える高血圧群で有意にADL、認知機能が高かった。超高齢者における血圧やコレステロールの管理は中高年と異なる基準を考慮する事を示唆する。

v) 百寿者におけるアデイポネクチンの研究—長寿と脂肪組織—  
百寿者ではアデイポネクチン濃度が若年群に

比較して2倍高かった。高齢者対照群における濃度は若年群と百寿者群の中間値を示しており、加齢に伴い上昇していくものと考えられる。アデイポネクチン遺伝子多型の頻度については若年群と有意差を認めなかった。このためアデイポネクチンが高い理由としては環境要因によると考えられる。アデイポネクチンの *in vivo* での作用を推測するために各種パラメーターとの相関を調べた。百寿者ではアデイポネクチン濃度は体格指数(BMI)、レプチン、HbA1C、CRP、E-selectin と有意の負の相関を示した。また HDL-C とは正の相関を認めた。一方、若年群では HDL-C とのみ正の相関を認めた。これらの結果より、百寿者においてはアデイポネクチンは抗炎症作用、抗動脈硬化作用(内皮マーカーの E-selectin と負、HDL と正の相関)、抗糖尿病作用(HbA1C と負の相関)を示す可能性があることが示唆された。

脂肪細胞の萎縮するリポジストロフィー、同様な病態を示す progeria、また大型化するメタボリック症候群を考えると、小型の機能の良い脂肪組織は抗老化作用を持つ可能性があり、今後脂肪組織の分化増殖を司る要因の検討が興味深い。

#### vi), 血液型と長寿との関連性について

百寿者 269 名(男性 67 名、女性 202 名、平均年齢 101.2 ± 1.8 歳)と対照群との間で ABO 血液型(表現型)の比較を行った。対照群は東京在住で平成 15 年度に慶應健康相談センターで人間ドックを受診した 7153 名(男性 5480 名、女性 1673 名、平均年齢 54.8 ± 11.0 歳)(対照群 1)、およびそれらのうちの 70 歳以上の群 740 名(対照群 2、平均年齢 74.8 ± 4.4 歳)、80 歳以上の群 118 名(対照群 3、平均年齢 82.8 ± 2.8 歳)の 3 つである。採血で確認した血液型の頻度分布は A、O、B、AB の順に、百寿者では 34.2%、28.3%、29.4%、8.2% であったが、対照群 1 は 38.6%、30.1%、21.9%、9.4% であり、百寿者では B 型の頻度が有意に多かった(カイ二乗値 8.41、 $p = 0.04$ )。この傾向は、対照群 2、対照群 3 との比較においても見られた。ABO 血液型は糖鎖により決定される。この結果は糖鎖が長寿に寄与する可能性を示唆し今後網羅的な糖鎖分析方法である glycomics が興味深い研究テーマとなろう。

#### vii) 性格と長寿

百寿者の性格については今までにも報告があったが我々は Neo-PI を用いて性格を、神経症傾向、外向性、開放性(新しいものを取り入れる)、調和性、誠実性(きめたことを実行する、几帳面)の五つ因子に分けて他の年齢層と違い

があるかについて検討した。その結果、1) 男性では開放性が高いこと、2) 女性では、外向性、開放性、誠実性が高いことが判明した。性格がなぜ長寿に関係するのかについて以下の仮説を考えた。1) 性格遺伝子の性格以外への影響、2) 性格と人間関係、3) 健康行動、4) 幸せ感、5) 依存症への影響。

#### viii) 性格遺伝子と長寿

セロトニントランスポーター(5HTT)プロモーター領域の多型性には L 型と S 型が存在する。S 型と神経症傾向、ストレス時のうつ状態へなりやすいこと、喫煙習慣を持つこととの関連が報告されている。

5HTT 多型を百寿者と若年群で比較した。1/1 の頻度が女性百寿者では有意に高く(8.5% vs. 3.9%,  $p < 0.01$ ) 男性でも高い傾向を示した(5.1 vs. 3.3%)。この多型性と長寿の関連を調べるため、1) 性格、2) 認知機能、3) ADL、4) 栄養状態(アルブミン値)との関係を調べた。栄養状態のみが多型性と関連した(1/1 vs. 1/s vs. s/s : 3.85 vs. 3.66 vs. 3.59 g/dl,  $P < 0.05$ )。この多型性は性格とも関連するが栄養状態に関連し長寿と結びつく可能性が考えられた。

#### ix), 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 304 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に加えて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。現在解析中であるが 100 歳者の予後を決める因子は単独ではなく、互いに関連している可能性が高い。互いに関連しながら変動することは根元的な原理(energy reserve depletion, physical fitness depletion などと呼ばれる状態)が存在するためなのかもしれない。

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。

#### x) Proteomics、glycomics の可能性—老化修飾因子、老化指標の同定—

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の検討はほとんどされていなかった。我々は SSC 血液を元に Proteomics が可能であることを示した。すなわち 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示唆される。また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 と若年で異なるということで glycomics が解析に利用できよう。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る。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。

## II、超百寿者調査の概略

百寿者は指数関数的に増加している。平成 17 年度の調査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平成 18 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。ではどの年齢が健康長寿モデルとして最適であろうか？

平成 14 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。

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 105 歳以降での死亡率の減少、統計解析に必要な対象数の確保の 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。

SSC と 100 歳者の生化学データを比較検討するとアルブミン、コレステロール、HDL コレステロールでは SSC で有意に低下していた。また CRP は有意に高値を示した。100 歳者に見られる所見がより増強されていることが示された。即ち究極の人の老化を調べるのに最適の年齢層と考えられる。

一方 SSC の遺伝背景について調べると、アポ E genotype では E2/E2 が SSC で高い傾向を示した。ミトコンドリア遺伝多型では高齢者に多い D4a ハプロタイプが SSC では 100 歳者に比較して有意に高かった(15.2 vs 8.3%,  $p=0.026$ )。このように SSC は 100 歳者と異なる遺伝素因を持つ可能性が示唆された。長寿遺伝子の探索に最適の年齢層と考えられる。

## <倫理面への配慮>

本調査に当たって、本人及び家族に文書により十分な説明をおこなった。その後本人(可能な場合は)または家族より同意文書への署名を得た。この調査は慶應義塾大学医学部倫理委員会、東京都老人総合研究所、理化学研究所、愛媛大学医学部の承認を得た。

## III、遺伝解析

i) 高密度 DNA マイクロアレイを用いたゲノムワイドコピー数解析システムの開発と百寿者、超百寿者のゲノムコピー数同定  
健康長寿モデルである 105 歳以上の超百寿者、東京百寿者を対象に、その遺伝的要因を解析した。本研究では、SNPs タイピング用高密度 DNA マイクロアレイを用いて超百寿者 180 検体、東京百寿者 257 検体(内 40 検体は超百寿者と東京百寿者で重複)、百寿者家族 81 検体のゲノム全域 50 万 SNPs タイピングを行った。若年対照群との比較を行い相関解析によりゲノム全域を網羅する健康長寿関連候補遺伝子リスト作成が可能となりつつある。また、ヒトゲノム解読完了後、急速に明らかになりつつあるヒトゲノムのコピー数多型についても、高密度 DNA マイクロアレイデータを用いて解析が可能であることが判明した。コピー数多型の長寿に及ぼす影響の検討が可能となりつつある。

ii), 機能に影響する遺伝素因の解析  
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SSC180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。

#### D. 将来展望

人長寿モデルであるSSCを対象にして、1)長寿遺伝子の同定、2)抗老化物質、老化指標の同定、3)究極の老化像を明らかにする事などが考えられる。

SSCと百寿者を対象にして超高齢期の機能関連遺伝子の検討が可能となると予想される。

以下に特に長寿遺伝子と抗老化物質、老化指標の同定について考察したい。

#### I. 長寿遺伝子

i) 長寿遺伝子はあるのか?あるとすればどのような作用を持つか

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するため特別に長寿遺伝子を想定する必要はない。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。現在では技術の進歩、長寿者の収集により実際に調べることが可能となってきている。

長寿遺伝子の働きについて、NadeauらはNature Genetics(2006,38.1095)に健康遺伝子の重要性に関する意見を発表した。それによると従来は疾患の遺伝子を同定しそれを元に治療を行ってきたが、副作用の問題、有効性が弱いことなどから、発想を転換して健康遺伝子(長寿遺伝子)を同定することにより様々な病態に対する防御因子が明らかに出来るのではないかとこの仮説を提唱している。この考えは我々の考えの通りで画期的な治療方法に結びつくと考えられる。

ii) 対象群と対照群(表1)

長寿モデルと考えられる超百寿者を対象としたい。

対照群については検討が必要である。

疾患関連遺伝子の同定におけるcase-control studyの有用性はすでに確立されている。長寿関連遺伝子および老化関連遺伝子の研究分野においてもこれまで多くの成果がcase-control studyからもたらされてきた。しかしながら、controlとしてどのようなpopulationを選定するかという明確な基準がこれまで示されてこなかったことが米国など移民や人口移動の大きい地域で問題となっている。例を以下に挙げる。

Geesamanらは米国の98歳以上の長寿者と同じく米国在住ではあるがさまざまな地域から募った50歳以下の健常者を対照群としたcase-control studyの結果、MTP遺伝子多型が長寿と関連することを報告した(Proc Natl Acad Sci U S A. 2003; 100, 14115)。しかし、この結果は他の研究では再現されなかった。この原因を検討したところGeesamanらが用いた対照群の設定に問題があり、MTP遺伝多型は長寿と関連していたのではなく、移民の多い米国における人種構成の変化を反映している可能性があることが指摘された(Proc Natl Acad Sci U S A. 2005; 102: 7906)。長寿者のサンプルを苦労して集めても、対照群の選定によっては研究結果が偏ったものになる可能性があることが示唆された。これを受けて長寿関連遺伝子、老化制御遺伝子の同定においては長寿者サンプル同様に対照サンプルにも明確な定義が必要であることが世界的な潮流となっている。

われわれは長寿遺伝子研究における世界的潮流を考慮し、1)長寿者(発端者 proband)、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。2)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した。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確実性の高い結果が得られると考えられる。

グループ名	調査群	特徴	年齢分布	解析(進行中及び予定)	調査終了数	調査目標数
対象群1	百寿者	正常老化の究極	100-104歳	相関解析、抗老化物質、老化指標	300	400
対象群2	超百寿者 (106歳以上者)	人長寿モデル	106歳以上	相関解析、抗老化物質、老化指標	180	250
対象群3	90歳以上の同胞	遺伝素因の強い長寿者	90歳以上	連鎖解析、抗老化物質の対照	45組	250組
対照群1	若年者	若年対照群	21-50歳	相関解析、抗老化物質、老化指標	200	300
対照群2	長寿者家族 (子孫とその配偶者)		46-85歳	相関解析、抗老化物質、老化指標の対照	70家系	200家系
A	子孫	長寿者と同じ環境で直系				
B	子孫の配偶者	長寿者と同じ環境で非直系				
剖検群	100歳に到達せず(ほぼ平均寿命で)亡くなった例		70-90歳	相関解析対照群	1800	1800+

表1 調査群

### iii) 解析方法 (図1)

遺伝解析には相関解析と、連鎖解析が知られている。二つの方法は原理が異なるためこの方法で疾患の原因遺伝子の同定が行われている。糖尿病などがその例である。相関解析を長寿関連遺伝子解析に初めて応用したのが Schacter らである (Nature Genetics 1994:6, 29)。これ以後相関解析が長寿遺伝子同定に常用されている。

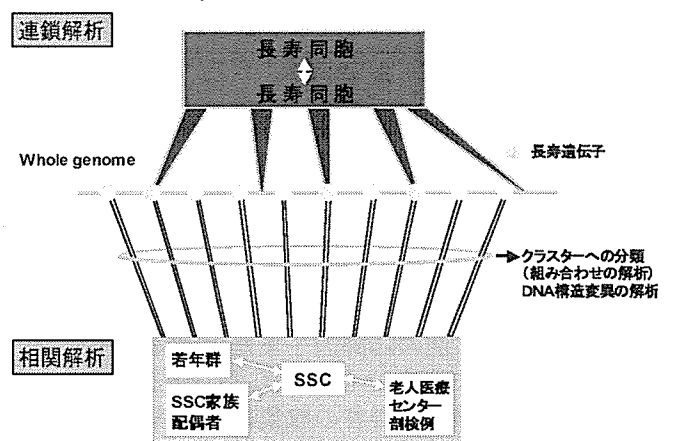
一方長寿遺伝子の同定に連鎖解析が応用できるとことは Perls らが初めて報告した (PNAS 2001 98: 10505)。

我々は SSC と比較対照群での whole genome scan による相関解析を行っている。現在比較対照群の収集と解析を開始した。数年以内に長寿関連遺伝子のリスト作成が終了する予定である。

連鎖解析には長寿同胞の収集が必要である。現在40組の収集を終了したが、今後200組の収集を行い連鎖解析を試みる。

相関解析と連鎖解析の結果を総合することで確実に長寿遺伝子を同定したい。

図1、長寿遺伝子同定の戦略



### III) 抗老化物質、老化指標の網羅的検索 (図2)

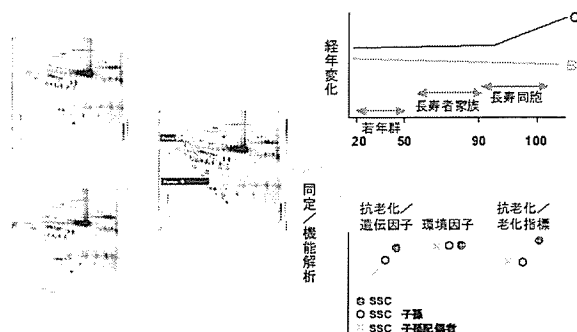
超百寿者をはじめとする長寿者は、長寿遺伝子の遺伝子産物としての抗老化物質を持っている可能性がある。この検討のために、proteomics、glycomics を用いて血漿中の糖タンパク質の包括的な検討が可能となってきた。生理的な現象である老化に proteomics が応用できるかは不明であったが、我々は SSC と若年対照群で proteomics を用いた包括的糖タンパク質のスクリーニングが可能であることを見いだした。これを用いれば、濃度に有意差のある物質をターゲットに、1) 年齢による変化、2)

長寿者家族内での違い（環境要因か遺伝要因かが明らかになる）、3) データベースの検索による機能の検討を行い、抗老化物質、老化指標を同定できる可能性がある。老化指標は抗老化医療に必須である。

## F. 謝辞

平成11年より続いた研究は、百寿者、超百寿者、その家族、介護提供者などのご協力がなければ不可能であり、深く感謝いたします。

図2. 抗老化物質と老化指標同定の戦略



## E. 結語

長寿科学は長寿者の増加、解析技術の急速な進歩、基礎老化学の進歩より今後急展開が期待できる分野である。

癌、動脈硬化の研究で常用される、相関解析 (Schacter)、連鎖解析 (Perls) が長寿科学で応用可能であることが示された。また我々の予備実験から proteomics が長寿科学に応用できることが示された。Glycomics の応用も長寿科学で強力な手段となると考えられる。

これまでに報告されている膨大な長寿科学研究の成果から、超高齢期の健康維持や長寿には複数の相互関連がかかわっていることが明らかにされている。特に長寿遺伝子研究において高密度 DNA マイクロアレイデータを解析することにより従来の遺伝子研究ではなしえなかった遺伝子-遺伝子相互関連、遺伝子-環境関連の検討が可能となる。

本報告書の各論に述べたごとく、われわれは研究開始当初から多面的に健康長寿のメカニズムを研究して来た。そしてようやくこれらのデータを統合して健康長寿を総体として解析しうる段階に到達した。どのような解析手法を用いて膨大なデータを統合するのが今後の課題であるが、こうしたアプローチは長寿科学のみならず medical science 全体にも大きな影響を与えられ、今後画期的な成果がでてくることが期待される。

### 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 semisuper centenarians	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
Gondo Y, Hirose N, et.al	Functional status of centenarians in Tokyo, Japan: developing better phenotypes of exceptional longevity	J Gerontology med sci	61A	305-310	2006
Arai Y, Nakazawa S, Kojima T, Gondo Y(10 番目), Hirose N(14 番目)	High adiponectin concentration and its role for longevity in female centenarians,	Geriatr Gerontol Int	6	32-39	2006
広瀬信義、権藤恭之(6 番/7 人)、小島俊男(7 番/7 人)	百寿者の抗老化機序—健康長寿達成に向けて—	日本内科学会雑誌	95	447-452	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
Gondo Y, Hirose N et al (11 人)	Contribution of an affect-associated gene to human longevity: Prevalence of the long-allele genotype of the serotonin transporter-linked gene in Japanese centenarians	Mechanism of aging and development	126	1178-1184	2005



発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
NiemiAK, Nirose N(8 番/9人)	A combination of three common inherited mitochondrial DNA polymorphisms promotes longevity in Finnish and Japanese subjects	Eur J hum genet	13	166-170	2005
広瀬信義、小島俊男、権藤恭之、その他	百寿者の遺伝背景—長寿遺伝子同定の戦略—	日本老年医学会雑誌	42	664-665	2005
Abe M, Wu Z, Yamamoto M, Jin JJ, Tabara Y, Mogi M, Kohara K, Miki T, Nakura J	Association of dopamine beta-hydroxylase polymorphism with hypertension through interaction with fasting plasma glucose in Japanese	Hypertens Res	28	215-221	2005
増井幸恵・権藤恭之・稲垣宏樹・広瀬信義	超高齢者用認知機能評価尺度の開発	老年精神医学雑誌	16	837-845	2005
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
Tanaka M, Hirose N(19番/28名)	Mitochondrial Genome Variation in Eastern Asia and the Peopling of Japan.	Genome Research	14	1832-1850	2004
Hirose et al	Tokyo centenarian study: aging inflammation hypothesis.	Geriatr Gerontol Int	4	S182-S185	2004
Kojima T, Hirose N(12番/12名)	Association analysis between longevity in the Japanese population and polymorphic variants of genes involved in insulin and insulin-like growth factor 1 signaling pathways	Exp Gerontol	39	1595-1598	2004
Shimizu K, Hirose N	Blood type B might imply longevity.	Exp Gerontol	39	1563-1565	2004
Arai Y, Hirose N	Aging and HDL metabolism in elderly people more than 100 years old	J atheroscler Thromb	11	246-252	2004

## 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  $24.4 \pm 0.9$  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  $(24.4 \pm 0.9)$  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](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  $62.1 \pm 8.9$  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  $76.5 \pm 9.7$  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  $21 \pm 2$  years; range, 18–25 years) with a body mass index (BMI) of  $30.1 \pm 2.5$ , 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  $20 \pm 3$  years; range, 18–25) with a BMI of  $20.2 \pm 2.3$ , 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  $58 \pm 5$  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  $65 \pm 10$  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 haplogroup 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
<b>*D4b2b</b>	<b>3</b>	<b>7</b>	<b>2</b>	<b>4</b>	<b>9</b>	<b>4</b>	<b>11</b>	<b>3.41</b>	<b>0.007</b>	<b>6.03</b>	<b>0.009</b>
D4b2a	3	1	6	1	2	2	6	0.00	0.500	1.00	0.617
<b>D4b2</b>	<b>6</b>	<b>8</b>	<b>8</b>	<b>5</b>	<b>11</b>	<b>6</b>	<b>17</b>	<b>4.17</b>	<b>0.003</b>	<b>2.36</b>	<b>0.042</b>
<b>*D5</b>	<b>2</b>	<b>5</b>	<b>1</b>	<b>4</b>	<b>4</b>	<b>5</b>	<b>7</b>	<b>3.65</b>	<b>0.005</b>	<b>7.41</b>	<b>0.032</b>
<b>*D4a</b>	<b>7</b>	<b>6</b>	<b>4</b>	<b>6</b>	<b>5</b>	<b>5</b>	<b>8</b>	<b>3.81</b>	<b>0.004</b>	<b>2.08</b>	<b>0.186</b>
D4b	9	9	12	6	14	10	19	2.54	0.022	1.72	0.120
<b>D4</b>	<b>25</b>	<b>25</b>	<b>28</b>	<b>27</b>	<b>28</b>	<b>24</b>	<b>36</b>	<b>4.64</b>	<b>0.002</b>	<b>1.45</b>	<b>0.142</b>
<b>D</b>	<b>33</b>	<b>39</b>	<b>34</b>	<b>35</b>	<b>36</b>	<b>38</b>	<b>47</b>	<b>5.61</b>	<b>0.001</b>	<b>1.74</b>	<b>0.040</b>
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  $t_{HN} = (\text{number of centenarians} - \text{number of healthy normals})/\text{sdev}$ , 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.

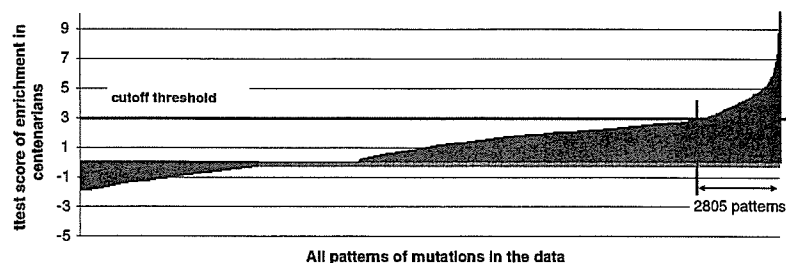
For all D samples taken as a single group, we observe the known (Tanaka et al. 1998) enrichment in centenarians correlated with two markers, a non-synonymous transversion 5178C > A (ND2: Leu237Met) and a synonymous transition 4883C > T. The subgroup D4b2a of D4b2 (which does not have the 9296C > T mutation), 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.

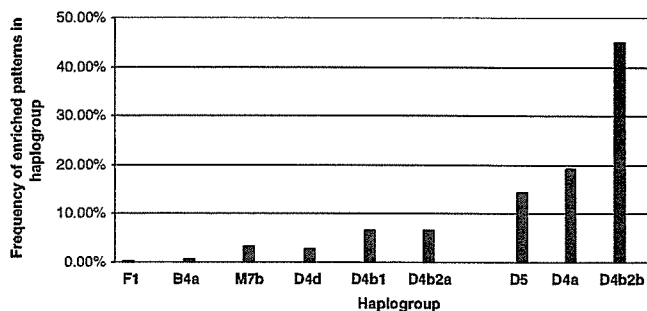
Figure 2 shows the haplogroups in which the significant patterns of Fig. 1 are found. Note again the appearance of the haplogroups D4a, D5, and D4b2b. Other haplogroups also show up in this figure but with lower frequency. Table 2 shows the mtSNPs in the enriched patterns. The mtSNPs are sorted by the 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  $\sigma_M$ . The time to most recent common ancestor (TMRCA) is then given by:

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

where  $\mu$  is the mutation rate in units of number of mutations per sample per year. Using the value  $1.7 \times 10^{-8}$  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  $\geq 3.00$ )

SNP	$t$ -test statistic	$t$ -test $P$ -value	Haplogroups with SNP <sup>b</sup>	Haplogroup specificity	rCRS	Mutant	Type	Gene	Sense
9,296 <sup>a</sup>	3.41	0.007	1	D4b2b	C	T	TS	CO3	Syn
14,605 <sup>a</sup>	3.87	0.004	1	D4b2b	A	G	TS	ND6	Syn
11,944 <sup>a</sup>	3.95	0.004	1	D5	T	C	TS	ND4	Syn
12,026 <sup>a</sup>	3.95	0.004	1	D5	A	G	TS	ND4	Nsyn
1,107 <sup>a</sup>	4.26	0.003	3	D5	T	C	TS	12S rRNA	
5,301 <sup>a</sup>	4.26	0.003	3	D5	A	G	TS	ND2	Nsyn
10,397 <sup>a</sup>	4.26	0.003	3	D5	A	G	TS	ND3	Syn
752 <sup>a</sup>	3.95	0.004	1	D5	C	T	TS	12S rRNA	
14,979 <sup>a</sup>	3.30	0.008	1	D4a	T	C	TS	Cytb	Nsyn
3,206 <sup>a</sup>	3.81	0.004	1	D4a	C	T	TS	16S rRNA	
8,473 <sup>a</sup>	3.30	0.008	1	D4a	T	C	TS	ATP8	Syn
5,178 <sup>b</sup>	6.37	0.001	15	D	C	A	TV	ND2	Nsyn
4,883 <sup>b</sup>	5.81	0.001	15	D	C	T	TS	ND2	Syn
8,414 <sup>b</sup>	4.08	0.003	13	D4	C	T	TS	ATP8	Nsyn
8,020 <sup>b</sup>	3.30	0.008	5	D4b	G	A	TS	CO2	Syn
1,382 <sup>b</sup>	3.99	0.004	2	D4b2	A	C	TV	12S rRNA	
8,964 <sup>b</sup>	4.17	0.003	2	D4b2	C	T	TS	ATP6	Syn
194 <sup>c</sup>	4.58	0.002	2	D4b2b, D4d1a	C	T	TS	MajorNCR1	
199 <sup>c</sup>	6.10	0.001	3	M7b, B4a, D4d1a	T	C	TS	MajorNCR1	
5,460 <sup>c</sup>	3.02	0.012	2	M7b	G	A	TS	ND2	Nsyn
204 <sup>c</sup>	4.95	0.001	1	B5	T	C	TS	MajorNCR1	
4,071 <sup>c</sup>	4.06	0.003	1	M7b	C	T	TS	ND1	Syn
13,651 <sup>c</sup>	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

<sup>a</sup> mtSNPs that are specific for haplogroups D4a, D4b2b, and D5

<sup>b</sup> mtSNPs that are specific for their ancestor megahaplogroup

<sup>c</sup> 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  $24.4 \pm 0.9$  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 4883 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  $62.1 \pm 8.9$  and  $76.5 \pm 9.7$  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 ( $24.4 \pm 0.9$  kYBP) and the smallest was between D4e and D4h ( $3.5 \pm 0.3$  kYBP). We find that the D4b2b haplogroup (with mtSNP 9296C > T) split off from D4b2 approximately ( $6.8 \pm 0.6$ ) 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*<sub>1</sub> 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, Herrnstadt 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

## Determination of genomic breakpoints in an epileptic patient using genotyping array

Toshio Kojima <sup>a,e,\*</sup>, Wakako Mukai <sup>a</sup>, Daisuke Fuma <sup>a,b</sup>, Yuto Ueda <sup>c,e</sup>, Motohiro Okada <sup>d,e</sup>, Yoshiyuki Sakaki <sup>a</sup>, Sunao Kaneko <sup>d,e</sup>

<sup>a</sup> Comparative Systems Biology Team, Genomic Sciences Center, RIKEN, Yokohama, Japan

<sup>b</sup> Department of Periodontology, School of Dentistry, Aichi-gakuin University, Nagoya, Japan

<sup>c</sup> Department of Psychiatry, Miyazaki Medical College, Miyazaki, Japan

<sup>d</sup> Department of Neuropsychiatry, Hirosaki University School of Medicine, Hirosaki, Japan

<sup>e</sup> The Epilepsy Genetic Study Group, Japan

Received 8 January 2006

Available online 19 January 2006

### Abstract

Recent advances in DNA microarray technology have enabled the identification of small alterations throughout the genome. We used standard karyotype analysis, followed by DNA microarray analysis and PCR to precisely map the chromosomal 4p deletion and determine the deletion breakpoints in the genome of an epileptic patient. The karyotype of the patient was 46,XY,del(4)(p15.2p15.3) as determined by G-banding analysis. We used a high-density oligonucleotide genotyping array to estimate the size of the deletion (4.5 Mb) and to locate the breakpoints within a 9-kb region on one side of the deletion and a 100-kb region on the other side. We amplified by PCR and sequenced the genomic region encompassing the breakpoints, and mapped the deletion to regions extending from 21648457 to 26164287 and from 26164505 to 26167493, respectively (chromosome 4 of NCBI Homo sapiens Genome Build 35.1). The deletion involves 18 genes, one of which (*CCKAR*) is partially deleted.

© 2006 Elsevier Inc. All rights reserved.

**Keywords:** Genomic breakpoints; DNA microarray; PCR; Epilepsy

Recent advances in DNA microarray technology have enabled the identification of small alterations throughout the genome. The DNA microarray techniques have been used to analyze gene dosage in cancer and congenital disorders [1–4]. The role of large-scale copy number variations in the evolution of the human genome has become a focus of research attention because of their recent discovery in the genome of healthy individuals [5,6]. Genome-wide screening for DNA copy number changes can be carried out by array-based comparative genomic hybridization. Its highest map resolution is less than 100 kb, with approximately 30,000 arrayed BAC clones spanning the entire human genome [7]. Another genome-wide DNA screening

approach uses high-density oligonucleotide microarrays, originally designed to genotype more than 10,000 SNPs in the entire human genome [8–10]. These array analyses have some clear advantages. They follow simple protocols that are easily reproducible. Moreover, analyses can be easily carried out with established assay systems that are commercially available. However, copy number analysis can be complicated by poor signal-to-noise ratio caused by various experimental biases. Although several effective algorithms/software for copy number analysis have been reported [11–13], the currently available ones do not support the analysis using the new oligonucleotide array, namely, GeneChip Mapping 500K Set. Here, we report a simple method that eliminates background noise from data obtained by the oligonucleotide microarray analysis of gene dosage. We have used this method to locate deletion breakpoints in the genome of an epileptic patient.

\* Corresponding author. Fax: +81 45 503 9170.  
E-mail address: [tkojima@gsc.riken.jp](mailto:tkojima@gsc.riken.jp) (T. Kojima).