



OPEN

Genetic variants associated with alcohol dependence co-ordinate regulation of *ADH* genes in gastrointestinal and adipose tissues

Rebecca Hibberd^{1,2,4,5}, Evgeniia Golovina^{1,3,5}, Sophie Farrow¹ & Justin M. O'Sullivan^{1,2,3}✉

GWAS studies have identified genetic variants associated with Alcohol Dependence (AD), but how they link to genes, their regulation and disease traits, remains largely unexplored. Here we integrated information on the 3D genome organization with expression quantitative loci (eQTLs) analysis, using CoDeS3D, to identify the functional impacts of single nucleotide polymorphisms associated with AD ($p < 1 \times 10^{-6}$). We report that 42% of the 285 significant tissue-specific regulatory interactions we identify were associated with four genes encoding Alcohol Dehydrogenase - *ADH1A*, *ADH1B*, *ADH1C* and *ADH4*. Identified eQTLs produced a co-ordinated regulatory action between *ADH* genes, especially between *ADH1A* and *ADH1C* within the subcutaneous adipose and gastrointestinal tissues. Five eQTLs were associated with regulatory motif alterations and tissue-specific histone marks consistent with these variants falling in enhancer and promoter regions. By contrast, few regulatory connections were identified in the stomach and liver. This suggests that changes in gene regulation associated with AD are linked to changes in tissues other than the primary sites of alcohol absorption and metabolism. Future work to functionally characterise the putative regulatory regions we have identified and their links to metabolic and regulatory changes in genes will improve our mechanistic understanding of AD disease development and progression.

Alcohol dependence (AD) is distinct from the responsible social use of alcohol. Rather, AD is a medical condition that is characterised by “craving, tolerance, a preoccupation with alcohol and continued drinking, in spite of harmful consequences” (NICE)¹. AD contributes to the 5.1% of world-wide disease burden with which alcohol is associated (WHO)². Common medical conditions associated with AD include physical conditions such as heart disease, hypertension and a range of gastrointestinal cancers³, as well as mental and psychiatric disorders such as major depressive disorder, schizophrenia and bipolar disorder⁴. In addition to its medical implications, AD has been linked to increases in antisocial behaviour in young adults⁵ and crime incidence⁶, adding to the economic costs of the condition.

AD is a clinical diagnosis and the severity of the condition presents as a continuum across populations. Both genetic and environmental factors impact an individual's likelihood of developing the condition, with twin-studies estimating the heritability of alcohol use disorders to be around 50%⁷. Genetic variants that are associated with, and therefore likely predispose to, AD have been identified^{8–13}. However, the impacts of these genetic variants on metabolism and alcohol detoxification, and the mechanisms driving such impacts, remain poorly understood.

Ethanol, the alcohol present in ‘alcoholic’ beverages, is a toxic substance which is primarily removed from the body via its oxidation in the liver. Ethanol is absorbed through the stomach and small intestine into the blood which is circulated to the liver via the hepatic portal vein for filtering¹⁴. The liver is generally recognised as the main site for alcohol metabolism¹⁵. First pass metabolism also occurs in the gastric system and depends on the

¹Liggins Institute, The University of Auckland, Auckland, New Zealand. ²MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, United Kingdom. ³A Better Start National Science Challenge, Auckland, New Zealand. ⁴Natural Sciences, Faculty of Environmental and Life Sciences, University of Southampton, Southampton, United Kingdom. ⁵These authors contributed equally: Rebecca Hibberd and Evgeniia Golovina. ✉e-mail: justin.osullivan@auckland.ac.nz

activity of enzymes and the rate of gastric emptying¹⁴. After first pass metabolism, ethanol is distributed around the body by the blood to various tissues where it can also be metabolised in the cytosol of cells¹⁴.

The main pathway of ethanol breakdown involves the oxidation of ethanol to acetaldehyde, mediated by alcohol dehydrogenase (ADH). Due to the toxicity of acetaldehyde at high or sustained concentrations¹⁶, a second reaction oxidises acetaldehyde to acetate, catalysed by aldehyde dehydrogenase (ALDH). There are seven genes coding for ADH isozymes, aligned head to tail on chromosome 4. ADH isozymes are very highly transcribed in the liver (excluding ADH7; GTE version 8¹⁷). Other enzymes have minor roles in alcohol metabolism, for example CYP2E1 is part of the liver cytochrome P450 enzyme system, and catalase^{14,18,19}.

Although the genes coding ADH and ALDH are known to be important in alcohol metabolism^{10,14,20,21}, the way in which their expression is regulated in different tissues is not fully understood. Studies have identified alleles in *ADH* and *ALDH* that have protective effects against developing AD^{22–24}. Similarly, comparative population studies have highlighted the roles of different alleles in AD disease risk (reviewed in Dasgupta 2015¹⁴).

Genome wide association studies (GWAS) search the entire human genome to identify genetic markers (i.e. Single Nucleotide Polymorphisms [SNPs]) that are associated with and may even be causal for a particular phenotype. GWAS incorporate no prior knowledge of the SNPs, meaning that there is no assigned functional information. Thus, all variants have an equal likelihood of being causally associated with the tested phenotype to start with. Despite GWAS studies into AD identifying SNPs that confirm a role for genetic variation^{8–13}, the individual and co-operative mechanism through which these variants impact an individual's disease risk remain largely unknown. The functional characterisation of these variants is complicated by the fact that most of the SNPs are present in non-coding (intergenic and intronic) regions of the genome^{25–27}. It is likely that many of these variants affect the regulation of metabolic genes consistent with observations for different phenotypes^{26,28–30}. Regulatory roles for these alcohol dependent variants could impact a wide range of biological pathways in different ways and to different extents in different tissues. The infinitesimal hypothesis (Fisher)³¹ argues that the impact of each variant is small, requiring numerous changes to affect a phenotype. Incorporating expression quantitative trait loci (eQTL) and chromatin structural data (i.e. Hi-C) into SNP analyses can help to inform on tissue specific regulatory impacts for SNPs^{28,32–34}. Recent studies have taken advantage of these novel methods in a range of pathophysiological contexts to determine the collective impacts of the genetic variation on tissue specific gene regulation^{33,35–38}. These novel approaches hold the key to understanding the role that genetic variation plays in AD, potential therapeutic and preventative treatments.

Here, we interrogate the long-range regulatory effects of 73 GWAS SNPs associated with AD³⁹. By integrating data on the functions of known genes, tissue specific expression, regulatory elements, and structural genome data, we identify how SNPs may individually and collaboratively impact the pathophysiology of AD.

Results

AD-associated GWAS SNPs mark spatial regulatory eQTLs. SNPs ($n = 73$) associated with AD ($p < 1 \times 10^{-6}$) were obtained from the NHGRI-EBI GWAS Catalog³⁹ (www.ebi.ac.uk/gwas/; Supplementary Table 1). Of the 73 SNPs, 6 were coding variants (4 missense, 2 synonymous) and 67 were non-coding variants (Supplementary Table 1). CoDeS3D⁴⁰ was used to interrogate 3D genome structure data (Supplementary Fig. 1; Supplementary Table 2) to identify tissue-specific spatial eQTLs (and their associated gene – “eGene”; GTE multi-tissue dataset version 7, Supplementary Table 3) for 34 SNPs. The 34 eQTLs (16 intronic, 9 intergenic, 3 exonic, 3 ncRNA intronic, 1 upstream, 2 UTR3; Supplementary Fig. 2) were involved in 274 cis- and 11 trans-acting regulatory interactions between 72 statistically significant (FDR < 0.05; Benjamini Hotchberg⁴⁰) SNP-eGene pairs across 48 different human tissues (Supplementary Table 4). Nine of the trans-regulated eGenes are loss of function tolerant (pLI = 0.00–0.10, gnomAD v2.1.1⁴¹), whilst *GREBL1L* is severely loss of function intolerant (pLI = 1, gnomAD v2.1.1⁴¹). *TPT1-AS1* had no data available. For example, rs2094081 was associated with trans-acting down-regulation of *CYP4B1* (pLI = 0.00, gnomAD v2.1.1⁴¹) in stomach tissue. *CYP4B1* is a member of the Cytochrome P450 superfamily of enzymes, which contribute to an alternative pathway for alcohol metabolism^{14,18}.

24 interactions we identified were associated with regulation of 12 eGenes (including *WDR5B*, *FABP3*, *KCNJ6* and *ADH1C*) in 9 different brain tissues, of which two were trans-acting (Supplementary Table 4). Notably, the trans-acting interaction with rs7913179 was associated with upregulation of *GFRA2* (pLI = 0.1, gnomAD v2.1.1⁴¹) in brain amygdala. *GFRA2* encodes glial cell line-derived neurotrophic factors (GDNF) that play a role in different biological processes including neurons cell survival and neurite outgrowth⁴². Signaling by GDNF promotes the survival of dopaminergic neurons⁴² and has been considered as a therapeutic target for the treatment of alcoholism⁴³.

We performed gene ontology (GO) and pathway analyses using g:Profiler (the g:GOST module)⁴⁴ on the genes affected by the eQTLs. We identified 17 molecular function categories (including “alcohol dehydrogenase activity”), 25 biological process categories (including “ethanol oxidation” and “ethanol metabolic process”) and 27 cellular components were significantly enriched ($p < 0.05$, adjusted using the SCS algorithm⁴⁴; Supplementary Table 5) for eGenes. Pathway analysis identified eight significantly enriched biological pathways (including fatty acid degradation and metabolic pathways) associated with AD ($p < 0.05$, adjusted by the SCS algorithm⁴⁴; Supplementary Table 6). These findings provide further evidence supporting the importance of ethanol oxidation, lipid storage and metabolism in the pathophysiology of alcohol disorders such as AD^{18,45}.

Despite the pivotal role of the liver in ethanol metabolism^{15,18}, only one eQTL exhibited significant impact on gene expression in the liver (rs4699741-*SCARB2*) (Supplementary Table 4, Supplementary Fig. 3). This was the only eQTL within the *ADH* region that exhibited a significant regulatory effect on a gene other than *ADH*. Notably the presence of the alternate allele upregulated *SCARB2* expression through trans-regulation. *SCARB2* is loss of function tolerant (pLI = 0.00; gnomAD v2.1.1⁴¹).

A 100 kb locus on chromosome 4 (chr4:100210000–100310000), encompassing the *ADH* genes, contained 38% (n = 13) of the eQTLs we identified (Supplementary Fig. 2). 12 of these eQTLs (3 intronic, 4 intergenic, 3 exonic, 1 upstream, 1 ncRNA intronic) and ~42% of regulatory interactions (n = 119) were associated with regulation of the *ADH* genes. In total there were 21 statistically significant eQTL-eGene pairs involving *ADH* genes with a total of 118 cis-acting regulatory interactions across 24 different human tissues (Table 1, Supplementary Fig. 4). These interactions involved four of the seven *ADH* genes within the locus: *ADH1A*, *ADH1B*, *ADH1C* and *ADH4*. Notably, the 3 exonic SNPs that were tested and had significant eQTLs were located within the *ADH* gene region (Supplementary Table 4, Supplementary Fig. 2). Two SNPs (rs1229984, rs17028615) located within the *ADH* locus that were not associated with altered *ADH* gene expression, also did not exhibit any significant regulatory effects on other genes. eQTLs within the *ADH* locus were associated with a range of tissue-specific regulatory interactions and effect sizes with the four *ADH* genes (Fig. 1a; Table 1). None of the SNPs that were eQTLs for the *ADH* genes exhibited significant eQTLs with any other genes. There were no AD-associated SNPs located within the *ADH1A* or *ADH4* genes (Fig. 1a). rs1826907 was the only eQTL associated with regulation of three *ADH* genes - *ADH1A*, *ADH1B* and *ADH1C* (Fig. 1b). By contrast, 11 SNPs either associated with regulation of *ADH1A*, *ADH1A* and *ADH1C*, or *ADH1C* and *ADH4* simultaneously – but not necessarily in the same tissues (Fig. 1a,b, and Table 1). Four eQTLs (rs1789891, rs1789924, rs1826907 and rs2066702) were associated with downregulation of at least one *ADH* gene in subcutaneous adipose tissue (Fig. 1a, Table 1).

ADH regulation in adipose tissue. Fatty Liver Disease has direct links to chronic alcohol consumption through the mobilisation of adipose tissue^{19,46,47}. Moreover, subcutaneous adipose tissue is thought to impact the secondary metabolism of ethanol after the liver. We observed variant specific impacts on adipose tissue at the *ADH* locus (Fig. 2 and Table 1). Subcutaneous adipose has the most eQTL interactions involving *ADH* genes of any tissue, with 13 regulatory interactions across 11 of the 12 SNPs. It is the only tissue in which we identified eQTLs for all four *ADH* genes (Table 1, Supplementary Fig. 4, Supplementary Table 4). By contrast, only one eQTL, involving *ADH1C* and rs1789924, was identifiable in visceral adipose.

Epigenetic modifications (e.g. H3K4ac, H3K27ac, H3K4me3 and H3K9me3) are recognised as marking promoter and enhancer regions⁴⁸. To further investigate the regulatory activity of the 12 significant eQTLs associated with *ADH* genes in subcutaneous adipose, we obtained data on promoter and enhancer histone marks, DNase marks and regulatory motif alterations from HaploReg (version 4.1)^{49,50} (Fig. 2, Supplementary Table 7). The locus marked by rs1826907 was associated with down regulation of *ADH1B* interaction in subcutaneous adipose (Fig. 2). HaploReg^{49,50} annotates the intergenic rs1826907 as modifying regulatory sequences within an enhancer region associated with adipose nuclei (Fig. 2, Supplementary Table 8). rs1789891 and rs1826907 are associated with the downregulation of different *ADH* genes (*ADH1B* and *ADH1A/1C*, respectively) in a variety of tissues (Table 1, Fig. 2), but are both associated with downregulatory interactions in subcutaneous adipose tissue. This may be explained by the fact that rs1789891 and rs1826907 alter regulatory binding motifs for many transcription factors. Notably, rs1789891 is predicted to decrease binding affinity for Ap-1, Pbx-3, Nf-Y, Rfx5 and Sp2 transcription factors, whereas rs1826907 is predicted to increase binding affinity for the same transcription factors (Supplementary Table 9). All of these transcription factors are transcribed in both subcutaneous and visceral adipose tissue (GTEx version 8¹⁷). Similarly, rs1789891 is predicted to increase binding affinity for Sox at one known motif, whereas rs1826907 is predicted to increase binding affinity at nine alternative Sox motifs (Supplementary Table 9). This provides a potential mechanistic explanation for the observed but coordinated differences in eQTL effects.

Rs2066702 is an exonic variant within *ADH1B* causes a missense mutation⁵¹ and is an eQTL associated with downregulation of *ADH1A* in subcutaneous fat. This is the only interaction for this eQTL in any gene or tissue (Supplementary Table 4, Supplementary Fig. 5). Notably, the locus marked by rs2066702 has both enhancer (H3K4me1, H3K27ac) and promoter (H3K9ac) histone marks in adipose nuclei (Supplementary Tables 7 and 8). However, rs2066702 only alters one regulatory motif, GATA, and the predicted binding affinity was not substantially changed (Supplementary Table 9).

rs1789882, rs1693457 and rs904092 are in high linkage disequilibrium (LD) and are associated with upregulation of *ADH1A* transcription in subcutaneous adipose (Fig. 2, Supplementary Fig. 4). rs1693457 alters a Sox binding motif, predicted to cause increased binding affinity with the transcription factor (Supplementary Table 9). rs1789882 has alterations in Bdp1 and Stat4 motifs that are predicted to substantially increase the binding affinity to these transcription factors (Supplementary Table 9). rs1693457 and rs1789882 are also associated with both enhancer (H3K4me1, H3K27ac) and promoter (H3K4me3, H3K9ac) adipose-specific histone marks (Supplementary Tables 7 and 8). These findings are suggestive of a significant regulatory region located within the *ADH1B* gene.

rs12639833, rs2173201, rs1614972 and rs2241894 are all significantly associated with the upregulation of *ADH1C* in subcutaneous adipose tissue. The four SNPs have a similar pattern of associated interactions with *ADH1C* across tissues, with rs12639833, rs2173201 and rs2241894 having interactions with the same 8 tissues, and rs1614972 having associated interactions with fewer (n = 3). All four of these SNPs are in high LD (Supplementary Fig. 5), which likely explains the similarity in the regulatory interactions we found. The exonic variant rs2241894 has the only annotated enhancer histone mark (H3K27ac) for these SNPs in adipose nuclei (Supplementary Tables 7 and 8) but does not alter any regulatory motifs, unlike the other three (Fig. 2, Supplementary Table 9).

When looking across all gene interactions, only 7 eQTLs were identified to regulate gene expression in visceral adipose, compared to 23 in subcutaneous adipose tissue. By contrast, more than 55% (13 out of 23) of the identified eQTL interactions within the subcutaneous adipose occur within the *ADH* region. Rs1789924 is the only eQTL that is associated with genetic regulation in visceral adipose regulation and downregulates *ADH1C* in both visceral and subcutaneous adipose (as well as other tissues; Table 1). The variant is annotated with enhancer

Tissue	Gene	SNP ID	Effect Size	SNP ID	Effect Size	SNP ID	Effect Size	SNP ID	Effect Size	SNP ID	Effect Size	SNP ID	Effect Size
Adipose Subcutaneous	<i>ADH1A</i>	rs1693457	0.638	rs1789882	0.638	rs1789891	-0.342	rs2066702	-0.696	rs904092	0.634		
	<i>ADH1B</i>	rs1826907	-0.174										
	<i>ADH1C</i>	rs12639833	0.252	rs1614972	0.255	rs1789891	-0.321	rs1789924	-0.459	rs2173201	0.252	rs2241894	0.256
	<i>ADH4</i>	rs1789924	-0.26										
Adipose Visceral Omentum	<i>ADH1C</i>	rs1789924	-0.409										
Adrenal Gland	<i>ADH1A</i>	rs1693457	0.539	rs1789882	0.54	rs904092	0.54						
Artery Aorta	<i>ADH1A</i>	rs1693457	0.691	rs1789882	0.691	rs1789891	-0.484	rs904092	0.692				
	<i>ADH1B</i>	rs1826907	-0.296										
Artery Tibial	<i>ADH1A</i>	rs1693457	0.479	rs1789882	0.49	rs904092	0.489						
	<i>ADH1B</i>	rs1826907	-0.298										
	<i>ADH1C</i>	rs1789891	-0.279	rs1789924	-0.26								
Brain Caudate basal ganglia	<i>ADH1C</i>	rs12639833	0.632	rs2173201	0.634	rs2241894	0.623						
Breast Mammary Tissue	<i>ADH1A</i>	rs1693457	0.491	rs1789882	0.491	rs904092	0.491						
	<i>ADH1C</i>	rs1789924	-0.385										
Cells Transformed fibroblasts	<i>ADH1A</i>	rs1693457	0.336	rs1789882	0.336	rs1826907	-0.272	rs904092	0.343				
	<i>ADH1B</i>	rs1826907	-0.166										
	<i>ADH1C</i>	rs1826907	-0.224										
Colon Sigmoid	<i>ADH1A</i>	rs1693457	0.636	rs1789882	0.636	rs904092	0.636						
	<i>ADH1C</i>	rs1789924	-0.383										
Colon Transverse	<i>ADH1A</i>	rs1693457	0.447	rs1789882	0.433								
Esophagus Gastroesophageal Junction	<i>ADH1A</i>	rs1693457	0.424										
Esophagus Mucosa	<i>ADH1C</i>	rs12639833	0.371	rs1614972	0.284	rs1789891	-0.285	rs1789924	-0.308	rs2173201	0.375	rs2241894	0.373
	<i>ADH1A</i>	rs1693457	0.438	rs1789882	0.438	rs904092	0.427						
	<i>ADH1C</i>	rs1789924	-0.273										
Esophagus Muscularis	<i>ADH4</i>	rs12639833	0.387	rs2173201	0.404	rs2241894	0.393						
	<i>ADH1A</i>	rs1693457	0.577	rs1789882	0.586	rs904092	0.586						
	<i>ADH1C</i>	rs12639833	0.328	rs1789891	-0.501	rs1789924	-0.508	rs2173201	0.332	rs2241894	0.332		
Heart Atrial Appendage	<i>ADH1A</i>	rs1693457	0.476	rs1789882	0.471	rs904092	0.458						
	<i>ADH1C</i>	rs1789924	-0.362										
Heart Left Ventricle	<i>ADH1A</i>	rs1693457	0.476	rs1789882	0.471	rs904092	0.458						
	<i>ADH1C</i>	rs1789924	-0.362										
Lung	<i>ADH1A</i>	rs1693457	0.646	rs1789882	0.648	rs904092	0.648						
	<i>ADH1C</i>	rs12639833	0.336	rs1614972	0.289	rs1789924	-0.266	rs2241894	0.337	rs9307239	0.216	rs2173201	0.338
Muscle Skeletal	<i>ADH1B</i>	rs1826907	-0.24										
	<i>ADH4</i>	rs12639833	0.31	rs1614972	0.256	rs2173201	0.305	rs2241894	0.305	rs9307239	0.237		
Nerve Tibial	<i>ADH1A</i>	rs1693457	0.372	rs1789882	0.372	rs904092	0.373						
	<i>ADH1B</i>	rs1826907	-0.301										
	<i>ADH1C</i>	rs1789891	-0.297	rs1789924	-0.221								
Pancreas	<i>ADH1C</i>	rs12639833	0.53	rs2173201	0.534	rs2241894	0.536						
Skin Not Sun Exposed Suprapubic	<i>ADH1A</i>	rs1693457	0.327	rs1789882	0.342	rs904092	0.344						
	<i>ADH1C</i>	rs1789924	-0.253										
Skin Sun Exposed Lower leg	<i>ADH1A</i>	rs1693457	0.461	rs1789882	0.463	rs904092	0.457						
	<i>ADH1C</i>	rs12639833	0.233	rs1789924	-0.257	rs2173201	0.236	rs2241894	0.239				
Small Intestine Terminal Ileum	<i>ADH1A</i>	rs1693457	0.727	rs1789882	0.727	rs904092	0.74						
Spleen	<i>ADH1C</i>	rs12639833	0.588	rs2173201	0.572	rs2241894	0.602						
Thyroid	<i>ADH1A</i>	rs1693457	0.441	rs1789882	0.44	rs904092	0.444						
	<i>ADH1B</i>	rs1826907	-0.217										
	<i>ADH1C</i>	rs1789924	-0.264										

Table 1. Significant eQTL-tissue regulatory interactions associated with the *ADH* genes located on Chromosome 4. SNP ID and effect size are shown for each interaction.

(H3K4me1, H3K27ac) and promoter (H3K4me3, H3K9ac) histone marks in adipose nuclei (Supplementary Tables 7 and 8).

Collectively, these results suggest that changes to the subcutaneous adipose tissue specific regulation of the *ADH* locus might be a predisposing factor for AD.

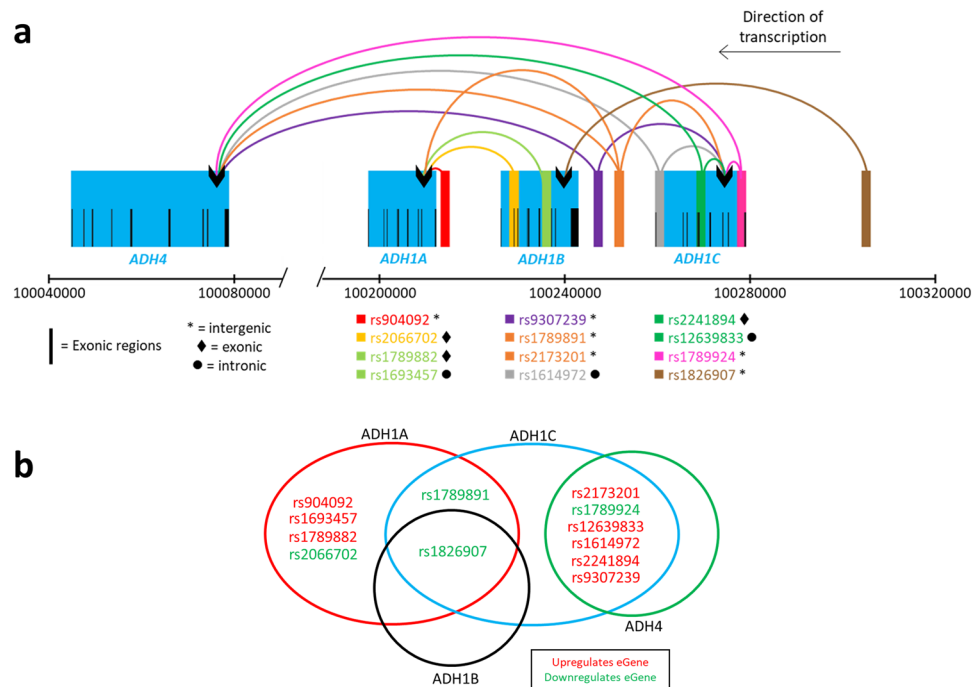


Figure 1. The 12 eQTLs identified as having regulatory interactions with *ADH* genes. Three SNPs located within the *ADH* gene cluster did not interact with *ADH* genes, so are excluded. Of these, one interacted with another gene, *SCARB2*, whilst the other two were not identified as eQTLs. (a) Spatial eQTL-eGene interactions within the *ADH* gene cluster (Chromosome 4). ♦♦*symbols represent functional annotations obtained from wANNOVAR^{59,60}. For simplicity, SNPs separated by <2 kbp in the linear sequence were grouped with the same colour. Genomic locations are according to human genome hg19. The eQTL-eGene interaction analysis used GTEx v7¹⁷. (b) Cross-over of eQTL-eGene interactions in the *ADH* gene region. Red text denotes the SNP is associated with upregulation of the eGene, whereas green text denotes associated downregulation. Multiple variants are associated with co-ordinated upregulation of *ADH1A* or *ADH4* and *ADH1C*.

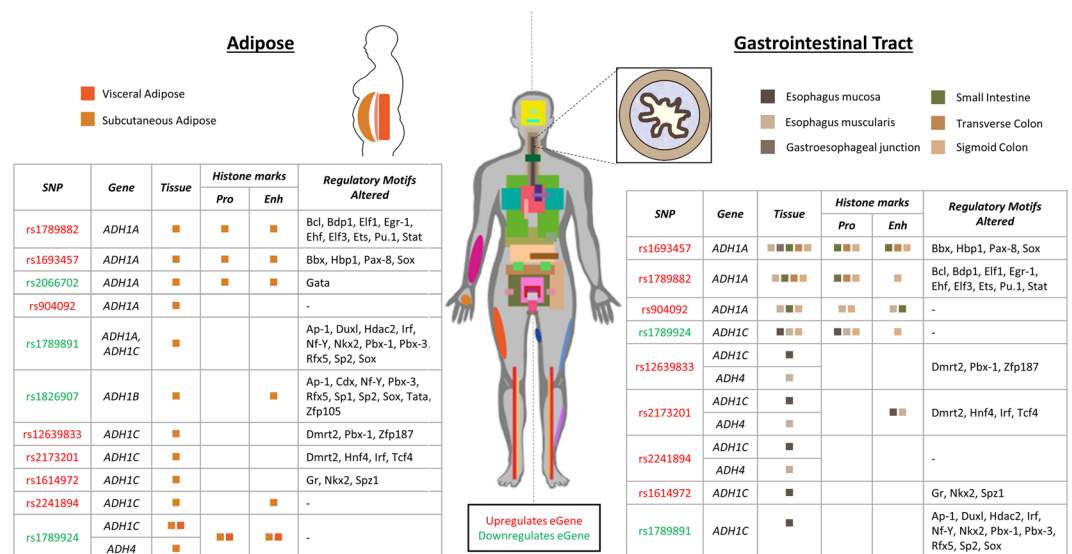


Figure 2. Adipose and gastrointestinal tract tissues are of interest due to their known roles in alcohol metabolism and have many significant eQTL interactions with *ADH* genes. SNPs affect regulatory motifs and are present in areas of the genome that have tissue-specific histone marks that infer some regulatory function. Coloured boxes represent the presence of an eQTL for the listed eGene in the tissue and found tissue-specific histone marks for the SNP. Tissues are represented by the same colour in the table and diagrams. Enh = enhancer histone marks; Pro = promoter histone marks. H3K4ac and H3K27ac are grouped as enhancer histone marks and H3K4me3 and H3K9ac as promoter histone marks. Histone marks and regulatory motif alterations were obtained from HaploReg (version 4.1)^{49,50}. Red text denotes the SNP is associated with upregulation of the eGene, whereas green text denotes associated downregulation.

ADH regulation in the Gastrointestinal system. The gastrointestinal (GI) tract plays a pivotal role in the absorption of alcohol after consumption, primarily in the stomach and small intestine (20% and 80% of absorption, respectively)¹⁴. These tissues also have a role in primary metabolism of the ethanol through ADH isozymes, of which *ADH1B*, *ADH1C* and *ADH5* are highly transcribed in stomach and small intestine (GTEx version 8¹⁷).

68% (49 out of 72) of eQTLs associated with regulation of *ADH* gene expression were found to have regulatory effects in tissues of the GI tract (8 sigmoid colon, 4 transverse colon, 2 gastroesophageal junction, 15 esophagus mucosa, 13 esophagus muscularis, 1 minor salivary gland, 4 small intestine-terminal ileum, 2 stomach; Fig. 2, Supplementary Fig. 5). *ADH1A* expression was modified by AD associated SNPs in 5 tissues across the GI tract; *ADH1C* in 3 and *ADH4* in 1. Notably, *ADH4* expression was regulated by eQTLs that also affect *ADH1C* expression in esophagus tissue (*ADH4* in muscularis, *ADH1C* in mucosa; Fig. 2). 3 eQTLs (rs2173201, rs2241894, rs12639833) affected *ADH4* expression in both esophagus muscularis tissue and skeletal muscle, suggestive of them acting as a co-ordinated set in muscular tissues. Of these variants, rs2173201 is the only one with H3K27ac (enhancer) histone marks in esophageal tissue (Fig. 2, Supplementary Table 7 and 8) and is associated with a predicted increase in binding affinity of Dmrt2 and Irf transcription factors to the regulatory motif (Supplementary Table 9).

Similar to our observations in adipose tissue, rs904092, rs1789882 and rs1693457 show a related pattern of regulatory interactions for upregulation of *ADH1A* within GI tissues (i.e. esophagus muscularis, small intestine and sigmoid colon tissues). Moreover, each of these variants is associated with both enhancer (H3K4me1 and/or H3K27ac) and promoter (H3K4me3 and/or H3K9ac) histone marks in at least one GI tissue (Fig. 2, Supplementary Table 7 and 8). rs1693457 is an intronic SNP within *ADH1B* that is associated with changes in *ADH1A* expression and has the most tissue interactions of any eQTL we identified. Compared to rs904092, rs1693457 has additional upregulatory associations with *ADH1A* in the gastroesophageal junction and transverse colon. This is the only interaction we identified in the gastroesophageal junction. As stated earlier, the similarity between rs904092, rs1789882 and rs1693457 is likely to reflect the high LD between the SNPs and the existence of a composite regulatory element. Notably, rs1789882 is an exonic SNP within *ADH1B* that is linked to differential expression of *ADH1A*. The rs1789882 variant alters Ehf and Elf3 regulatory motifs (Fig. 2, Supplementary Table 9), transcription factor paralogs which are reported to have epithelial-specific expression in the intestine⁵². However, the change to the minor allele does not appear to substantially change the predicted binding affinity of these proteins. Rather, alterations to Bdp1 and Stat4 binding sites are predicted to increase binding affinity and these transcription factors are transcribed in all studied tissues in the GI tract (GTEx version 8¹⁷). rs2173201 and rs1789891 are intergenic SNPs that are located <2kbp apart but have opposing regulatory effects on *ADH* gene expression in esophagus tissues. rs2173201 is associated with *ADH1C* upregulation in esophagus mucosa; *ADH4* upregulation in *ADH4*; H3K27ac (enhancer) histone marks in esophagus cells; and is predicted to alter binding motifs as previously mentioned (Fig. 2). Conversely, rs1789891 is associated with down regulation of *ADH1C* in esophagus mucosa and lacks associated histone marks in esophagus cells (Fig. 2, Supplementary Table 8). However, rs1789891 substantially decreases binding affinity for many transcription factors including Ap-1, Irf, Nf-Y, Pbx-3, Rfx5, Sp2 (Supplementary Table 9). All of these are transcribed in a variety of GI tract tissues, as well as many other tissues in the body (GTEx version 8¹⁷).

rs1614972 is an example of an intronic SNP within *ADH1C*, which is associated with upregulation of *ADH1C* expression. rs1614972 only has association in a single tissue for the GI system, esophagus mucosa, and has no histone marks for this tissue. The SNP is however associated with a predicted increase in Nkx2 transcription factor binding affinity and decrease in Gr (aka Nr3c1) and Spz1 transcription factor binding affinity (Supplementary Table 9). Nkx2 and Spz1 are usually lowly transcribed in the GI tissues, whereas Nr3c1 is highly transcribed (GTEx version 8¹⁷).

ADH regulation is altered in the cardiovascular system and lungs. Alcoholism is known to have links with cardiovascular disease³. To investigate if a genetic predisposition to change *ADH* regulation contributes mechanistically to the risk of alcohol linked cardiovascular disease, we investigated eQTLs with interactions in cardiovascular tissues (aorta, atrial appendage, left ventricle, tibial artery) and the lung, integrating HaploReg (version 4.1)^{49,50} data on regulatory motifs and histone marks (Fig. 3). Consistent with our earlier findings, we observed specific eQTLs that affected expression of *ADH1A*, *ADH1B* and *ADH1C*. Again, these variants modify known transcription factor binding sites and were associated with epigenetic modifications that indicate the presence of enhancer elements (Fig. 3, Supplementary Tables 8 and 9).

Discussion

In this study, we integrated information on AD-associated GWAS SNPs with data on 3D genome organization, eQTLs and regulatory annotations to identify the tissue-specific regulatory roles of these variants, and thus how they potentially contribute to AD. Our findings expand our understanding of AD-associated variants, highlighting the regulatory impacts of these variants on gene expression in tissues not conventionally associated with AD.

We have shown that these variants seem to impact *ADH* expression in tissues not generally considered important for first pass metabolism of ethanol, including the lower GI tract and subcutaneous adipose tissue. Whilst these variants were not found to directly impact pathways involved in behavioural aspects of AD (i.e. in brain tissue), they suggest a potential increase in the ability of individuals to tolerate alcohol consumption through metabolic changes, therefore making them more susceptible to developing AD. This has yet to be proven clinically.

Alcohol usage is associated with many cardiovascular disorders³. Consistent with this, we identified associated *ADH* regulatory interactions in arteries, heart and lung tissues. This suggests that these disorders may not just be a side product of alcohol consumption, but rather they may be multimorbid due to tissue specific impacts of a sub-set of the AD associated genetic variants. This multimorbidity may extend to the obesity that is often

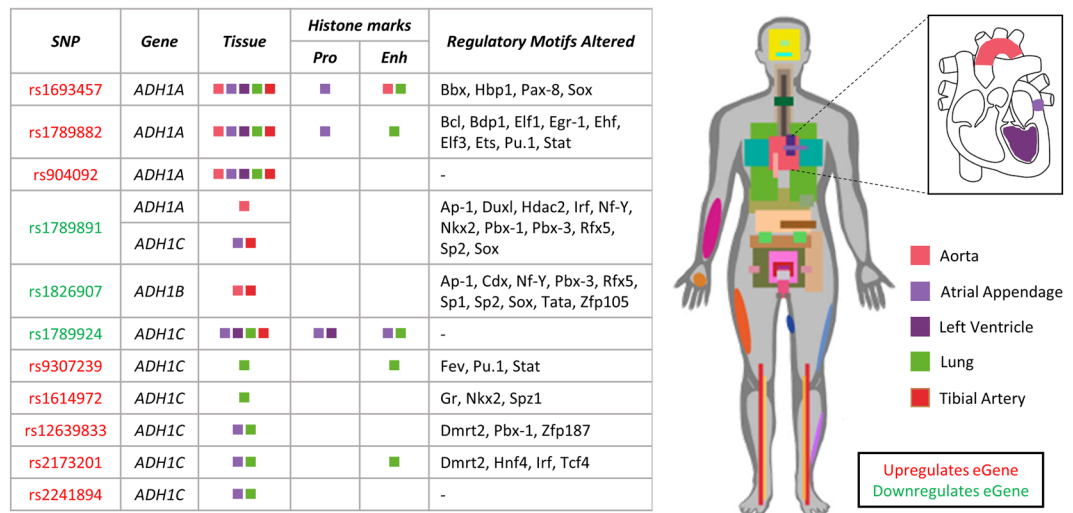


Figure 3. eQTL-tissue regulatory interactions associated with tissues in the cardiovascular system and the *ADH* genes. SNPs affect regulatory motifs and are present in areas of the genome that have tissue-specific histone marks that infer some regulatory function. Coloured boxes represent a significant eQTL-eGene interaction was found in the tissue and any tissue-specific histone marks, in respective columns. Tissues are represented by the same colour in the table and body diagrams. Enh = enhancer histone marks; Pro = promoter histone marks. H3K4ac and H3K27ac are grouped as enhancer histone marks and H3K4me3 and H3K9ac as promoter histone marks. Red text denotes the SNP is associated with upregulation of the eGene, whereas green text denotes associated downregulation. Histone marks and regulatory motif alterations were obtained from HaploReg (version 4.1)^{49,50}.

associated with AD. Future analyses that compare the impacts of obesity/BMI associated variants with those of AD will help to resolve whether there are shared regulatory mechanisms between AD and obesity.

We identified just one association in liver tissue, a trans-acting upregulatory interaction between rs4699741 (on chromosome 4, just beyond the *ADH* locus) and *SCARB2*. *SCARB2* encodes LIMP-2 which functions as part of lysosome and endosome assembly and organisation⁵³ and has not previously been linked to AD. Notably, we observed upregulation of *CYP4B1* was associated with rs2094081 in the stomach. *CYP2E1* has previously been linked to AD⁵⁴ and is known to be activated in high concentrations of blood alcohol^{14,18}. *CYP4B1*, is also a member of the cytochrome p450 family of proteins (known to play a role in response and toxicity to many drugs) and functions as part of arachidonic or fatty acid metabolism⁵⁵.

It is well known that alleles in coding regions of *ADH* genes are associated with AD^{10,14,22}. Alcoholism risk is thought to involve elevation of acetaldehyde either through increased metabolism of ethanol, or slower conversion by *ALDH*¹⁵. Our work indicates that non-coding regions within the *ADH* locus play a role in coordinating increased regulation of *ADH* genes, associated with increased AD risk. Our work also shows that upregulation of *ADH* genes can occur through a set of variants, not just a single variant, causing a co-ordinated response across multiple genes and tissues. Our results suggest that 3D chromatin structure, epigenetic regulatory histone marks and altered transcription factor binding motifs could all be playing a role in regulation of *ADH* genes. We observed upregulatory interactions with associated histone marks consistent with enhancer regions, and down-regulatory interactions with associated predictions of decreased binding affinity of transcription factors due to motif alterations. Collectively these results are consistent with the existence of composite regulatory elements.

The regulatory interactions we have identified associated with SNPs in AD individuals remain putative until functionally confirmed including whether the alternative or reference alleles at the SNPs are the cause. Future GWAS studies that complete the identification of the AD associated SNPs and expanded eQTL data sets (which include analysis of immediate ancestry and gender) will enable greater refinement of the predictions we have made. Expanding Hi-C and protein expression sets will also strengthen our findings.

Some SNPs that are strongly associated by GWAS had no regulatory impacts in our assay. Our pipeline works on the assumption that the eQTLs rely on physical contact between the variant and gene it controls. However, there are other methods through which a genetic variant can impact on gene regulation that do not require this assumption be fulfilled. For example, variants in intronic regions may affect splicing activity⁵⁶. Our approach will not identify these alternate regulatory interactions. Moreover, studying other genetic variation, including the incorporation of other genetic variants (e.g. indels, CNVs), would also contribute to strengthening our findings.

We identified eQTLs for 34 not all 73 of the AD associated SNPs. The CoDeS3D pipeline⁴⁰ we used in this study only identifies eQTLs for a genetic variant if there are physical interactions between the variant and the gene for which it is an eQTL. However, eQTLs can impact on the expression of the gene through mechanisms that do not require a direct physical connection between the genetic variant and the gene. For example, some variants may affect splicing activity (so called splicing or sQTL variants⁵⁶), or trans-acting factors (e.g. long non-coding RNAs). Therefore, the lack of a finding in this study does not mean that the variant is not important, rather it is consistent with these variants impacting on AD through a different mechanism of action.

In conclusion we have identified the potential functional impacts of specific AD-associated variants on ADH regulation and show that sets of variants can produce co-ordinated effects across ADH genes and body tissues. Knowing the positions of loci that are associated with AD does not make them clinically useful. Rather, clinical utility requires an understanding of the functional or biological impact of the genetic change. Our results emphasise the insight to be gained from detailed study of eQTL, epigenetic histone mark and motif annotation in a functional and spatial context to advance the interpretation of GWAS SNPs. These results are consistent with the growing evidence that genetic variants exert phenotypic effects through tissue-specific gene enhancers and promoters^{38,57,58}. Future clinical use of this information could occur through a) development of novel tissue-specific therapeutic strategies that impact the biological and not behavioural aspects of AD; and b) stratification of patient cohorts in clinical trials. Furthermore, the integration of these data into clinical studies of AD will contribute to individualized mechanistic understandings of AD prognosis, disease development and progression.

Methods

GWAS SNPs. Single-nucleotide polymorphisms (SNPs) associated with alcohol dependence with p values $< 1 \times 10^{-6}$ were downloaded from NHGRI-EBI GWAS Catalog³⁹ (www.ebi.ac.uk/gwas/; 07/12/2018) (see Data and code availability). The wANNOVAR tool^{59,60} was used to perform functional annotation of SNPs. Genomic positions and annotations of SNPs were according to the human genome build hg19 release 75 (GRCh37) (see Data and code availability).

CoDeS3D pipeline. The CoDeS3D⁴⁰ pipeline was used to identify tissue-specific spatial interactions between regulatory regions (marked by SNPs associated with alcohol dependence) and their target genes. The CoDeS3D method we used here was previously outlined in³⁶ (Supplementary Fig. 1). Briefly, 28 Hi-C chromatin interaction libraries were used in this study (Supplementary Table 2). To identify DNA fragments, the hg19 reference genome was digested with the same restriction enzyme as used in Hi-C library preparation (i.e. MboI or HindIII). First, the location of SNPs within the restriction fragments was identified. Next, the algorithm identified the fragments that interact with the SNP-containing fragments in each of 28 Hi-C chromatin interaction libraries. These fragments were further overlapped with gene regions to identify only spatial SNP-gene pairs (where SNP-containing fragments spatially interact with gene-overlapping fragments). GENCODE transcript model version 19 was used as the reference for gene annotations. Then, we queried the GTEx database (<https://gtexportal.org/>), GTEx multi-tissue dataset v7, Supplementary Table 3) with the SNP-gene pairs, to identify cis- and trans-acting eQTL SNP interactions with eGenes (i.e. genes, whose tissue-specific expression changes are associated with eQTL SNP). Lastly, the Benjamini-Hochberg FDR control algorithm⁴⁰ was applied to adjust the p values of the identified eQTL associations and output only significant tissue-specific eQTL SNP-eGene interactions (FDR < 0.05).

LD analysis. Analysis of linkage disequilibrium (LD) was performed using LDlink 3.7⁶¹. The LDmatrix module was used to calculate LD statistics for AD-associated eQTL SNPs within *ADH* locus (GRCh37/hg19 genome assembly; SNP RS numbers based on dbSNP151; genotyping data from phase 3 (version 5) of the 1000 Genome Project; European population).

GO enrichment and pathway analyses. Gene enrichment and pathway analyses were performed using g:Profiler⁴⁴ (the g:GOST module). Three Gene Ontology terms (i.e. biological process, molecular function and cellular component) were used to identify functional categories that are statistically overrepresented in the spatially regulated AD-associated eGene set. Kyoto Encyclopedia of Genes and Genomes (KEGG) database⁶² was used to identify the most impacted biological pathways. All known human genes were chosen as a statistical domain scope. The significance of the overrepresented GO terms and pathways were corrected by the SCS algorithm⁴⁴ (adjusted $p < 0.05$).

Regulatory annotation. Promoter-enhancer histone marks and affected regulatory motifs were obtained from the HaploReg (version 4.1)^{49,50} website for the SNP-eQTLs identified as associating with the *ADH* gene region. The presence of H3K4ac, H3K27ac, H3K4me3 and H3K9me3 epigenetic histone marks and DNase marks were recorded for all cell lines. The cell lines were then filtered for those that closely matched tissue types from the GTEx database and histone marks were recorded for the corresponding tissues for eQTLs identified. This was done for eQTLs that had significant interactions with tissues in the cardiovascular, gastrointestinal and adipose tissues (See Supplementary Tables 7 & 8). Altered regulatory motifs and reference and alternative allele scores were obtained from HaploReg (version 4.1)^{49,50} based on constructed position weight matrices (PWMs)^{63,64}.

URLs. NHGRI-EBI GWAS Catalog: <https://www.ebi.ac.uk/gwas/>
gnomAD v2.1.1: <https://www.gnomad.broadinstitute.org/>
wANNOVAR: <http://wannovar.wglab.org/>
CoDeS3D pipeline: <https://github.com/Genome3d/codes3d-v1>
GTEx Portal: <https://gtexportal.org/home/>
LDlink 3.7.25: <https://ldlink.nci.nih.gov/>
g:Profiler (version e95_eg42_p13_f6e58b9): <https://biit.cs.ut.ee/gprofiler/>
HaploReg v4: https://pubs.broadinstitute.org/mammals/haploreg/haploreg_v4.php

Data and code availability

Human genome build hg19 release 75 (GRCh37) was downloaded from ftp://ftp.ensembl.org/pub/release-75/fasta/homo_sapiens/dna/Homo_sapiens.GRCh37.75.dna.primary_assembly.fa.gz.

SNP genomic positions (CoDeS3D SNP database) were obtained from ftp://ftp.ncbi.nih.gov/snp/organisms/human_9606_b151_GRCh37p13/.

Gene annotations were downloaded from https://storage.googleapis.com/gtex_analysis_v7/reference/genecode.v19.transcripts.patched_contigs.gtf.

All Python and R scripts used for data analysis and visualisation are available at <https://github.com/Genome3d/alcohol-dependence>.

Received: 24 January 2020; Accepted: 13 May 2020;

Published online: 18 June 2020

References

- Alcohol-use disorders: diagnosis, assessment and management of harmful drinking (high-risk drinking) and alcohol dependence. <https://www.nice.org.uk/guidance/cg115> (2011).
- Harmful use of alcohol. Available at https://www.who.int/health-topics/alcohol#tab=tab_1, (Accessed: 18th January 2020).
- Shield, K. D., Parry, C. & Rehm, J. Chronic diseases and conditions related to alcohol use. *Alcohol Res.* **35**, 155–173 (2013).
- Jeanblanc, J. Comorbidity Between Psychiatric Diseases and Alcohol Use Disorders: Impact of Adolescent Alcohol Consumption. *Current Addiction Reports* **2**, 293–301 (2015).
- Hammerton, G. *et al.* Effects of Excessive Alcohol Use on Antisocial Behavior Across Adolescence and Early Adulthood. *J. Am. Acad. Child Adolesc. Psychiatry* **56**, 857–865 (2017).
- Blankfield, A. Women, alcohol dependence and crime. *Drug Alcohol Depend.* **27**, 185–190 (1991).
- Verhulst, B., Neale, M. C. & Kendler, K. S. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychol. Med.* **45**, 1061–1072 (2015).
- Clarke, T.-K. *et al.* Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N = 112 117). *Mol. Psychiatry* **22**, 1376–1384 (2017).
- Bierut, L. J. *et al.* A genome-wide association study of alcohol dependence. *Proc. Natl. Acad. Sci. USA* **107**, 5082–5087 (2010).
- Park, B. L. *et al.* Extended genetic effects of ADH cluster genes on the risk of alcohol dependence: from GWAS to replication. *Hum. Genet.* **132**, 657–668 (2013).
- Edenberg, H. J. *et al.* Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcohol. Clin. Exp. Res.* **34**, 840–852 (2010).
- Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
- Gelernter, J. *et al.* Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol. Psychiatry* **19**, 41–49 (2014).
- Dasgupta, A. Genetic Aspects of Alcohol Metabolism and Drinking Behavior. *Alcohol and its Biomarkers*, 37–63, <https://doi.org/10.1016/b978-0-12-800339-8.00002-x> (2015).
- Edenberg, H. J. & Foroud, T. Genetics of alcoholism. *Handb. Clin. Neurol.* **125**, 561–571 (2014).
- Dubinkina, V. B. *et al.* Links of gut microbiota composition with alcohol dependence syndrome and alcoholic liver disease. *Microbiome* **5**, 141 (2017).
- GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
- Song, B.-J., Akbar, M., Jo, L., Hardwick, J. P. & Abdelmegeed, M. A. Translational Implications of the Alcohol-Metabolizing Enzymes, Including Cytochrome P450-2E1, in Alcoholic and Nonalcoholic Liver Disease. *Adv. Pharmacol.* **74**, 303–372 (2015).
- Kema, V. H., Mojerla, N. R., Khan, I. & Mandal, P. Effect of alcohol on adipose tissue: a review on ethanol mediated adipose tissue injury. *Adipocyte* **4**, 225–231 (2015).
- Edenberg, H. J. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res. Health* **30**, 5–13 (2007).
- Vaswani, M. Chapter 4 - ADH and ALDH Polymorphisms in Alcoholism and Alcohol Misuse/Dependence. In *Neuroscience of Alcohol* (ed. Preedy, V. R.) 29–38 (Academic Press, 2019).
- Ehlers, C. L., Liang, T. & Gizer, I. R. ADH and ALDH polymorphisms and alcohol dependence in Mexican and Native Americans. *Am. J. Drug Alcohol Abuse* **38**, 389–394 (2012).
- Li, D., Zhao, H. & Gelernter, J. Strong protective effect of the aldehyde dehydrogenase gene (ALDH2) 504lys (*2) allele against alcoholism and alcohol-induced medical diseases in Asians. *Hum. Genet.* **131**, 725–737 (2012).
- Luczak, S. E., Glatt, S. J. & Wall, T. L. Meta-analyses of ALDH2 and ADH1B with alcohol dependence in Asians. *Psychol. Bull.* **132**, 607–621 (2006).
- Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
- Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–110 (2015).
- Bryzgalov, L. O. *et al.* Detection of regulatory SNPs in human genome using ChIP-seq ENCODE data. *PLoS One* **8**, e78833 (2013).
- Schierding, W., Antony, J., Cutfield, W. S., Horsfield, J. A. & O'Sullivan, J. M. Intergenic GWAS SNPs are key components of the spatial and regulatory network for human growth. *Hum. Mol. Genet.* **25**, 3372–3382 (2016).
- Li, H. *et al.* Novel disease syndromes unveiled by integrative multiscale network analysis of diseases sharing molecular effectors and comorbidities. *BMC Med. Genomics* **11**, 112 (2018).
- Rearick, D. *et al.* Critical association of ncRNA with introns. *Nucleic Acids Res.* **39**, 2357–2366 (2011).
- Barton, N. H., Etheridge, A. M. & Véber, A. The infinitesimal model: Definition, derivation, and implications. *Theor. Popul. Biol.* **118**, 50–73 (2017).
- Fadason, T., Schierding, W., Lumley, T. & O'Sullivan, J. M. Chromatin interactions and expression quantitative trait loci reveal genetic drivers of multimorbidities. *Nat. Commun.* **9**, 5198 (2018).
- Fadason, T. *et al.* Reconstructing the blood metabolome and genotype using long-range chromatin interactions, <https://doi.org/10.1101/656132>.
- Mizuno, A. & Okada, Y. Biological characterization of expression quantitative trait loci (eQTLs) showing tissue-specific opposite directional effects. *Eur. J. Hum. Genet.* **27**, 1745–1756 (2019).
- Jung, I. *et al.* A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nature Genetics* **51**, 1442–1449 (2019).
- Golovina, E., Vickers, M. H., Erb, C. D. & O'Sullivan, J. M. Decoding the multimorbidities among psychiatric disorders and cognitive functioning. *bioRxiv* (2019).
- Fagny, M., Platig, J., Kuijjer, M. L., Lin, X. & Quackenbush, J. Nongenetic cancer-risk SNPs affect oncogenes, tumour-suppressor genes, and immune function. *Br. J. Cancer* **122**, 569–577 (2020).
- Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366**, 1134–1139 (2019).
- MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).

40. Fadason, T., Ekblad, C., Ingram, J. R., Schierding, W. S. & O'Sullivan, J. M. Physical Interactions and Expression Quantitative Traits Loci Identify Regulatory Connections for Obesity and Type 2 Diabetes Associated SNPs. *Front. Genet.* **8**, 1–12 (2017).
41. Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210, <https://doi.org/10.1101/531210> (2019).
42. Airaksinen, M. S. & Saarma, M. The GDNF family: signalling, biological functions and therapeutic value. *Nat. Rev. Neurosci.* **3**, 383–394 (2002).
43. He, D.-Y. & -Y. He, D. Glial Cell Line-Derived Neurotrophic Factor Mediates the Desirable Actions of the Anti-Addiction Drug Ibogaine against Alcohol Consumption. *Journal of Neuroscience* **25**, 619–628 (2005).
44. Reimand, J. *et al.* g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Research* **44**, W83–W89 (2016).
45. Steiner, J. L. & Lang, C. H. Alcohol, Adipose Tissue and Lipid Dysregulation. *Biomolecules* **7** (2017).
46. Kim, K. H. *et al.* Alcohol consumption and its relation to visceral and subcutaneous adipose tissues in healthy male Koreans. *Ann. Nutr. Metab.* **60**, 52–61 (2012).
47. Hannukainen, J. C. *et al.* Higher Free Fatty Acid Uptake in Visceral Than in Abdominal Subcutaneous Fat Tissue in Men. *Obesity* **18**, 261–265 (2010).
48. Pradeepa, M. M. Causal role of histone acetylations in enhancer function. *Transcription* **8**, 40–47 (2017).
49. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Research* **40**, D930–D934 (2012).
50. Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44**, D877–81 (2016).
51. Justice, A. C. *et al.* Validating Harmful Alcohol Use as a Phenotype for Genetic Discovery Using Phosphatidylethanol and a Polymorphism in ADH1B. *Alcoholism: Clinical and Experimental Research* **41**, 998–1003 (2017).
52. Jedlicka, P. & Gutierrez-Hartmann, A. Ets transcription factors in intestinal morphogenesis, homeostasis and disease. *Histol. Histopathol.* **23**, 1417–1424 (2008).
53. Eskelinen, E.-L., Tanaka, Y. & Saftig, P. At the acidic edge: emerging functions for lysosomal membrane proteins. *Trends in Cell Biology* **13**, 137–145 (2003).
54. Zeng, T. & Xie, K.-Q. Cytochrome P450E1 Gene Polymorphisms and the Risks of Ethanol-Induced Health Problems in Alcoholics. *Molecular Aspects of Alcohol and Nutrition*, 231–245, <https://doi.org/10.1016/b978-0-12-800773-0.00019-7> (2016).
55. Mittal, B., Tulsyan, S., Kumar, S., Mittal, R. D. & Agarwal, G. Chapter Four - Cytochrome P450 in Cancer Susceptibility and Treatment. In *Advances in Clinical Chemistry* (ed. Makowski, G. S.) vol. 71 77–139 (Elsevier, 2015).
56. Chiang, H.-L., Wu, J.-Y. & Chen, Y.-T. Identification of functional single nucleotide polymorphisms in the branchpoint site. *Hum. Genomics* **11**, 27 (2017).
57. Huang, D. & Ovcharenko, I. Identifying causal regulatory SNPs in ChIP-seq enhancers. *Nucleic Acids Res.* **43**, 225–236 (2015).
58. Corradin, O. & Scacheri, P. C. Enhancer variants: evaluating functions in common disease. *Genome Med.* **6**, 85 (2014).
59. Chang, X. & Wang, K. wANNOVAR: annotating genetic variants for personal genomes via the web. *J. Med. Genet.* **49**, 433–436 (2012).
60. Yang, H. & Wang, K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature Protocols* **10**, 1556–1566 (2015).
61. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).
62. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
63. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
64. Kheradpour, P. & Kellis, M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.* **42**, 2976–2987 (2014).

Acknowledgements

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Author contributions

J.O. and E.G. conceptualised the project. E.G. and R.H. performed data analysis and created figures. S.F. created figures and commented on the manuscript. E.G., R.H. and J.O. co-wrote the manuscript. J.O. lead the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-66048-z>.

Correspondence and requests for materials should be addressed to J.M.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020