Published in final edited form as:
Nat Genet. 2009 February ; 41(2): 192–198. doi:10.1038/ng.305.

Genetic variation in *PCDH11X* is associated with susceptibility to late onset Alzheimer's disease

Minerva M. Carrasquillo[1], **Fanggeng Zou**[1], **V. Shane Pankratz**[2], **Samantha L. Wilcox**[1], **Li. Ma**[1], **Louise P. Walker**[1], **Samuel G. Younkin**[1], **Curtis S. Younkin**[1], **Linda H. Younkin**[1], **Gina D. Bisceglio**[1], **Nilufer Ertekin-Taner**[1,3], **Julia E. Crook**[4], **Dennis W. Dickson**[1], **Ronald C. Petersen**[5], **Neill R. Graff-Radford**[1,3], and **Steven G. Younkin**[1]

[1]Department of Neuroscience, Mayo Clinic College of Medicine, Jacksonville, FL 32224, USA

[2]Division of Biomedical Statistics and Informatics, Mayo Clinic and Mayo Foundation, Rochester, MN 55905, USA

[3]Department of Neurology, Mayo Clinic College of Medicine, Jacksonville, FL 32224, USA

[4]Biostatistics Unit, Mayo Clinic College of Medicine, Jacksonville, FL 32224, USA

[5]Department of Neurology and the Mayo Alzheimer Disease Research Center, Mayo Clinic College of Medicine, Rochester, MN 55905, USA

Abstract

By analyzing late onset Alzheimer's disease (LOAD) in a genome wide association study (313,504 SNPs, 3 series, 844 cases/1,255 controls) and evaluating the 25 SNPs with most significant allelic association in 4 additional series (1,547 cases/1,209 controls), we identified a SNP (rs5984894) on Xq21.3 in *PCDH11X* that is strongly associated with LOAD in American Caucasians. Analysis of rs5984894 by multivariable logistic regression adjusted for sex gave global *P* values of 5.7×10^{-5} in stage I, 4.8×10^{-6} in stage II, and 3.9×10^{-12} in the combined data. Odds ratios were 1.75 (95% CI 1.42-2.16) for female homozygotes ($P=2.0\times10^{-7}$) and 1.26 (95% CI 1.05-1.51) for female heterozygotes (*P=0*.01) compared to female non-carriers. For male hemizygotes (*P*=0.07) compared to male non-carriers the odds ratio was 1.18 (95% CI 0.99-1.41).

Late onset Alzheimer's disease (LOAD) is a neurodegenerative disease characterized by large numbers of senile plaques and neurofibrillary tangles in the brain. LOAD is the most common cause of dementia in the elderly, affecting approximately 10% of those aged 65 years or older[1]. Multiple rare mutations in the genes encoding the amyloid ß protein precursor, presenilin 1, and presenilin 2 cause an early onset familial form of AD with autosomal dominant inheritance, but the only well established susceptibility allele for LOAD is the *APOE* ε4,

To whom correspondence should be address: younkin.steven@mayo.edu.

Author Contributions: M.M.C. spearheaded and participated in all aspects of this study, and drafted the manuscript along with Steven G. Younkin who is the lead investigator of this study. F.Z., S.L.W., L.M. and L.P.W. participated in the SEQUENOM genotyping. F.Z., L.M., L.H.Y. and G.D.B. were responsible for DNA sample preparation and quality control. L.M. also generated all DNA replica plates. Samuel G. Younkin and C.S.Y were instrumental in data management and analysis. N.E.T. participated in critical revisions of the manuscript. V.S.P and J.E.C. provided statistical expertise. N.R.G. and R.C.P. are the neurologists who diagnosed and provided samples for the Mayo Clinic Jacksonville (JS) and Mayo Clinic Rochester (RS) series, respectively. D.W.D. is the pathologist who diagnosed and provided brain samples for the autopsy-confirmed (AUT) series.

URLs. PLINK, http://pngu.mgh.harvard.edu/purcell/plink/

Accession codes. RefSeq: *PCDH11X* mRNA isoform a precursor, NM_014522.1; *PCDH11X* mRNA isoform b precursor, NM_032967.1; *PCDH11X* mRNA isoform c, NM_032968.2; *PCDH11X* mRNA isoform d precursor, NM_032969.2. **Entrez Gene:** *PCDH11X*, 27328; *PCDH11Y*, 83259.

allele[2-4]. Twin studies indicate that susceptibility alleles contribute to as much as 80% of LOAD cases[5] but definitive identification of other genes with LOAD susceptibility alleles has proven difficult.

To identify novel LOAD susceptibility genes, we performed a two-stage genome-wide association study (GWAS) using Illumina HumanHap300 BeadChips. In stage I, after stringent quality control, we analyzed 313,504 SNPs in 844 cases and 1,255 controls (Supplementary Methods online, Supplementary Tables 1 and 2). Stage I subjects had ages at diagnosis/entry of 60-80 years and were drawn from three series. Two series were assembled from cases and controls ascertained clinically at the Mayo Clinic in Jacksonville, FL (JS: 353 AD, 331 control) and Rochester, MN (RS: 245 AD, 701 control). A third, autopsy-confirmed series (AUT: 246 AD, 223 control) was assembled from the Mayo brain bank (see Supplementary Methods online for details). In stage I, SNPs were tested for allelic association using the χ^2 test implemented in PLINK[6] (Supplementary Table 3 online). After adjustment for population stratification using EIGENSTRAT[7] and Bonferroni correction for the 313,504 SNPs tested, only six *APOE*-linked SNPs showed genome-wide significance in stage I (Supplementary Table 3 and 4, and Supplementary Methods online).

In stage II, we employed SEQUENOM iPLEX technology to genotype the 25 SNPs with the most significant association in stage I (Supplementary Table 3 online) in an additional 845 cases and 1,000 controls drawn from the same three series but with ages at diagnosis/entry of over 80 years (JS: 237 AD, 260 control; RS: 276 AD, 624 control; AUT: 332 AD, 116 control) and from a National Cell Repository for AD series of 702 cases (1/family) and 209 controls with ages at diagnosis/entry of over 60 years. The top 25 SNPs tested in stage II included 10 SNPs in the *APOE* region of chromosome 19 and 15 SNPs on other chromosomes. The allelic association results for these 25 SNPs in stages I, II, and I + II are shown in Supplementary Table 3 online. All 10 SNPs in the *APOE* region showed significant association in stage II with *P* values ranging from 9.5×10^{-79} to 0.05. One of the two SNPs on the X chromosome, rs5984894, also replicated well in the stage II follow-up series with a *P* value of 0.0006 that retained significance (*P*=0.015) even after conservative Bonferroni correction for 25 SNPs tested in stage II. None of the other SNPs replicated in stage II. The second SNP on the X chromosome, rs1279795, had a *P* value of 0.43 and the 13 additional SNPs had *P* values of 0.28-0.98 (Supplementary Table 3 online).

In stages I + II, rs5984894 was genotyped successfully in 2,356 of the 2,391 cases (99%) and 2,384 of the 2,464 controls (97%). Because our stage I and stage II subjects came from multiple American Caucasian series, PLINK was employed to test rs5984894 for allelic association in the combined stage I, II, and I + II datasets using the Mantel-Haenszel method (Table 1) in addition to the χ^2 test on combined allele counts (Supplementary Table 3 online). In the combined stage I dataset, the *P* values were 1.2×10^{-5} and 1.5×10^{-5} for the allelic association and the Mantel-Haenszel method respectively. Both methods showed highly significant replication in the stage II dataset where the unadjusted *P* values were 0.0006 and 0.002 respectively and the Bonferroni-adjusted *P* values were 0.015 and 0.05. The overall unadjusted *P* value for allelic association in stage I + II was 3.8×10^{-8} with an OR of 1.29 (95% CI 1.18-1.41), and 2.2×10^{-7} with an OR of 1.30 (95% CI 1.18-1.43) using the Mantel-Haenszel method (Table 1). The Breslow Day *P* values calculated by PLINK to test for series to series heterogeneity were 0.95, 0.22, and 0.43 in stages I, II, and I + II, indicating a lack of statistical evidence for series to series heterogeneity among the seven series tested.

rs5984894 is within the gene (*PCDH11X*) encoding protocadherin 11, X-linked (Fig. 1). *PCDH11X is* located in the hominid-specific non-pseudoautosomal homologous region Xq21.3/Yp11.2[8]. It has been proposed that known coding and expression level differences between *PCDH11X* and *PCDH11Y* may have functional consequences that could lead to

sexually dimorphic traits[9]. To explore this possibility, we analyzed rs5984894 by multivariable logistic regression with sex as a covariate (Table 2). Using this approach, which specifically models each carrier group, the global *P* value in the combined series improved substantially to 3.9×10^{-12} as compared to 3.8×10^{-8} for allelic association (Supplementary Table 3) and 2.2×10^{-7} using the Mantel-Haenszel method (Table 1). In the combined series, odds ratios were 1.75 (95% CI 1.42-2.16) for female homozygotes ($P=2.0\times10^{-7}$) and 1.26 (95% CI 1.05-1.51) for female heterozygotes ($P=0.01$) compared to female non-carriers. For male hemizygotes ($P=0.07$) compared to male non-carriers, the odds ratio was 1.18 (95% CI 0.99-1.41) (Table 2). Male sex, which had an OR of 0.86 (95% CI 0.71-1.05) was not a significant covariate ($P=0.14$) in the combined data. Female homozygotes in the combined series were at significantly increased risk not only when compared to female non-carriers ($P=2.0\times10^{-7}$) but also when compared to female heterozygotes ($P=0.0005$) or male hemizygotes ($P=1.4\times10^{-7}$) (Supplementary Table 5 online, model 3). The OR for female homozygotes in stage I was 1.92 (95% CI 1.36-2.70) with a *P* value of 0.0002. This association replicated well in stage II where the OR was 1.70 (95% CI 1.29-2.24) with a *P* value of 0.0002 (Table 2). The global *P* value of 5.7×10^{-5} in stage I also replicated well, improving to 4.8×10^{-6} on follow-up.

Replication for female homozygotes and heterozygotes was highly consistent when subjects with ages at diagnosis/entry of 60-80 years were compared to subjects with ages at diagnosis/entry of over 80 years; the ORs were 1.74 (95% CI 1.31-2.32) and 1.25 (95% CI 0.98-1.60) vs. 1.76 (95% CI 1.29-2.40) and 1.26 (95% CI 0.97-1.65) respectively (Supplementary Table 5 online, Stage I + II, model 1). Although male hemizygotes showed significant risk ($P=0.04$) in stage I with an OR of 1.33 (95% CI 1.02-1.74), in stage II ($P=0.74$) or in all subjects with ages at diagnosis/entry of over 80 years ($P=0.66$) male hemizygotes showed no statistically significant evidence for increased risk with ORs of 1.04 (95% CI 0.82-1.33) and 1.07 (95% CI 0.80-1.42) respectively (Table 2 and Supplementary Table 5 online, model 1; see section on Power considerations in Supplementary Methods for additional discussion).

Logistic regression models that included sex, age at diagnosis/entry (years over 60) and the presence of an *APOE* ε4 allele as covariates were also evaluated (Supplementary Methods and Supplementary Table 5 online, model 4). In the combined data, age ($P=4.9\times10^{-7}$) and *APOE* ε4 ($P<2.2\times10^{-16}$) were significant covariates with ORs of 1.02 (95% CI 1.01-1.03) and 6.21 (95% CI 5.45-7.08) respectively. When these two covariates were included (Supplementary Table 5 online, model 4), the significant associations for female heterozygotes and homozygotes persisted with ORs of 1.23 (95% CI 1.01-1.51) and 1.68 (95% CI 1.33-2.12) respectively. We also investigated series-to-series heterogeneity by examining series × genotype interactions in our logistic regression analyses. Consistent with the results of the Breslow Day tests described above, these analyses provided no significant evidence for differences among series in the associations observed (data not shown).

Using stage I GWAS data, population stratification was evaluated using the principal components approach implemented in EIGENSTRAT. Adjustment for population substructure was performed by including the top ten axes of variation generated by EIGENSTRAT as additional covariates in logistic regression analyses using an allelic dosage model and in multivariable logistic regression analyses of rs5984894. These adjustments to the allelic dosage (Supplementary Table 4 online) and multivariable logistic regression (Supplementary Table 5 online, model 5 vs. model 1) analyses had essentially no effect on the results obtained for the stage I GWAS data. Thus, population substructure did not inflate the significance of stage I GWAS results, and given the similarity in the populations included in stages I and II it is unlikely that it inflated the highly significant associations observed in stage II and in the combined data.

rs5984894 maps to a 102kb linkage disequilibrium (LD) block on chromosome Xq21.3 that lies entirely within the gene (*PCDH11X*) encoding protocadherin 11, X-linked. This LD block encompasses part of intron 2, exon 3 and part of intron 3 of *PCDH11X isoforms* c and d (Fig. 1). In the stage I GWAS, 2 of the 6 additional SNPs within this 102kb block (rs5941047 and rs4568761) showed strong association with LOAD. Both SNPs had *P* values for allelic association of 0.0023 and both are in strong LD with rs5984894 (Fig. 1). The 7 SNPs on the 102kb block form 7 haplotypes with frequencies above 1% that account for 98% of all haplotypes. In the stage I GWAS, a χ^2 test gave a global *P* value for haplotypic association of 0.0007. The most common haplotype (H1), which had major alleles at all 7 sites, showed highly significant association ($P=3.3\times10^{-6}$) with a protective OR of 0.72 (95% CI 0.62-0.83). The minor allele of rs5984894 is included in haplotypes H2, H3, H4, and H6. Of these, H3 ($P=0.04$) and H4 ($P=0.01$) showed significant association with ORs of 1.21 (95% CI 1.01-1.46) and 1.36 (95% CI 1.07-1.74) respectively (see legend of Table 3 for additional discussion).

To extend our analysis of *PCDH11X*, three *PCDH11X* SNPs (rs5941047 and rs4568761 and rs2573905) that reside on the same haplotype block as rs5984894 were genotyped in all stage I + II subjects (2,524 AD, 2,698 control) from the JS (635 AD, 698 control), RS (577 AD, 1418 control), AUT (610 AD, 373 control), and NCRAD (702 AD, 209 control) series. rs5941047 and rs4568761 were followed-up in the stage II subjects because both had nominally significant *P* values of 0.0023 for allelic association in stage I. rs2573905 is located 8,483 bp 3′ of rs5984894. Both rs5984894 and rs2573905 reside deep in intron 2 of *PCDH11X* isoforms c and d, over 54 kb and 62 kb upstream of exon 3 respectively. rs2573905 was genotyped in the combined series because it is in a 100 bp region that is 70% conserved between the human and mouse sequence and therefore likely to be functionally relevant. All three SNPs were analyzed for association with LOAD using the Mantel-Haenszel method (Table 4, see legend for results of χ^2 tests). In the combined dataset, highly significant associations were observed for all three SNPs (Table 4) with *P* values for rs2573905, rs5941047, and rs4568761 of 1.6×10^{-7}, 8.0×10^{-5} and 0.001 respectively. Breslow Day *P* values for rs2573905, rs5941047, and rs4568761 were 0.55, 0.35, and 0.23 respectively indicating a lack of statistical evidence for series to series heterogeneity among the seven series tested.

rs2573905 is in strong linkage disequilibrium with rs5984894 ($r^2=0.98$, D'=0.99) and the minor alleles of these two SNPs occur on the same haplotypes (H2, H3, H4, and H6 in Table 3). Thus functional changes caused by rs2573905 may account for the strong association of rs5984894 with LOAD. Table 5 shows the results obtained when rs2573905 was analyzed by multivariable regression with sex as a covariate. Because of the strong LD between rs2573905 and rs5984894, this analysis of rs2573905 (Table 5) gave results for female heterozygotes, female homozygotes and male hemizygotes that were nearly the same as those for rs5984894 (Table 4) although the global *P* value for rs2573905 was more significant in the combined series (5.4×10^{-13} vs. 3.9×10^{-12}) where rs2573905 was genotyped successfully in more subjects (5,010 vs. 4,740).

Lopes *et al*. have proposed that known coding and expression level differences between *PCDH11X* and *PCDH11Y* may have functional consequences that could lead to sexually dimorphic traits[9]. Durand. *et al*. tested this idea with respect to common psychiatric disorders such as autism, ADHD, OCD and schizophrenia in which differences in risk/age of onset between females and males have been observed[10]. Although they found no statistically significant association with any of these traits, our data provide substantial evidence for an association between genetic variation in the *PCDH11X* gene and increased risk of LOAD in females.

PCDH11X/Y belong to the protocadherin gene subfamily of the cadherin superfamily of cell surface receptor molecules. The cadherins mediate cell-cell adhesion and play a role in cell

signaling that is critical in the development of the central nervous system[11]. The most recent studies of the *PCDH11X/Y* gene structure and expression report that these genes consist of at least 17 exons spanning over 700 kb. Alternative splicing produces multiple isoforms that are mainly expressed in the brain[12]. Expression is particularly strong in the cortex and hippocampus and weaker in the cerebellum[10]. Based on their splicing patterns and functional domains, it has been proposed that *PCDH11X/Y* resemble cadherin related neural receptors[12] which are known to localize at the synaptic junction[13]. Interestingly, some protocadherins are known to undergo presenilin-dependent processing[14].

In summary, the results of our two-stage GWAS provide the first evidence that genetic variation in *PCDH11X* is strongly associated with LOAD susceptibility in a combined American Caucasian sample of 2,391 cases and 2,464 controls. The SNP identified, rs5984894, resides in a haplotype block that falls entirely within *PCDH11X*, and it is in strong linkage disequilibrium with rs2573905, which is more likely to alter *PCDH11X* function since it resides in a conserved region. To date, however, no functional variants have been identified in this gene. Further study to determine how risk for LOAD is mediated by specific genetic variation in *PCDH11X* should improve understanding of the molecular basis of LOAD and open new therapeutic possibilities for this devastating disease.

Methods

Subjects

All case-control series consisted of Caucasian subjects from the United States ascertained at the Mayo Clinic or through the National Cell Repository for Alzheimer's Disease (NCRAD). This study was approved by the appropriate institutional review board and appropriate informed consent was obtained from all participants. A complete description of the study subjects can be found in the section on **Study populations and ascertainment** in the Supplementary Methods.

Our stage I GWAS was performed on JS, RS, and AUT subjects with an age at diagnosis/entry of 60-80 years. We genotyped 970 AD cases and 1,495 controls (JS: 381 AD, 350 control; RS: 291 AD, 787 control, AUT 298 AD, 358 CON). After stringent quality control (see **Stage I Quality Control** section below), we analyzed 844 AD cases and 1,255 controls (JS: 353 AD, 331 control; RS: 245 AD, 701 control, AUT 246 AD and 223 control).

Our stage II follow-up analysis of the 25 SNPs with the most significant allelic association in stage I was performed on JS, RS, and AUT subjects with an age at diagnosis/entry of over 80 years, and on additional samples obtained through the National Cell Repository for Alzheimer's Disease (NCRAD) with an age at diagnosis/entry of over 60 years. In stage II, we genotyped and analyzed 1,547 AD cases and 1,209 controls (JS: 237 AD, 260 control, RS: 276 AD, 624 control, AUT: 332 AD, 116 control, NCRAD: 702 AD, 209 control). One AD case from each of the 702 late-onset NCRAD families was analyzed. NCRAD AD cases were selected based on strength of diagnosis (autopsy-confirmed: 32% > probable: 45% > possible: 8% > family report: 15%); the case with the earliest age at diagnosis was taken when several cases had equally strong diagnoses. The 209 NCRAD controls that we employed are unrelated Caucasian subjects from the United States with a Clinical Dementia Rating of 0, specifically collected for inclusion in case-control series.

Age and gender data for the cases and controls in each series included in the stage I and stage II analyses are shown in Supplementary Table 1.

Sample collection, DNA isolation, and DNA amplification

Blood samples were collected in 10 ml EDTA tubes from subjects in the Mayo JS and RS series, and genomic DNA was isolated from whole blood using an AutoGenFlex STAR instrument (AutoGen, Inc, Holliston, MA). Genomic DNA from the cerebellum of subjects in the AUT series was obtained by Wizard® Genomic DNA Purification Kit (Promega Corp., Madison, WI). DNA from the RS and AUT series was scarce, so samples from these two series were subjected to whole genome amplification (WGA) using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Bio-Sciences Corp., Piscataway, NJ). To attenuate random amplification errors, we performed four 5 ul reactions for each sample, rather than a single 20 ul reaction. Each 5 ul reaction contained 5-15 ng of genomic DNA as template, according to the quality of the genomic DNA. These four reactions were then combined. To evaluate the quality of each WGA DNA sample, a TaqMan® SNP Genotyping Assay (Applied Biosystems, Foster City, CA) was used to obtain genotypes for SNP rs2830072 in both the original genomic (non-WGA) DNA and in the WGA DNA. Only WGA DNA samples that fell within well defined genotype clusters and that had genotype calls for rs2830072 that were in agreement with their non-WGA DNA genotypes were included in the series. In our hands, pooling four 5 ul reactions gave better genotype clusters and fewer miscalls than a single 20 ul reaction.

Genotyping Methods

The genotype data from stage I samples (n=2,465) was generated using HumanHap300-Duo Genotyping BeadChips processed with an Illumina BeadLab station (Illumina, SanDiego, CA) at the Mayo Clinic Genotyping Shared Resource (Rochester, Minnesota) according to the manufacturer's protocols. The HumanHap300-Duo chips allow simultaneous genotyping of two independent samples for 318,237 SNPs across the genome. Genotype calls were made using the auto-calling algorithm in Illumina's BeadStudio 2.0 software.

The genotype data from stage II samples (n=2,756) was generated using SEQUENOM's MassArray iPLEX technology (SEQUENOM Inc, San Diego, CA) following the manufacturers instructions. The follow up genotypes obtained for three SNPs (rs5941047 and rs4568761 and rs2573905) in the combined stage I + II series (n=5,222) were also generated using SEQUENOM's MassArray iPLEX technology. Genotype calls were made using the default post-processing calling parameters in SEQUENOM's Typer 4.0 software, followed by visual inspection to remove genotype calls that were obviously erroneous, based on the presence or absence of allele peaks in an individual sample's spectrogram.

Stage I Quality Control

In our stage I GWAS, we genotyped 318,237 SNPs in samples from 2,465 subjects. Genotype clusters were determined using Illumina's BeadStudio 2.0 software after first eliminating 240 samples (9.7%) with call rates of <90% on the first pass. This initial quality control measure eliminated a higher percentage of the WGA DNA samples. Of the 1,734 RS/AUT samples, all of which were WGA DNA, 213 (12.3%) had call rates of < 90%. Of the 731 JS samples, all of which were non-WGA DNA, 27 (3.7%) had call rates <90%.

We also eliminated 87 AUT samples (3.5%) with Braak stages of 3.0 or 3.5, so that all AUT AD samples had a Braak stage of 4.0 or greater, and all AUT control samples had a Braak stage of 2.5 or lower. Using filters available in PLINK[6], we eliminated all SNPs with call rates <90%, minor allele frequencies <0.01, and/or Hardy-Weinberg *P* values <0.001. Using the sex check option provided by PLINK, we identified and removed 21 additional samples (0.9%) with a mismatch between the recorded sex and the sex deduced by evaluating the heterozygosity of SNPs on the X chromosome. We also checked for cryptic relatedness by using the --genome option in PLINK to evaluate paired identity by descent in all samples genotyped in stage I. This check revealed 16 pairs with PI_HAT over 99% thereby identifying 16 subjects for which

two samples had been genotyped. Of these 32 samples, 14 were retained and 18 (0.8%) were eliminated. We eliminated one sample from 14 subjects where all samples had identical subject information and where we were able to confirm independently that the paired samples came from the same subject. We eliminated four samples (two pairs) where key subject information (e.g. gender, age) associated with the two samples was in conflict. Two DNA samples from different blood draws were genotyped in one RS subject where only one of the two samples was retained. The other 13 subjects that were retained had one DNA sample derived from the brain at autopsy (AUT), which was retained, and one derived from blood taken during life (JS or RS), which was eliminated. The duplicates in these 13 subjects went undetected because the identifiers for samples in the AUT samples, which mostly came from the brains of subjects who were not seen at Mayo, were not linked to the identifiers in the JS or RS series.

These quality control measures left 2,099 subjects (85.2%) in whom 313,504 SNPs (98.5%) were analyzed. Since rs5984894 was not successfully genotyped in all samples that met quality control criteria, this SNP was analyzed in a total of 2,024 subjects (96.4%) in stage I.

Of the 2,099 samples that met our quality control criteria in stage I, 1,415 were RS or AUT (WGA DNA) samples of good quality. The average call rate of 99.2% in the WGA samples was essentially identical to the average call rate of 99.3% in the 684 DNA samples from the JS series (non-WGA DNA). Thus the call rates using BeadChips were comparable for WGA and non-WGA DNA once WGA samples of poor quality were identified and eliminated.

The genotype clusters for the 25 SNPs with the most significant *P* values in stage I (Supplementary Table 3) were visually inspected as an additional quality control check. This check showed that the three SNPs noted with an asterisk in Supplementary Table 3 (rs3858095, rs2318144, and rs3007421) had unsatisfactory clusters that caused inaccurate genotyping of many heterozygotes and minor allele homozygotes. This is evidenced by the much higher minor allele frequencies observed in controls of the follow-up series where all three SNPs were genotyped well using SEQUENOM iPLEX technology. Two of these SNPs (rs3858095 and rs3007421) were eliminated by increasing the stage I call rate cut-off for samples and SNPs from 90% to 95%, but rs2318144 was not eliminated even when the call rate cut-off for samples and SNPs was increased to 98%. rs2318144 had a Hardy-Weinberg *P* value of 0.05 and therefore also failed to be eliminated by our cut-off *P* value of 0.001. These results underscore the importance of checking SNPs with highly significant association by visually inspecting their genotype clusters and by genotyping follow-up series on a different platform. The genotype clusters for rs5984894, which had a 96.4% call rate in the samples that met quality control criteria in stage I, are shown in Supplementary Figure 1a.

A subset of the stage I samples (total n=347: JS n=84, RS n=183, AUT n=80), for which the HumanHap300 call rate was >0.90, were also genotyped using the iPLEX method employed with the stage II samples, in order to test for genotype concordance between the two genotyping platforms that were utilized. The genotype call concordance rate between iPLEX and BeadChip for the 22 SNPs followed-up in stage II that had satisfactory BeadChip genotype clusters was 99.8%. The genotype call discordance rates for the JS samples (non-WGA) and the RS+AUT samples (WGA) were essentially identical (JS=0.0022 vs. RS+AUT=0.0009).

The effect of eliminating samples and SNPs using call rate cut-offs of 95% as compared to 90% (Supplementary Table 2) is discussed in Supplementary Methods online.

Stage II Quality Control

To be sure that each subject was sampled one time only, stage II samples were checked for cryptic relatedness using 138 SNPs genotyped in all of those samples. Among the samples chosen initially for analysis in stage II (1,594 ADs and 1,221 controls), there were 52 subjects

in whom multiple samples had been genotyped (51 had two samples, one had three samples). Of these 105 samples, 46 were retained and 59 were eliminated. We retained one sample from 46 subjects where all samples had identical subject information and where we were able to confirm independently that all samples came from the same subject. For 6 subjects with duplicate samples, key subject information (e.g. gender, age) associated with the two samples was in conflict, so both samples were eliminated. Thus we assured that a single sample was analyzed in the stage II subjects on which we report (1,547 ADs and 1,209 controls).

All of the iPLEX genotype cluster plots for the variants genotyped in the follow-up series were visually inspected to be sure that each genotype that was called fell within a well defined cluster. The overall call rate for the 25 SNPs was 98.3%. Nineteen SNPs including rs5984894, the SNP in *PCDH11X* that showed highly significant association, had call rates of 99%. The remaining SNPs had call rates of 93-98%. Call rates were similar in WGA (98.2%) and non-WGA (98.9%) DNA samples. The genotype clusters for rs5984894 in stage II are shown in Supplementary Figure 1b.

Statistical Analyses

Genotype reports produced by Illumina BeadStudio 2.0 software (stage I data) or SEQUENOM Typer 4.0 software (stage II data) were used to generate lgen, map, and fam files that were imported into PLINK. The SNP genotypes in stage I, stage II, and the combined datasets were analyzed for allelic association with AD using the allelic association χ^2 test implemented in PLINK. With α=0.05 and Bonferroni correction for the 313,504 SNPs tested in stage I, a *P* value of 1.6×10^{-7} is required for "genome-wide" significance. Using this criterion, the only SNPs to achieve genome-wide significance in stage I were six *APOE*-linked SNPs. In the combined stage I + II data, rs5984894 and one additional *APOE*-linked SNP also achieved this level of significance (Supplementary Table 3).

Since rs5984894 showed highly significant association with LOAD in stages I, II, and I + II, and this novel LOAD SNP was analyzed in seven distinct American Caucasian case-control series, we analyzed it in PLINK using not only the χ^2 test on combined allele counts but also the Mantel-Haenszel method in which the Breslow Day option was employed to test for series to series heterogeneity (Table 1). We also analyzed rs5984894 by multivariable logistic regression (Table 2) as described in the Supplementary Methods.

The solid spine haplotype block definition in Haploview 4.0[15] was employed to generate a linkage disequilibrium plot of the genomic region encompassing *PCDH11X* and to evaluate the 7 haplotypes formed by the SNPs included in the haplotype block containing rs5984894. Using haplotype counts provided by Haploview 4.0, ORs were calculated for each of the 7 haplotypes, and *P* values were obtained with a χ^2 test. A χ^2 test was also employed to calculate a global *P* value for haplotypic association.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Support for this research was provided by the NIH grants: NIA R01 AG18023 (N.R.G-R, S.G.Y); Mayo Alzheimer's Disease Research Center: P50 AG16574 (R.C.P, D.W.D, N.R.G-R, S.G.Y); Mayo Alzheimer's Disease Patient Registry: U01 AG06576 (R.C.P); NIA AG25711, AG17216, AG03949 (D.W.D). Samples from the National Cell Repository for Alzheimer's Disease (NCRAD), which receives government support under a cooperative agreement grant (U24 AG21886) awarded by the National Institute on Aging (NIA), were used in this study. We thank contributors, including the Alzheimer's Disease Centers who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible. This project was also generously supported

by the Robert and Clarice Smith Postdoctoral Fellowship (M.M.C.); Robert and Clarice Smith and Abigail Van Buren Alzheimer's Disease Research Program (R.C.P., D.W.D., N.R.G-R; S.G.Y) and by the Palumbo Professorship in Alzheimer's Disease Research (S.G.Y.).

References

1. Evans DA, et al. Prevalence of Alzheimer's disease in a community population of older persons. Higher than previously reported. Jama 1989;262:2551–6. [PubMed: 2810583]
2. Corder EH, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261:921–3. [PubMed: 8346443]
3. Corder EH, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet 1994;7:180–4. [PubMed: 7920638]
4. Farrer LA, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. *APOE* and Alzheimer Disease Meta Analysis Consortium. Jama 1997;278:1349–56. [PubMed: 9343467]
5. Gatz M, et al. Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 2006;63:168–74. [PubMed: 16461860]
6. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75. [PubMed: 17701901]
7. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9. [PubMed: 16862161]
8. Williams NA, Close JP, Giouzeli M, Crow TJ. Accelerated evolution of Protocadherin11X/Y: a candidate gene-pair for cerebral asymmetry and language. Am J Med Genet B Neuropsychiatr Genet 2006;141:623–33. [PubMed: 16874762]
9. Lopes AM, et al. Inactivation status of *PCDH11X*: sexual dimorphisms in gene expression levels in brain. Hum Genet 2006;119:267–75. [PubMed: 16425037]
10. Durand CM, et al. Expression and genetic variability of *PCDH11Y*, a gene specific to Homo sapiens and candidate for susceptibility to psychiatric disorders. Am J Med Genet B Neuropsychiatr Genet 2006;141:67–70. [PubMed: 16331680]
11. Blanco P, Sargent CA, Boucher CA, Mitchell M, Affara NA. Conservation of *PCDHX* in mammals; expression of human X/Y genes predominantly in brain. Mamm Genome 2000;11:906–14. [PubMed: 11003707]
12. Blanco-Arias P, Sargent CA, Affara NA. Protocadherin X (*PCDHX*) and Y (*PCDHY*) genes; multiple mRNA isoforms encoding variant signal peptides and cytoplasmic domains. Mamm Genome 2004;15:41–52. [PubMed: 14727141]
13. Senzaki K, Ogawa M, Yagi T. Proteins of the CNR family are multiple receptors for Reelin. Cell 1999;99:635–47. [PubMed: 10612399]
14. Haas IG, Frank M, Veron N, Kemler R. Presenilin-dependent processing and nuclear function of gamma-protocadherins. J Biol Chem 2005;280:9313–9. [PubMed: 15611067]
15. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263–5. [PubMed: 15297300]

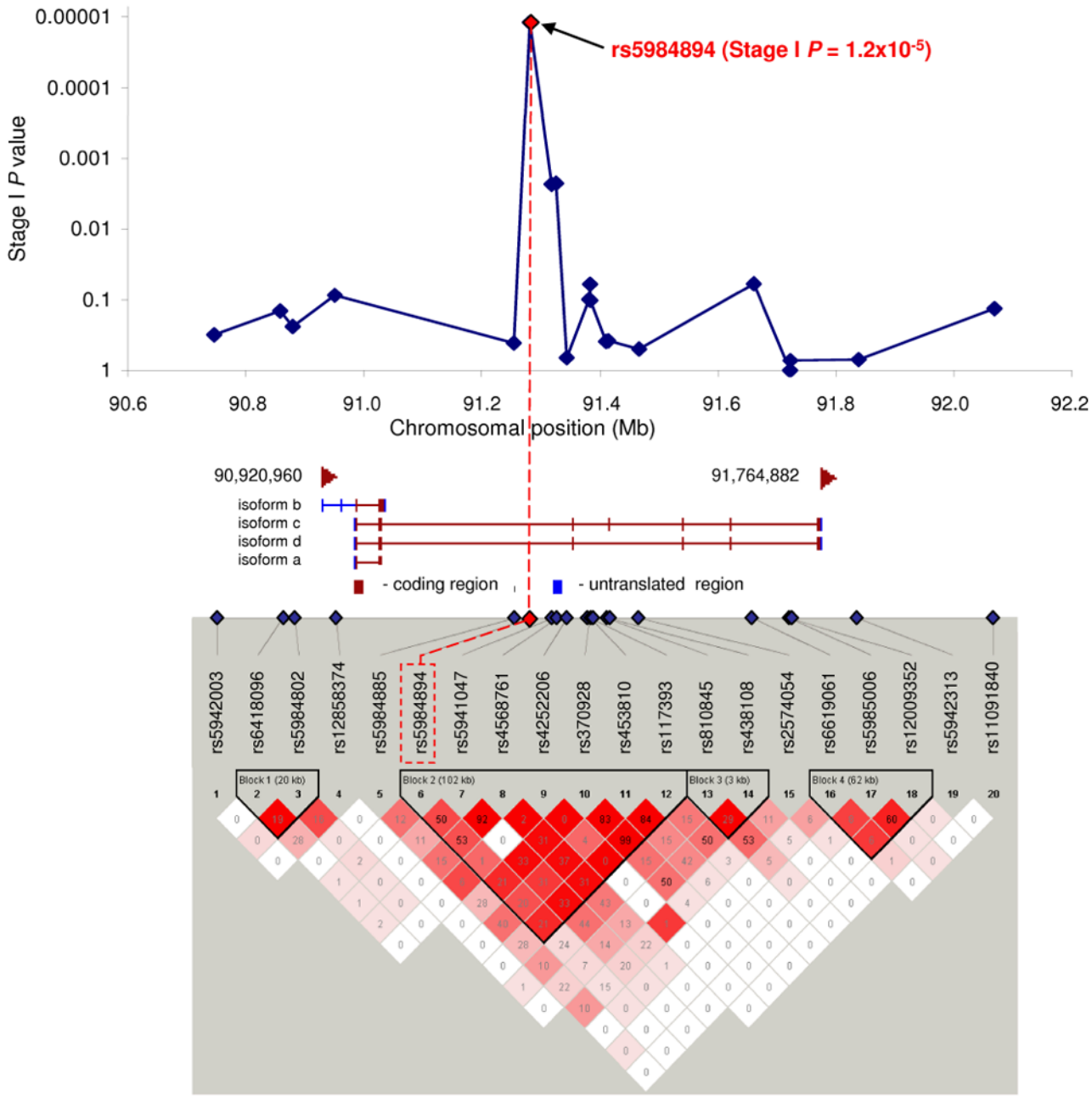

Figure 1.

Schematic overview of *PCDH11X* and LD plot showing *PCDH11X* haplotype blocks. Unadjusted allelic association *P* values from stage I for variants encompassing the *PCDH11X* locus are plotted over physical distance above the *PCDH11X* gene diagram. The four *PCDH11X* RefSeq isoforms and their chromosomal positions are depicted as in Entrez Gene (build 36.3). The LD plot shown is for variants in the *PCDH11X* locus (stage I data in Haploview 4.0, solid spine haplotype block definition, r^2 values with D'color scheme).

Table 1

Descriptive statistics and allelic association results for SNP rs5984894.

	N		MAF[a]		HWE[b]			
Series	**Cases**	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**	***P* value[c]**	**OR (95% CI)[d]**
Stage I								
JS 60-80	350	323	0.52	0.44	0.89	0.19	0.006	1.40 (1.10-1.77)
RS 60-80	235	669	0.53	0.46	0.40	0.24	0.01	1.35 (1.06-1.71)
AUT 60-80	239	208	0.55	0.46	1.00	0.66	0.02	1.44 (1.05-1.96)
Stage I combined[e]	824	1200	0.53	0.45	0.85	1.00	1.5×10^{-5}	1.39 (1.20-1.61)
Stage II								
JS 80+	232	254	0.50	0.47	0.41	0.42	0.50	1.10 (0.83-1.46)
RS 80+	275	615	0.54	0.45	0.05	0.52	0.001	1.45 (1.16-1.81)
AUT 80+	328	106	0.52	0.52	0.51	0.28	0.83	0.96 (0.68-1.37)
NCRAD 60-80	697	209	0.51	0.46	0.19	0.86	0.10	1.23 (1.08-1.57)
Stage II combined[e]	1532	1184	0.51	0.46	0.31	0.76	0.002	1.23 (1.08-1.40)
Stage I + II combined[e]	2356	2384	0.52	0.46	0.47	0.82	2.2×10^{-7}	1.30 (1.18-1.43)

[a] Minor allele frequency in cases and controls for each series. MAF was not different between males and females in controls.

[b] Hardy-Weinberg equilibrium *P* values for female cases and female controls in each population.

[c] *P* values were calculated for each individual series using a χ^2 test on allele counts.

[d] Odds ratios (OR) were calculated for the minor allele in each series; 95% confidence intervals are shown in parentheses.

[e] *P* values and ORs using data from multiple series were calculated using the Mantel-Haenzel method.

Table 2

Logistic regression results for rs5984894 comparing male hemizygotes, female heterozygotes, and female homozygotes to the female non-carriers, using male sex as covariate. For the effect of age and *APOE* ε4 as covariates see Supplementary Table 5 online.

	Sex		Male Hemizygotes		Female Heterozygotes		Female Homozygotes		
Series	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***	**Global *P***
Stage I									
JS 60-80	1.28 (0.78-2.11)	0.33	1.28 (0.79-2.09)	0.31	1.66 (1.04-2.63)	0.03	1.96 (1.14-3.36)	0.01	0.09
RS 60-80	1.00 (0.58-1.72)	0.99	1.20 (0.76-1.90)	0.43	1.46 (0.87-2.44)	0.16	2.02 (1.12-3.64)	0.02	0.04
AUT 60-80	0.79 (0.41-1.53)	0.48	1.40 (0.85-2.32)	0.19	1.55 (0.77-3.12)	0.22	2.00 (0.91-4.40)	0.09	0.03
Stage I combined	0.98 (0.72-1.33)	0.90	1.33 (1.02-1.74)	0.04	1.43 (1.06-1.92)	0.02	1.92 (1.36-2.70)	0.0002	5.7×10^{-5}
Stage II									
JS 80+	1.33 (0.73-2.44)	0.35	0.96 (0.54-1.71)	0.89	1.58 (0.91-2.72)	0.10	1.33 (0.70-2.55)	0.38	0.60
RS 80+	0.75 (0.46-1.22)	0.25	1.19 (0.74-1.91)	0.48	1.04 (0.67-1.62)	0.86	2.28 (1.39-3.73)	0.001	1.0×10^{-4}
AUT 80+	0.44 (0.19-1.01)	0.05	0.97 (0.49-1.92)	0.94	0.76 (0.36-1.61)	0.47	0.97 (0.39-2.39)	0.95	0.05
NCRAD 60+	1.11 (0.67-1.84)	0.67	0.95 (0.58-1.58)	0.86	1.19 (0.75-1.88)	0.47	1.71 (0.98-2.97)	0.06	0.35
Stage II combined	0.89 (0.69-1.15)	0.37	1.04 (0.82-1.33)	0.74	1.19 (0.94-1.50)	0.15	1.70 (1.29-2.24)	0.0002	4.8×10^{-6}
Stage I + II combined	0.86 (0.71-1.05)	0.14	1.18 (0.99-1.41)	0.07	1.26 (1.05-1.51)	0.01	1.75 (1.42-2.16)	2.0×10^{-7}	3.9×10^{-12}

Nat Genet. Author manuscript; available in PMC 2010 May 19.

Table 3

Stage I association results for block 2 haplotypes. The global *P* value for haplotype association was 0.0007. From left to right (5′ to 3′), the SNPs in each haplotype are rs5984894, rs5941047, rs4568761, rs4252206, rs370928, rs453810 and rs117393. Minor alleles are underlined. As noted in the text, the minor A allele of rs5984894 occurs on H2, H3, H4, and H6. H3 and H4 show significant association that is stronger than the association of H2 and H6 with LOAD. This could suggest that there is an untyped functional allele associated with H3 and H4, but the ORs for the four haplotypes are not significantly different from each other. Thus the differences observed may have occurred by chance alone.

		Stage I Haplotype Counts (Frequency)			
Block 2 Haplotypes	**Haplotype Name**	**AD**	**Control**	**OR (95% CI)**	***P* value**
GAAAGCG	H1	554.7/768.3 (0.419)	949.0/940.0 (0.502)	0.72 (0.62 to 0.83)	3.3E-06
AGGAAAA	H2	266.7/1056.3 (0.202)	346.7/1542.3 (0.184)	1.12 (0.94 to 1.35)	0.20
AGGAGCG	H3	264.9/1058.1 (0.200)	323.0/1566.0 (0.171)	1.21 (1.01 to 1.46)	0.04
AAAAGCG	H4	146.9/1176.1 (0.111)	158.4/1730.6 (0.084)	1.36 (1.07 to 1.74)	0.01
GGGAGCG	H5	22.0/1301.0 (0.017)	38.5/1850.5 (0.020)	0.81 (0.46 to 1.42)	0.44
AAGCGAG	H6	16.7/1306.3 (0.013)	25.5/1863.5 (0.014)	0.93 (0.49 to 1.86)	0.83
GGGAGAG	H7	16.1/1306.9 (0.012)	19.0/1870.0 (0.010)	1.21 (0.57 to 2.47)	0.57

Table 4

Descriptive statistics and allelic association results for SNPs rs2573905, rs5941047 and rs4568761 in the combined stage I + II series.

	n		MAF[a]		HWE[b]			
SNP	Cases	Controls	Cases	Controls	Cases	Controls	*P* value[c]	OR (95% CI)[c]
rs2573905	2449	2561	0.52	0.46	0.33	0.67	1.6×10^{-7}	1.29 (1.17-1.42)
rs5941047	2461	2576	0.44	0.39	0.18	1.00	8.0×10^{-5}	1.21 (1.10-1.34)
rs4568761	2456	2572	0.46	0.42	0.24	0.55	0.001	1.17 (1.07-1.29)

[a] Minor allele frequency in cases and controls. MAFs were not different between males and females in controls.

[b] Hardy-Weinberg equilibrium *P* values for female cases and female controls in each population.

[c] *P* values and odds ratios (OR) were calculated for the minor allele using the Mantel-Haenzel method; 95% confidence intervals are shown in parentheses. Analysis of allelic association using a χ^2 test gave *P* values of 6.6×10^{-8}, 4.4×10^{-5}, and 0.0001 for rs2573905, rs5941047, and rs4568761 respectively with odds ratios (95% CI) of 1.28 (1.17-1.39), 1.20 (1.10-1.32), and 1.19 (1.09-1.30).

Nat Genet. Author manuscript; available in PMC 2010 May 19.

Table 5

Logistic regression results for rs2573905 comparing male hemizygotes, female heterozygotes and female homozygotes to the female non-carriers, using male sex as covariate.

	Sex		Male Hemizygotes		Female Heterozygotes		Female Homozygotes		
Series	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***	**Global *P***
Stage I									
JS 60-80	1.18 (0.73-1.92)	0.50	1.27 (0.79-2.04)	0.33	1.48 (0.94-2.33)	0.09	2.01 (1.19-3.42)	0.009	0.08
RS 60-80	0.82 (0.51-0.32)	0.42	1.22 (0.81-1.86)	0.34	1.17 (0.74-1.85)	0.49	1.61 (0.96-2.71)	0.07	0.06
AUT 60-80	0.89 (0.47-1.68)	0.72	1.39 (0.88-2.20)	0.16	1.71 (0.89-3.30)	0.11	2.48 (1.15-5.36)	0.02	0.01
Stage I combined	0.93 (0.69-1.24)	0.62	1.30 (1.01-1.67)	0.04	1.32 (1.00-1.75)	0.05	1.85 (1.34-2.55)	0.0002	3.2×10^{-5}
Stage II									
JS 80+	1.28 (0.70-2.33)	0.42	0.96 (0.54-1.70)	0.88	1.6 (0.92-2.76)	0.09	1.41 (0.74-2.69)	0.29	0.54
RS 80+	0.76 (0.47-1.23)	0.26	1.15 (0.72-1.86)	0.55	1.01 (0.65-1.57)	0.95	2.16 (1.32-3.53)	0.002	2.5×10^{-4}
AUT 80+	0.46 (0.21-1.02)	0.06	1.06 (0.54-2.07)	0.86	0.80 (0.38-1.66)	0.55	1.03 (0.43-2.47)	0.94	0.07
NCRAD 60+	1.09 (0.66-1.80)	0.75	0.97 (0.58-1.61)	0.91	1.19 (0.75-1.89)	0.47	1.69 (0.97-2.93)	0.06	0.35
Stage II combined	0.87 (0.68-1.12)	0.29	1.05 (0.82-1.34)	0.69	1.17 (0.93-1.48)	0.17	1.68 (1.27-2.20)	0.0002	3.9×10^{-6}
Stage I + II combined	0.84 (0.70-1.02)	0.07	1.17 (0.98-1.39)	0.08	1.22 (1.02-1.45)	0.03	1.72 (1.40-2.12)	0.0001	5.4×10^{-13}

Nat Genet. Author manuscript; available in PMC 2010 May 19.