Evidence that the gene encoding insulin degrading enzyme influences human lifespan

Mun-Gwan Hong, Chandra Reynolds, Margaret Gatz, Boo Johansson, Jennifer C. Palmer, Harvest F. Gu, Kaj Blennow, Patrick G. Kehoe, Ulf de Faire, Nancy L. Pedersen and Jonathan A. Prince

Studies in model organisms have demonstrated that components of insulin and insulin-like signaling pathways are involved in the regulation of lifespan but the relevance of those findings to humans has remained obscure. Here we provide evidence suggesting that variants of the gene encoding insulin-degrading enzyme (IDE) may be influencing human lifespan. We have employed a variety of models and diverse samples that reproducibly indicate the relative change in IDE genotype frequency across the age spectrum as well as allow the detection of association with age-at-death. A tenable molecular basis of this is suggested by the observation of genetic association with both fasting plasma insulin levels and IDE mRNA expression. Across populations the emergent genetic model is indicative of over-dominance, where heterozygotes of critical markers have increased IDE mRNA expression and insulin levels, and this is reflected in diminished heterozygosity at advanced age. A critical and replicating feature of this study is that change in IDE genotype frequency with advancing age appears to be occurring only in men, and this is supported in that insulin levels are only associated with IDE in men. Results suggest a relationship between a gene that is intimately involved in insulin metabolism and the determination of lifespan in humans, but over-dominance and gender specificity will be important parameters to consider clarifying the biological importance of these findings.

INTRODUCTION

Human lifespan is determined to a large degree by the cumulative effects of age-related pathologies such as cardiovascular disease, cancer, neurodegenerative disorders and diabetes, all of which are influenced by both environmental and genetic factors. To date, the only widely reproduced example of a lifespan-modifying gene is that encoding apolipoprotein E (APOE) (1,2), likely deriving from its association with both coronary heart disease (CHD) (3) and Alzheimer disease (AD) (4). The study of APOE is illustrative of a standard approach to the detection of ‘longevity’ genes, specifically a comparison of elderly individuals versus younger populations. There are variations on this theme (5), but on a fundamental level the initial evidence of association will reflect genotype or allele frequencies of a particular gene that are changing with age at a population level. The search for genes that exhibit this phenomenon may provide an indication of biological processes that contribute to a decline in health during aging. The association of APOE, for example, highlights the involvement of lipoprotein metabolism in lifespan, and this is reinforced by genetic evidence that APOC3 may also...
associate with longevity phenotypes (6) as well as studies on lipoprotein particle size across age groups (7).

Insulin metabolism is one of the biological pathways that has dominated longevity research in recent years (8,9). This stems primarily from evidence that insulin signaling is central to lifespan regulation in experimental organisms (10–14). Specifically, mutation in daf-2, which encodes an insulin/insulin-like growth factor receptor, affects survival in C. elegans (11) and in Drosophila, and mutation in the insulin-like receptor Inr has been shown to increase lifespan (12). In mammals, extended lifespan has been more recently seen in the fat-specific insulin receptor knockout mouse (13) and in heterozygous Igf1r knockout mice (14). Genetic association studies of insulin-related genes and human longevity are sparse but have been conducted, although the evidence is still equivocal (15).

In humans, the gene encoding insulin-degrading enzyme (IDE) is one of the central regulators of insulin metabolism, but also plays an integral role in the degradation of a variety of proteins that includes insulin-like growth factor-2 (IGF2) and β-amyloid (16,17). Genetic association of IDE has been indicated with several heritable quantitative traits and diseases (18–22). This includes evidence of association with AD (18,19) and type II diabetes (21). For the latter in particular, despite modest initial evidence, widespread replication has been obtained by whole genome association studies, which all highlight a region that includes the IDE gene (23–26). The capacity of IDE to influence many phenotypes (pleiotropy), while likely reflecting its broad substrate specificity, is reminiscent of APOE. These observations, together with prior evidence of association of IDE with plasma insulin levels (20), prompted us to investigate if genetic sequence variation in IDE might detectably impact lifespan in human populations.

RESULTS

With haplotype and linkage disequilibrium (LD) mapping efforts as well as previous association studies (20–22) as a guide, a set of five genetic markers was selected for initial genotyping in the present study (see Materials and Methods). As IDE previously appeared to impact plasma insulin levels only in men (20), gender was regarded as an important covariate for all analyses. To begin, this five marker set was tested in four populations representing control samples for diverse diseases (Table 1; samples 1–4). For this analysis, the dependence of genotype on age was modeled using multinomial logistic regression, which has been presented as an intuitive means to detect longevity genes (5) (see Materials and Methods). In the combined analyses of all four samples, this provided robust evidence of association for rs1887922 and rs2251101 in men ($\chi^2(2) = 30.6, P = 0.00000022$ and $\chi^2(2) = 19.8, P = 0.000051$, respectively), with no evidence for any marker in women. To illustrate what is giving rise to this significance level, mean ages-at-sampling according to rs1887922 genotype in men and women for these four samples are shown in Figure 1 (rs2251101 exhibited a similar pattern and is not shown). In this figure, the emergent genetic model suggests that heterozygotes are in some way disadvantaged, and this can be seen by their narrowly lower average sampling ages in men. The central theme, in our view, is that this phenomenon is replicating across samples. We note that ANOVA models provide equivalent evidence for rs1887922 and rs2251101 in men ($F_{2,874} = 15.5, P = 0.00000024$ and $F_{2,838} = 10.0, P = 0.000052$, respectively). The genotype by gender interaction term in the whole sample was highly significant for rs1887922 ($F_{2,1772} = 14.9, P = 0.00000038$), confirming that it was an appropriate covariate to stratify. An expanded set of 32 genetic markers (see Materials and Methods) around IDE was tested against age distributions in sample 4, but only rs2251101 was significant in men ($F_{2,76} = 3.5, P = 0.035$). As an essential validation of this approach, marker rs429358 in APOE (which distinguishes e4-positive from e4-negative individuals) was tested against age-at-sampling distributions using ANOVA in samples 1 and 4, and this was significant in the combined samples ($F_{2,769} = 6.0, P = 0.0025$). The rs429358 genotype by gender interaction term was not significant in this analysis.

The above strategy provided initial evidence that IDE genotype frequencies may vary across the age spectrum. As a second complementary approach, the five IDE marker set (see Materials and Methods) was tested in two additional population-based samples to test directly for an impact upon age-at-death (AAD) (Table 1; samples 5–6). AAD was obtained through population registries and modeled as a function of genotype using survival analyses. A critical feature of these samples was that they lacked any disease-related phenotypic selection criteria. Marker rs429358 in APOE was included as a covariate in all analyses. Significant associations were observed for marker rs2251101 in men and in women. In men, a significant interaction between rs2251101 genotype and age-at-sampling was observed for longevity where those with an earlier age-at-sampling and who were heterozygous were at a lower risk of mortality, whereas heterozygotes with a later age-at-sampling showed a greater risk of earlier mortality relative to those with homozygous genotypes. In women, significant additive and dominance effects were observed irrespective of age-at-sampling with homozygous rare individuals showing the greatest risk of mortality and with statistically indistinguishable survival curves between heterozygous and homozygous common individuals. Kaplan–Meier survival plots for rs2251101 for two ages-at-sampling are shown in Figure 2. This arbitrary division point was selected for the

### Table 1. Brief descriptions of human samples

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample size (M/F)</th>
<th>Origin</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>601 (290/311)</td>
<td>Swedish</td>
<td>AD-free controls</td>
</tr>
<tr>
<td>Sample 2</td>
<td>321 (116/205)</td>
<td>Australian</td>
<td>PD-free controls</td>
</tr>
<tr>
<td>Sample 3</td>
<td>724 (411/313)</td>
<td>Swedish</td>
<td>Non-diabetic controls</td>
</tr>
<tr>
<td>Sample 4</td>
<td>178 (77/101)</td>
<td>Swedish</td>
<td>AD-free controls</td>
</tr>
<tr>
<td>Sample 5</td>
<td>539 (181/358)</td>
<td>Swedish</td>
<td>Random population</td>
</tr>
<tr>
<td>Sample 6</td>
<td>590 (249/341)</td>
<td>Swedish</td>
<td>Random population</td>
</tr>
<tr>
<td>Sample 7</td>
<td>2703 (1862/841)</td>
<td>Swedish</td>
<td>MI case–control</td>
</tr>
<tr>
<td>Sample 8</td>
<td>40 (14/26)</td>
<td>Swedish</td>
<td>Sequencing</td>
</tr>
<tr>
<td>Sample 9</td>
<td>178 (83/95)</td>
<td>English</td>
<td>DNA/RNA brain</td>
</tr>
</tbody>
</table>

M, male; F, female; AD, Alzheimer disease; MI, myocardial infarction; PD, Parkinsons disease.

Sample size represents the total sample for which genotyping was performed. See Supplementary Text S1 for detailed descriptions of samples.
purpose of presentation only, since there was some evidence that an interaction with age-at-sampling might be present. No significant effects of rs1887922 on longevity were observed for men or women in combined analyses of samples 5 and 6. Separate survival analyses in samples 5 and 6 are provided as Supplementary Text S2.

As an alternative to survival analysis and in further support of results from samples 1–4, AAD distributions across the genotype classes of rs2251101 were found to differ significantly in men from sample 5 ($F_{2,163} = 7.6$, $P = 0.0007$), where heterozygous males exhibited a lower AAD compared with opposite homozygous individuals (Fig. 3). We noted that AAD and age-at-sampling were strongly correlated in this sample ($r^2 = 0.37$). As was seen for samples 1–4, age-at-sampling distributions across the genotype classes for rs2251101 and rs1887922 were found to differ significantly in men from sample 5 ($F_{2,174} = 9.5$, $P = 0.0001$ and $F_{2,177} = 5.0$, $P = 0.0081$, respectively), the pattern being similar to what was seen for AAD (data not shown). There was no evidence of this in women or men from sample 6.

Since the most commonly applied approach for detecting potential longevity genes involves the comparison of genotype frequencies between groups with large differences in their age distributions (6), we also considered it important to apply a similar strategy here. For this, we elected to simply contrast our most extreme samples in terms of age, these being samples 3 and 5, where the difference in mean ages is $>30$ years with no overlap. We provide the results of this analysis in Table 2, where differences are again noted for men for marker rs1887922, further supporting the relative change of IDE genotypes with age, in a manner consistent with observations in previous models. The pattern for rs2251101 is similar, but did not attain significance (Table 2).

To further complement the above analyses, survival at follow-up in a case–control study was modeled as a function of IDE genotype. For this, a CHD sample (sample 7) was obtained, originally collected to evaluate risk factors for incident myocardial infarction (MI). An analysis of men and women was performed examining the independent effects of age together with IDE genotypes for either rs1887922 or rs2251101. Both additive (linear) and heterozygosity genetic models were assessed. Significance for both markers was noted in men, although after stratification this was evident only in cases (Table 3). The best model fit was obtained in which heterozygotes exhibited increased mortality. For male cases, the linear models for rs1887922 and rs2251101 were both significant (OR 1.56, 95% CI 1.13–2.16, $P = 0.0064$ for rs1887922 and OR 1.39, 95% CI 1.06–1.84, $P = 0.018$ for rs2251101). Neither rs1887922 or rs2251101 alone, nor haplotypes constructed with previously defined tag markers (18) were associated with incident MI risk in men or women (Supplementary Material, Table S1).

Towards affirming that IDE, and not a nearby gene, is giving rise to the above findings, its central role in insulin metabolism was considered (16). Our reasoning was that if functional sequence variation exists that directly impacts IDE, detecting association with metrics of one of its primary substrates should provide evidence to this effect. To address this question, men from sample 3 were revisited after 10 years and new fasting insulin measurements were obtained.
against which rs1887922 and rs2251101 were tested [there was no evidence of association for any IDE marker with insulin levels in women in the original report (20) and they were consequently not followed]. Although the sample size was diminished, the evidence of association for rs2251101 is stronger than originally reported (20) and this is shown in Figure 4. Importantly, the proportion of variance explained by IDE in the 10-year follow-up sample exceeds the original measure (6.4 versus 3.0%), suggesting that the influence of IDE on insulin levels may be increasing with age.

Towards understanding the molecular basis of these findings, we investigated if functional polymorphism in IDE might be regulatory, since previous sequencing efforts did not reveal common non-synonymous variation in coding regions (27). Two principal isoforms of IDE, generated with either of the alternative exons 15a or 15b, have been shown to exist in brain with different catalytic activities (28). Their presence in various brain regions was confirmed (data not shown). However, we also confirmed the existence of a spliceform containing both exons 15a and 15b (we term this 15ab), not previously reported in brain (GenBank accession no. BC036188). The five IDE marker set was genotyped in an independent brain sample in which both genomic DNA and mRNA were obtained (Table 1; sample 9), and the levels of the three IDE spliceforms were selectively measured and used in quantitative trait analyses. The results of this are shown in Figure 5, where over-dominance for the 15a and 15b spliceforms is again evident for marker rs2251101. The 15ab spliceform was not significantly affected and there was no evidence for the effects of the other IDE markers (data not shown). At this fundamental level, a genetic effect appears to be present in both men and women and this was supported in second-order factorial ANOVA models, where the genotype by sex interaction term was not significant for either the 15a or 15b spliceforms ($P = 0.33$ and $P = 0.55$, respectively). Complete sequencing of the region (including introns) around the 15a and 15b exons was also conducted in 40 Swedish individuals (sample 8), primarily since the 15b exon had not been previously examined (27). No previously unreported polymorphisms were discovered.

Since neither rs1887922 nor rs2251101 have obvious functional effects on gene regulation, we considered it important to examine other markers that might be in strong LD for functional candidature. We searched for SNPs across a 1 Mb region around IDE in the CEU population using HapMap database and Haploview v3.2. This region was represented by 901 SNPs from rs11186799 to rs3736937 (999038 bp). Among these, 557 SNPs from rs833372 to rs3736936 (998294 bp) had a minor allele frequency of more than 0.1%. There were nine SNPs with an LD of $r^2 = 0.5$ or more with either of rs1887922 or rs2251101. Only one additional SNP emerged when the threshold was reduced to $r^2 = 0.3$ (0.35 for

Figure 2. Survival analysis Kaplan–Meier survival plots for IDE marker rs2251101 in samples 5 and 6 in Swedish men (A and B) and women (C and D) for first age-at-sampling of 75 years or younger (A and C) and 76 years or older (B and D): black, rare-allele homozygotes; dotted, common-allele homozygotes; gray, heterozygotes. For men, Cox regression analyses suggested that the dominance contrast was at trend significance [$\chi^2 (1) = 2.86, P = 0.09]$ while the interaction between age-at-sampling and the dominance contrast was significant [$\chi^2 (1) = 8.53, P = 0.0035$]. Heterozygous males with an earlier first age-at-sampling showed a delayed mortality risk while heterozygous males with a later age-at-sampling showed the greatest risk of earlier mortality relative to homozygous rare or common males (A versus B). For women, the additive and dominance contrasts were significant [$\chi^2 (1) = 9.51, P = 0.0020$ and $\chi^2 (1) = 4.72, P = 0.030$, respectively] while the interaction between age-at-sampling and the additive contrast was at trend significance [$\chi^2 (1) = 3.56, P = 0.059$]. Homozygous rare females were at greatest risk of earlier mortality while heterozygous and homozygous common females were similar (C and D); the homozygous rare effect may lessen at later ages-at-sampling (C versus D). The main effect of first age-at-sampling was significant in all models fit for men and women (all $P < 0.0001$). A main effect of APOE e4 in women was observed [$\chi^2 (1) = 12.23, P = 0.00050$] but not for men.
There were no perfect proxies ($r^2 = 1$) for rs2251101. These markers consisted of (from 5' to 3') of rs551266, rs7093418, rs1855916, rs7084090, rs7898493, rs11187061, rs7908111, rs11187074 and rs7100623, all residing in introns and all residing within IDE. Each of these markers including rs1887922 and rs2251101 were assessed for occurrence in regions with (i) high conservation across species, (ii) presence at an intron-exon boundary, (iii) evidence of affecting transcription factor binding or (iv) evidence of affecting a potential poly(A) site. Across markers, there was no evidence for any such effect.

### DISCUSSION

We have attempted to present a continuum of evidence that stems from the initial observation that IDE genotype frequencies may be varying with age, to association with AAD, association with an intermediate trait (plasma insulin) and ultimately to a potentially primary effect upon IDE mRNA levels. While we regard this as a relatively strong case implicating IDE as a modifier of human lifespan, there are a number of central issues that deserve discussion.

First, it was considered vital to employ multiple models and different samples for replication. This was particularly important given that age-at-sampling used in samples 1–4 is not an ideal surrogate for lifespan, and we acknowledge this as a weakness of this study. We highlight nonetheless that there is equivalent replication across those samples, and the validity of the approach suggests that age-at-sampling distributions also vary depending upon the genotypes of APOE, which is at present the only widely recognized gene that can affect AAD (1,2).

Contrasting our most extreme samples in terms of age-at-sampling provided an independent replication of what was observed in samples 1–4. The observation of association with AAD in an equivalent manner in sample 5 also suggests that heterozygotes are being lost more rapidly with age from the population and provides yet another form of independent replication. While there was no evidence of this in sample 6, we note that samples 5 and 6 have very different age distributions, with sample 5 being the oldest. Thus, if IDE does impact AAD, it may have a more pronounced effect at later ages. Because of this we were cautious about providing any evaluation of effect size or power estimate for replication efforts, since different age ranges appear to be important. The different approaches used here have strengths and weaknesses, although we consider them complementary. Issues such as different immigration patterns between age sub-sets within a sample could confound any of the above models. This might have a larger effect on the age-at-sampling analysis since the age ranges are quite broad. In contrast, AAD in sample 5 has a very narrow range, and might thus be less susceptible to this. We regard this nevertheless as a difficult balance between, on the one hand, obtaining samples from different age ranges and, on the other, accounting for potential demographic differences.

Secondly, the prevailing genetic model that has emerged in this study is indicative of over-dominance, which can arise due to true over-dominance at a single polymorphic site, or pseudo-overdominance due to dominance effects at multiple sites (29). The key markers exhibiting association, rs2251101 and rs1887922 are in very strong LD ($r^2 = 0.5$).

---

**Table 2.** Genotype frequencies for markers rs1887922 and rs2251101 in old (sample 5) and younger (sample 3) individuals

<table>
<thead>
<tr>
<th>rsID/sample</th>
<th>Group</th>
<th>Genotypes</th>
<th>CC</th>
<th>CT</th>
<th>TT</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1887922</td>
<td>Men</td>
<td>Old</td>
<td>4  (0.02)</td>
<td>33 (0.18)</td>
<td>142 (0.79)</td>
<td>0.0052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Younger</td>
<td>11 (0.03)</td>
<td>127 (0.31)</td>
<td>270 (0.66)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>Old</td>
<td>8  (0.02)</td>
<td>89 (0.25)</td>
<td>259 (0.72)</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Younger</td>
<td>6  (0.02)</td>
<td>81 (0.26)</td>
<td>226 (0.72)</td>
<td></td>
</tr>
<tr>
<td>rs2251101</td>
<td>Men</td>
<td>Old</td>
<td>110 (0.63)</td>
<td>61 (0.34)</td>
<td>5  (0.03)</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Younger</td>
<td>207 (0.55)</td>
<td>141 (0.38)</td>
<td>22 (0.06)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>Old</td>
<td>214 (0.61)</td>
<td>112 (0.32)</td>
<td>24 (0.07)</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Younger</td>
<td>183 (0.62)</td>
<td>96 (0.33)</td>
<td>14 (0.05)</td>
<td></td>
</tr>
</tbody>
</table>

*Significance is reported from $\chi^2$ analyses in $3 \times 2$ tables.

---

**Figure 3.** AAD according to rs2251101 genotype in samples 5 and 6. AAD distributions (mean ± SEM) in the three genotype classes of rs2251101 in women (A and B) and men (C and D) from samples 5 and 6 (from left to right). The contrast is significant for men from sample 5 (C) yielding $F_{2,163} = 7.6, P = 0.0007$, providing support for analyses performed in samples 1–4.
adjacent genes, KIF11 and IDE, around the synonymous coding sequence variants in proxies for rs2251101. There are no validated non-West Africans in relation to ACE genotypes (30).

Of these two markers, we consider the evidence for rs2251101 to be more consistent across the various phenotypes examined. Varying degrees of LD with yet a third functional variant is thus possible, though this may be an example allelic heterogeneity, with two or more different functional sites. Deciphering the character of functional polymorphism in the IDE region is a daunting task if allelic heterogeneity is present. Since strong LD is present in this region, solutions to this question might include exploring IDE expression or insulin levels in relation to IDE genotypes in a population with much more haplotype diversity, such as has been done in West Africans in relation to ACE genotypes (30).

We explored LD relationships among all known SNPs around the IDE region, and surprisingly found no perfect proxies for rs2251101. There are no validated non-synonymous coding sequence variants in IDE or in the two adjacent genes, KIF11 and HHEX, in the LD block of which IDE is a part. While the molecular basis of functional polymorphism in IDE remains incompletely resolved, our data lead us to conclude that variable mRNA expression and thus gene regulation is important. This possibly involves a selective increase in the 15b-isoform which has been demonstrated to be catalytically less-active than the 15a-isoform (28). Here we focused primarily on the previously characterized isoforms of IDE, but acknowledge that additional uncharacterized isoforms may exist. An intriguing question for us that will need to be explored in future studies is whether rs2251101 or an as yet undiscovered variant has any effect upon 3′-UTR function. For example, the study of Farris et al. (28) highlights multiple length 3′-UTR transcripts that might indicate this as an important site of gene regulation. We note that the 3′-UTR can play a major role in mRNA expression itself (31) and so one option for further study that we are considering is further quantification of transcripts with different 3′-UTR lengths.

Thirdly, the indication here is that if IDE does indeed contribute to AAD, it does so only in men. In our view, the strongest support for the veracity of this is that there is fairly consistent replication across samples; additionally, a gender-specific association of IDE with plasma insulin level was seen, which supports this. A weakness, however, in our approach is that we only chose to follow-up men after the initial study, whereby effect size appeared to be increasing. We are unable to resolve if association might be evident in women at more advanced age. In considering how a gender-specific effect might arise, we noted that there was a considerable difference in fasting plasma insulin levels between men and women in the original report for sample 3, whereby men had the highest levels (20). This is also apparent in other studies, such as in participants of the Framingham Heart Study (32). Thus, if men are already at an increased risk of pathologies associated with hyperinsulinemia relative to women, IDE might make a sufficient genetic contribution to exceed a threshold for disease onset. Yet another factor that could contribute to the difference between men and women here is that women undergo menopause which may be associated with a decrease in plasma insulin levels (33). Given the ages of the samples here, most women being postmenopausal, this might confound the detection of gene effects. Gender specific effects of genes, while not common, have been demonstrated previously (34–36). Our findings here also highlight the potential importance gender may have in genetic association studies, but as with any covariate that requires stratification, further replication is still warranted in even larger samples.

Finally, while insulin metabolism in our view provides the most cogent link to lifespan across these analyses, we acknowledge that the spectrum of substrates for IDE spans a number of proteins critical to cellular metabolism, which includes insulin, β-amyloid, glucagon, atrial natriuretic peptide (ANP), insulin-like growth factor 2 (IGF2) and transforming growth factor-α (TGFα) (reviewed in 16). Common polymorphism affecting IDE may have repercussions for many cellular processes that if compromised.

### Table 3. Independent predictors of mortality in MI cases and controls (sample 7)

<table>
<thead>
<tr>
<th>Group (alive/deceased)</th>
<th>Source of variation</th>
<th>OR (95% CI)</th>
<th>Pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men-cases (676:153)</td>
<td>Age</td>
<td>1.064 (1.034–1.095)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>rs2251101</td>
<td>1.883 (1.288–2.754)</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>rs1887922</td>
<td>2.021 (1.365–2.991)</td>
<td>0.0004</td>
</tr>
<tr>
<td>Men-controls (940:93)</td>
<td>Age</td>
<td>1.096 (1.056–1.138)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>rs2251101</td>
<td>1.015 (0.633–1.626)</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>rs1887922</td>
<td>1.037 (0.636–1.692)</td>
<td>0.88</td>
</tr>
<tr>
<td>Women-cases (293:55)</td>
<td>Age</td>
<td>1.039 (0.990–1.092)</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>rs2251101</td>
<td>0.744 (0.385–1.439)</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>rs1887922</td>
<td>0.747 (0.369–1.511)</td>
<td>0.42</td>
</tr>
<tr>
<td>Women-controls (465:28)</td>
<td>Age</td>
<td>1.075 (0.999–1.156)</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>rs2251101</td>
<td>1.006 (0.434–2.336)</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>rs1887922</td>
<td>1.247 (0.547–2.841)</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Logistic regression was used to assess the significance of the various independent variables. Both linear (additive) and heterozygosity (heterozygotes against the combined homozygotes) models were assessed, the significance of the latter being shown in the table above. Odds ratios presented for markers represent the increase in odds given heterozygosity versus both opposite homozygous genotype classes. The number of individuals in each group is shown in parenthesis.
might collectively impact lifespan, and this will need to be illuminated in further replication efforts in other human populations.

**MATERIALS AND METHODS**

**Human samples**

Human samples are summarized in Table 1 and detailed descriptions are provided as Supplementary Text S1. There, we also document where and how tissue sampling was conducted. Local ethical approval for this study was obtained through the Karolinska Institutet Ethical Review Board. Informed written consent was obtained for samples 1–8 during the studies described in Supplementary Text S1. All samples from the Bristol data set (sample 9) were donations following appropriate Local Research Ethics Committee approved consent or care-giver assent procedures, or its contemporaneous equivalent for archival samples, to the South West Dementia Brain Bank which is a fully licensed resource under the United Kingdom Human Tissue Act 2004.

**Marker selection and genotyping**

Regional LD around IDE in Swedes has been previously described based upon 32 genetic markers (22). This full set was genotyped in sample 4. A five marker set was selected for genotyping in all additional samples in the present study. This included three variants capable of delineating common haplotypes from a block-like region across IDE (rs2251101, rs1832196 and rs1544210) (18). In addition, markers rs1887922 and rs967878 were included since they have previously elicited multiple trait associations in independent populations.

**IDE spliceform analysis and quantitation**

DNA reference sequences were extracted from the Human May 2004 (hg17) assembly from the UCSC genome browser (http://genome.ucsc.edu/). Sequences of primers and dual-labeled fluorogenic probes were designed using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). All selected sequences were blasted to check specificity. Potential secondary structure was examined with the web application on the DINAMelt server (http://www.bioinfo.rpi.edu/applications/hybrid/). Primers and probes were generated by Biomers.net GmbH. To identify the 15ab-spliceform, PCR using two primers that bind to the 15a- and 15b-exons, was applied in eight cDNA samples of temporal brain tissue (these samples were from the same source as sample 9, but are non-overlapping. They are not further described here). For quantitative real-time PCR (qPCR), the reverse primer for the 15a-spliceform spans the 3’ end of 15a-exon and the 5’ end of 16th exon to avoid amplification of the 15b- or 15ab-spliceforms. Likewise, the forward primer for the 15b-spliceform spans the boundary between 14th exon and 15b-exon to achieve specificity. The amplicon for the 15ab-spliceform spans the 15a- and 15b-exons. All primers were designed to bind far from the highly conserved region between 15a- and 15b-exons.

Two levels of standardization were performed, one prior to qPCR of IDE spliceforms, and an additional GAPDH estimation at the same time as the primary IDE quantitation experiment. Initially, GAPDH in 1 μl (sample 9) was quantified by qPCR. According to the measure, each sample was normalized so that it would provide the same Ct (PCR cycle number at threshold—defined automatically) of 26 with 1 μl of stock sample. Some samples that were of too low concentration to provide the Ct of 26 (35 samples total) were diluted to a volume sufficient for all following qPCR reactions. With the normalized cDNA template, quantities of 15a-, 15b- and 15ab-spliceforms were measured on separate 96-well plates along with GAPDH, using an ABI PRISM 7000 Sequence Detection System (Applied Biosystems). Reactions for IDE were run in duplicate, and only once for GAPDH. The measured Ct values were averaged, and the
relative quantities of each spliceform were calculated from standard curves. The standard curve for each assay was achieved in triplicate runs of cDNA templates on separate plates. Details of the PCR and qPCR protocols and sequences of primers and probes are described in Supplementary Text S3.

Sequencing and polymorphism discovery

Primers were selected as in the above section (see Supplementary Text S3 for sequences and reaction conditions). PCR products used as templates for DNA sequencing covered a repetitive sequence upstream of the 15a-exon that includes an SNP (rs3831274), 15a- and 15b-exons, and the flanking intronic regions (hg17, chr10:94,229,523-94,228,006). The PCR products were amplified in two overlapping fragments, 961 and 746 bp, that include the 15a- and 15b-exons, respectively. Concentration was approximated on agarose gels and PCR products including the 15a-exon were purified by PCR purification kit (QIAGEN) to reduce noise. Five sequencing primers were applied, one for the repetitive sequence, two for flanking region of 15a-exon toward opposite direction, two for 15b-exon neighbors. Electrophoresis and sequencing detection was performed using a MegaBACETM 1000 with DYEnamic™ ET Terminator kit (Amersham Biosciences). Base-calls and assembly of reads was facilitated using the PhredPhrap package (38).

Statistical analyses

Hardy–Weinberg equilibrium for individual loci was assessed using the Pearson $\chi^2$ statistic. Normality of age-at-sampling, AAD, and metabolic trait distributions was assessed using a Kolgomorov–Smirnov test. To quantify genotype effects in samples 1–4, logistic regression was used to model the dependency of genotype frequency as a function of age (5). Multinomial regression, an extension of logistic regression to the polycotomous case, was used to model all di-allele genotypes without a priori assumptions as to the genetic model, e.g. additive or non-additive. We report the significance of the likelihood ratio tests provided by SPSS 11.0 for the age effect versus an intercept-only model in single sample tests. For tests combining clinical materials, a sample identifier was included as a covariate, thus age effects reported were adjusted for sample source. Additionally, differences in age distributions among genotype classes were assessed using ANOVA (normality of age distributions was confirmed for all samples in which this was used). Haplotypes were probabilistically reconstructed from unphased genotypes at multiple individual loci using the PHASE v2.1 program (39). Logistic regression was used to assess a possible effect on mortality in sample 7 by including age and IDE genotype as independent variables and treating ‘deceased’ or ‘alive’ at follow-up as a dependent variable.

To facilitate analyses of the population-based twin samples, Cox regression models were fit to samples 5 and 6 combined using SAS PROC PHREG. (Separate analyses for each sample are presented in Supplementary Text S2). IDE genotype (rs2251101 or rs1887922) was recoded into typical contrasts to estimate additive and dominance effects: the additive contrast coded as $-1$ for those who were homozygous common, 0 for heterozygotes and 1 for those who were homozygous rare. The dominance contrast was created by subtracting the squared values of the additive contrast values from 1; thus, the heterozygote genotype was coded as 1 and the two homozygous genotypes were coded as 0. The main effect of first age-at-sampling (centered on age 65 years) and its interaction with the additive and dominance contrasts were specified. Marker rs429358 in APOE was considered as a covariate in all analyses where the number of rare alleles was counted. To obtain appropriate standard errors and $P$-values, we accounted for the twin sibling structure. Specifically, each twin pair was assigned a unique pair number and models were fit specifying pair number as the identification number to account for pair similarity. Furthermore, zygosity was designated as a stratum variable to adjust for any differences in the hazard function.

Spliceform analyses were conducted using ANOVA treating qPCR Ct value directly as a quantitative trait. All models included GAPDH level as a covariate. We note that Ct values for all spliceforms followed normal distributions. Samples included both AD cases and controls and an identifier was included for those two classes as a covariate in analyses. Estimation of relative expression from Ct data from standard curves was performed for the purpose of presentation only (Fig. 5).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

Conflict of Interest statement. None declared.

FUNDING

Financial support from National Institutes of Health (NIH, grant no. R01 AG08724), The Swedish Medical Research Council, The Swedish Alzheimer Foundation (Alzheimerfonden), Loo and Hans Ostermans Foundation, King Gustav the Vth and Queen Victoria’s Foundation, The Sigmund Gestetner Foundation, a Reverend Williams PhD studentship and The Swedish Heart and Lung Foundation.

REFERENCES


