



Published in final edited form as:

Mol Psychiatry. 2021 June ; 26(6): 2200–2211. doi:10.1038/s41380-020-00989-9.

Long Tracks of Homozygosity Predict the Severity of Alcohol Use Disorders in an American Indian Population

Qian Peng, Ph.D.^{1,*}, Cindy L. Ehlers, Ph.D.^{1,*}

¹Department of Neuroscience, The Scripps Research Institute, La Jolla, CA 92037 USA

Abstract

Runs of homozygosity (ROH) arise when an individual inherits two copies of the same haplotype segment. While ROH are ubiquitous across human populations, Native populations—with shared parental ancestry arising from isolation and endogamy—can carry a substantial enrichment for ROH. We have been investigating genetic and environmental risk factors for alcohol use disorders (AUD) in a group of American Indians (AI) who have higher rates of AUD than the general U. S. population. Here we explore whether ROH might be associated with incidence and severity of AUD in this admixed AI population ($n=742$) that live on geographically contiguous reservations, using low-coverage whole genome sequences. We have found that the genomic regions in the ROH that were identified in this population had significantly elevated American Indian heritage compared with the rest of the genome. Increased ROH abundance and ROH burden are likely risk factors for AUD severity in this AI population, especially in those diagnosed with severe and moderate AUD. The association between ROH and AUD was mostly driven by ROH of moderate lengths between 1–2Mb. An ROH island on chromosome 1p32.3 and a rare ROH pool on chromosome 3p12.3 were found to be significantly associated with AUD severity. They contain genes involved in lipid metabolism, oxidative stress and inflammatory responses; and *OSBPL9* was found to reside on the consensus part of the ROH island. These data demonstrate that ROH are associated with risk for AUD severity in this AI population.

INTRODUCTION

Alcohol use disorders (AUD) are highly prevalent worldwide. However, the incidence varies across populations and ethnic groups, with particularly high rates found in some indigenous populations such as American Indians (AI) (1, 2). Like many other complex diseases, the differences in incidence between ethnic groups are likely due to both environmental and genetic factors (3).

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Corresponding authors: Cindy L. Ehlers, 10550 North Torrey Pines Road, La Jolla, CA 92037, Mail SR-311, Tel: 858-784-7055; Fax: 858-784-7475; cindy@scripps.edu, Qian Peng, 3366 North Torrey Pines Court, Suite 200, La Jolla, CA 92037, Tel: 858-784-7215; qpeng@scripps.edu.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Supplementary information is available at MP's website.

In general, the most replicable genetic findings for AUD traits have been found for variants in the genes that code for differences in the major alcohol-metabolizing enzymes (4). The allele frequencies of these genetic variants differ substantially between populations (5–8) including American Indians (9, 10). Recent genome-wide association studies (GWAS) and meta-analyses with increasingly large sample sizes have identified an additional small and diverse set of single nucleotide polymorphisms (SNPs) associated with alcohol dependence, alcohol consumption, and related traits in genes not associated with alcohol metabolism (11–13). These large GWAS findings are primarily in populations of European descent. Variants independent of the alcohol metabolizing enzymes have also been identified for AUD traits in American Indians. For instance, distinct rare variants in a potassium (K_{2p}) channel gene *KCNK2* and a pro-inflammatory mediator gene of the phosphodiesterase family *PDE4C* have been linked to AUD severity in American Indians and Euro-Americans (14). An interleukin subunit gene *EBI3* (*IL-27B*) and a serine/threonine protein kinase family gene *PRKG2* were found to be uniquely associated with alcohol-induced affective symptoms in AI (14). These studies demonstrate that genetic variants underlying AUD likely vary across populations. American Indians who have elevated rates of AUD, especially severe AUD, may carry distinct genetic risk for the disorder.

It is important to identify genetic and environmental risk factors that are specific to AI populations that account for the high rates of AUD seen in many tribes as they may lead to better approaches to prevention and treatment (3). One theoretical assumption concerning Native people is that isolation and endogamy over many generations coupled with a long history of dependence on foraging and subsistence agriculture may have led to selective enrichment of traits that increase food consumption when highly caloric food is available—the so-called “thrifty” or “fat-sparing” genes hypothesis. There have been examples of rare variant effects under positive selection such as what has been observed in Samoan islanders where a thrifty variant, extremely rare in general populations, is found to be highly common among the islanders (15). We have suggested that this same selective pressure may have caused the enrichment for genetic variants that influence the risk for consumption of other high-salience substances, such as alcohol (16, 17). In addition to consumption related traits, other systems such as the stress and immune response systems could also be under selection pressure.

American Indians share certain genetic characteristics with other indigenous population isolates (18), which may give clues to understanding the genetics of AUD in these populations. For instance, many aboriginal people with a history of isolation, such as Australian aboriginals, Maori of New Zealand, Pacific islanders, and North and South American Indian tribes, have highly increased risk for substance abuse (19). One characteristic that population isolates share is the enrichment of long runs of homozygosity in individuals' genomes that arise when an individual inherits two copies of the same haplotype segment from each parent. Autozygous tracks across genome can be estimated as long stretches of homozygous SNPs in a row, which are referred to as runs of homozygosity (ROH). Distributions of ROH may reflect the processes of population size reduction, consanguinity, admixture, and natural selection (20, 21). Thus, studying ROH provides insight into a population's genetic events over time and their potential impact on diseases. While ROH are ubiquitous across human populations, American Indian populations have

been shown to have a high burden (total length) of ROH (20). ROH can theoretically be more likely to harbor recessive deleterious variants, especially in long ROH. Recent inbreeding events tend to generate long ROH that enable rare deleterious variants to occur as homozygotes (22, 23). This is often the motivation for studying ROH in the context of complex disorders. Genomic regions under selective pressure may also show characteristic of ROH (24–29), since selection purges deleterious variants or elevates haplotype frequencies around a favored allele thereby increasing homozygosity surrounding the target loci (30). For this to happen, the ROH regions need to be sufficiently stable across generations to allow the selection effect to accumulate (22). ROH that survived many generations tend to be shorter. What benefited the population in the past as a result of selection could also contribute to increased risk for present-day disorders. Both inbreeding effects and selective sweeps (partial or complete) can be amplified in population isolates especially during population reduction. This further motivates us to search in ROH for alleles that may potentially confer increased risk for AUD in AI. In human populations, ROH have been associated with traits such as stature (31), cognitive functions (32–35), complex disorders such as schizophrenia (36, 37), autism (38) and dementias (39, 40), and more recently with a broad range of other phenotypes (41). However, while ROH have been shown to have health consequences due to inbreeding depression or possibly selection, it is not known whether areas of the genome with ROH can harbor genes that confer risk for AUD traits in AI.

In the present study we focused on a group of AI that have been demonstrated to have high rates of AUD (42). Using whole genome sequence data we investigated whether the total amount of ROH, or similarly, the proportion of genomes that are in ROH—referred to as F_{ROH} —may predict the severity levels of AUD in this population. We further investigated whether specific ROH segments can predict AUD severity in this AI population. Since American Indians are typically admixed populations, and the AI cohort under study has extended pedigrees, we used a linear mixed model to accommodate both population structure and potential relatedness. Studies have suggested that socioeconomic factors, such as: education attainment, religiosity, and socioeconomic status, can possibly bias ROH studies (43, 44). Since these environmental factors may also contribute to the development of AUD, we also included socioeconomic factors in our models. This study represents the first investigation, to date, into the potential consequences of ROH with respect to the risk for alcohol use disorders.

MATERIALS AND METHODS

Participants

Nine hundred and three (903) American Indians (AI) from extended pedigrees participated in the study. The population characteristics and the recruitment procedures were previously described (42, 45). Briefly, participants who had at least one-sixteenth self-reported American-Indian heritage and aged between 18 and 70 were targeted and recruited from geographically contiguous Indian reservations with a total population of about 3,000 individuals for the study (46). The recruitment was conducted using a combination of a venue-based method for sampling hard-to-reach populations (47, 48) and a respondent-

driven procedure (49). Seven hundred-fifty (750) individuals had their whole genome sequenced. Removing individuals with missing phenotypes or covariates for the present study, 742 individuals remained. Their demographics are characterized in Supplementary Table 1. The protocol for the study of this American Indian cohort was approved by the Institutional Review Board of The Scripps Research Institute (TSRI-IRB) and the Indian Health Council, a tribal review group overseeing health issues for the reservations where recruitment was undertaken. Written informed consent was obtained from each participant after the procedures had been fully explained.

Phenotypes and genotypes

All participants were interviewed and assessed with Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (50, 51), which was used to gather demographic information and make life-time alcohol use disorder (AUD) diagnoses according to the Diagnostic and Statistical Manual (DSM-5) (52). The SSAGA instrument has undergone both reliability and validity testing (50, 51), and has been successfully used in other American Indian populations previously (53, 54). All interviewers were trained by personnel from the Collaborative Study on the Genetics of Alcoholism (COGA). A research psychiatrist/addiction specialist made all best final diagnoses (42, 55). The SSAGA interview also retrospectively asks about the occurrence of alcohol-related life events and the ages when the problems first occurred, from which the main quantitative phenotype for this study, the severity level of AUD, was derived. The severity level of AUD was indexed by the 36 alcohol-related life events in the clinical course of the disorder, as listed in Supplementary Table 2 (56). The clinical course of AUD in this AI cohort has been previously described (42, 57). The order of the alcohol-related life events was based on the mean age of occurrence over the person's lifetime. The life events were given a severity weight of 1 for events 1–12, 2 for 13–24, and 3 for 25–36. The quantitative severity phenotype was then defined as the sum of the severity weights of the 36 life events for AUD (14). Supplementary Figure 1 illustrates the distribution of the AUD severity and its relations to AUD diagnoses in the AI. 70% individuals were diagnosed with DSM-5 AUD. Supplementary Table 3 contrasts the demographics between the group of individuals who were diagnosed with AUD and the group who did not meet the criteria for AUD.

The AI participants had low-coverage whole genome sequencing (LCWGS) on their blood-derived DNAs (58). Reads from the whole genome sequencing were mapped to the GRCh37/hg19 human reference genome using BWA and GATK (59). Variants were called and imputed with GATK and Thunder (60, 61). For quality control, variants were removed if they had >5% missing rate, or >5% Mendel error rate, or were out of Hardy-Weinberg equilibrium with $p < 0.001$; individuals were removed if they missed 2% genotypes. For ROH detection, variants having allele frequency <5% were excluded. Further details are given in the Supplementary Methods.

Detections of runs of homozygosity (ROH)

The homozygosity function in PLINK (62) was used to obtain the ROH for each individual and to derive consensus ROH pools across individuals. Using the low-coverage whole genome sequencing, we restricted ROH segments to longer than 1Mb to avoid shorter

ROH resulting from link disequilibrium (LD) effects. We allowed for maximum three heterozygous variants per ROH scanning window for reliable ROH callings as recommended for LCWGS (63). An ROH pool is made of overlapping ROH across individuals that also have matching alleles. The minimum overlapping ROH across all individuals in an ROH pool is referred to as a *consensus* ROH. We considered pools that had at least five individuals with the minimum consensus ROH length of 100Kb and 10 SNPs. Detailed parameter settings are listed in Supplementary Methods.

An individual's total number of ROH is referred to as *ROH abundance*, and their sum total length of ROH as *ROH burden*. One individual was considered as an outlier and removed, leaving 741 individuals for subsequent analyses: the individual's ROH abundance was beyond 5 standard deviations and F_{ROH} was at 10.4%; but we had no parental information to determine whether it resulted from higher level of inbreeding such as avuncular union. F_{ROH} is defined as the sum total of ROH above a certain length as a proportion of the autosomal genome length. We used all ROH above 1Mb to derive F_{ROH} thus it's proportional to the ROH burden in our study. F_{ROH} is also considered as a measure of the genomic inbreeding coefficient.

Assessing ancestral admixtures

Since patterns of ROH can be intricately related to the ancestral background of the genomes (23), and the ancestral makeup may confound the relationship of ROH to a trait (AUD severity is correlated with the degree of AI ancestry in this cohort), we first examined the relationship between ROH and the AI ancestry. We estimated ancestral admixtures for each AI individuals as shown in Supplementary Figure 2. The AI cohort was predominantly admixed between American Indian and European ancestries. We further estimated the local ancestries for each ROH segment, and computed for each individual the averaged ancestral admixtures in all of this individual's ROH segments, or across a subset of ROH segments that are part of the consensus pools. Details are given in the Supplementary Methods.

Associations between ROH and AUD severity

To account for the population admixture and the relatedness in the AI cohort, we used a liner mixed model (LMM) to assess the relationships between AUD severity and ROH measurements (ROH abundance or F_{ROH}). The genetic relationship matrix (GRM) was used to accommodate both family and population structures. We estimated GRM using the genotypes as implemented in GCTA (64), and carried out the LMM regressions using the generalized LMM association tests package (*GMMAT*) in R (65). We further stratified ROH segments by their length into four groups: 1–2Mb, 2–4Mb, 4–8Mb, and >8Mb, and modified the LMM regression model to include four length groups as predictors. Details are given in Supplementary Methods.

We considered two models for covariates: (i) a main model that initially included sex, age, age-squared and the global AI ancestry; (ii) an extended model incorporating additional socioeconomic factors such as gross income, years of education, employment status, marriage status, religion, and frequency of religious service attendance. We first assessed whether the covariates were associated with AUD severity prior to ROH association analyses

using the LMM modeling, and dropped the factors with negligible effect ($p > 0.1$). As a result, the main model retained sex, age, and age-squared as significant covariates; and the extended model additionally controlled for years of education, employment, and religion (Supplementary Table 4). Due to missing data, the sample size of the extended model reduced to 718 while the main model had 741 individuals. For the significant values, we adjusted for two models and two ROH measurements. However, since ROH abundance and F_{ROH} are correlated, we adjusted for the effective number of independent variables 1.05 instead.

For each ROH consensus segment that is shared by at least five individuals, an association using LMM modeling was tested between the AUD severity and whether an individual has the ROH segment. The extended model incorporating socioeconomic factors was used in this analysis. The Benjamini-Hochberg false discovery rate procedure was used to adjust for multiple comparisons (66). Furthermore, a binomial test was used to examine whether there was an aggregated burden of increased ($\beta > 0$, positive effects from association tests) or decreased ($\beta < 0$, negative effects) risk for AUD severity from the consensus ROH pools.

To identify whether there might exist subpopulations in the AI cohort that differed in characteristics with respect to the relation between global ROH and the AUD severity, we carried out an unsupervised clustering analysis. We used a mixture model of linear regression and estimated the parameters using the R package *flexmix* (67). Details are given in the Supplementary Methods.

Functional analysis

Once ROH pools were identified and found to be associated with AUD, we used Combined Annotation Dependent Depletion (CADD) (68, 69) to examine the deleteriousness of the variants in these regions. CADD integrated multiple functional annotations to produce one PHRED-scaled C-score. The higher the C-score, the more likely a variant is deleterious (for instance, 10 and 20 indicated 10% and 1% most deleterious mutations respectively). We chose 15 as the threshold representing roughly top 3% most deleterious variants. Functional and pathway analysis of genes in these regions were performed using GENE2FUNC in FUMA, in which enriched biological functions were extracted by testing against gene sets from MsigDB (70) and WikiPathways (71) using hyper-geometric tests (72). We further searched in the GWAS catalog for traits and disorders that have been associated with the genes in the identified ROH pools (73). The enrichment of GWAS catalog associations for the gene sets was also carried out in FUMA.

Data and code availability

In accordance with the wishes of the tribes no sharing of the AI data are possible. All analysis codes were written in R and are available upon requests.

RESULTS

ROH profiles of the AI population and their relations to the AI ancestry

The AI individuals had on average 41 ± 18 ROH segments of at least 1Mb long, with a range between 2 and 132 segments. The sum total length of ROH in each individual had a mean of 61.3 ± 31.2 Mb, and ranged between 2.6 and 206.7 Mb. The distributions of ROH burden stratified by the ROH length are illustrated in Supplementary Figure 3. F_{ROH} averaged at $2.12\% \pm 1.08\%$ with a range from 0.09% to 7.17%. ROH distributed along the genome unevenly, as illustrated in Figure 1, with predominately short to medium sized segments concentrating on particular regions—known as ROH islands (74)—punctuated by a small number of randomly placed long ROH, reflecting more recent inbreeding. 863 ROH pools were identified. Some pools coincided with known ROH islands. For instance, a pool with consensus segments falling between 49.64 and 50.05 Mb on chromosome 1 (Figure 1) is on an known ROH island (between 48.8–51.5 Mb) in European populations (74). The averaged European ancestry of this pool was indeed relatively high at 48.6% and in the highest 4th quartile. Supplementary Figure 4 shows the frequencies of the identified ROH pools across the whole genome.

The inbreeding coefficient F_{ROH} was related to an individual's degree of American Indian ancestry in a quadratic fashion as shown in Figure 2A. F_{ROH} reaches the minimum value at about 23% of AI ancestry in the fitted curve. As illustrated in Figure 2B, at individuals' level, genomic regions in the ROH had significantly higher American Indian heritage on average (52.8%) than the whole genome (mean=47.5%, adjusted $p=6.9E-6$), while the regions in ROH pools had even more significantly elevated AI ancestry (mean=54.8%, adjusted $p=2.54E-10$). The ancestry component of the ROH pools was predominately AI when there were fewer than 100 individuals in the pools. As a pool grew larger with more individuals, other ancestral components especially European increased (Supplementary Figure 5), although the length of the consensus segments became shorter invariably (data not shown). The patterns reflected that the population admixtures of different ancestries happened in the distant past.

Associations between AUD severity and ROH

The severity level of AUD was positively associated with both F_{ROH} (adjusted $p=0.015$ and 0.023 for the main and the extended models respectively, see Table 1) and ROH abundance (adjusted $p=0.0086$ in the main model and 0.0077 in the extended model, see Supplementary Table 5). The estimated effect size β in the extended model for F_{ROH} was 2.04, indicating that with a 1% increase in F_{ROH} , we would predict that the AUD severity increases by 2.04 points. None of the socioeconomic factors in the extended model was associated with F_{ROH} (Supplementary Table 6). When ROH segments were stratified by their lengths, only the ROH of 1–2 Mb long were found to be significantly associated with AUD severity (adjusted $p=0.013$ for F_{ROH} and 0.0088 for ROH abundance in the extended model, Supplementary Tables 7 & 8).

The unsupervised clustering analysis detected two subgroups in the AI cohort that significantly differed in AUD severity levels ($p=9E-155$, see Figure 3). The two subgroups

were not found to be significantly different for any of the demographic variables (see Figure 3B). In the higher severity group ($n=484$, 67%), the AUD severity level was significantly positively associated with F_{ROH} ($p=0.023$ after controlling all covariates in the extended model); 52% individuals in this group had been diagnosed with severe AUD and 22% with moderate AUD. In the lower severity group, AUD severity was not associated with F_{ROH} ; this subgroup had a number of mild AUD, one moderate AUD but no severe AUD diagnosis (Figure 3). The estimated effect size β for F_{ROH} was 2.38 ($p=0.023$) and 0.03 ($p=0.95$) for the higher and the lower severity subgroups respectively (Figure 3A). No significant interaction effect was detected between F_{ROH} and AUD subgroup (Supplementary Table 9). The identification of the two subgroups suggested that the effect of ROH on AUD was more pronounced and detectable in populations with high prevalence of severe AUD.

ROH pools associated with AUD severity in AI

ROH pool association analysis identified one common ROH group and one rare ROH group that were significantly associated with the severity level of AUD (Figure 4). The common ROH group is on 1p32.3 as shown in Figure 1 (adjusted $p=0.028$). 130 individuals (17.5%) had this ROH (ranges from 1.0–4.5 Mb long). Having this ROH increased the expected severity level of AUD from 22.6 to 28.7. Supplementary Figure 6 contrasts the distributions of AUD severity between individuals with or without this ROH segment. The consensus part of the pool is 169kb long on chr1: 52.1–52.2Mb. The maximum boundary of this ROH group spans 4.9Mb (Figure 4B). The consensus segment encompasses the oxysterol binding protein-like 9 gene *OSBPL9*. This gene is part of the bile acid and bile salt metabolic pathway. The maximum span (union of all ROH segments) of this ROH pool consists of 44 genes. The enriched functional groups of these genes are detailed in Supplementary Table 10. There were two enriched immunologic signatures, and several enriched functional groups including adipogenesis, TGF- β 1 targets and neuroblastoma. Several GWAS catalog traits were enriched in these genes; the top two being hippocampal tail volume, and cerebrospinal fluid amyloid beta 1–42 levels. The complete list of traits and disorders that were found to be associated with genes in this ROH region is listed in Supplementary Table 11.

Over 100 variants in this region had CADD scores of at least 15, among which 17 had scores over 20 representing the 1% most deleterious mutations, including nonsynonymous variants on the following genes: epidermal growth factor receptor pathway substrate 15 (*EPS15*), cytochrome C oxidase assembly factor 7 (*COA7*), solute carrier family 1 member 7 (*SLC1A7*), carnitine palmitoyltransferase 2 (*CPT2*), LDL receptor related protein 8 (*LRP8*) and GLIS family zinc finger 1 (*GLIS1*) (see Supplementary Table 12 for the full list). Nonsynonymous variant rs5174 (1:53712727) on *LRP8* had the highest CADD score of 35 (top 0.03% most deleterious) in the region. The variant also found eQTL for *LRP8* in tissues including cerebellar hemisphere and adipose (75). *LRP8* has been associated with educational attainment (76), risk-taking behavior (77), as well as body mass index (BMI) and alcohol intake interaction in a Hispanic population (78).

Since there is presently no publicly available replication sample for AI population with both AUD phenotypes and genome sequencing data, we searched for GWAS results in this

genomic region for AUD or alcohol consumption traits in other populations (12, 78–84) and listed the top SNPs in Supplementary Table 13. Variant rs12116501 on gene *LRP8* was found to be associated with AUD diagnosis ($p=9.1E-5$) in the Euro-American population of the Million Veteran Program (MVP) (12). Variants on Fas associated factor 1 gene *FAF1* were associated with comorbid alcohol dependence and depression ($p=3.9E-6$) in Euro-American samples in COGA (82) and with drinks-per-week ($p=2.1E-5$) in the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN) study (80). Variants on ATP/GTP binding protein-like 4 gene *AGBL4* have been associated with substance dependence (alcohol, heroin, methamphetamine) ($p=3E-12$) in Chinese population (83) and with aspartate aminotransferase (AST) level in excessive alcohol consumption ($p=2E-7$) in Australians (84).

The rare ROH pool that was associated with AUD severity (adjusted $p=0.028$) is on 3p12.3 with the consensus segment spanning 222kb long on chr3: 76.3–76.5Mb (Figure 4). Only five individuals had this ROH (ranges from 1.1–1.7Mb long), which elevated the expected AUD severity from 22.6 to 56.6 (Supplementary Figure 6). The consensus segment is located on roundabout guidance receptor 2 gene *ROBO2*, while the maximum span (of 2.6Mb long) covers nine genes. 60 variants in this region had CADD scores of at least 15. Two intronic variants on *ROBO2* in the region had CADD scores over 20 (Supplementary Table 14). *ROBO2* has been associated with a number of traits and disorders including BMI, visceral fat, eating disorder, smoking behaviors, and sleep (Supplementary Table 11).

Additionally, of the 863 ROH pools, if we consider those with $\beta > 0$ from the association test as potentially increasing risk for AUD severity and those $\beta < 0$ as decreasing risk, aggregated burden analysis showed that there was an over-representation of ROH pools increasing the risk (58%) than those that were protective (42%). The difference was significant whether the ROH pools that were significantly associated with AUD severity were excluded ($p=1.3E-6$) or not ($p=1.7E-6$).

DISCUSSION

The present study estimated ROH in an admixed American Indian population with elevated rates of AUD, using low-coverage whole genome sequence data. These data were used in order to investigate the relationship between ROH and ancestry admixture, and to determine whether ROH represent potential risk factors for severe AUD in AI.

Enrichment of American Indian heritage in the ROH

ROH burden and F_{ROH} were found to decrease as an individual's degree of American Indian ancestry increased from nearly 0 to about 23%, and then increased quadratically. This is the expected relationship considering that the AI cohort is an admixed population, primarily between American Indian and European ancestries. Admixed populations usually have fewer ROH than their parental populations. Individuals within an admixed population tend to have a different burden of short to medium sized ROH, which was also reflected in the AI cohort (Supplementary Figure 3). Specific patterns of ROH lengths may depend on when the admixture occurred as well as the ROH burden of the parental populations. In general, the higher the American Indian ancestry component, the greater the probability of

ROH formation (20). We found that the genomic regions in the ROH also had significantly elevated AI heritage compared with the rest of the genome.

ROH segments associated with AUD severity in AI contain genes involved in lipid metabolism, oxidative stress and inflammatory responses

The severity level of AUD was positively associated with both ROH abundance and F_{ROH} (or ROH burden). The association was dominated by the moderate sized ROH. Unsupervised clustering identified two subpopulations that significantly differed in AUD severity. Individuals in the higher severity group had a significantly increased risk of AUD associated with their ROH burden, with over half of them having been diagnosed with severe AUD in their lifetime.

An ROH pool residing on chromosome 1p32.3 was found to be significantly associated with AUD severity in this AI population. The consensus part of the ROH pool contains gene *OSBPL9*. *OSBPL9* encodes for an oxysterol-binding protein that belongs to a group of intracellular lipid receptors. Oxysterols are products of cholesterol metabolism and markers of lipid-related oxidative stress and have been suggested to play a role in inflammation and neuroinflammation. It has also been shown recently that oxysterols can modulate glial cell activation (85).

The full span of this ROH pool encompasses 44 genes. Functional enrichment analysis of these genes identified several significantly enriched traits in the GWAS catalog: the top two being hippocampal tail volume, and cerebrospinal fluid amyloid beta 1–42 levels (CSF A β 42). Reduced hippocampal tail volume has been reported in persons with AUD in a recent study (86), as well as in individuals with major depressive disorder (MDD) (87). CSF A β 42 is one of the biomarkers for mild cognitive impairment and dementia (88), and may also be related to AUD. A recent animal study has shown that alcohol intake can reduce the uptake of A β 42 by primary microglia (89). Within these genes on the top ROH pool, there were also two enriched immunologic signatures, and several enriched functional groups including neuroblastoma, TGF- β 1 targets and adipogenesis. The significantly enriched immunologic signatures were sets of genes that were up regulated when stimulated with a toll-like receptor 4 (TLR4) agonist or down-regulated when stimulated with a TLR3 agonist in dendritic cells (90, 91). This immune signaling cascade is thought to play an important role in neurodegeneration and the development of AUD (92). A recent study found that activation of TLR3 increased alcohol intake in mice (93). A significant number of genes in this ROH pool region were also down regulated by TGF- β 1 via TGF- β 1 receptors (94). TGF- β 1 is a cytokine mediating the pathogenesis of chronic inflammatory processes. It has been shown that alcohol, and its metabolite acetaldehyde, could increase TGF- β 1 expression; and elevated TGF- β 1 signaling may mediate alcohol induced deleterious effects such as hepatic fibrosis by involving extracellular matrix (ECM) deposition (95) and suppressed neuronal development in patients with AUD (96). TGF- β 1 alters cell migration in the developing cortex through regulating cell adhesion proteins thus TGF- β 1 system is a target of alcohol toxicity (97).

Genes in this ROH region were also found to contain pathways for adipogenesis. Evidence suggests that chronic alcohol intake can perturb lipid metabolism leading to the promotion

of an inflammatory environment. Alcohol induced adipose tissue dysfunction and the resulting inflammatory environment may contribute to the injury progression of other organs and systems, and other pathological states that accompany chronic AUD (98). Our data supports this finding as we found that organ damage associated with drinking, memory problems and neuropathy were significantly associated with the presence of this ROH segment (Supplementary Table 15).

A rare ROH pool on chromosome 3p12.3 was also significantly associated with AUD severity. The consensus part of this ROH pool resides on an axon guidance receptor gene *ROBO2* that encodes a roundabout (ROBO) family protein that is highly conserved from fly to human. This gene has been linked to multiple traits including smoking initiation (80), sleep measurements such as chronotype (99), BMI (100), education attainment (76), and unipolar depression (101).

In summary, this study represents the first investigation into the potential consequences of runs of homozygosity associated with alcohol use disorders. We have established that increased ROH are likely risk factors for severe AUD in this American Indian population. ROH of relatively moderate length of 1–2Mb produced the most significant associations. The moderate sized ROH most likely resulted from inbreeding of distant common ancestors, which may have happened when the population experienced a size reduction and the subsequent bottleneck and admixture. An ROH island harboring genes involved in lipid metabolism, oxidative stress and inflammatory responses was significantly associated with AUD severity, and was enriched for adipogenesis and immune signatures. These results are consistent with the thrifty gene related hypothesis; however, to draw further inferences requires rigorous tests for evolutionary signatures. At the moderate sample size of 741, we had limited statistical power for genome-wide significant associations. However, by focusing on the ROH segments across the genome, we dramatically reduced the number of multiple comparisons. Our findings could also be unique to this AI population and may not be found in other AI populations or non-AI populations. In fact, a recent study has found that increased F_{ROH} is associated with *reduced* risk-taking behaviors, including weekly alcohol consumption, in a primarily European population (41). While alcohol consumption and use disorders may have some overlapping but largely differing genetic underpinnings (12), it is also possible that ROH affect disorders differentially depending on the population background due to their distinct coalescent histories. For instance, endogamy within isolated groups with very low drinking levels may result in a negative association between ROH and drinking. Further studies in other population isolates should inform this hypothesis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We would like to acknowledge and thank all of our American Indian participants, and the following people for their roles in 1) the genotyping effort: Kirk Wilhelmsen, Scott Chasse, Piotr Mieczkowski, Ewa Patrycja Malc, Joshua Sailsbery, Phil Owens, and Chris Bizon; and 2) recruiting participants, and collection and preparation of the clinical data: David Gilder, Corinne Kim, Evie Phillips, Phillip Lau, and Derek Wills.

This work was supported by the National Institutes of Health (NIH): National Institute on Alcohol Abuse and Alcoholism (NIAAA) K25 AA025095 to QP; NIAAA R01 AA027316 to CLE; National Institute on Drug Abuse (NIDA) R01 DA030976 to CLE. NIAAA and NIDA had no further role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

REFERENCES

1. Grant BF, Goldstein RB, Saha TD, Chou SP, Jung J, Zhang H, et al. Epidemiology of DSM-5 alcohol use disorder: Results from the national epidemiologic survey on alcohol and related conditions III. *JAMA Psychiatry*. 2015;72(8):757–66. [PubMed: 26039070]
2. Compton WM, Thomas YF, Stinson FS, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: results from the national epidemiologic survey on alcohol and related conditions. *Archives of General Psychiatry*. 2007;64(5):566–76. [PubMed: 17485608]
3. Ehlers CL, Gizer IR. Evidence for a genetic component for substance dependence in Native Americans. *American Journal of Psychiatry*. 2013;170(2):154–64.
4. Edenberg HJ, McClintick JN. Alcohol Dehydrogenases, Aldehyde Dehydrogenases, and Alcohol Use Disorders: A Critical Review. *Alcoholism: Clinical and Experimental Research*. 2018;42(12):2281–97.
5. Edenberg HJ, Foroud T. Genetics and alcoholism. *Nature Reviews Gastroenterology & Hepatology*. 2013;10(8):487–94.
6. Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, et al. Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry*. 2014;19(1):41–9. [PubMed: 24166409]
7. Jorgenson E, Thai KK, Hoffmann TJ, Sakoda LC, Kvale MN, Banda Y, et al. Genetic contributors to variation in alcohol consumption vary by race/ethnicity in a large multi-ethnic genome-wide association study. *Molecular Psychiatry*. 2017;22:1359. [PubMed: 28485404]
8. Hart AB, Kranzler HR. Alcohol Dependence Genetics: Lessons Learned From Genome-Wide Association Studies (GWAS) and Post-GWAS Analyses. *Alcoholism: Clinical and Experimental Research*. 2015;39(8):1312–27.
9. Peng Q, Gizer IR, Libiger O, Bizon C, Wilhelmsen KC, Schork NJ, et al. Association and ancestry analysis of sequence variants in ADH and ALDH using alcohol-related phenotypes in a Native American community sample. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2014;165(8):673–83.
10. Peng Q, Gizer IR, Wilhelmsen KC, Ehlers CL. Associations Between Genomic Variants in Alcohol Dehydrogenase Genes and Alcohol Symptomatology in American Indians and European Americans: Distinctions and Convergence. *Alcoholism: Clinical and Experimental Research*. 2017;41(10):1695–704.
11. Sanchez-Roige S, Palmer AA, Fontanillas P, Elson SL, Adams MJ, Howard DM, et al. Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *American Journal of Psychiatry*. 2018;176(2):107–18.
12. Kranzler HR, Zhou H, Kember RL, Vickers Smith R, Justice AC, Damrauer S, et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nature Communications*. 2019;10(1):1499.
13. Clarke TK, Adams MJ, Davies G, Howard DM, Hall LS, Padmanabhan S, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). *Molecular Psychiatry*. 2017;22:1376. [PubMed: 28937693]
14. Peng Q, Bizon C, Gizer IR, Wilhelmsen KC, Ehlers CL. Genetic loci for alcohol-related life events and substance-induced affective symptoms: indexing the “dark side” of addiction. *Translational Psychiatry*. 2019;9(1):71. [PubMed: 30718457]
15. Minster RL, Hawley NL, Su C-T, Sun G, Kershaw EE, Cheng H, et al. A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nature Genetics*. 2016;48(9):1049–54. [PubMed: 27455349]

16. Ehlers CL, Wilhelmsen KC. Genomic scan for alcohol craving in Mission Indians. *Psychiatr Genet.* 2005;15(1):71–5. [PubMed: 15722961]
17. Ehlers CL, Wilhelmsen KC. Genomic screen for substance dependence and body mass index in southwest California Indians. *Genes, Brain and Behavior.* 2007;6(2):184–91.
18. Mooney JA, Huber CD, Service S, Sul JH, Marsden CD, Zhang Z, et al. Understanding the Hidden Complexity of Latin American Population Isolates. *The American Journal of Human Genetics.* 2018;103(5):707–26. [PubMed: 30401458]
19. Leske S, Harris MG, Charlson FJ, Ferrari AJ, Baxter AJ, Logan JM, et al. Systematic review of interventions for Indigenous adults with mental and substance use disorders in Australia, Canada, New Zealand and the United States. *Australian & New Zealand Journal of Psychiatry.* 2016;50(11):1040–54.
20. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics.* 2018;19:220.
21. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics.* 2006;15(5):789–95. [PubMed: 16436455]
22. Szpiech Zachary A, Xu J, Pemberton Trevor J, Peng W, Zöllner S, Rosenberg Noah A, et al. Long Runs of Homozygosity Are Enriched for Deleterious Variation. *The American Journal of Human Genetics.* 2013;93(1):90–102. [PubMed: 23746547]
23. Szpiech ZA, Mak ACY, White MJ, Hu D, Eng C, Burchard EG, et al. Ancestry-Dependent Enrichment of Deleterious Homozygotes in Runs of Homozygosity. *The American Journal of Human Genetics.* 2019;105(4):747–62. [PubMed: 31543216]
24. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, et al. A Simple Genetic Architecture Underlies Morphological Variation in Dogs. *PLOS Biology.* 2010;8(8):e1000451. [PubMed: 20711490]
25. Kim E-S, Cole JB, Huson H, Wiggans GR, Van Tassell CP, Crooker BA, et al. Effect of Artificial Selection on Runs of Homozygosity in U.S. Holstein Cattle. *PLOS ONE.* 2013;8(11):e80813. [PubMed: 24348915]
26. Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, et al. A genome-wide scan for signatures of recent selection in Holstein cattle. *Animal Genetics.* 2010;41(4):377–89. [PubMed: 20096028]
27. Ablondi M, Viklund Å, Lindgren G, Eriksson S, Mikko S. Signatures of selection in the genome of Swedish warmblood horses selected for sport performance. *BMC Genomics.* 2019;20(1):717. [PubMed: 31533613]
28. Metzger J, Karwath M, Tonda R, Beltran S, Águeda L, Gut M, et al. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics.* 2015;16(1):764. [PubMed: 26452642]
29. Purfield DC, McParland S, Wall E, Berry DP. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS one.* 2017;12(5):e0176780–e. [PubMed: 28463982]
30. Peripolli E, Munari DP, Silva MVGB, Lima ALF, Irgang R, Baldi F. Runs of homozygosity: current knowledge and applications in livestock. *Animal Genetics.* 2017;48(3):255–71. [PubMed: 27910110]
31. McQuillan R, Eklund N, Pirastu N, Kuningas M, McEvoy BP, Esko T, et al. Evidence of Inbreeding Depression on Human Height. *PLOS Genetics.* 2012;8(7):e1002655. [PubMed: 22829771]
32. Hu H, Kahrizi K, Musante L, Fattahi Z, Herwig R, Hosseini M, et al. Genetics of intellectual disability in consanguineous families. *Molecular Psychiatry.* 2018.
33. Gandin I, Faletra F, Faletra F, Carella M, Pecile V, Ferrero GB, et al. Excess of runs of homozygosity is associated with severe cognitive impairment in intellectual disability. *Genetics In Medicine.* 2014;17:396. [PubMed: 25232855]
34. Howrigan DP, Simonson MA, Davies G, Harris SE, Tenesa A, Starr JM, et al. Genome-wide autozygosity is associated with lower general cognitive ability. *Molecular Psychiatry.* 2015;21:837. [PubMed: 26390830]

35. Howrigan DP, Simonson MA, Davies G, Harris SE, Tenesa A, Starr JM, et al. Genome-wide autozygosity is associated with lower general cognitive ability. *Molecular Psychiatry*. 2016;21(6):837–43. [PubMed: 26390830]
36. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proceedings of the National Academy of Sciences*. 2007;104(50):19942.
37. Keller MC, Simonson MA, Ripke S, Neale BM, Gejman PV, Howrigan DP, et al. Runs of Homozygosity Implicate Autozygosity as a Schizophrenia Risk Factor. *PLOS Genetics*. 2012;8(4):e1002656. [PubMed: 22511889]
38. Casey JP, Magalhaes T, Conroy JM, Regan R, Shah N, Anney R, et al. A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Human Genetics*. 2012;131(4):565–79. [PubMed: 21996756]
39. Ghani M, Reitz C, Cheng R, et al. Association of long runs of homozygosity with alzheimer disease among african american individuals. *JAMA Neurology*. 2015;72(11):1313–23. [PubMed: 26366463]
40. Nalls MA, Guerreiro RJ, Simon-Sanchez J, Bras JT, Traynor BJ, Gibbs JR, et al. Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. *neurogenetics*. 2009;10(3):183–90. [PubMed: 19271249]
41. Clark DW, Okada Y, Moore KHS, Mason D, Pirastu N, Gandin I, et al. Associations of autozygosity with a broad range of human phenotypes. *Nature Communications*. 2019;10(1):4957.
42. Ehlers CL, Wall TL, Betancourt M, Gilder DA. The clinical course of alcoholism in 243 Mission Indians. *American Journal of Psychiatry*. 2004;161(7):1204–10.
43. Abdellaoui A, Hottenga J-J, Willemsen G, Bartels M, van Beijsterveldt T, Ehli EA, et al. Educational Attainment Influences Levels of Homozygosity through Migration and Assortative Mating. *PLOS ONE*. 2015;10(3):e0118935. [PubMed: 25734509]
44. Johnson EC, Evans LM, Keller MC. Relationships between estimated autozygosity and complex traits in the UK Biobank. *PLOS Genetics*. 2018;14(7):e1007556. [PubMed: 30052639]
45. Ehlers CL, Gizer IR, Gilder DA, Wilhelmsen KC. Linkage analyses of stimulant dependence, craving, and heavy use in American Indians. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2011;156B(7):772–80. [PubMed: 21812097]
46. Ehlers CL, Spence JP, Wall TL, Gilder DA, Carr LG. Association of ALDH1 promoter polymorphisms with alcohol-related phenotypes in southwest California Indians. *Alcohol Clin Exp Res*. 2004;28(10):1481–6. [PubMed: 15597079]
47. Kalton G, Anderson DW. Sampling Rare Populations. *Journal of the Royal Statistical Society: Series A (General)*. 1986;149(1):65–82.
48. Muhib FB, Lin LS, Stueve A, Miller RL, Ford WL, Johnson WD, et al. A Venue-Based Method for Sampling Hard-to-Reach Populations. *Public Health Reports*. 2001;116(1_suppl):216–22. [PubMed: 11889287]
49. Heckathorn DD. Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations*. *Social Problems*. 2014;44(2):174–99.
50. Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr., et al. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *Journal of Studies on Alcohol*. 1994;55(2):149–58. [PubMed: 8189735]
51. Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V. A validity study of the SSAGA--a comparison with the SCAN. *Addiction*. 1999;94(9):1361–70. [PubMed: 10615721]
52. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5 (5th ed.)*. Arlington, VA: American Psychiatric Publishing; 2013.
53. Wall TL, Carr LG, Ehlers CL. Protective association of genetic variation in alcohol dehydrogenase with alcohol dependence in Native American Mission Indians. *American Journal of Psychiatry*. 2003;160(1):41–6.
54. Hesselbrock VM, Segal B, Hesselbrock MN. Alcohol dependence among Alaska Natives entering alcoholism treatment: a gender comparison. *Journal of Studies on Alcohol*. 2000;61(1):150–6. [PubMed: 10627109]

55. Gilder DA, Wall TL, Ehlers CL. Comorbidity of select anxiety and affective disorders with alcohol dependence in southwest California Indians. *Alcohol Clin Exp Res*. 2004;28(12):1805–13. [PubMed: 15608596]
56. Schuckit MA, Smith TL, Anthenelli R, Irwin M. Clinical course of alcoholism in 636 male inpatients. *American Journal of Psychiatry*. 1993;150(5):786–92.
57. Ehlers CL, Stouffer GM, Corey L, Gilder DA. The clinical course of DSM-5 alcohol use disorders in young adult native and Mexican Americans. *The American Journal on Addictions*. 2015;24(8):713–21. [PubMed: 26346282]
58. Bizon C, Spiegel M, Chasse SA, Gizer IR, Li Y, Malc EP, et al. Variant calling in low-coverage whole genome sequencing of a Native American population sample. *BMC Genomics*. 2014;15:85. [PubMed: 24479562]
59. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*. 2011;43(5):491–8. [PubMed: 21478889]
60. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*: John Wiley & Sons, Inc.; 2002.
61. Li Y, Sidore C, Kang HM, Boehnke M, Abecasis GR. Low-coverage sequencing: implications for design of complex trait association studies. *Genome Research*. 2011;21(6):940–51. [PubMed: 21460063]
62. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75. [PubMed: 17701901]
63. Ceballos FC, Hazelhurst S, Ramsay M. Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. *BMC genomics*. 2018;19(1):106–. [PubMed: 29378520]
64. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*. 2011;88(1):76–82. [PubMed: 21167468]
65. Chen H, Wang C, Conomos Matthew P, Stilp Adrienne M, Li Z, Sofer T, et al. Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *The American Journal of Human Genetics*. 2016;98(4):653–66. [PubMed: 27018471]
66. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57(1):289–300.
67. Leisch F. FFlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R. *Journal of Statistical Software*; Vol 1, Issue 8 (2004). 2004.
68. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*. 2014;46(3):310–5. [PubMed: 24487276]
69. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*. 2018;47(D1):D886–D94.
70. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739–40. [PubMed: 21546393]
71. Kutmon M, Riutta A, Nunes N, Hanspers K, Willighagen Egon L, Bohler A, et al. WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic Acids Research*. 2016;44(D1):D488–D94. [PubMed: 26481357]
72. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*. 2017;8(1):1826.
73. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*. 2013;42(D1):D1001–D6. [PubMed: 24316577]
74. Nothnagel M, Lu TT, Kayser M, Krawczak M. Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Human Molecular Genetics*. 2010;19(15):2927–35. [PubMed: 20462934]

75. GTEx Consortium T, Aguet F, Brown AA, Castel SE, Davis JR, He Y, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550:204. [PubMed: 29022597]
76. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*. 2018;50(8):1112–21. [PubMed: 30038396]
77. Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*. 2019;51(2):245–57. [PubMed: 30643258]
78. Velez Edwards DR, Naj AC, Monda K, North KE, Neuhaus M, Magvanjav O, et al. Gene-environment interactions and obesity traits among postmenopausal African-American and Hispanic women in the Women’s Health Initiative SHARe Study. *Human Genetics*. 2013;132(3):323–36. [PubMed: 23192594]
79. Walters RK, Polimanti R, Johnson EC, McClintick JN, Adams MJ, Adkins AE, et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature Neuroscience*. 2018;21(12):1656–69. [PubMed: 30482948]
80. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*. 2019;51(2):237–44. [PubMed: 30643251]
81. Matoba N, Akiyama M, Ishigaki K, Kanai M, Takahashi A, Momozawa Y, et al. GWAS of 165,084 Japanese individuals identified nine loci associated with dietary habits. *Nature Human Behaviour*. 2020;4(3):308–16.
82. Edwards AC, Aliev F, Bierut LJ, Bucholz KK, Edenberg H, Hesselbrock V, et al. Genome-wide association study of comorbid depressive syndrome and alcohol dependence. *Psychiatric genetics*. 2012;22(1):31–41. [PubMed: 22064162]
83. Sun Y, Chang S, Liu Z, Zhang L, Wang F, Yue W, et al. Identification of novel risk loci with shared effects on alcoholism, heroin, and methamphetamine dependence. *Molecular Psychiatry*. 2019.
84. Whitfield JB, Zhu G, Madden PAF, Montgomery GW, Heath AC, Martin NG. Biomarker and Genomic Risk Factors for Liver Function Test Abnormality in Hazardous Drinkers. *Alcoholism, clinical and experimental research*. 2019;43(3):473–82.
85. Mutemberezi V, Buisseret B, Masquelier J, Guillemot-Legris O, Alhouayek M, Muccioli GG. Oxysterol levels and metabolism in the course of neuroinflammation: insights from in vitro and in vivo models. *Journal of neuroinflammation*. 2018;15(1):74–. [PubMed: 29523207]
86. Sawyer KS, Adra N, Salz DM, Kempainen MI, Ruiz SM, Harris GJ, et al. Hippocampal Subfield Volumes in Abstinent Men and Women with a History of Alcoholism. *bioRxiv*. 2019:715292.
87. Maller JJ, Broadhouse K, Rush AJ, Gordon E, Koslow S, Grieve SM. Increased hippocampal tail volume predicts depression status and remission to anti-depressant medications in major depression. *Molecular Psychiatry*. 2018;23(8):1737–44. [PubMed: 29133948]
88. Skillbäck T, Farahmand BY, Rosén C, Mattsson N, Nägga K, Kilander L, et al. Cerebrospinal fluid tau and amyloid- β 1–42 in patients with dementia. *Brain*. 2015;138(9):2716–31. [PubMed: 26133663]
89. Kalinin S, González-Prieto M, Scheiblich H, Lisi L, Kusumo H, Heneka MT, et al. Transcriptome analysis of alcohol-treated microglia reveals downregulation of beta amyloid phagocytosis. *Journal of Neuroinflammation*. 2018;15(1):141. [PubMed: 29759078]
90. Olex AL, Hiltbold EM, Leng X, Fetrow JS. Dynamics of dendritic cell maturation are identified through a novel filtering strategy applied to biological time-course microarray replicates. *BMC Immunology*. 2010;11(1):41. [PubMed: 20682054]
91. Amit I, Garber M, Chevrier N, Leite AP, Donner Y, Eisenhaure T, et al. Unbiased Reconstruction of a Mammalian Transcriptional Network Mediating Pathogen Responses. *Science*. 2009;326(5950):257. [PubMed: 19729616]
92. Lawrimore CJ, Crews FT. Ethanol, TLR3, and TLR4 Agonists Have Unique Innate Immune Responses in Neuron-Like SH-SY5Y and Microglia-Like BV2. *Alcoholism: Clinical and Experimental Research*. 2017;41(5):939–54.

93. Warden AS, Azzam M, DaCosta A, Mason S, Blednov YA, Messing RO, et al. Toll-like receptor 3 activation increases voluntary alcohol intake in C57BL/6J male mice. *Brain, Behavior, and Immunity*. 2019;77:55–65.
94. Karlsson G, Liu Y, Larsson J, Goumans M-J, Lee J-S, Thorgeirsson SS, et al. Gene expression profiling demonstrates that TGF- β 1 signals exclusively through receptor complexes involving Alk5 and identifies targets of TGF- β signaling. *Physiological Genomics*. 2005;21(3):396–403. [PubMed: 15769904]
95. Breitkopf K, Haas S, Wiercinska E, Singer MV, Dooley S. Anti-TGF- β Strategies for the Treatment of Chronic Liver Disease. *Alcoholism: Clinical and Experimental Research*. 2005;29(s2):121S–31S.
96. Kim Y-K, Lee BC, Ham BJ, Yang B-H, Roh S, Choi J, et al. Increased transforming growth factor-beta1 in alcohol dependence. *J Korean Med Sci*. 2009;24(5):941–4. [PubMed: 19794996]
97. Siegenthaler JA, Miller MW. Transforming Growth Factor β 1 Modulates Cell Migration in Rat Cortex: Effects of Ethanol. *Cerebral Cortex*. 2004;14(7):791–802. [PubMed: 15084492]
98. Steiner JL, Lang CH. Alcohol, Adipose Tissue and Lipid Dysregulation. *Biomolecules*. 2017;7(1):16.
99. Jones SE, Lane JM, Wood AR, van Hees VT, Tyrrell J, Beaumont RN, et al. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nature Communications*. 2019;10(1):343.
100. Hoffmann TJ, Choquet H, Yin J, Banda Y, Kvale MN, Glymour M, et al. A Large Multiethnic Genome-Wide Association Study of Adult Body Mass Index Identifies Novel Loci. *Genetics*. 2018;210(2):499. [PubMed: 30108127]
101. Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics*. 2018;50(7):920–7. [PubMed: 29942085]

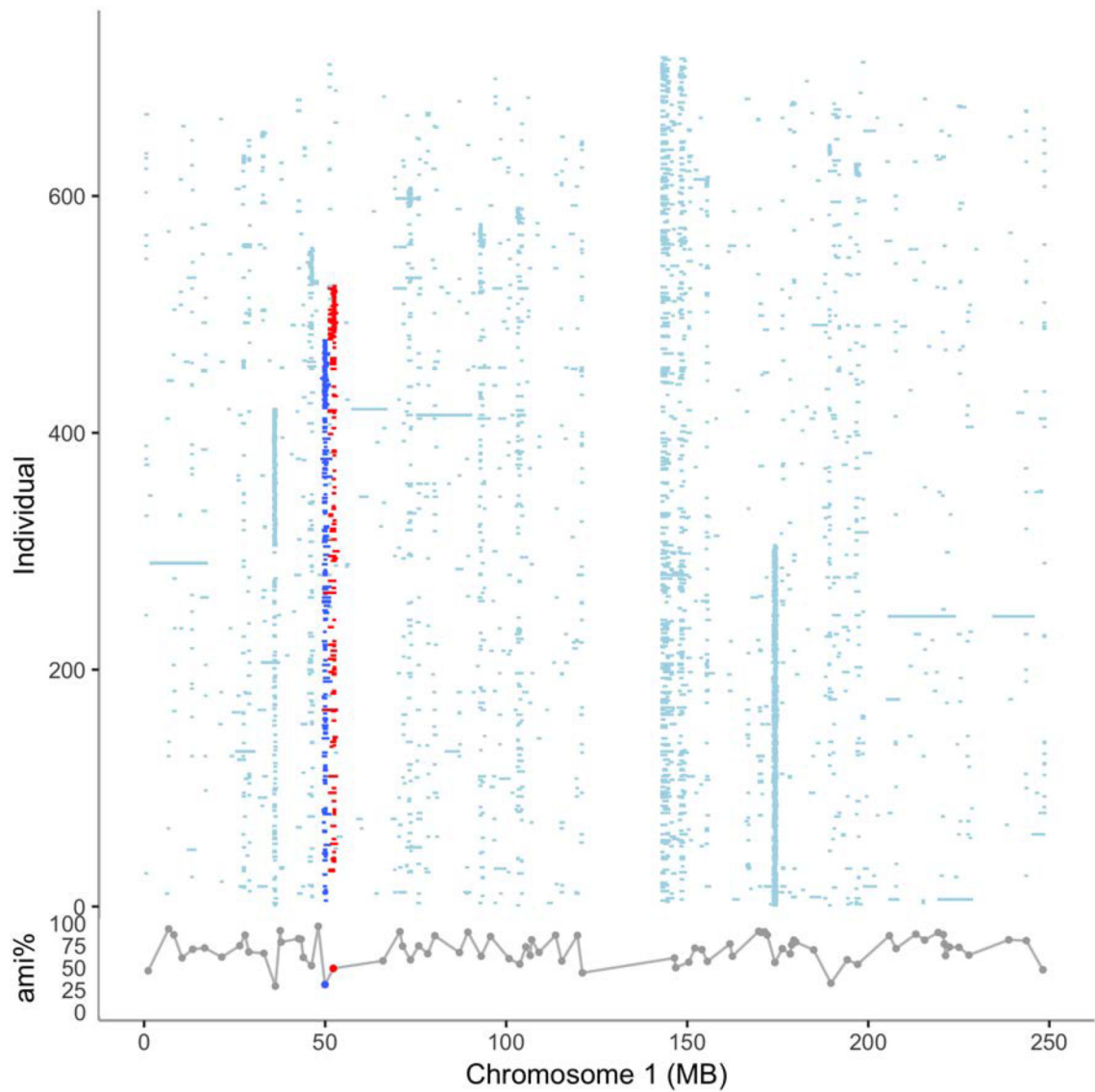


Figure 1. ROH distribution on chromosome 1 in the AI population.

Gray curve at the bottom: average American Indian ancestry of each ROH pool.

Dark blue color: A pool of ROH with consensus falling on 49.64–50.05Mb coincided with an ROH island in European populations. The mean local American Indian ancestry of this ROH pool is at a much lower level than average.

Red color: The top ROH pool on chromosome 1p32.3 that was significantly associated with AUD severity in the AI.

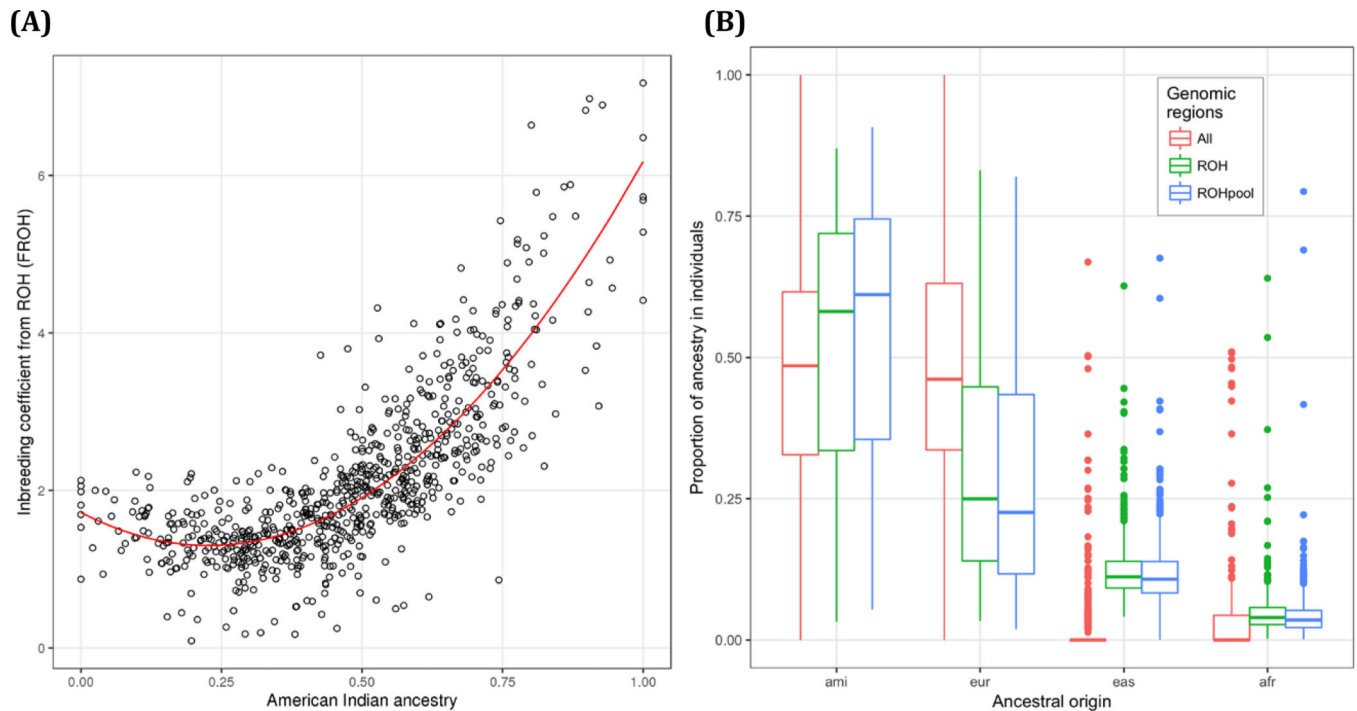
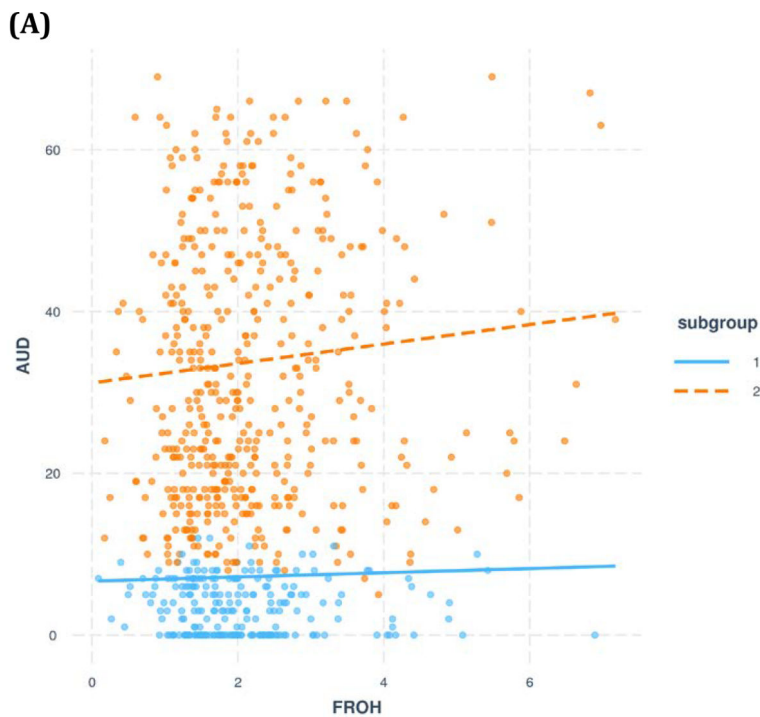


Figure 2. Relationships between ROH and American Indian ancestry.

(A) Inbreeding coefficient F_{ROH} derived from ROH (>1Mb) as a function of American Indian ancestry. The fitted red curve: $F_{ROH} = 8.15AMI^2 - 3.70AMI + 1.72$, and $AMI_{ext} = 0.227$.

(B) The genomic regions falling on ROH have significantly elevated levels of American Indian ancestry components as illustrated by the distributions of individuals' ancestral proportions across the whole genome (All: red), within ROH (ROH: green), or within ROH segments that were part of an ROH pool (ROHpool: blue). All-versus-ROH has adjusted $p=6.9E-6$. All-versus-ROHpool has adjusted $p=2.54E-10$. The difference of ancestry components between genomic regions in the ROH pools and in all ROH was not significant ($p=0.16$). Ancestral origins correspond to four major continental populations including American Indian (ami), European (eur), East Asian (eas) and African (afr).



(B)

Mean or count	Subgroup 1	Subgroup 2	Welch's <i>t</i> or Odds ratio ¹	<i>P</i> value
Individuals (<i>n</i>)	234	484		
AUD severity measure	3.51	31.91	-37.414	8.9E-155
AI ancestry	0.45	0.48	-1.871	0.062
F _{ROH}	2.04	2.17	-1.560	0.119
Age	31.00	31.31	-0.277	0.782
Sex (Male Female)	88 146	216 268	0.748	0.077
Education (years)	11.73	11.52	1.685	0.093
Employed (No Yes)	140 94	313 171	0.814	0.217
Religion (N A C1 C2) ²	57 11 65 101	90 41 137 216		0.125
AUD diagnosis				
None	187	24		
Mild AUD	46	100		
Moderate AUD	1	106		
Severe AUD	0	254		

¹ *t*-test for quantitative traits, or odds ratio (Fisher's exact test) for categorical traits.

² Religion: N=None, A=American Indian religions, C1=Christianity, C2=Catholicism.

Figure 3. Two subgroups were identified by the unsupervised clustering analysis with respect to the relationships between AUD severity and F_{ROH}.

(A) Two groups significantly differed in their severity levels of AUD. It is in the subgroup with higher severity on average (subgroup 2, orange line and dots) that AUD severity was significantly associated with F_{ROH} ($p=0.023$) as illustrated in the figure. The regression lines in the figure were plotted with continuous covariates mean-centered solely for illustration purpose. The actual analysis was conducted with the extended model controlling all covariates.

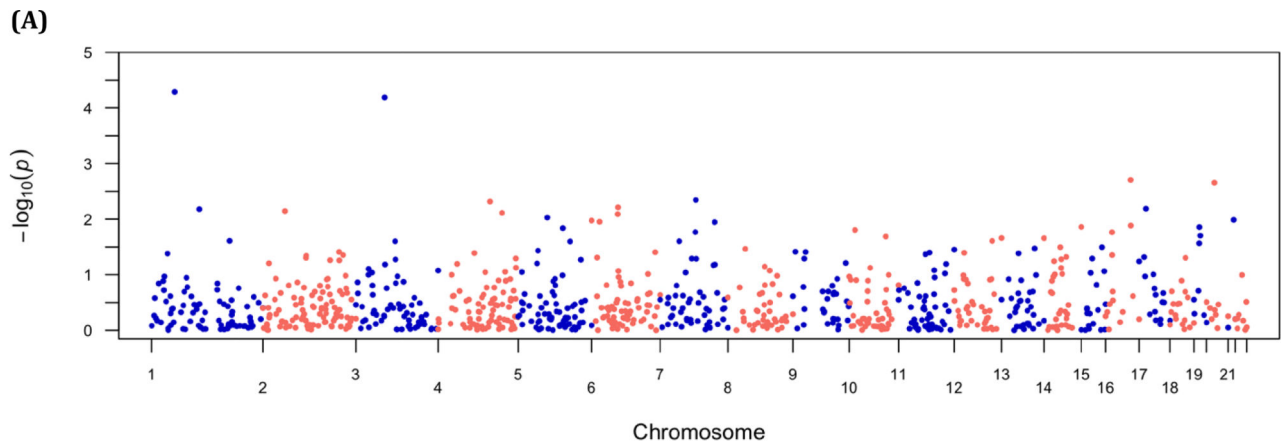
(B) While subgroup 2 consisted of much more individuals with AUD especially moderate and severe AUD diagnoses, two subgroups had similar demographics.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



(B)

Chr	Consensus segment of ROH			Union of ROH in the pool			Samples	β	p -value	adj. p
	Positions	KB	#SNP	Positions	KB	#SNP				
1	52065411 - 52234339	168.9	100	49082023 - 54005913	4923.9	5180	130	7.05	5.17E-05	0.028
3	76271735 - 76493619	221.9	529	74777436 - 77368358	2590.9	5674	5	31.01	6.47E-05	0.028
16	69965592 - 70164741	199.2	135	68840025 - 72893697	4053.7	5360	46	8.53	1.97E-03	0.383
16	70180088 - 70414495	234.4	82	68840025 - 72893697	4053.7	5360	46	8.53	1.97E-03	0.383
20	31232940 - 31359574	126.6	306	29426956 - 38950579	9523.6	13222	34	9.59	2.22E-03	0.383

Figure 4. Associations between ROH pools and AUD severity.

(A) Manhattan plot of associations between ROH pool and AUD severity.

(B) Top ROH pools associated with AUD severity. A common ROH pool on chromosome 1p32.3 and a rare ROH pool on chromosome 3p12.3 were significantly associated with AUD severity after multiple comparisons adjustments.

AUD severity is positively associated with F_{ROH} in the AI. Effects were assessed with a linear mixed model accounting for both population and family structures.

Table 1.

	Main model ($n = 741$)				Extended model ($n = 718$)					
	Estimate	SE	χ^2	p -value	Estimate	SE	χ^2	p -value	adj. p	
(Intercept)	-12.49	4.49	7.75	0.0054	0.011	5.31	6.84	0.60	0.44	0.92
Sex	-6.04	1.31	21.36	3.82E-06	8.01E-06	-5.61	1.32	17.95	2.26E-05	4.75E-05
Age	1.99	0.24	70.13	<2E-16	<2E-16	2.06	0.25	68.45	1.11E-16	2.33E-16
Age-squared	-0.02	0.00	63.94	1.33E-15	2.80E-15	-0.03	0.00	61.51	4.44E-15	9.33E-15
Years of education						-1.72	0.44	15.53	8.12E-05	1.71E-04
Currently employed						-3.22	1.40	5.31	0.021	0.045
Religion (American Indian)						6.84	2.85	5.74	0.017	0.035
Religion (Christianity)						1.49	1.94	0.59	0.44	0.93
Religion (Catholicism)						2.46	1.79	1.89	0.17	0.36
F_{ROH}	2.17	0.80	7.25	0.0071	0.015	2.04	0.80	6.47	0.011	0.023

I p -values were adjusted for two models and two ROH measurements (F_{ROH} and ROH abundance, see Supplementary Table 5). Since two ROH measurements are correlated, we adjusted for the effective number of independent variables 1.05.