Intronic rs2147363 Variant in ATP7B Transcription Factor-Binding Site Associated with Alzheimer’s Disease

Serena Bucossia, Renato Polimanti, Mariacarla Ventriglia, Stefania Mariani, Mariaristina Siotto, Francesca Ursini, Laura Trotta, Federica Serracchia, Antonio Callea, Fabrizio Vernieri and Rosanna Squitti

*Department of Neurology, “Campus Bio-Medico” University, Rome, Italy
**Department of Neuroscience, AFaR - Fatebenefratelli Hospital “San Giovanni Calibita”, Rome, Italy
***Department of Biology, “Tor Vergata” University, Rome, Italy
†Don Carlo Gnocchi Foundation ONLUS, Italy
‡Laboratory of Neurodegeneration, IRCSS San Raffaele Pisana, Rome, Italy

Handling Associate Editor: Roberta Ghidoni

Accepted 28 May 2013

Abstract. Copper homeostasis abnormalities have been shown to be associated with Alzheimer’s disease (AD), possibly by accelerating amyloid-β toxicity and plaque formation. The ATP7B gene plays a key role in controlling body copper balance. Our previous studies showed an association between ATP7B variants and AD risk. Among these variants, an intronic single nucleotide polymorphism, rs2147363, was associated with AD risk. In order to understand this intronic association, we screened a population of 286 AD patients and 283 healthy controls, and verified the presence of other functional coding variants in linkage disequilibrium (LD). Then we searched for a regulatory function region close to rs2147363. An LD analysis revealed the presence of an LD between rs2147363 and a Wilson’s disease-causing variant, rs7334118. However, this mutation did not explain the observed genetic association. Conversely, in silico analyses of rs2147363 functionality highlighted that this variant is located in a binding site of a transcription factor, and is, consequently, associated with regulatory function. These data suggest that the genetic variation in cis-regulatory elements located in non-coding regions can have a role in determining ATP7B functionality and account for some of the AD missing hereditability.

Keywords: Alzheimer’s disease, ATP7B, cis-regulatory element, copper, intronic variant

INTRODUCTION

Alzheimer’s disease (AD) is the most common form of dementia, and it has a high heritability. Risk of developing AD is related to several biological and genetic factors. Recent studies have demonstrated that abnormal disposition of certain metals, notably copper, may also influence the risk for AD. In particular, a number of studies have demonstrated abnormalities in copper [1] as well as in the portion of copper non-bound to ceruloplasmin (also called ‘free’ or labile copper). This occurs in general circulation [2–5] or in the AD brain [6–8], which can accelerate amyloid-β (Aβ) precipitation in plaques and toxicity [9].

We recently performed a study on informative single-nucleotide polymorphisms (SNPs) associated with AD risk [10]. Among these variants, one intronic SNP (rs2147363) showed a significant association with AD
454
S. Buess et al. / Intronic ATP7B SNP in Alzheimer’s Disease

risk, suggesting a potential role of this non-coding variant. A recent concept in genetics is the role of the non-coding regions of the genome in the regulation of gene function [11]. Several cis-regulatory elements (i.e., promoters, enhancers, repressors, and insulators) are present in non-coding genomes that finely control gene expression through the recruitment of transcription factors (TFs) [12]. Genetic variations in these regulatory non-coding regions are, in some cases, involved in the alteration of gene expression. Additionally, they strongly contribute to the onset of the diseases [13]. Scientific literature pertaining to non-coding regulatory variants is increasing, and a number of studies have reported functional variation within introns, far upstream regions, and at splicing sites or within microRNA sites, associating them with a wide spectrum of disease phenotypes [14–16].

Regarding the genetic association of rs2147363, two hypotheses can be taken into account: i) rs2147363 may be in linkage disequilibrium (LD) with coding variants with high functional impact on ATP7B protein; ii) rs2147363 may be located in an intronic region with a regulatory role on the ATP7B gene, and the variants may have a direct or indirect effect on these cis-regulatory elements.

The aim of the present study is to understand how strong the genetic association between intronic rs2147363 and the risk of AD is and to explore its pathogenetic mechanism.

METHODS

Subjects

A total of 286 AD patients and 283 healthy controls were recruited through two specialized dementia care centers located in Rome, Italy, the Department of Neuroscience of San Giovanni Calibita - Fatebenefratelli Hospital and the Department of Neurology of Campus Bio-Medico University. Both care centers used the same standardized clinical protocol to recruit the study population [17]. The study was approved by the local IRB, and all participants or legal guardians signed an informed consent form.

All AD patients had been diagnosed as ‘probable AD’, according to NINCDS-ADRDA (National Institute of Neurosciences of San Giovanni Calibita - Fatebenefratelli Hospital and the Department of Neurology of Campus Bio-Medico University. Both care centers used the same standardized clinical protocol to recruit the study population [17]. The study was approved by the local IRB, and all participants or legal guardians signed an informed consent form.

All AD patients underwent general medical, neurologic, and psychiatric assessments. Neuroimaging diagnostic procedures (MRI or computed tomography) and complete laboratory analyses were performed to exclude other causes of progressive or reversible dementia.

The control sample consisted of healthy volunteers with no clinical evidence of neurological or psychiatric disease. To assess the absence of neurologic conditions, controls were evaluated by neurologists and geriatricians.

Exclusion criteria for both patients and controls were conditions known to affect copper metabolism and biological variables of oxidative stress (e.g., diabetes mellitus, inflammatory diseases, recent history of heart or respiratory failure, chronic liver or renal failure, malignant tumors, and a recent history of alcohol abuse). 60% of our study population partially overlaps the population in a previous study [10].

SNP genotyping

Genomic DNA extraction was carried out from peripheral blood through standardized salting-out method [20]. Genotyping of rs2147363 and rs7334118 was performed by the TaqMan allelic discrimination assay as previously described [21]. The predesigned SNP genotyping assay ID for rs2147363 is C_25473601, whereas for rs2147363, a TaqMan Custom Assay was used. Direct DNA bidirectional sequencing was performed for 15% of the PCR products, which were randomly selected and analyzed to confirm the genotypes. Apolipoprotein E (APOE) genotyping was performed in accordance with the established methods [22].

Statistical analyses

Demographic and clinical characteristics in our patient and control samples were described either in terms of mean ± SD if quantitative, or in terms of proportions. Student’s t-test and the chi-square test were used to compare the characteristics of AD patients and controls using the statistical analysis software package SPSS 15.0. An estimation of sample power to detect associations between rs2147363 and AD was evaluated using the Power of Genetic Analysis (PGA) package [23]. The minimum detectable effects with Odds Ratios (ORs) was calculated in a general (codominant) model, based on an alpha of 0.05 and AD prevalence of 1.5% in the general population. With this sample size, we have 80% power to detect an OR of 1.44 if the minor allele frequency is 25%.
Weinberg equilibrium of ATP7B SNPs were evaluated using SNPStats [24]. The differences in the genotype distributions among AD patients and healthy individuals were assessed by χ² tests. A logistic regression was used to calculate adjusted ORs and 95% confidence intervals (CIs) for the association between rs2147363 and AD. To estimate the ORs, different genetic models were considered: dominant (one copy of the allele is sufficient to increase the disease risk), recessive (two copies of the allele are necessary to increase the disease risk), and log-additive (r-fold increased risk for one copy of the allele, r² increased risk for two copies of the allele). To evaluate the best model, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used. Both these selection criteria were used to obtain a more reliable choice. FastSNP (Function Analysis and Selection Tool for Single Nucleotide Polymorphisms), Polyphen2, and SIFT (Sorting Intolerant From Tolerant) were used to predict the functional impact of the SNP on the coding protein [25–27]. GERP++ (Genomic Evolutionary Rate Profiling) was used to calculate nucleotide-level conservation scores, and a score threshold of 2 was considered to determine nucleotide conservation [28]. The is-rSNP algorithm was used to predict the TF-binding sites in the rs2147363 region [29].

**RESULTS**

**Demographic and clinical characteristics**

The demographic and clinical characteristics for AD and control groups are shown in Table 1. AD patients and controls did not differ in gender ratio, but differed in age, mean MMSE score, and APOE E4 status. As expected, the mean MMSE score was lower in patients than in controls and the presence of at least one APOE E4 allele was more frequent in patients than in controls (p < 0.001). To avoid confounding bias in genetic association analysis, age, gender, and APOE genotype were taken as covariates in all the statistical analyses.

<table>
<thead>
<tr>
<th>Characteristics of AD patients and elderly healthy controls</th>
<th>AD patients (n = 286)</th>
<th>Controls (n = 283)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years) [mean (SD)]</td>
<td>75.4 ± 6.0</td>
<td>69.5 ± 10.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Gender, % females</td>
<td>67</td>
<td>69</td>
<td>0.717</td>
</tr>
<tr>
<td>MMSE score [mean (SD)]</td>
<td>18.5 ± 5.5</td>
<td>27.9 ± 2.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>APOE, % E4 carriers</td>
<td>36</td>
<td>12</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

MMSE, Mini-Mental State Examination.

**Genetic association analysis between rs2147363 and AD risk**

The distributions of allele and genotype frequencies in AD patients and healthy controls are reported in Table 2. The genotype frequencies of ATP7B SNP rs2147363 were in the Hardy–Weinberg equilibrium. All genotype frequencies in our controls were within the ranges reported previously in the European origin populations of the HapMap project (available on http://hapmap.ncbi.nlm.nih.gov).

The significant association between rs2147363 and AD was achieved by crude and logistic regression analyses (Table 2). When data were adjusted for confounding variables (i.e., age, gender, and APOE genotype), significant results were obtained for the recessive model (OR: 1.63, 95% CI: 1.03–2.57; p = 0.035) and Log-additive model (OR: 1.51, 95% CI: 1.05–2.16; p = 0.025). To select the best model for our data, we used two standard model selection criteria: AIC and BIC. The log-additive model achieved the best scores in AIC and BIC.

**LD and association analysis of rs7334118**

To verify whether the rs2147363 has any functional variants in LD, we analyzed the SNPs that showed a complete LD (D² = 1) with this ATP7B variant in Tuscans in Italy (TSI), which is the HapMap population most genetically related to ours. We identified ten SNPs in complete LD with the rs2147363. We prioritized these variants using FastSNP. This analysis highlighted that one variant is a non-synonymous SNP (medium-high risk), four are intronic enhancers (very low-low risk), four are intronic with no known function (no effect), and one is a downstream variant with no known function (no effect) (Table 3). To predict the functional impact of the non-synonymous SNP (rs7334118), we used two different bioinformatic tools: SIFT and Polyphen2. The application of both the SIFT algorithm using orthologous sequences and the Polyphen2 tool highlighted that this coding variant may have an adverse effect on the ATP7B protein.

To verify the hypothesis that genetic association of rs2147363 is due to LD with the rs7334118 SNP, 176 AD subjects and 169 healthy controls among the study population were genotyped for the SNP rs7334118. Two AD patients were carriers of the rs7334118G allele, whereas no healthy individual with this ATP7B mutation was identified. Although this coding variant was identified only in AD patients, its allele frequency in AD group is in line with the minor allele frequency.
Table 2
Allele and genotype distribution of AD patients and controls

<table>
<thead>
<tr>
<th>SNP (rs214736)</th>
<th>All subjects (%)</th>
<th>AD patients (%)</th>
<th>Controls (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>305 (27)</td>
<td>140 (25)</td>
<td>165 (29)</td>
</tr>
<tr>
<td>C</td>
<td>833 (73)</td>
<td>432 (75)</td>
<td>401 (71)</td>
</tr>
<tr>
<td>Genotype</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>38 (7)</td>
<td>17 (6)</td>
<td>21 (7)</td>
</tr>
<tr>
<td>AC</td>
<td>229 (40)</td>
<td>106 (37)</td>
<td>123 (44)</td>
</tr>
<tr>
<td>CC</td>
<td>302 (53)</td>
<td>163 (57)</td>
<td>139 (49)</td>
</tr>
</tbody>
</table>

Table 3
Prioritization and functional analysis of ATP7B SNPs in LD with rs214736

<table>
<thead>
<tr>
<th>SNP ID (rs)</th>
<th>Possible Functional Effects</th>
<th>Lower Risk Region</th>
<th>Upper Risk Region</th>
<th>SIFT</th>
<th>PolyPhen2</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7334141</td>
<td>Missense (non-conservative); Splicing regulation</td>
<td>Medium</td>
<td>High</td>
<td>Coding</td>
<td>Damaging</td>
</tr>
<tr>
<td>rs4943053</td>
<td>Intronic enhancer</td>
<td>Very low</td>
<td>Low</td>
<td>intronic</td>
<td>-</td>
</tr>
<tr>
<td>rs955803</td>
<td>Intronic enhancer</td>
<td>Very low</td>
<td>Low</td>
<td>intronic</td>
<td>-</td>
</tr>
<tr>
<td>rs2147362</td>
<td>Intronic enhancer</td>
<td>Very low</td>
<td>Low</td>
<td>intronic</td>
<td>-</td>
</tr>
<tr>
<td>rs1859458</td>
<td>Intronic enhancer</td>
<td>Very low</td>
<td>Low</td>
<td>intronic</td>
<td>-</td>
</tr>
<tr>
<td>rs950682</td>
<td>Intronic enhancer</td>
<td>Very low</td>
<td>Low</td>
<td>intronic</td>
<td>-</td>
</tr>
<tr>
<td>rs953582</td>
<td>Intronic enhancer</td>
<td>Very low</td>
<td>Low</td>
<td>intronic</td>
<td>-</td>
</tr>
<tr>
<td>rs9535794</td>
<td>Downstream with no known function</td>
<td>No effect</td>
<td>No effect</td>
<td>3-UTR</td>
<td>-</td>
</tr>
<tr>
<td>rs9535806</td>
<td>Intronic enhancer</td>
<td>Very low</td>
<td>Low</td>
<td>intronic</td>
<td>-</td>
</tr>
</tbody>
</table>

observed in a general population of European origin (information available at dbSNP).

Functional prediction analysis of rs2147363

To test whether the intronic region, in which rs2147363 is located, plays a role in the regulation of ATP7B gene function, the GERP++ and is-rSNP algorithms were used. In Table 4, we report the outcomes of the GERP analysis. Regarding rs2147363, nucleotide position resulted non-conserved, whereas, in the intronic region around this SNP, 7 nucleotide positions on the 16 analyzed (43%) achieve the GERP threshold. Table 5 highlighted the findings of is-rSNP algorithm. This analysis predicted the presence of binding sites of 8 TFs, and, in two cases (i.e., Zfp423 and PLAG1), the prediction results were significant (adjusted p < 0.05).

DISCUSSION

The main result of this study is that the significant association between rs2147363 SNP and AD can be explained on the basis of the presence of binding sites in the region, in which this variant is located. Thereby, rs2147363 may be associated with cis-regulatory function. Specifically, we observed two of the genetic models of rs2147363 SNP associated with AD: a
intronic association. Candidate variants that could be tested to explain our database of 1,000 Genomes Project, may furnish new present. Using other genomic databases, such as the coding variants, in which only few rare variants are we used the HapMap database to identify functional effects may explain this intronic association. Indeed, exclude additional rare variants with large functional disease mutation (99%). However, this result does not rs2147363, were negative for the rs7334118 Wilson’s other 166 AD patients, carriers of the C allele in rs2147363 (1%). Whereas, heterozygous for the rs7334118G mutation were also observed for rs2147363. Specifically, two AD patients tion could not alone explain the genetic association ATP7B can be involved in molecular pathways linked to brain activity regulation, suggesting new perspectives in the interpretation of neurologic signs of Wilson’s disease and AD. This conclusion fits well with our original hypothesis that ATP7B can harbor variants that may account for some of the missing hereditability of AD [32]. In particular, the present data furnished a new insight into the copper hypothesis: non-coding regions can play a role in ATP7B function, and, thereby, genetic variation in ATP7B cis-regulatory elements within non-coding regions may be associated with AD risk. Our study certainly has some limitations. First, the two cohorts included in this study were not matched for age, so we showed our results as age-adjusted data. However, we consider our estimation conservative, since some controls could possibly convert to AD, making our controls and cases closer to one in agreement with the case-control study design, which assumes that controls should have the possibility to “become cases” [33]. Furthermore, genetic data adjusted for age confounding effect confirmed the association between rs2147363 and AD risk.

Finally, although in silico analyses support our hypothesis, further functional studies are necessary to confirm the role of non-coding variants in ATP7B gene function and in AD risk.

ACKNOWLEDGMENTS
This study was partially supported by the following grants: 1) European Community’s Seventh Framework Programme Project MEGMRI (n. 200859); 2) FISM – Fondazione Italiana Sclerosi Multipla –

### Table 5

<table>
<thead>
<tr>
<th>Database</th>
<th>Transcription factor</th>
<th>Adjusted p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>JASPAR</td>
<td>Zipf23</td>
<td>0.005</td>
</tr>
<tr>
<td>JASPAR</td>
<td>PLAG1</td>
<td>0.041</td>
</tr>
<tr>
<td>JASPAR</td>
<td>ERSS1</td>
<td>0.007</td>
</tr>
<tr>
<td>JASPAR</td>
<td>Inv2</td>
<td>0.138</td>
</tr>
<tr>
<td>JASPAR</td>
<td>Mafu</td>
<td>0.170</td>
</tr>
<tr>
<td>JASPAR</td>
<td>GTGNNNYNGNASSA</td>
<td>0.181</td>
</tr>
<tr>
<td>JASPAR</td>
<td>GTRBCATER</td>
<td>0.190</td>
</tr>
<tr>
<td>JASPAR</td>
<td>RXRA::VDR</td>
<td>0.190</td>
</tr>
</tbody>
</table>

This study was partially supported by the following grants: 1) European Community’s Seventh Framework Programme Project MEGMRI (n. 200859); 2) FISM – Fondazione Italiana Sclerosi Multipla –

Conversely, in silico analyses on rs2147363 revealed that this intronic SNP can have a regulatory role on ATP7B function. In particular, the GERP analysis highlighted that the intronic region around rs2147363 resulted significantly conserved, and the is-rSNP algorithm significantly predicted the presence of two TF binding sites in the rs2147363 region. These independent findings suggested that this genomic region has a regulatory function, and consequently rs2147363 can be associated with clinical phenotypes related to ATP7B dysfunction. Specifically, Zipf23 and PLAG1 have been predicted to have a binding-site in the genetic region of rs2147363. Both these TFs are involved in complex metabolic processes. For example, Zipf23 has been reported to regulate neural activity regulation, suggesting new perspectives in the interpretation of neurologic signs of Wilson’s disease and AD. This conclusion fits well with our original hypothesis that ATP7B can harbor variants that may account for some of the missing hereditability of AD [32]. In particular, the present data furnished a new insight into the copper hypothesis: non-coding regions can play a role in ATP7B function, and, thereby, genetic variation in ATP7B cis-regulatory elements within non-coding regions may be associated with AD risk. Our study certainly has some limitations. First, the two cohorts included in this study were not matched for age, so we showed our results as age-adjusted data. However, we consider our estimation conservative, since some controls could possibly convert to AD, making our controls and cases closer to one in agreement with the case-control study design, which assumes that controls should have the possibility to “become cases” [33]. Furthermore, genetic data adjusted for age confounding effect confirmed the association between rs2147363 and AD risk.

Finally, although in silico analyses support our hypothesis, further functional studies are necessary to confirm the role of non-coding variants in ATP7B gene function and in AD risk.

REFERENCES


