Global and Local Ancestry in African Americans: Implications for Alzheimer's Disease Risk

A full list of authors and affiliations appears at the end of the article.

Abstract

African American (AA) individuals have a higher risk for late-onset Alzheimer's disease (LOAD) than Americans of primarily European ancestry (EA). Recently, the largest genome-wide association study in AAs to date confirmed that six of the AD-related genetic variants originally discovered in EA cohorts are also risk variants in AA; however, the risk attributable to many of the loci (e.g., *APOE, ABCA7*) differed substantially from previous studies in EA. There likely are risk variants of higher frequency in AAs that have not been discovered. We performed a comprehensive analysis of genetically determined local and global ancestry in AAs with regard to LOAD status. Compared to controls, LOAD cases showed higher levels of African ancestry, both globally and at several LOAD relevant loci, which explained risk for AD beyond global differences. Exploratory post-hoc analyses highlight regions with greatest differences in ancestry as potential candidate regions for future genetic analyses.

Keywords

local admixture; local ancestry; Alzheimer's disease; genome-wide association analysis (GWAS); African American; admixture mapping

Background

Late-onset Alzheimer's disease (LOAD) is a debilitating neurodegenerative disease with 4.7 million cases reported in the United States in 2010, a number that is projected to increase three-fold by the year 2050.[1] The strongest genetic risk factor for LOAD – the ε4 variant of the *APOE* gene on chromosome 19 – was identified in 1993 and increases risk for LOAD in a dose-dependent manner.[2] Over the past ten years a number of genome-wide association studies (GWAS) have identified and replicated effects in 20 other loci that explain variance in LOAD risk.[3–7] Taken together, these loci are estimated to explain approximately 30–40% of the total heritability for LOAD,[8,9] and yet this still falls substantially below the 60–80% heritability expected based on prior estimates from twin studies.[10] Multiple strategies, including the identification of rare variants and gene-gene interactions, will be needed to successfully explain all genetic variation associated with LOAD.[8]

While the number of GWAS has increased substantially in recent years, the majority of these studies have focused on individuals of mostly western European ancestry. This is particularly relevant because previous work has suggested the prevalence of LOAD may be higher in African American (AA) individuals than in European Americans (EA) within the same community,[11] although findings have been somewhat variable depending on the geographic location from which the sample was ascertained.[12] Recently, a GWAS of LOAD in a large sample of AA individuals replicated many of the previous risk loci identified in EA individuals (*APOE, ABCA7, CR1, BIN1, EPHA1, CD33*).[13] Perhaps more importantly, however, was the discovery that in this AA dataset, the amount of risk attributable to *APOE* and *ABCA7* differed substantially from previous studies in EA. This is interesting because the association between *APOE* genotype and AD differs by ancestral background. For example, previous work suggested that there is no effect of *APOE* genotype in Nigerian populations.[14] More recent work has suggested that a homozygous effect is in fact present in Nigerian populations, but the overall effect of *APOE* on AD and cognitive performance is attenuated relative to African American populations (AA).[15] Therefore, the finding by Reitz and colleagues that the effect of *APOE* is reduced and that of *ABCA7* is increased in AAs relative to EAs further suggests that ancestral background, particularly in African Americans, might be relevant to calculations of AD risk.[13]

Other recent work has reported that a higher percentage of genetically determined African ancestry in Brazilian individuals is associated with lower levels of LOAD related neuropathology,[16] however it is unclear whether such a finding would extend to African Americans given the skew toward a moderate to low percentage of African ancestry in the Brazilian cohort. Thus, while the total genetic risk for LOAD may be comparable between EA and AA individuals, the findings to date suggest the risk profiles of specific genetic loci might vary by ancestral genomic background. A comprehensive analysis of genetically determined ancestry in African Americans could both explain some of the differences in genetic risk profiles across ancestral groups and significantly improve our understanding of the pathogenesis of the disease in general.

Genetic ancestry can be estimated in two different ways. "Global" genetic ancestry is an estimation of the percentage of markers across the entire genome that are inherited from a

given ancestral population. This is often estimated using ancestry-informative markers that are known to differentiate one population group from another. "Local" ancestry is an estimate of the percentage of ancestry at a given genetic locus based on genomic inheritance across ancestral blocks. Our study design takes advantage of the recent admixture in AA to search for loci relevant to LOAD and uses information about both levels of analysis to better understand ancestral differences in AD risk. Our first analysis highlights global differences in genetically determined African ancestry between cases and controls (differences in genetic ancestry calculated across the entire genome). Secondly, we demonstrate that there is localized variation in ancestry, particularly at regions known to contain LOAD-relevant markers, that at least in part drive this global difference. Finally, we present a comprehensive picture of the differences in African ancestry across the genome, highlighting those loci that show the largest differences in ancestry between cases and controls and, therefore, are likely to harbor novel candidate loci with risk profiles that differ by ancestry.

Methods

Subjects

Data were provided by the Adult Changes in Thought study (ACT), the Chicago Health and Aging Project, the National Institute on Aging-Late-Onset Alzheimer's Disease/National Cell Repository for Alzheimer's Disease (NIA-LOAD), Indianapolis University, the Mount Sinai School of Medicine, the Religious Orders Study/Rush Memory and Aging Project/ Minority Aging Research Study/Clinical Minority Core at Rush University, the University of Miami/Vanderbilt University, the University of Pittsburgh, the Washington Heights Columbia Aging Project, and Washington University. Complete details on this sample have been published previously (dbGaP accession phs000372.v1.p1).[13]

For this analysis, we included 6,250 African American subjects with genotype data from 10 datasets that were contributed to the Alzheimer's Disease Genetics Consortium (ADGC).[13] Of those, 342 subjects were removed because they did not have phenotype data available, 143 were removed because they did not have covariate data available or were under the age of 60 at diagnosis, and 60 were removed because they were outliers in global ancestry (greater than three standard deviations beyond the mean, global ancestry is explained in detail below) leaving a total of 5,644 subjects for local ancestry analysis. Demographic data are presented in Table 1. Diagnostic status was determined following the National Institute of Neurological and Communicative Disorders and Stroke—Alzheimer's Disease and Related Disorders Association criteria.[17] Demographic characteristics are presented for the combined dataset used in this analysis, however additional demographic information stratified by dataset has been published previously.[13]

Genotyping

Genotyping was performed on various Illumina platforms including the 1M, 660k, 610k,Omni Express, and 300k chips (additional details about genotyping quality control by dataset has been published previously).[13] Quality control (QC) procedures were performed using PLINK v.2 (*http://pngu.mgh.harvard.edu/~purcell/plink/*). SNPs with MAF less than

1%, call rates less than 98%, or not in Hardy-Weinberg equilibrium ($p < 10^{-6}$ in controls) were excluded. Participants whose genetically determined sex did not match their reported sex, or for whom relatedness to another sample was established (PI_HAT 0.4), were excluded. Additional details about these QC procedures in this sample have been published.[13] All QC procedures were completed within each cohort, and when multiple chips were used for genotyping within a cohort QC was performed within each chip. We then merged all QC'd data into one combined dataset which included those markers present in any dataset (1,281,137 markers). However, all files were in Illumina's A/B SNP format, so we updated these alleles using the Illumina specification file, which reduced our total number of SNPs to 968,202. Then we selected only those SNPs that were available in Hapmap because we had to have all SNPs on our reference panels to perform the analysis (767,929 SNPs). Finally, we filtered down to those SNPs that were available in 98% of samples across all datasets (269,379 SNPs). The total genotyping rate in this final merged dataset was 98.4%.

Global Ancestry Analysis

Global ancestry was calculated using Admixture.[18] HapMap phase III data from CEU and YRI populations were used as reference populations in the global and local ancestry analysis. Ancestry was calculated using markers that were present in our genotyped sample and in the HapMap sample. The genotyped and reference populations were recoded to ensure all files had the same reference allele. Individual estimates of global ancestry were imported to R (http://www.r-project.org/) for group statistical analyses. First, we tested for differences in the variance of global ancestral estimates across cases and controls using the Levene test.[19] Next, we performed an independent samples t-test assuming unequal variance to test for differences in global ancestry between cases and controls. Follow-up analyses, in which covariates were included, were run using a binary logistic regression model utilizing the generalized linear model in SPSS v.22. Case/control status was set as the outcome, global ancestry was set as a predictor, and covariates were entered into the model as outlined in the result. The site covariate was entered as a categorical predictor, and the site x global ancestry test statistic was evaluated for whether the effect of global ancestry on case/control status differed across datasets.

Local Ancestry Analysis

Local ancestry was calculated using LAMP-LD.[20,21] First, the genotyped and reference populations were recoded to ensure all files had the same reference allele. Next, reference population files were phased using Shape-It version 2.[22] After running Lamp-LD, African ancestry estimates were calculated in R as the number of YRI-derived alleles at a given locus (0–2) divided by the total number of alleles (2 * number of subjects), for cases and controls separately. Difference in ancestry was then tested for significance at each locus using a Wilcoxon rank sum test. Plotting was performed using the R package ggplot2 (http://ggplot2.org/). Statistical correction was performed using the FDR procedure as outlined below in the two analysis subsections. Follow-up analyses, in which covariates were included, were again run using binary regression utilizing the generalized linear model in SPSS.

SNP Selection for Disease Relevant Loci

LOAD relevant SNPs were selected based on previously performed meta-analyses in European American[5,23] and African American[13] subjects. Out of the 34 SNPs implicated in those meta-analyses, nine were present in the current dataset: rs3764650 (*ABCA7*), rs3865444 (*CD33*), rs11771145 (E*PHA1*), rs10498633 (SLC24A4), rs3851179 (*PICALM*), rs3818361 (*CR1*), rs17125944 (*FERMT2*), rs744373 (*BIN1*), rs610932 (*MS4A6A*). We used proximal SNPs for the other the loci based on the nearest SNP to the target location in our dataset (< 28 Kb, see Table 2 for additional details). The relevance of the proximal SNP was evaluated using D` because, unlike $R^2$, D` is a metric of historical recombination and ensures that proximal SNP selected is within the same unbroken ancestral block, and thus would be most likely to fall within the same peak region in ancestry analyses. D` ranges from 0 to 1 where 0 indicates no linkage and 1 indicates perfect linkage. FDR correction was performed for the total number of unique SNPs analyzed (FDR < 0.05).

Post-hoc Genome Wide Scan

For the post-hoc genome-wide analysis, the same local ancestry analysis procedures were applied across the genome. We report all peaks which include SNPs where FDR < 0.05. To help identify more focal candidate regions within the large ancestral peaks, we calculated the difference in MAF between cases and controls at each SNP within a given peak. These differences were intended only to highlight potential markers within the peak that may be driving the observed differences in disease risk, and to provide hints at potential mechanism of such an ancestral effect. For SNPs that passed correction for multiple comparisons, we also reran analyses controlling for age, sex, and global ancestry in a binary logistic regression model with diagnosis set as the outcome. Additional annotation information including GENCODE annotations and dbSNP functional annotations were pulled using Haploreg (v. 3).[24]

Results

Differences in Global Ancestry

Demographic characteristics of the sample are presented in Table 1. Cases and controls did not have equal variance (p = 0.027). For that reason, we performed an independent sample t-test assuming unequal variance and found cases showed higher levels of African ancestry than controls, t(3469.66) = 2.46, p = 0.013 (Figure 1). When including covariates in the model, the effect of global ancestry remained statistically significant (OR = 1.16, p = 0.003). Results also remained significant when including a site covariate (OR = 1.75, p = 0.021). The dataset x site interaction term showed no significant difference between the datasets (Wald χ^2 = 13.2 [df = 8], p = 0.10) suggesting site effects were not driving our result.

Differences in Local Ancestry at Known Disease Relevant Loci

We tested whether differences in ancestry existed at risk loci confirmed in previously published EA GWAS of LOAD.[3–5,7] We chose to use the genotyped data rather than 1000 Genomes imputed data so as to ensure that we did not bias our estimates of local ancestry during the imputation process. Therefore, of the 34 previously published associated SNPS, 9

were present in our dataset. For the SNPs that were not genotyped in both patient and reference panels, we used the closest SNP in this evaluation (LD statistics presented in Table 2). SNPs within *ABCA7, CD33,* and *GRIN3B* showed statistically significant differences in ancestry when correcting for multiple comparisons (Table 2). Only *ABCA7* and *GRIN3B* survived a more conservative correction using the Bonferroni procedure. The *ABCA7* and *CD33* SNPs had been genotyped directly, so we further evaluated whether local ancestry at these loci explained the difference in clinical status using logistic regression and covarying for global ancestry, *APOE* genotype, age, sex, and the given genotype. *APOE* genotype was modeled first using an ε4 additive coding (0 – no ε4, 1- one copy ε4, 2 – two copies of ε4), and then reran using a full allelic model (ε2/ε2, ε2/ε3, ε2/ε4, ε3/ε3, ε3/ε4, ε4/ε4) with ε3/ε3 coded as the referent group. The breakdown of *APOE* genotypes by diagnosis is presented in Table 1. For the *ABCA7* SNP (rs3764650), local ancestry was significantly associated with disease status when covarying for *APOE* genotype, age, and sex (OR = 1.36, p = 0.007) and remained significant when including global ancestry and the genotype at rs3764650 in the model (OR = 1.32, p = 0.031). The full allelic coding of *APOE* did not alter our results (OR = 1.30, p = 0.042). Again, to ensure there were no site differences, we also included site as a covariate, which had no effect on our result (OR = 1.33, p = 0.009), and the site x local ancestry interaction term was not statistically significant (Wald χ^2 = 3.76 [df = 8], p = 0.878), suggesting site effects were not driving our result.

Similarly, for the *CD33* SNP (rs3865444), local ancestry predicted disease status when covarying for *APOE (*additive coding), age, and sex (OR = 1.31, p = 0.013), and remained significant when including global ancestry and the genotype at rs3865444 in the model (OR = 1.32, p = 0.035). The full allelic coding of *APOE* did not alter our results (OR = 1.36, p = 0.019). Including site as a covariate did not alter this result (OR = 1.36, p=0.003), and the site x local ancestry interaction term was not statistically significant (Wald χ^2 = 7.986 [df = 8], p = 0.435), suggesting that site effects were not driving this result.

Differences in Local Ancestry across the Genome

In exploratory posthoc analyses, we analyzed differences in ancestry across the entire genome. The largest differences in local ancestry were seen on chromosome 1 (lowest p = 1.3×10^{-4}, FDR = 0.037), chromosome 2 (lowest p = 2.9×10^{-5}, FDR = 0.037), chromosome 4 (lowest p = 1.5×10^{-5}, FDR= 0.037), chromosome 12 (lowest p = 3.3×10^{-4}, FDR = 0.042), and chromosome 19 (lowest p = 8.9×10^{-5}, FDR = 0.037). Additional details for all significant loci are presented in Supplementary Table 1. Given the range of observed effect sizes (0.09 – 0.12), our sample size, and our nominal p-value when performing an FDR correction (p = 0.0005), we achieved between 45% and 83% power.

On chromosome 1, the strongest signal came from rs12142787 in the cytochrome P450m family 4, subfamily B, polypeptide 1 gene (*CYP4B1*), although the region passing FDR correction included two small peaks approximately 12 Mb apart with the strongest signal in the second region coming from rs2806403 annotated to Dab reelin signal transducer homolog 1 (*DAB1;* Figure 2a). To identify potential genes of interest within these large peaks, we looked for SNPs that passed FDR correction and also showed a large deviation in minor allele frequency (MAF) between cases and controls (difference > 0.025). Five SNPs

showed such a difference, with four falling in or around *CYP4B1* (rs837401, rs6679068, rs863915, rs4646484), one was 5kb downstream of the Myb-like, SWIRM and MPN domains 1 gene (*MYSM1*). The two SNPs with the next highest difference in MAF between cases and controls (difference > 0.024) included a missense mutation within *CYP4B1* (rs2297810) and a SNP within the 3`-UTR of *MYSM1* (rs232777).

On chromosome 2, the strongest signal came from rs2286250 within the plakophilin 4 gene (*PKP4*). The region passing FDR correction was approximately 17 Mb in length (Figure 2b). Within this region, 39 SNPs showed large differences in MAF between cases and controls (difference > .025). Of these 39 SNPs, seven SNPs were in or around enhancer of polycomb homolog 2 (*EPC2*), seven SNPs were in or around formin-like 2 (*FMNL2*), seven were in or around UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 5 (*GALNT5*), five SNPs were in or around activin A receptor, type I (*ACVR1*), four SNPs were in or around calcium channel, voltage-dependent, beta 4 subunit (*CACNB4*), four were in or around LY6/PLAUR domain containing 6B (*LYPD6B*), two were in RNA binding motif, single stranded interacting protein 1 (*RBMS1*), and two were in or around cytohesin 1 interacting protein (*CYTIP*) including one SNP in the 3`-UTR of the gene (rs267992).

On chromosome 4, there were two large regions of difference in local ancestry and three smaller peaks at the end of the chromosome (Figure 2c). The strongest signal was in an 11 Mb region clustered around intergenic SNP rs6858204; the second large peak was a 10 Mb region clustered around intergenic SNP rs2301134 within the alpha synuclein gene (*SNCA*). The first small peak was a 2 Mb region clustered around intergenic SNP rs13102201. The second was a three SNP cluster around the ring finger protein 150 gene (*RNF150*), and the final was a two SNP cluster within the nuclear receptor subfamily 3, group C, member 2 gene (*NR3C2*). Within the region of the strongest peak, the largest deviation in MAF between cases and controls (difference = 0.043) was at a SNP within the *CAMK2D* gene (rs6845568). 15 additional SNPs showed large differences in MAF between cases and controls (difference > 0.025), three fell within an intergenic region previously implicated in GWAS of longevity (Lunetta et al., 2007) and triglyceride levels (Kathiresan et al., 2007), two fell in or around dickkopf WNT signaling pathway inhibitor 2 (*DKK2*), two around tet methylcytosine dioxygenase 2 (*TET2*), two around ankyrin 2 (*ANK2*), 2 around ELOVL fatty acid elongase 6 (*ELOVL6*), one upstream of tachykinin receptor 3 (*TACR3*) and one within coiled-coil domain containing 109B (*CCDC109B*).

On chromosome 12 (Figure 2d), the observed peak was a 279Kb region around rs11066726 annotated to uncharacterized LOC100506465 (*LOC100506465*). The region also included SNPs in and around the RNA binding motif protein 19 gene (*RBM19*).

On chromosome 19, there were two regions of large differences in ancestry between cases and controls (Figure 2e). The first region was 1 Mb in length and encompassed the *ABCA7* locus, mentioned above. However, its strongest signal came from rs17685286 annotated to the zinc finger protein 554 gene (*ZNF554*). The second region was 2 Mb in length and encompassed the *CD33* locus previously mentioned. Its strongest signal came from rs3852863 which annotated to ER membrane protein complex subunit 10 (*EMC10*).

DISCUSSION

The results of this study suggest that, among AA subjects, the percentage of African ancestry is higher in LOAD cases than in controls. Moreover, this global difference appears to be driven in part by local differences in ancestry at disease relevant loci. While ancestral differences exist and can be assessed globally, our exploratory posthoc analyses highlight large variability across the genome, with some regions showing strong deviations in ancestry between cases and controls. These results suggest differences in local ancestry may underlie differential risk for LOAD within African Americans and that local ancestral considerations may be useful when evaluating disease risk within any admixed population. Future fine mapping analyses within the peaks identified in the current manuscript may clarify the mechanism of heightened risk for LOAD in AA subjects.

Differences in Global Ancestry

The overall levels of African ancestry observed in African Americans in the current study are consistent with previous reports of 75% – 80% African ancestry.[25] As predicted, cases had a higher proportion of African ancestry than controls, highlighting the polygenic nature of AD etiology and suggesting that there may be many loci with small effect sizes driving the ancestral differences in disease risk. Although this is the first study to demonstrate higher levels of African ancestry in AA LOAD cases, studies in other complex diseases have shown similar effect sizes suggesting that among AA, the proportion of African ancestry is related to a risk of type 2 diabetes,[26] insulin resistance, and levels of total cholesterol and LDL-cholesterol.[25] Our results are in contrast to recent reports that African ancestry may be protective against the accumulation of neuritic plaques.[16] In both vascular health and LOAD, it appears the relationship between genetic ancestry and disease risk is quite complex, and it might vary by geographic region based on the genetic context of a given admixed population. We also cannot rule out the possibility that other non-genetic factors that might be correlated with degree of African ancestry in our sample, such as social, environmental or economic effects based on skin color or other physical features, might explain some of the observed association with ancestry. However, the present findings may partially explain the observation that within AA families, the risk of LOAD to first degree relatives of a person with LOAD is higher than it is in EA families.[27] To better understand how differences in global ancestry might contribute to disease risk, we chose to investigate differences in ancestry at disease relevant loci.

Differences in Local Ancestry at Disease Relevant Loci

Differences in local ancestry were observed at disease relevant loci, most strongly on chromosome 19, including *ABCA7, GRIN3B*, and *CD33*. Moreover, the difference in local ancestry at each of these loci was associated with disease status in a binary logistic regression model even when age, sex, *APOE*, and global ancestry were included as predictors. These results suggest that local ancestry mapping in unrelated case/control cohorts can identify positional candidate regions for further investigation, in a similar way linkage analysis can for family-based studies. Moreover, the variance explained by local ancestry above and beyond genotype, *APOE*, and global ancestral effects in these candidate

analyses highlights the potential benefit in performing a full GWAS analysis while considering global and local ancestral effects.

Differences in Local Ancestry across the Genome

Although we were somewhat under powered given our sample size and the effect sizes observed, we were still able to detect some differences in ancestry relevant to AD in the genome wide analysis. It should also be noted that while the effects observed were relatively small, they are actually comparable or stronger than the observed genomic effects in GWAS of AD (odds ratios around 1.30, or roughly Cohen's D of 0.06). Our primary results highlight differences in African ancestry between AA cases and controls, so we explored the peak differences in local ancestry across the genome in a posthoc analysis. Large deviations in local ancestry between cases and controls were observed in two small peaks on chromosome 1. Within these peaks were a few SNPs that also showed large deviations in MAF between cases and controls, including a missense mutation in *CYP4B1*. Interestingly, the CYP superfamily has been implicated in the pharmacogenetic response to cholinesterase inhibitors such as donepezil and tacrine.[28] Other members of the CYP superfamily have shown a weak association with AD risk and pathogenesis, perhaps through alterations in cholesterol metabolism,[29] and onset of AD in women with Down's Syndrome perhaps through alterations in the bioavailability of estrogen.[30]

On chromosome 2, 39 SNPs annotated to eight genes within the region that showed a statistically significant deviation in local ancestry also showed large deviations in allele frequency between cases and controls, some of which have biological functions with known relevance to AD pathogensis. *EPC2* has been associated with levels of cerebrospinal fluid tau in a recent GWAS study using tau as a quantitative endophenotype.[31] Similarly, *ACVR1* has been implicated in the phosphorylation of tau at AD relevant regions within the cortex of hTau mutant mice.[32] *CACNB4* has not been associated with AD, however interactions among calcium channel genes have previously shown an association with amyloid load measured in humans,[33] and calcium channel expression has been associated with the formation of plaques in AD mouse models.[34] Similarly, *LYPD6B* modulates calcium conductance in nicotinic acetylcholine receptors and may also be relevant through modulating calcium homeostasis.[35] *FMNL2* has not been implicated in AD, however its protein product interacts with the amyloid precursor protein.

Within the local ancestry peak on chromosome 4, the largest deviation in MAF between cases and controls was at a SNP within the *CAMK2D* gene (rs6845568). The gene product of *CAMK2D* phosphorylates tau at about 1/4 of its phosphorylation sites, potentially contributing to the hyperphosphorylation of tau observed in the AD brain.[36]

The local ancestry peak on chromosome 12 encompassed *RBM19* and *LOC10050646*, neither of which have been associated with AD in previous work. However, *RBM19* has an altered expression pattern in hippocampal cells of AD patients compared to cognitively normal controls.[37]

The two local ancestry peaks on chromosome 19 encompassed two known disease relevant loci: *ABCA7* and *CD33* (discussed in detail above). It is also interesting to note that the peak

difference around *CD33* did not appear to stretch to the *APOE* locus, suggesting that the *APOE* effect is not confounded with local ancestry in this sample. This is particularly relevant given the varying effects of *APOE* on AD risk reported in African and African American populations,[13–15] and further suggests that while local ancestral differences at disease relevant loci may play a role in the differences in AD risk observed between AAs and EAs, such effects are unlikely to be driven by differences in the *APOE* effect between AA and EA populations.

Strengths and Weaknesses

This manuscript has numerous strengths including the large sample evaluated, the methods which highlight both global and local differences in African ancestry that relate to AD risk, and the hierarchical models of risk which can place ancestral risk within the larger context of AD risk and resilience. However, this manuscript is not without weaknesses. We did not have covariates related to socioeconomic status or educational attainment available for regression analyses, which limits our ability to speak to how potential differences within these critical factors may moderate or mediate the observed ancestral effect. Additional work focused on ancestral differences within the context of educational and socioeconomic factors will be needed to tease apart these often interrelated factors. We also restricted our African reference population to the Yoruban population from Hapmap, which has the advantage of making the interpretation of ancestral estimates quite simple, but also may not account for all markers of African ancestry across the genome, particularly given the high levels of genetic diversity across African populations. Future studies incorporating multiple African reference populations may help clarify if differences in ancestry related to various regions of genetic inheritance exist.

Conclusions

The present study has demonstrated that differences in global and local ancestry are relevant to Alzheimer's disease risk within an African American cohort. The genomic regions implicated in our genome wide scan of local ancestry highlight a few potential mechanisms of the observed effects including genes that regulate the complex interplay between the acetylcholine system and calcium homeostasis, and genes that may increase vulnerability to tau hyperphosphorylation. Future fine mapping work may help to clarify how ancestral differences in these genomic regions relate specifically to disease risk and progression. The observed ancestral differences observed in this study also begin to shed some light on the complex genomic differences that may underlie the observed disparity in AD risk between African American and European American populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Timothy J. Hohman, Ph.D[1], Jessica N. Cooke-Bailey, Ph.D[1,2], Christiane Reitz, MD, PhD[3], Gyungah Jun, PhD[4,7,30], Adam Naj, PhD[5], Gary W. Beecham, PhD[6], Zhi Liu,

MS[6], Regina M. Carney, MD[6,32], Jeffrey M. Vance, MD, PhD[6,33], Michael L. Cuccaro, PhD[6,34], Ruchita Rajbhandary, MPH[6], Badri Narayan Vardarajan, PhD[3], Li-San Wang, PhD[8], Otto Valladares, MS[8], Chiao-Feng Lin, PhD[8], Eric B. Larson, MD, MPH[9], Neill R. Graff-Radford, MD[10], Denis Evans, MD[11], Philip L. De Jager, MD, PhD[12,13], Paul K. Crane, MD, MPH[9], Joseph D. Buxbaum, PhD[14], Jill R. Murrell, PhD[15], Towfique Raj, PhD[12,13], Nilufer Ertekin-Taner, MD, PhD[10], Mark W. Logue, PhD[7,30], Clinton T. Baldwin, PhD[7,31], Robert C. Green, MD, MPH[16], Lisa L. Barnes, PhD[17], Laura B. Cantwell, MPH[8], M. Daniele Fallin, PhD[18], Rodney C. P. Go, PhD[19], Patrick Griffith, MD[20], Thomas O. Obisesan, MD[21], Jennifer J. Manly, PhD[3], Kathryn L. Lunetta, PhD[30], M. Ilyas Kamboh, PhD[22], Oscar L. Lopez, MD[22], David A. Bennett, MD[17,23], John Hardy, Ph.D.[36], Hugh C. Hendrie, MB, ChB, DSc[24], Kathleen S. Hall, PhD[24], Alison M. Goate, PhD[25], Rosalyn Lang, PhD[26], Goldie S. Byrd, PhD[26], Walter A. Kukull, PhD[27], Tatiana M. Foroud, PhD[28], Lindsay A. Farrer, PhD[4,7,29,30], Eden R. Martin, PhD[6,35], Margaret A. Pericak-Vance, PhD[6,33], Gerard D. Schellenberg, PhD[8], Richard Mayeux, MD, MSc[3], Jonathan L. Haines, Ph.D[1,2], Tricia A. Thornton-Wells, Ph.D[1], and for the Alzheimer Disease Genetics Consortium

Affiliations

1

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

ACKNOWLEDGEMENTS

REFERENCES

1. 2013 Alzheimer's disease facts and figures. Alzheimer's & Dementia. 9:208–245.

2. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science. 1993; 261:921–923. [PubMed: 8346443]

3. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nature genetics. 2009; 41:1088–1093. [PubMed: 19734902]

4. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, et al. Common variants at ABCA7 , MS4A6A / MS4A4E , EPHA1 , CD33 and CD2AP are associated with Alzheimer's disease. Nature genetics. 2011; 43:429–435. [PubMed: 21460840]

5. Jean-Charles L, Carla AI-V, Denis H, Adam CN, Rebecca S, Celine B, et al. Extended meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature genetics. 2013; 45:1452–1458. [PubMed: 24162737]

6. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, et al. Common variants at MS4A4 / MS4A6E , CD2AP , CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nature genetics. 2011; 43:436–441. [PubMed: 21460841]

7. Seshadri S, Fitzpatrick AL, Ikram M. Genome-wide analysis of genetic loci associated with alzheimer disease. JAMA. 2010; 303:1832–1840. [PubMed: 20460622]

8. Ridge PG, Mukherjee S, Crane PK, Kauwe JSK. Alzheimer's disease: analyzing the missing heritability. PloS one. 2013; 8:e79771. [PubMed: 24244562]

9. So H-C, Gui AHS, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genetic epidemiology. 2011; 35:310–317. [PubMed: 21374718]

10. Gatz M, Reynolds CA, Fratiglioni L. ROle of genes and environments for explaining alzheimer disease. Archives of General Psychiatry. 2006; 63:168–174. [PubMed: 16461860]

11. Tang MX, Cross P, Andrews H, Jacobs DM, Small S, Bell K, et al. Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. Neurology. 2001; 56:49–56. [PubMed: 11148235]

12. Fillenbaum GG, Heyman A, Huber MS, Woodbury MA, Leiss J, Schmader KE, et al. The Prevalence and 3-Year Incidence of Dementia in Older Black and White Community Residents. Journal of Clinical Epidemiology. 1998; 51:587–595. [PubMed: 9674666]

13. Reitz C, Jun G, Naj A. Variants in the atp-binding cassette transporter (abca7), apolipoprotein e-4, and the risk of late-onset alzheimer disease in african americans. JAMA. 2013; 309:1483–1492. [PubMed: 23571587]

14. Gureje O, Ogunniyi A, Baiyewu O, Price B, Unverzagt FW, Evans RM, et al. APOE e4 is not associated with Alzheimer's Disease in elderly Nigerians. Annals of neurology. 2006; 59:182–185. [PubMed: 16278853]

15. Hendrie HC, Murrell J, Baiyewu O, Lane KA, Purnell C, Ogunniyi A, et al. APOE e4 and the risk for Alzheimer disease and cognitive decline in African Americans and Yoruba. International Psychogeriatrics. 2014; 26:977–985. [PubMed: 24565289]

16. Schlesinger D, Grinberg LT, Alba JG, Naslavsky MS, Licinio L, Farfel JM, et al. African ancestry protects against Alzheimer's disease-related neuropathology. Molecular psychiatry. 2013; 18:79–85. [PubMed: 22064377]

17. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 1984; 34:939–939. [PubMed: 6610841]

18. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Research. 2009; 19:1655–1664. [PubMed: 19648217]

19. Levene, H. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Olkin, I., editor. Stanford University Press; 1960. p. 278-292.

20. Baran Y, Pasaniuc B, Sankararaman S, Torgerson DG, Gignoux C, Eng C, et al. Fast and accurate inference of local ancestry in Latino populations. Bioinformatics. 2012; 28:1359–1367. [PubMed: 22495753]

21. Pasaniuc B, Sankararaman S, Kimmel G, Halperin E. Inference of locus-specific ancestry in closely related populations. Bioinformatics. 2009; 25:i213–i221. [PubMed: 19477991]

22. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nature methods. 2012; 10:5–6. [PubMed: 23269371]

23. Jun G, Naj AC, Beecham GW, Wang LS, Buros J, Gallins PJ, et al. Meta-analysis confirms CR1 , CLU , and PICALM as Alzheimer disease risk loci and reveals interactions with APOE genotypes. Archives of Neurology. 2010; 67:1473–1484. [PubMed: 20697030]

24. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Research. 2011; 40:D930–D934. [PubMed: 22064851]

25. Reiner A, Carlson C, Ziv E, Iribarren C, Jaquish C, Nickerson D. Genetic ancestry, population sub-structure, and cardiovascular disease-related traits among African-American participants in the CARDIA Study. Human Genetics. 2007; 121:565–575. [PubMed: 17356887]

26. Cheng CY, Reich D, Haiman CA, Tandon A, Patterson N, Elizabeth S, et al. African ancestry and its correlation to type 2 diabetes in African Americans: a genetic admixture analysis in three US population cohorts. PloS one. 2012; 7:e32840. [PubMed: 22438884]

27. Green RC, Cupples LA, Go R, Benke KS, Edeki T, Griffith PA, et al. Risk of dementia among white and African American relatives of patients with Alzheimer disease. JAMA: The Journal of the American Medical Association. 2002; 287:329–336. [PubMed: 11790212]

28. Cacabelos R, Llovo R, Fraile C, Fernandez-Novoa L. Pharmacogenetic Aspects of Therapy with Cholinesterase Inhibitors: The Role of CYP2D6 in Alzheimers Disease Pharmacogenetics. Current Alzheimer Research. 2007; 4:479–500. [PubMed: 17908053]

29. Kölsch D, L tjohann D, Jessen F, Popp J, Hentschel F, Kelemen P, et al. *CYP46A1* variants influence Alzheimer's disease risk and brain cholesterol metabolism. European Psychiatry. 2009; 24:183–190. [PubMed: 19286353]

30. Chace C, Pang D, Weng C, Temkin A, Lax S, Silverman W, et al. Variants in CYP17 and CYP19 cytochrome P450 genes are associated with onset of Alzheimer's disease in women with down syndrome. Journal of Alzheimer's Disease. 2012; 28:601–612.

31. Cruchaga C, Kauwe John S, Harari O, Jin Sheng C, Cai Y, Karch Celeste M, et al. GWAS of Cerebrospinal Fluid Tau Levels Identifies Risk Variants for Alzheimers Disease. Neuron. 2013; 78:256–268. [PubMed: 23562540]

32. Cavallini A, Brewerton S, Bell A, Sargent S, Glover S, Hardy C, et al. An Unbiased Approach to Identifying Tau Kinases That Phosphorylate Tau at Sites Associated with Alzheimer Disease. Journal of Biological Chemistry. 2013; 288:23331–23347. [PubMed: 23798682]

33. Koran MI, Hohman TJ, Thornton-Wells TA. Genetic interactions found between calcium channel genes modulate amyloid load measured by positron emission tomography. Human Genetics. 2014; 133:85–93. [PubMed: 24026422]

34. Daschil N, Obermair GJ, Flucher BE, Stefanova N, Hutter-Paier B, Windisch M, et al. Ca V 1.2 calcium channel expression in reactive astrocytes is associated with the formation of amyloid-á plaques in an Alzheimer's disease mouse model. Journal of Alzheimer's Disease. 2013; 37:439–451.

35. Darvas M, Morsch M, Racz I, Ahmadi S, Swandulla D, Zimmer A. Modulation of the Ca^{2+} conductance of nicotinic acetylcholine receptors by Lypd6. European Neuropsychopharmacology. 2009; 19:670–681. [PubMed: 19403274]

36. Yamauchi T. Neuronal Ca2+/Calmodulin-Dependent Protein Kinase II—Discovery, Progress in a Quarter of a Century, and Perspective: Implication for Learning and Memory. Biological and Pharmaceutical Bulletin. 2005; 28:1342–1354. [PubMed: 16079472]

37. Ravetti MG, Rosso OA, Berretta R, Moscato P. Uncovering molecular biomarkers that correlate cognitive decline with the changes of hippocampus' gene expression profiles in Alzheimer's disease. PloS one. 2010; 5:e10153. [PubMed: 20405009]

Research in Context

Systematic Review

We performed a comprehensive review of existing literature investigating the relationship between African ancestry and Alzheimer's disease (AD) risk. Previous work has highlighted increased risk for AD in African Americans (AAs) and differences in the observed effect size of risk genes at various loci when comparing AAs to European Americans (EAs), although there are inconsistencies among these reports. This manuscript is the largest systematic evaluation of ancestral differences among AAs across the genome with regard to AD risk.

Interpretation

Our results suggest that in AAs, higher levels of African ancestry—at the whole genome level and at specific AD-related genetic loci—are associated with an increased risk for AD.

Future Directions

Our results highlight peaks in local ancestry differences between AD cases and controls. Future fine mapping analyses within the peaks identified in the current manuscript may clarify the mechanism of heightened risk for LOAD in AA subjects.
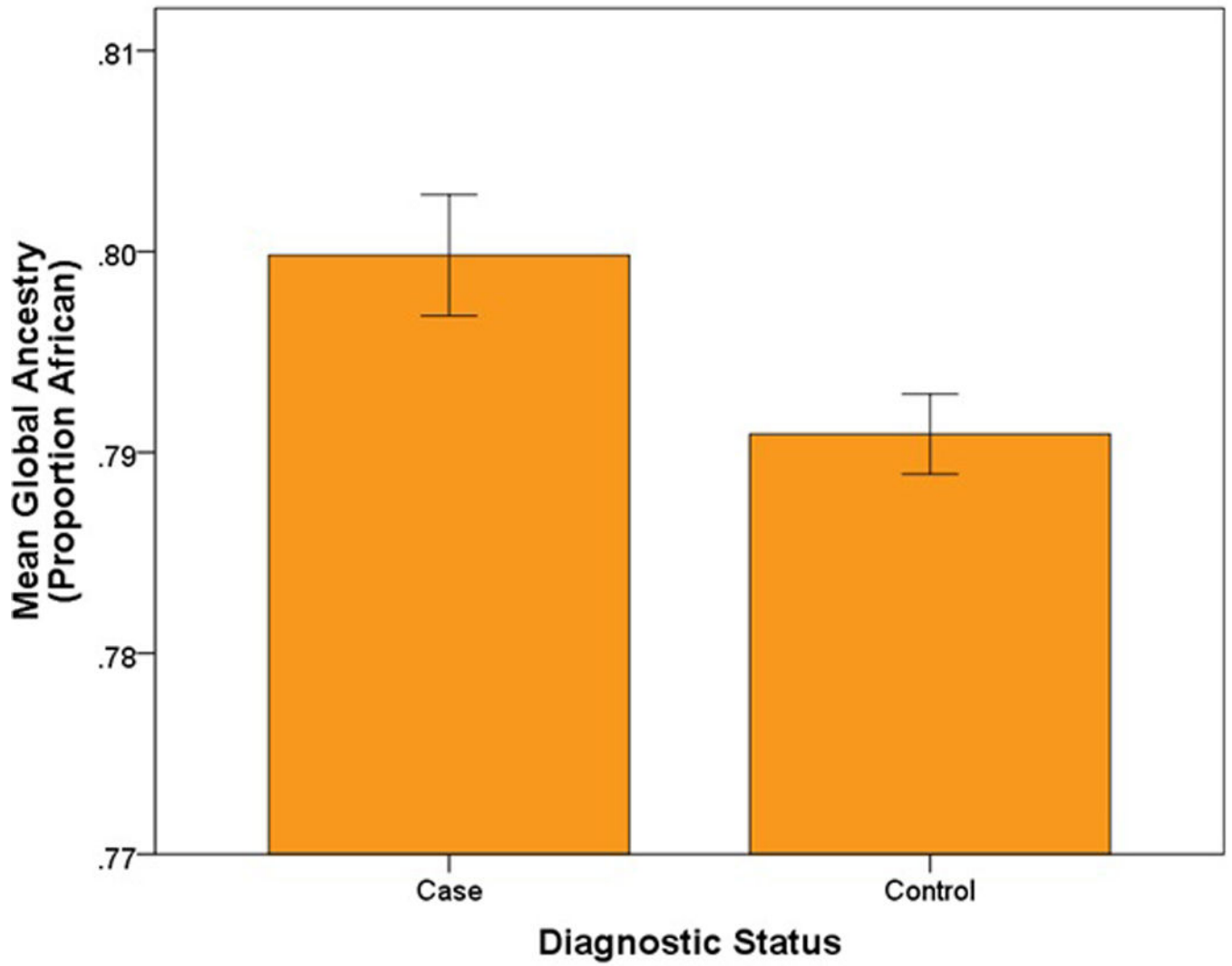
Figure 1. Differences in global ancestry between Alzheimer's cases and controls
Error bars represent the standard error of the mean. Group difference is statistically significant at p = 0.008.
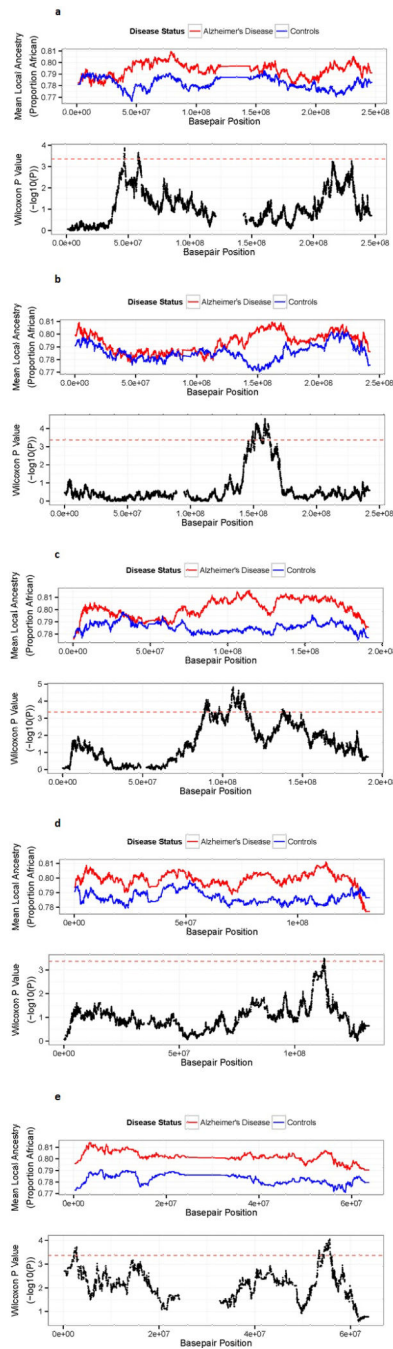
Figure 2. Differences in Local Ancestry by Chromosome: (a) Chr. 1, (b) Chr. 2, (c) Chr. 4, (d) Chr. 12 and (e) Chr. 19

For each sub-figure, the top panel illustrates the proportion of African ancestry in cases (red) and controls (blue) across chromosome 1. The bottom panel presents the –log 10 p-value for a Wilcoxon rank sum test performed at each locus. The red dotted line signifies the threshold for statistical significance (FDR < 0.05).

Table 1

Sample Demographic Characteristics

	Clinical Diagnosis	
	Normal Control	Alzheimer's Disease
Number of Patients	3,804	1,840
APOE Genotype [a]		
Number of *APOE* ε2/ε2 carriers (%)	36 (0.97%)	10 (0.61%)
Number of *APOE* ε2/ε3 carriers (%)	625 (16.88%)	122 (7.46%)
Number of *APOE* ε2/ε4 carriers (%)	169 (4.57%)	65 (3.97%)
Number of *APOE* ε3/ε3 carriers (%)	1733 (46.81%)	515 (31.48%)
Number of *APOE* ε3/ε4 carriers (%)	1030 (27.82%)	710 (43.4%)
Number of *APOE* ε4/ε4 carriers (%)	109 (2.94%)	214 (13.08%)
Number of Females (Percent)	2,677 (70.37 %)	1,273 (69.18%)
Mean Age (SD)	77.16 (8.23)	78.56 (7.91)
Mean African Ancestry (SD)	0.79 (0.12)	0.80 (0.12)

[a] 102 controls and 204 cases did not have *APOE* genotype available

Table 2

Local Ancestry Differences at Loci Previously Implicated in LOAD

SNP	GENE	EFFECT SIZE (Cohen's D)	P-VALUE (Wilcoxon)	Correted P (FDR)	Correted P (Bonferroni)	Correted P (Hochberg)
rs3764650	ABCA7	0.09	0.001	0.015	0.043	0.041
rs115550680[a]	ABCA7	0.09	0.001	0.015	0.043	0.041
rs115553053[b]	ABCA7	0.09	0.001	0.015	0.043	0.041
rs4147929[c]	ABCA7	0.09	0.001	0.015	0.043	0.041
rs115882880[d]	GRIN3B	0.09	0.002	0.015	0.046	0.043
rs3865444	CD33	0.08	0.003	0.022	0.089	0.08
rs10498633	SLC24A4	0.07	0.019	0.113	0.567	0.491
rs11767557[e]	EPHA1	0.05	0.026	0.119	0.787	0.656
rs11771145	EPHA1	0.05	0.026	0.119	0.787	0.656
rs145848414[f]	MSX2	0.06	0.028	0.119	0.831	0.665
rs72745818[g]	CASS4	0.05	0.057	0.163	1	0.966
rs8093731[h]	DSG2	0.05	0.060	0.163	1	0.966
rs3818361	CR1	0.05	0.063	0.163	1	0.966
rs6656401[i]	CR1	0.05	0.063	0.163	1	0.966
rs10792832[j]	PICALM	0.05	0.065	0.163	1	0.966
rs3851179	PICALM	0.05	0.065	0.163	1	0.966
rs9271192[k]	HLA-DRB1	0.05	0.090	0.207	1	0.966
rs17125944	FERMT2	0.04	0.142	0.265	1	0.966
rs2718058[l]	NME8	0.04	0.144	0.265	1	0.966
rs190982[m]	MEF2C	0.05	0.155	0.265	1	0.966
rs12989701[n]	BIN1	0.04	0.165	0.265	1	0.966
rs6733839[o]	BIN1	0.04	0.165	0.265	1	0.966
rs744373	BIN1	0.04	0.165	0.265	1	0.966

SNP	GENE	EFFECT SIZE (Cohen's D)	P-VALUE (Wilcoxon)	Correted P (FDR)	Correted P (Bonferroni)	Correted P (Hochberg)
rs11218343[p]	SORL1	0.03	0.167	0.265	1	0.966
rs1476679[q]	ZCWPW1	0.04	0.182	0.273	1	0.966
rs35349669[r]	INPP5D	0.03	0.302	0.432	1	0.966
rs9331896[s]	CLU	−0.01	0.602	0.818	1	0.966
rs11136000[t]	CLU	−0.01	0.627	0.818	1	0.966
rs28834970[u]	PTK2B	−0.01	0.657	0.821	1	0.966
rs10948363[v]	CD2AP	0.01	0.929	0.966	1	0.966
rs9349407[w]	CD2AP	0.01	0.944	0.966	1	0.966
rs1038725[x]	CELF1	0	0.952	0.966	1	0.966
rs610932	MS4A6A	0	0.966	0.966	1	0.966
rs670139[v]	MS4A4E	0	0.966	0.966	1	0.966

[a] proximal SNP rs3764650; 3,900bp (D` = 1.00)

[b] proximal SNP rs757232; 6,865bp (D` = 1.00)

[c] proximal SNP rs757232; 12,536bp (SNP pair not available in 1000 Genomes)

[d] proximal SNP rs4807395; 1,070bp (D` = 1.00)

[e] proximal SNP rs11771145; 1,873bp (D` = 0.67)

[f] proximal SNP rs13179274; 2,604bp (SNP pair not available in 1000 Genomes)

[g] proximal SNP rs6069746; 2,479bp (D` = 1.00)

[h] proximal SNP rs1031729; 2,660bp (D` = 1.00)

[i] proximal SNP rs11117959; 7,530bp (SNP pair not available in 1000 Genomes)

[j] proximal SNP rs3851179; 765bp (D` = 1.00)

[k] proximal SNP rs532098; 228bp (SNP pair not available in 1000 Genomes)

[l] proximal SNP rs2722363; 5,681bp (D` = 1.00)

[m] proximal SNP rs304132, 7,576bp (D` = 1.00)

[n] proximal SNP rs7561528; 1,902bp (D` = 0.915)

[o] proximal SNP rs7561528; 3,173 (D` = 1.00)

[p] proximal SNP rs7124060; 629bp (D` = 1.00)

[q] proximal SNP rs5015755; 9,206bp (SNP pair not available in 1000 Genomes)

[r] proximal SNP rs11673739; 4,534bp (D` = 1.00)

[s] proximal SNP rs492638; 11,543 (SNP pair not available in 1000 Genomes)

[t] proximal SNP rs10503814; 9,694bp (D` = 0.08)

[u] proximal SNP rs1879188; 948bp (D` = 1.00)

[v] proximal SNP rs9296562; 2,431bp (D` = .74)

[w] proximal SNP rs9296559; 858bp (D` = 1.00)

[x] proximal SNP rs7124681; 27,674bp (SNP pair not available in 1000 Genomes)

[y] proximal SNP rs7929589; 3,533bp (D` = 1.00)