2014

Disease variants in genomes of 44 centenarians

Y. Freudenberg-Hua  
Hofstra Northwell School of Medicine

J. Freudenberg  
Northwell Health

D. Oschwald

E. Christen  
Northwell Health

J. Koppel  
Hofstra Northwell School of Medicine

See next page for additional authors

Follow this and additional works at: https://academicworks.medicine.hofstra.edu/articles

Part of the Psychiatry Commons

Recommended Citation


This Article is brought to you for free and open access by Donald and Barbara Zucker School of Medicine Academic Works. It has been accepted for inclusion in Journal Articles by an authorized administrator of Donald and Barbara Zucker School of Medicine Academic Works. For more information, please contact academicworks@hofstra.edu.
Authors
Y. Freudenberg-Hua, J. Freudenberg, D. Oschwald, E. Christen, J. Koppel, B. Greenwald, R. B. Darnell, S. Germer, G. Atzmon, P. Davies, and +5 additional authors
Disease variants in genomes of 44 centenarians

Yun Freudenberg-Hua1,2, Jan Freudenberg3, Vladimir Vacic4, Avinash Abhyankar4, Anne-Katrin Emde4, Danny Ben-Avraham5, Nir Barzilai5, Dayna Oschwald4, Erika Christen1, Jeremy Koppel1,2, Blaine Greenwald2, Robert B. Darnell4,6, Soren Germer4, Gil Atzmon5 & Peter Davies1

1The Litwin-Zucker Research Center for the Study of Alzheimer’s Disease and Memory Disorders, The Feinstein Institute for Medical Research, North Shore-LIJ, Manhasset, New York 11030
2Division of Geriatric Psychiatry, Zucker Hillside Hospital, North Shore-LIJ, Glen Oaks, New York 11040
3Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, North Shore-LIJ, Manhasset, New York 11030
4New York Genome Center, 101 Avenue of the Americas, New York, New York 10013
5Institute for Aging Research Departments of Medicine and Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461
6Department of Molecular Neuro-Oncology, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10065

Abstract

To identify previously reported disease mutations that are compatible with extraordinary longevity, we screened the coding regions of the genomes of 44 Ashkenazi Jewish centenarians. Individual genome sequences were generated with 30× coverage on the Illumina HiSeq 2000 and single-nucleotide variants were called with the genome analysis toolkit (GATK). We identified 130 coding variants that were annotated as “pathogenic” or “likely pathogenic” based on the ClinVar database and that are infrequent in the general population. These variants were previously reported to cause a wide range of degenerative, neoplastic, and cardiac diseases with autosomal dominant, autosomal recessive, and X-linked inheritance. Several of these variants are located in genes that harbor actionable incidental findings, according to the recommendations of the American College of Medical Genetics. In addition, we found risk variants for late-onset neurodegenerative diseases, such as the APOE e4 allele that was even present in a homozygous state in one centenarian who did not develop Alzheimer’s disease. Our data demonstrate that the incidental finding of certain reported disease variants in an individual genome may not preclude an extraordinarily long life. When the observed variants are encountered in the context of clinical sequencing, it is thus important to exercise caution in justifying clinical decisions.

Keywords
Aging, Ashkenazi, centenarian, disease gene, incidental finding, whole genome sequencing

Correspondence
Yun Freudenberg-Hua, The Litwin-Zucker Research Center for the Study of Alzheimer’s Disease and Memory Disorders, The Feinstein Institute for Medical Research, North Shore-LIJ, 350 Community Drive, Manhasset, NY 11030. Tel: (516) 562 3492; Fax: (516) 5620401; E-mail: yfreuden@nshs.edu

Funding Information
This study is supported by the Mildred and Frank Feinberg Family Foundation. J. F. is supported by NIAMS/NIH under Award Number R03 AR063340. N. B., G. A., and D. B. A. are supported by NIH/NIA grants R01 AG168381, R01 AG042188, R01 AG046949, and P01 AG021654, the Einstein Nathan Shock Center (P30 AG038072), and the Einstein Glenn Center for the Biology of Human Aging.
more recently, ClinVar (Landrum et al. 2014) and the Genome Wide Association Studies (GWAS) catalog (Hindorff et al. 2009). However, the penetrance of these recorded variants spans a broad spectrum, ranging from complete penetrance for a set of monogenic mutations to the very small effect sizes of many GWAS hits. While it is widely known that most GWAS hits have only limited clinical prognostic relevance, the penetrance and prognostic value of many previously reported monogenic mutations is less clear. Recently, a significant percentage of such putative “disease mutations” were claimed to be of questionable pathogenicity (Cassa et al. 2013; Dorschner et al. 2013; Flannick et al. 2013; Kenna et al. 2013; Riggs et al. 2013). Accordingly, it is an important aim of current human genetic research to systematically assess the clinical significance of genetic variants (Duzkale et al. 2013). As one way to identify those mutations that may require reinterpretation, we use a sample of individuals with exceptional longevity to identify reported disease mutations for which the pathogenicity status should be preferentially reevaluated.

If a mutation for a dominant late-onset disease is encountered in the genome of a younger individual, a reduced penetrance is difficult to assign without longitudinal follow-up. In addition to the uncertain significance of many reported mutations for dominant diseases, the traditional assumption that heterozygous carriers for recessive disease alleles are unaffected (Nussbaum et al. 2007) may not hold true. For instance, Gaucher’s disease alleles in the GBA gene have been found to strongly increase the risk of developing Parkinson’s disease (Sidransky et al. 2009) as well as dementia with Lewy bodies (Nalls et al. 2013). Similarly, heterozygous carriers for autosomal recessive disease alleles for Alport syndromes have increased risk of renal failure (Temme et al. 2012). Therefore, different lines of evidence need to be integrated to assign a clinical significance score to each variant. For mutations previously linked to dominant diseases that negatively affect life expectancy, their observation in centenarian genomes may be viewed as evidence for unclear pathogenicity or at least incomplete penetrance. To a certain extent, reduced penetrance could result from a variety of protective genetic, epigenetic, environmental, and random factors (Cooper et al. 2013) or other buffering mechanisms (Bergman et al. 2007).

In the present study, we search the coding regions of 44 Ashkenazi Jewish (AJ) centenarians for variants that were previously reported as causal mutations for medically relevant phenotypes. To this end, we use the recently established and publicly accessible ClinVar database (Riggs et al. 2013) which not only provides highly structured data access but also includes community-based data curation (Landrum et al. 2014). We chose ClinVar because of the quality of its content. Our study is meant to be part of the community effort to further improve its quality, which is eased by its free availability. Such databases are gaining in importance, because the increased utilization of next-generation sequencing technology magnifies the challenge to interpret genetic testing results (Lyon and Wang 2012; Lohn et al. 2013; Manolio et al. 2013; Rehm et al. 2013). For example, in August 2013 the ClinVar database recorded 14,746 variants classified as pathogenic and 1672 as probably pathogenic. We focus on the description of those variants that have been previously reported as Mendelian disease mutations that would increase the risk of mortality, such as degenerative diseases, neoplasm, and cardiac diseases. The observation of previously reported disease variants in centenarian genomes may aid clinical geneticists who are confronted with the challenge to evaluate their pathogenicity.

Materials and Methods

Study population

DNA samples of 44 AJ centenarians were collected as part of a longevity study at the Albert Einstein College of Medicine that was described elsewhere (Barzilai et al. 2003). The sample includes eight male (95–103 years old) and 36 female (95–106 years old) subjects. The mean and median age of the subjects was 99.6 and 100 years (standard deviation $\sigma = 3.1$). Informed written consent was obtained in accordance with the policy of the Committee on Clinical Investigations of the Albert Einstein College of Medicine. Genomic DNA extracted from blood cells using standard procedure was sent for sequencing. Mini-mental status examinations (MMSE) (Folstein et al. 1975) and selected health status information such as hearing, vision, and whether a subject had a history of cancer or myocardial infarction (MI) are available for the majority of the centenarians. The level of education was not available. Given that median MMSE scores are found to be a function of age and level of education (Crum et al. 1993; Bravo and Hebert 1997) and can be as low as 20 for people 85 years and older with $<4$ years of education (Crum et al. 1993), we use the reference score of $\geq 24$ (one standard deviation of the mean MMSE score for people 85 and older; Bravo and Hebert 1997) to exclude the presence of dementia in our sample. To our knowledge, no reference MMSE scores specifically for centenarians are available.

Genome sequencing and variant calling

Whole genome sequencing was performed on Illumina HiSeq 2000 platform (Illumina, San Diego, CA). Paired-end $2 \times 100$ reads were aligned to the GRCh37 human reference using the Burrows-Wheeler Aligner (BWA) (Li
and Durbin 2009) and processed using the best-practices pipeline that includes marking of duplicate reads by the use of Picard tools (http://picard.sourceforge.net), and realignment around DNA insertions and deletions (indels) and base recalibration via Genome Analysis Toolkit (GATK) Lite ver. 2.3-9 (McKenna et al. 2010). Single-nucleotide variants (SNVs) were jointly called on the 44 unrelated AJ centenarians together with HapMap trio NA12877, NA12878, and NA12882 using the UnifiedGenotyper module of GATK Lite. The HapMap trio was sequenced at Illumina and released as part of the PlatinumGenomes project (http://www.illumina.com/platinumgenomes). We used variant quality score recalibration and Mendelian inconsistencies on the HapMap trio to determine the optimal variant quality score (VQSLOD) cutoff.

Joint calling for transcribed gene regions of 44 AJ centenarians identified a total of 6.3 million genetic variants. A total of 87,899 coding SNVs were found (100% call rate) in coding regions defined by the longest transcript according to the UCSC hg19 database (Meyer et al. 2012). For variants with a VQSLOD score >2, the concordance rate for SNVs between Illumina 2.5M chip and GATK variant calls was above 99.7% for all subjects. Due to the comparatively lower performance of the UnifiedGenotyper in calling short insertions and deletions, in the present study we decided not to analyze insertion/deletion variants.

Database annotation of variants

Using custom scripts and the SNP & Variation Suite software (Version 7.7.8; Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com), we annotated coding and splice variants based on the knownGene and the refGene tracks in the UCSC hg19 database (Meyer et al. 2012). To find variants of potential clinical relevance, variants were evaluated for their possible pathogenicity based on the August 2013 version of the ClinVar database (http://www.ncbi.nlm.nih.gov/clinvar/) (Riggs et al. 2013). This database provides free highly structured public access to clinically relevant sequence variants and it also provides evidence-based interpretation of clinical significance for each variant (Landrum et al. 2014). All ClinVar accession numbers for the identified variants are linked to the respective NCBI (The National Center for Biotechnology Information) database for reference sequences (RefSeq) as well as OMIM accession numbers. Population allele frequencies of variants were retrieved from the NHLBI Exome Sequencing Project (ESP) database (Exome Variant Server, NHLBI GO ESP, Seattle, WA) (http://evs.gs.washington.edu/EVS). To our knowledge, a similar large-scale population allele frequency data specific for the AJ population are currently not available.

Quality control filtering of putative disease variants

A total of 225 autosomal and seven X-chromosomal coding SNVs were found to be annotated as “pathogenic” or “likely pathogenic” in the ClinVar database. Eighteen of these variants had a VQSLOD score <4, most often due to poor mapping quality indicating genomic segmental duplications. However, genomic regions that are difficult to map are also known to produce false-negative variant calls (Lee and Schatz 2012). Therefore, we performed manual evaluation of the 18 variants. These include two mutations in the gene GBA that are known to cause Gaucher disease as well as being associated with late-onset Parkinson’s disease (most frequently referred to in the literature as L444P [uc001fjh.2.c:T1449C:p.L483P] and N370S [uc001fjh.2.c:A1227G:p.N409S]). Both mutations are known to be relatively frequent in the AJ population (Scott et al. 2010). Using BLAT (Kent 2002), N409S (rs76763715) was uniquely mapped to chromosome 1:155205634 and L483P (rs421016) was uniquely mapped to chr1:155205043. Thus, both variants are likely to be true SNVs in our sample and were therefore retained. The remaining 16 variants with low VQSLOD scores were removed.

Results

In the coding regions of 44 centenarian genomes, we found 210 autosomal and six X-chromosomal SNVs that passed quality control. Among these, 207 variants were classified as pathogenic and nine as likely pathogenic based on the ClinVar database. Excluding the two APOE risk variants, 86 of these SNVs were found to be common defined as minor allele frequency (MAF) of 5% or higher in either the European or African American populations in the ESP database (Table S1). Due to their high allele frequency in the general population, these 86 variants are considered unlikely to be strongly pathogenic with high penetrance and therefore not further discussed. According to the OMIM database, of the remaining 130 variants (Tables 1–5 and S2), 39 were reported to cause autosomal dominant diseases, 72 to cause autosomal recessive diseases, 6 were associated with X-chromosomal inheritance, and 13 with other modes of inheritance such as digenic, imprinting, complex, or unclear mode of inheritance.

Variants reported as causal for degenerative diseases of advanced age

Variants associated with age-related degenerative diseases were found in the genes APOE, GBA, UBQLN2, SEMA4A, RPI, MYO1A, CYP1B1, OPTN, VSXI, and WDR36 (Table 1).
### Table 1. Putative disease variants reported to cause degenerative diseases.

<table>
<thead>
<tr>
<th>CHR POS ID</th>
<th>RS</th>
<th>CHROMOSOME</th>
<th>POSITION</th>
<th>ID</th>
<th>REF</th>
<th>ALT</th>
<th>AC</th>
<th>GENE</th>
<th>EFFECT</th>
<th>CLNDBN</th>
<th>CLNACC</th>
<th>OMIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>155205043</td>
<td>rs421016</td>
<td>A</td>
<td>G</td>
<td>43/1/0</td>
<td>GBA</td>
<td>uc001fjh.2:c.T1449C:p.L483P</td>
<td>Gaucher's disease; late-onset Parkinson's disease; susceptibility to dementia with Lewy body</td>
<td>RCV000004511; RCV000004512</td>
<td>606463</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>155205634</td>
<td>rs76763715</td>
<td>T</td>
<td>C</td>
<td>43/1/0</td>
<td>GBA</td>
<td>uc001fjh.2:c.A1227G:p.N409S</td>
<td>Gaucher's disease; late-onset Parkinson's disease; susceptibility to dementia with Lewy body</td>
<td>RCV000004515; RCV000004516</td>
<td>606463</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>156146640</td>
<td>rs41265017</td>
<td>G</td>
<td>A</td>
<td>34/1/0</td>
<td>SEMA4A</td>
<td>uc001fnm.2:c.G2139A:p.R713Q</td>
<td>Retinitis pigmentosa 35</td>
<td>RCV000003528</td>
<td>607292</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>110441839</td>
<td>rs35703638</td>
<td>G</td>
<td>A</td>
<td>43/1/0</td>
<td>WDR36</td>
<td>uc003kd.2:c.G1346A:p.A449T</td>
<td>Glaucoma 1, open angle</td>
<td>RCV000001649</td>
<td>609669</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>55537560</td>
<td>rs77775126</td>
<td>C</td>
<td>T</td>
<td>43/1/0</td>
<td>RPI</td>
<td>uc003xsd.1:c.C1119T:p.T373I</td>
<td>Retinitis pigmentosa 1</td>
<td>RCV000006334</td>
<td>603937</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>13178766</td>
<td>rs75654767</td>
<td>G</td>
<td>A</td>
<td>43/1/0</td>
<td>OPTN</td>
<td>uc001lix.1:c.G1635A:p.R545Q</td>
<td>Glaucoma 1, open angle</td>
<td>RCV000007515</td>
<td>602432</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>45411941</td>
<td>rs429358</td>
<td>T</td>
<td>C</td>
<td>39/4/1</td>
<td>APOE</td>
<td>uc002pab.2:c.T389C:p.C130R</td>
<td>Hyperlipoproteinemia, type 3, autosomal dominant; Alzheimer's disease associated with APOE4 variant</td>
<td>RCV000019438; RCV000019448</td>
<td>107741</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>45412079</td>
<td>rs7412</td>
<td>C</td>
<td>T</td>
<td>33/1/0</td>
<td>APOE</td>
<td>uc002pab.2:c.C527T:p.R176C</td>
<td>Familial type 3 hyperlipoproteinemia, associated with APOE2</td>
<td>RCV000019428</td>
<td>107741</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>25060096</td>
<td>rs74315433</td>
<td>C</td>
<td>T</td>
<td>43/1/0</td>
<td>VSX1</td>
<td>uc002wuf.2:c.G480Ap.G60D</td>
<td>Keratoconus 1</td>
<td>RCV000024251</td>
<td>605020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>56591879</td>
<td>rs369947678</td>
<td>C</td>
<td>T</td>
<td>43/1/0</td>
<td>UBQCN2</td>
<td>uc004dus.2:c.C1574T:p.S255S</td>
<td>Amyotrophic lateral sclerosis 15 with or without frontotemporal dementia</td>
<td>RCV000022846</td>
<td>30264</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CHR, chromosome; POS, position on chromosome based on GRCh37; ID, dbSNP ID; REF, reference sequence; ALT, alternative sequence; AC, counts of subjects with homozygous reference alleles/heterozygous/homozygous nonreference alleles; Effect, position of nucleotide change on UCSC transcript and amino acid change; CLNDBN and CLNACC, selected disease condition and accession ID from ClinVar; OMIM, OMIM accession numbers for the genes.
### Table 2. Putative disease variants reported to cause neoplastic diseases.

<table>
<thead>
<tr>
<th>CHR</th>
<th>POS</th>
<th>ID</th>
<th>REF</th>
<th>ALT</th>
<th>AC</th>
<th>Gene</th>
<th>Effect</th>
<th>CLNDBN</th>
<th>CLNACC</th>
<th>OMIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17354297</td>
<td>rs33927012</td>
<td>A</td>
<td>G</td>
<td>43/1/0</td>
<td>SDHB</td>
<td>uc001 ba.2.c.T488C:p.S163P</td>
<td>Cowden-like syndrome</td>
<td>RCV000013633</td>
<td>185470</td>
</tr>
<tr>
<td>3</td>
<td>18255149</td>
<td>rs74313564</td>
<td>C</td>
<td>A</td>
<td>43/1/0</td>
<td>RNASEL</td>
<td>uc001 tgp.1.c.G794T:p.E265*</td>
<td>Prostate cancer, hereditary</td>
<td>RCV000013878</td>
<td>180435</td>
</tr>
<tr>
<td>4</td>
<td>11175240</td>
<td>rs1801166</td>
<td>G</td>
<td>C</td>
<td>43/1/0</td>
<td>APC</td>
<td>uc003 kpy.3.c.G3950C:p.E1317Q</td>
<td>Adenomatous polyposis coli</td>
<td>RCV00000872</td>
<td>611731</td>
</tr>
<tr>
<td>5</td>
<td>16012594</td>
<td>rs41341748</td>
<td>G</td>
<td>A</td>
<td>43/1/0</td>
<td>MSL1</td>
<td>uc003 wwu.2.c.G878T:p.R293*</td>
<td>Malignant tumor of prostate</td>
<td>RCV000015431</td>
<td>153622</td>
</tr>
<tr>
<td>6</td>
<td>43613908</td>
<td>rs77724903</td>
<td>A</td>
<td>T</td>
<td>42/2/0</td>
<td>RET</td>
<td>uc001 jal.2.c.A2373T:p.Y791 F</td>
<td>Familial medullary thyroid carcinoma</td>
<td>RCV000014962</td>
<td>164761</td>
</tr>
<tr>
<td>7</td>
<td>67258382</td>
<td>rs104894190</td>
<td>G</td>
<td>A</td>
<td>41/3/0</td>
<td>MP</td>
<td>uc001 lov.2.c.G912A.p.R304Q</td>
<td>Pituitary-dependent hypercortisolism</td>
<td>RCV000005171</td>
<td>605555</td>
</tr>
<tr>
<td>8</td>
<td>41226488</td>
<td>rs1800744</td>
<td>C</td>
<td>A</td>
<td>42/2/0</td>
<td>BRCA1</td>
<td>uc002 cct.2.c.G4599T:p.S1533I</td>
<td>Familial cancer of breast</td>
<td>RCV000048588</td>
<td>113705</td>
</tr>
</tbody>
</table>

CHR, chromosome; POS, position on chromosome based on GRCh37; ID, dbSNP ID; REF, reference sequence; ALT, alternative sequence; AC, counts of subjects with homozygous reference alleles/heterozygous/homozygous nonreference alleles; Effect, position of nucleotide change on UCSC transcript and amino acid change; CLNDBN and CLNACC, selected disease condition and accession ID from ClinVar; OMIM, OMIM accession numbers for the genes.

### Table 3. Putative disease variants reported to cause autosomal dominant cardiac diseases.

<table>
<thead>
<tr>
<th>CHR</th>
<th>POS</th>
<th>ID</th>
<th>REF</th>
<th>ALT</th>
<th>AC</th>
<th>Gene</th>
<th>Effect</th>
<th>CLNDBN</th>
<th>CLNACC</th>
<th>OMIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>236918491</td>
<td>rs19392635</td>
<td>C</td>
<td>T</td>
<td>43/1/0</td>
<td>ACTN2</td>
<td>uc001 hyf.2.c.C2148T:p.T716M</td>
<td>Cardiomyopathy</td>
<td>RCV000029298</td>
<td>102573</td>
</tr>
<tr>
<td>3</td>
<td>14180731</td>
<td>rs13449357</td>
<td>C</td>
<td>T</td>
<td>43/1/0</td>
<td>TMEM43</td>
<td>uc003 byk.2.c.G935T:p.R312W</td>
<td>Cardiomyopathy</td>
<td>RCV000030555</td>
<td>612048</td>
</tr>
<tr>
<td>4</td>
<td>114294537</td>
<td>rs45454496</td>
<td>G</td>
<td>A</td>
<td>42/2/0</td>
<td>ANK2</td>
<td>uc003 ide.3.c.G11792A.p.E3931K</td>
<td>Cardiac arrhythmia, ankyrin B-related</td>
<td>RCV000019677</td>
<td>106410</td>
</tr>
<tr>
<td>12</td>
<td>2659186</td>
<td>rs1219127775</td>
<td>G</td>
<td>A</td>
<td>43/1/0</td>
<td>CACNA1C</td>
<td>uc001 qkl.2.c.G1469A:p.G490R</td>
<td>Brugada syndrome 3</td>
<td>RCV000019201</td>
<td>114205</td>
</tr>
<tr>
<td>12</td>
<td>22017410</td>
<td>rs1668134</td>
<td>C</td>
<td>T</td>
<td>42/2/0</td>
<td>ABCC9</td>
<td>uc001 rfh.2.c.G2201 A:p.V734I</td>
<td>Myocardial infarction 1</td>
<td>RCV000029274</td>
<td>601439</td>
</tr>
<tr>
<td>12</td>
<td>111356964</td>
<td>rs10384636</td>
<td>C</td>
<td>T</td>
<td>43/1/0</td>
<td>MYL2</td>
<td>uc001 try.3.c.G388A:p.A13T</td>
<td>Familial hypertrophic cardiomyopathy 10</td>
<td>RCV000015108</td>
<td>160781</td>
</tr>
<tr>
<td>20</td>
<td>42744802</td>
<td>rs140740776</td>
<td>C</td>
<td>T</td>
<td>42/2/0</td>
<td>JPH2</td>
<td>uc002 xli.1.c.G1514A:p.G5055</td>
<td>Familial hypertrophic cardiomyopathy 17</td>
<td>RCV000023411</td>
<td>605267</td>
</tr>
<tr>
<td>21</td>
<td>35742938</td>
<td>rs74315447</td>
<td>T</td>
<td>C</td>
<td>43/1/0</td>
<td>KCNE2</td>
<td>uc002 ytt.1.c.T162C.p.M54T</td>
<td>Long QT syndrome 6</td>
<td>RCV000006425</td>
<td>603796</td>
</tr>
</tbody>
</table>

CHR, chromosome; POS, position on chromosome based on GRCh37; ID, dbSNP ID; REF, reference sequence; ALT, alternative sequence; AC, counts of subjects with homozygous reference alleles/heterozygous/homozygous nonreference alleles; Effect, position of nucleotide change on UCSC transcript and amino acid change; CLNDBN and CLNACC, selected disease condition and accession ID from ClinVar; OMIM, OMIM accession numbers for the genes.
<table>
<thead>
<tr>
<th>CHR</th>
<th>POS</th>
<th>ID</th>
<th>REF</th>
<th>ALT</th>
<th>AC</th>
<th>Gene</th>
<th>Effect</th>
<th>CLNDBN</th>
<th>CLNACC</th>
<th>OMIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>152285861</td>
<td>rs61816761</td>
<td>G</td>
<td>A</td>
<td>42/2/0</td>
<td>FLG</td>
<td>uc001 ezu.1:c.C1502T:p.R501* Ichthyosis vulgaris</td>
<td>RCV000017712</td>
<td>603415</td>
<td>135940</td>
</tr>
<tr>
<td>2</td>
<td>167141109</td>
<td>rs41268673</td>
<td>G</td>
<td>T</td>
<td>36/7/1</td>
<td>SCN9A</td>
<td>uc010flp.2:c.C1829A:p.P610T Primary erythromelalgia</td>
<td>RCV000004717</td>
<td>603415</td>
<td>606268</td>
</tr>
<tr>
<td>2</td>
<td>219755011</td>
<td>rs121908120</td>
<td>T</td>
<td>A</td>
<td>41/3/0</td>
<td>WNT10A</td>
<td>uc002ydi.1:c.T683A:p.F228I Odontoonychodermal dysplasia</td>
<td>RCV000004717</td>
<td>603415</td>
<td>606268</td>
</tr>
<tr>
<td>5</td>
<td>172662014</td>
<td>rs28936670</td>
<td>G</td>
<td>A</td>
<td>43/1/0</td>
<td>NKX2-5</td>
<td>uc003mcm.1:c.C74T:p.R25C Tetralogy of Fallot</td>
<td>RCV000009104</td>
<td>603415</td>
<td>606268</td>
</tr>
<tr>
<td>5</td>
<td>172662026</td>
<td>rs104893904</td>
<td>C</td>
<td>G</td>
<td>43/1/0</td>
<td>NKX2-5</td>
<td>uc003mcm.1:c.G62C:p.E21Q Tetralogy of Fallot</td>
<td>RCV00000974</td>
<td>603415</td>
<td>606268</td>
</tr>
<tr>
<td>6</td>
<td>32052313</td>
<td>rs121912575</td>
<td>C</td>
<td>T</td>
<td>42/2/0</td>
<td>TNXB</td>
<td>uc003nzl.2:c.G3323A:p.V1108M Ehlers–Danlos syndrome type 3</td>
<td>RCV000009083</td>
<td>603415</td>
<td>606268</td>
</tr>
<tr>
<td>8</td>
<td>11405576</td>
<td>rs5758736</td>
<td>G</td>
<td>A</td>
<td>42/2/0</td>
<td>BLK</td>
<td>uc003wty.2:c.G212A:p.A71T Maturity-onset diabetes of the young type 11</td>
<td>RCV000013112</td>
<td>603415</td>
<td>191305</td>
</tr>
<tr>
<td>8</td>
<td>18080001</td>
<td>rs4987076</td>
<td>G</td>
<td>A</td>
<td>39/5/0</td>
<td>NAT1</td>
<td>uc003wys.2:c.G632A:p.V2111 NAT1+7 ALLELE</td>
<td>RCV00019836</td>
<td>603415</td>
<td>108345</td>
</tr>
<tr>
<td>8</td>
<td>55372085</td>
<td>NA</td>
<td>T</td>
<td>A</td>
<td>42/2/0</td>
<td>SOX17</td>
<td>uc003xsb.3:c.T776A:p.Y259N Vesicoureteral reflux 3</td>
<td>RCV000001140</td>
<td>603415</td>
<td>610928</td>
</tr>
<tr>
<td>8</td>
<td>106431420</td>
<td>rs121908601</td>
<td>A</td>
<td>G</td>
<td>42/2/0</td>
<td>ZFPM2</td>
<td>uc003ymd.2:c.A90G:p.E30G Tetralogy of Fallot</td>
<td>RCV000006502</td>
<td>603415</td>
<td>603693</td>
</tr>
<tr>
<td>12</td>
<td>6442643</td>
<td>rs4149584</td>
<td>C</td>
<td>T</td>
<td>41/3/0</td>
<td>TNRFS1A</td>
<td>uc001qnu.2:c.G363A:p.R121Q TNF receptor-associated periodic fever syndrome (TRAPS)</td>
<td>RCV000013134</td>
<td>603415</td>
<td>191910</td>
</tr>
<tr>
<td>13</td>
<td>33251535</td>
<td>rs121918303</td>
<td>A</td>
<td>C</td>
<td>38/6/0</td>
<td>RXFP2</td>
<td>uc001 utt.2:c.A665C:p.T222P Cryptorchidism, unilateral or bilateral</td>
<td>RCV000003476</td>
<td>606655</td>
<td>606655</td>
</tr>
<tr>
<td>14</td>
<td>54418579</td>
<td>NA</td>
<td>T</td>
<td>C</td>
<td>43/1/0</td>
<td>BMP4</td>
<td>uc010aoh.2:c.A363G:p.H212R Microphthalmia syndromic 6</td>
<td>RCV000002458</td>
<td>606655</td>
<td>112262</td>
</tr>
<tr>
<td>15</td>
<td>74704267</td>
<td>rs56001514</td>
<td>G</td>
<td>A</td>
<td>43/1/0</td>
<td>SEMA7A</td>
<td>uc002avx.2:c.C1382T:p.R461C Blood group John Milton Hagen system</td>
<td>RCV000002925</td>
<td>606655</td>
<td>607961</td>
</tr>
<tr>
<td>16</td>
<td>1129586</td>
<td>rs121917877</td>
<td>C</td>
<td>T</td>
<td>42/2/0</td>
<td>SSTR5</td>
<td>uc002ckq.2:c.C719T:p.R240W Resistance to somatostatin analog</td>
<td>RCV000013734</td>
<td>606655</td>
<td>182455</td>
</tr>
<tr>
<td>18</td>
<td>580309060</td>
<td>rs121913563</td>
<td>C</td>
<td>T</td>
<td>43/1/0</td>
<td>MC4R</td>
<td>uc002lie.1:c.G524A:p.A175T Obesity</td>
<td>RCV000015406</td>
<td>606655</td>
<td>155541</td>
</tr>
<tr>
<td>19</td>
<td>11240278</td>
<td>rs137853964</td>
<td>G</td>
<td>A</td>
<td>43/1/0</td>
<td>LDLR</td>
<td>uc002moh.3:c.G2480A:p.V827I Familial hypercholesterolemia</td>
<td>RCV000003135</td>
<td>606655</td>
<td>609945</td>
</tr>
</tbody>
</table>

CHR, chromosome; POS, position on chromosome based on GRCh37; ID, dbSNP ID; REF, reference sequence; ALT, alternative sequence; AC, counts of subjects with homozygous reference alleles/heterozygous/homozygous nonreference alleles; Effect, position of nucleotide change on UCSC transcript and amino acid change; CLNDBN and CLNACC, selected disease condition and accession ID from ClinVar; OMIM, OMIM accession numbers for the genes.
Given the important role of APOE variants in late-onset Alzheimer’s disease, we first looked at the APOE e4 allele as defined by the ancestral alleles rs429358-C and rs7412-C (130Arg and 176Arg) and APOE e2 allele as defined by rs429358-T and rs7412-T (130Cys and 176Cys). In our centenarian sample, the allele frequencies for APOE e4, e3, and e2 were 6.8%, 80.7%, and 12.5%, respectively. One of the two e2/e4 and one of the two e3/e4 heterozygous carriers were found to have advanced dementia of unknown etiology and the other two subjects were cognitively intact. Notably, the centenarian homozygous for the e4 allele had an MMSE score of 25 at age 97. Neither this centenarian nor any other subject in our study carried the protective variant A673T in the APP gene (Jonsson et al. 2012). We also observed one carrier for each of the known GBA mutations L483P (also known as L444P) (rs421016) and N409S (N370S) (rs76763715), which is in line with the reported frequency of 3% for either mutation in AJ controls (Sidransky et al. 2009).

On the X chromosome we identified a female centenarian with a previously described mutation in UBQLN2 P525S (Deng et al. 2011), which was found to cause an X-linked dominant type of familial amyotrophic lateral sclerosis (ALS) and ALS/dementia with an estimated penetrance of approximately 90% (Table 1). This subject had normal MMSE (score = 28) at age 102 without any neurological symptoms.

Furthermore, disease alleles were observed for other autosomal dominantly inherited degenerative diseases of sensory function, including retinitis pigmentosa (SEMA4A, RP1), deafness (MYO1A), glaucoma (CYP1B1, OPTN, WDR36), and keratoconus (VSX1). Of note, none of the 10 centenarians carrying the SEMA4A R713Q had vision impairment but the single subject carrying the RP1 T373I was blind. No hearing impairment was noticed in one of two subjects homozygous and four of ten heterozygous carriers of the MYO1A G662E variant that was linked to an autosomal dominant form of deafness.

### Variants for neoplastic diseases

We found variants in the five genes APC, BRCA1, RET, RNASEL, and STK11 that were linked to autosomal dominant forms of cancer or neoplasm as well as four complex risk variants in ELAC2, MSR1, AIP, and SDHB (Table 2). The clinical relevance of these variants has been discussed in the literature and their presence in the genomes of centenarians indicates that these variants are compatible with exceptional longevity.

The tumor suppressor gene STK11 variant F355L (rs59912467) linked to Peutz–Jeghers syndrome was seen in one centenarian. It was reported to affect both the AMPK pathway and cell polarity, thus contributing to the
development of malignancies (Forcet et al. 2005). Notably, the same subject carried two additional variants linked to thyroid carcinoma (RET Y791F, rs77724903) and Cowden-like syndrome (SDHB S163P, rs33927012), but was free of any type of neoplasia at age 97. The Y791F variant in the RET protooncogene was further present in another centenarian with a history of cancer of unknown origin. This variant was first described in two German families with familial medullary thyroid carcinoma (Berndt et al. 1998). Six centenarians carried the A541T variant in the gene ELAC2 (rs5030739) that was reported to be associated with prostate cancer (Rebeck et al. 2000; Tavtigian et al. 2001; Camp and Tavtigian 2002). One of the two male carriers for A541T had no cancer at age 103 years and the other had an unknown type of cancer.

Variants for autosomal dominant forms of cardiac disease

We found variants in ABCC9, ACTN2, ANK2, CACNA1C, JPH2, KCNE2, MYL2, and TMEM43 that were linked to autosomal dominant phenotypes affecting cardiac function (Table 3).

Three of the variants in these genes are annotated as causes of cardiac arrhythmia with increased risk of sudden cardiac death. Two centenarians were heterozygous for the E3931K variant in ANK2 (rs104894363) variant, also known as E1813K, which was reported as a loss-of-function mutation in the ankyrin-B regulatory domain (Mohler et al. 2003, 2004). One centenarian carried the G490R variant in CACNA1C (rs121912775), a loss-of-function change that was linked to Brugada syndrome 3 (Antzelevitch et al. 2007). Another carried the M54T (rs74315447) in KCNE2, which was reported to alter the transmembrane domain of MirP1 that reduces potassium currents leading to long QT syndrome and ventricular fibrillation (Abbott et al. 1999; Splawski et al. 2000).

Five variants linked to cardiomyopathy were found in our centenarian sample. The MYL2 A13T (rs104894363) variant located in the regulatory light chain of myosin was reported to be causative for a subtype of familial hypertrophic cardiomyopathy with onset of clinical symptoms around middle age (Poetter et al. 1996; Andersen et al. 2001). The gene ABCC9 encodes the regulatory SUT2A subunit of the cardiac ATP-sensitive potassium (K\textsubscript{ATP}) channel. Two centenarians carried the ABCC9 V734I variant (rs61688134), which was reported to increase the risk of MI by 6.40-fold. (Minoretti et al. 2006). The G505S (rs140740776) variant in JPH2 was reported to be associated with hypertrophic cardiomyopathy in a relatively small sample of Japanese patients (Matsushita et al. 2007). One centenarian carried the TMEM43 R312W variant linked to autosomal dominant arrhythmogenic right ventricular cardiomyopathy/dysplasia (Haywood et al. 2013). Interestingly, one centenarian was found to be heterozygous for both JPH2 G505S and ABCC9 V734I and another was heterozygous for both JPH2 G505S and ANK2 E3931K.

Variants for other autosomal dominant and X-chromosomal diseases

We found 18 rare ClinVar variants for other autosomal dominant diseases (Table 4) and six mutations for X-chromosomal diseases (Table 5). These include variants that increase the risk of metabolic disorders, including hypercholesterolemia (LDLR, V827I, rs137853964), maturity-onset diabetes of the young (MODY) (BLK, A71T, rs55758736), and obesity (MC4R, A175T, rs121913563). One male and one female centenarian carried the V1108M variant in TNXB (rs121912575) linked to the dominant form of Ehlers–Danlos syndrome type 3 (Zweers et al. 2005). The remaining autosomal mutations are annotated to cause a wide range of dominantly inherited phenotypes including metabolic, genitourinary, and skin conditions as well as pediatric conditions, including developmental syndromes, for example, the tetralogy of Fallot, a severe malformation that would reduce life span. Three centenarians presented a variant in TNFRSF1A (R121Q, rs4149584) in the heterozygous state that was linked to TNF-receptor associated periodic fever. This was the only variant for dominantly inherited immune disease in this dataset.

Aside from the above mentioned mutation in UBQLN2, the observed X-chromosomal variants include three variants in the DMD gene that have been discussed as cause of Duchenne and Becker muscular dystrophy. One frequent missense variant rs1800279 (H2921R) in the DMD gene was observed in one male centenarian and another male centenarian carried both rare variants rs1800278 (N2912D) and rs41305353 (E2910V). The carriers did not have any documented muscle diseases.

Variants for recessive diseases and complex diseases

Among the 72 variants for recessive traits, we found three variants that were observed in the homozygous state in at least one centenarian (Table S2). These are variants in ADA, ALG6, and HPS5. The ALG6 variant Y131H (rs55383149) has been annotated to cause congenital disorder of glycosylation, type Ic, a childhood onset metabolic disorder accompanied by severe neurological symptoms (Miller et al. 2011). We found no literature evidence supporting the pathogenicity of the ADA variant.
D8N (rs73598374) or the HPS5 variant T1098I (rs61884288).

Three subjects were heterozygous carriers for Factor V Leiden (F5, Q534R, rs6025) (Bertina et al. 1994). Heterozygous carriers have elevated risks of deep venous thrombosis, pulmonary embolism (Juul et al. 2004), and stroke (Casas et al. 2004). Note that the GRCh37 genome assembly presents the risk allele as the reference allele and no individuals homozygous for the reference allele were seen in our sample. One carrier had a history of MI and another had a history of both MI and stroke, but the third carrier had no known thrombotic diseases by the age of 106 years.

Finally, we also identified homozygous individuals for putative recessive disease alleles with relevance for longevity regardless of their population allele frequency. Of note, we found three missense variants in the ACADS gene R107C (rs61732144), R171W (rs1800556), and G209S (rs1799958) (Tables S1 and S2) that were initially reported (Corydon et al. 2001; Pedersen et al. 2008) in 10 patients with ethylmalonic aciduria and short-chain acyl-CoA dehydrogenase (SCAD) deficiency, a mitochondrial fatty acid oxidation disorder causing neuromuscular phenotypes with hypotonia and developmental delay as the prominent features of the disease. Homozygosity for R107C as well as compound heterozygosity for R107C and G209S have been described as disease causing in AJ patients with SCAD with reduced penetrance (Tein et al. 2008) depending on other genetic modifiers or environmental stressors. Variants R171W and G209S have both high allele frequencies (5.5% and 23.5%, respectively) in the general population (U.S. and the Netherlands) and they are considered to confer susceptibility for clinical disease (van Maldegem et al. 2006). We found five AJ centenarians who are homozygous for the variant G209S (MAF = 36.4%) and they do not carry any other missense variants in the ACADS gene. The carrier frequency for the R107C variant in our centenarians (3/44) is identical to the reported carrier frequency in AJ population (Tein et al. 2008).

**Discussion**

In this study we observed many previously reported Mendelian mutations that are sufficiently benign to allow the individual carriers to achieve exceptional longevity. The presence of these specific variants in the genomes of centenarians can be helpful for clinical geneticists who are challenged with the evaluation of their putative pathogenicity as incidental findings. More generally, our findings support the notion that for many putative disease variants it is not straightforward to decide whether they should be regarded and acted upon as incidental findings, when they are observed in healthy individuals (Kingsmore 2013). Several genes harboring putatively pathogenic variants in our centenarians are on the list of 57 genes with reportable findings according to the ACMG recommendations (Green et al. 2013b). These genes comprise 4 of the 24 genes for cancer (BRCA1, APC, RET, STK11), 1 of the 20 cardiac disease genes (TMEM43), 1 of the remaining 13 genes (LDLR). This demonstrates the challenge to identify actionable mutations even in well-established disease genes. The variants found in four of the cancer genes APC, BRCA1, RET, and SDHB in our AJ centenarian sample had also been labeled as “almost certainly benign” in another sequencing study (ClinSeq) of 572 middle aged participants (17% of which are of AJ ancestry) (Johnston et al. 2012). Surprisingly, we also observed many rare variants linked to dominantly inherited forms of diseases that would clearly affect life expectancy. These disease traits include cardiac diseases such as cardiomyopathy and arrhythmia that increase the risk of sudden cardiac death as well as metabolic diseases such as diabetes, hypercholesterolemia, and obesity. One explanation for the presence of disease alleles in centenarians might be that these variants have incomplete penetrance as a result of complex interplay between modifying genetic and environmental factors (Bergman et al. 2007; Cooper et al. 2013). However, for some variants the published evidence may be viewed as too weak to uphold their classification as pathogenic. For example, the three missense variants in the DMD gene were all found in at least one male centenarian. Given the markedly reduced life expectancy of patients affected by Duchenne muscular dystrophy (Kieny et al. 2013), it is unlikely that these variants are pathogenic. In other cases, the resulting phenotypes of mutations could have little impact on longevity, which may primarily apply to diseases such as deafness, glaucoma, retinitis pigmentosa, premature ovarian failure, or ichthyosis. In other instances the possible influence on life expectancy is not clear, such as the variant V1108M in the TNXB gene, which was reported to cause Ehlers–Danlos syndrome type 3 (Zweers et al. 2005) and shown to affect the protein function (Zhuang et al. 2010).

We note that many presumable disease mutations are present in centenarians with fairly high frequency. We did not discuss these variants in detail because they are common in the ESP data. With the exception of the APOE ε4 allele, these common variants are from the beginning more likely to constitute polymorphisms than any clinically significant mutations. Despite its high population frequency, the APOE ε4 allele was suggested to follow semi-dominant inheritance (Genin et al. 2011). APOE is also the only reported gene reaching GWAS significance threshold for longevity (Beekman et al. 2013). In our centenarian sample, the APOE ε4 allele frequency is very
similar to the previously reported allele frequencies in 325 French Caucasian centenarians (Schachter et al. 1994), namely 5.2% for APOE e4 and 12.8% for APOE e2. In the ESP dataset, the respective allele frequencies for e4 and e2 are 11.7% and 5.6% in European Americans and 18.9% and 8.7% in African Americans. This is congruent with the previous finding that e4 was associated with reduced e2 with increased life span (Deelen et al. 2011). However, earlier studies did not observe any homozygote centenarians for APOE e4. As would be expected, the nonaffected APOE e4 homozygote had rare coding variants in APOE modifier candidate genes, but testing their significance needs to be left to subsequent studies. The future identification of such protective genetic factors in centenarians may increase the clinical utility of the APOE e4 allele. Due to the relatively small sample size and lack of a control data set from a matched population, we cannot search for novel variants with impact on longevity here.

In future, personalized genomic sequencing is likely to generate many incidental findings. Rigorous investigations for cosegregation in family members, large ancestry-matched population screening, and functional studies will be required to identify those variants that are sufficiently penetrant to be considered clinically relevant. Genetic counseling provided by trained clinicians will be necessary to communicate the potential impact of genetic variants to avoid unnecessary and far reaching burden. This is especially true in considering genetic screening in the young, where the demarcation of putatively actionable mutations needs to be approached with considerable caution, allowing appropriate balance to be conveyed in risk/benefit discussions with parents. Our study provides a list of variants that are currently flagged as pathogenic, but are compatible with exceptional longevity. Thereby, our study contributes to the genetic community’s efforts to refine the set of variants with strong clinical significance and mark previously reported disease mutations with unclear clinical significance.

**Acknowledgment**

We thank the Mildred and Frank Feinberg Family Foundation for their dedicated support for this study. We thank the centenarians and their families as well as the members of the Longevity Genes Project at the Albert Einstein College of Medicine for their contribution to this study. J. F. is supported by NIAMS/NIH under Award Number R03 AR063340. N. B., G. A., and D. B. A. are supported by NIH/NIA grants R01 AG618381, R01 AG042188, R01 AG046949, and P01 AG021654, the Einstein Nathan Shock Center (P30 AG038072), and the Einstein Glenn Center for the Biology of Human Aging.

**Conflict of Interest**

None declared.

**References**


Cassa, C. A., M. Y. Tong, and D. M. Jordan. 2013. Large numbers of genetic variants considered to be pathogenic are common in asymptomatic individuals. Hum. Mutat. 34:1216–1220.


© 2014 The Authors. Molecular Genetics & Genomic Medicine published by Wiley Periodicals, Inc.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Putative disease alleles found in AJ Centenarians that have minor allele frequency $\geq$5% in either Europeans or African Americans from the Exome Sequencing Project (ESP).
Table S2. Putative disease variants reported to cause autosomal recessive diseases and diseases with other mode of inheritance.