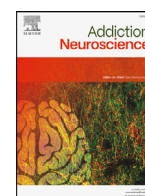


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Addiction Neuroscience

journal homepage: www.elsevier.com/locate/addicn

Context-dependant enhancers as a reservoir of functional polymorphisms and epigenetic markers linked to alcohol use disorders and comorbidities

Alasdair MacKenzie*, Elizabeth A. Hay, Andrew R. McEwan

School of Medicine, Medical Sciences and Nutrition, Institute of Medical Sciences, Foresterhill, University of Aberdeen, Aberdeen, AB25 2ZD, Scotland.



ARTICLE INFO

Keywords:

Complex disease
Alcohol use disorder
Gene regulation
Genome wide association studies
Chromatin modification
Comparative genomics
CRISPR genome editing
Promoter
Enhancer
Polymorphisms
Mental health

ABSTRACT

Alcohol use disorder (AUD) is one of the major causes of mortality and morbidity world-wide. It is estimated that 50% of the causes of AUD are heritable. Efforts to determine the genetic determinants governing AUD using genome wide association studies (GWAS) show that the most strongly associated SNPs occur within, or in the vicinity of, genes encoding enzymes that metabolise ethanol. However, these studies were not so conclusive in identifying the genes that influenced the choice to drink ethanol or why a proportion of the population become addicted. Most importantly, these studies also found that over 98% of the 1292 SNPs associated with AUD ($p < 1 \times 10^{-6}$) were found outside of coding regions and within the poorly understood non-coding genome. Many years of study have shown that functional components of the non-coding genome include enigmatic enhancer elements whose biological role is to modulate levels of gene expression in specific cells, in specific amounts and in response to the correct stimuli. The current short review introduces the functional components of the non-coding genome, such as promoters and enhancers, and critically assesses the latest methods of identifying and characterising their context dependant roles in AUD and mental health disorders. We then go on to examine what is known about the roles of enhancers, such as the GAL5.1 enhancer, in alcohol intake and explore how enhancers are affected by polymorphic variation and epigenetic markers such as DNA-methylation and may influence susceptibility to AUD. The review finishes by discussing the future of AUD genetics and what technologies will need to be brought to bear to understand how genetic and environmentally induced changes in enhancer structure may contribute to the need to drink alcohol to excess.

The problem at hand

Alcohol use disorders (AUD) continue to be a major problem in western countries like the UK partly because of its increased affordability, where alcohol is 74% more affordable in 2019 than it was in 1987 [1]. Although there are several reports that suggest the positive effects of moderate alcohol intake on cardiovascular health [2] the general effect of problematic alcohol use on society is overwhelmingly negative with 24% of adults in the UK drinking over the recommended 14 units a week, a level of alcohol intake that increases the risks of developing cancers, cardiovascular diseases and liver disease [3]. Worldwide, alcohol causes 5.3% of all deaths (>3 million people per year) and accounts for 132.6 million disability adjusted life years [3]. Thus, AUD presents a major health, societal and economic burden to countries worldwide. The challenge is to understand the factors which contribute to AUD so that preventative or therapeutic strategies can be designed and implemented.

Genetic studies of AUD

Based on adoption and twin studies, the genetic liability for alcohol abuse is estimated to be around 50% [4]. Genome wide association studies (GWAS, Fig. 1B) have indicated that the greatest risk loci for AUD centre on two genes, alcohol dehydrogenase (*ADH1B*) and aldehyde dehydrogenase (*ALDH2*), which encode enzymes that metabolise ethanol [5,6]. These studies suggest that a component in the development of AUD [7] may involve a change in the expression or function of these enzymes although the precise mechanisms of their involvement in the development of AUD remains to be established. However, the involvement of metabolising enzymes, that are primarily expressed in the liver, does not address why many people choose to drink alcohol and why alcohol use may develop into an addiction. It is widely understood that the decision to drink alcohol is modulated by regions of the brain, that include the hypothalamus and that addiction involves regions that include the nucleus accumbens [8,9]. Aside from the aforementioned metabolic genes, the majority of genetic risk seems to be spread amongst a large number of variants each with small effects, a known common feature of the genetics of complex diseases [10]. In addition to identifying the ethanol metabolising enzymes discussed above,

* Corresponding author.

E-mail address: mbi167@abdn.ac.uk (A. MacKenzie).

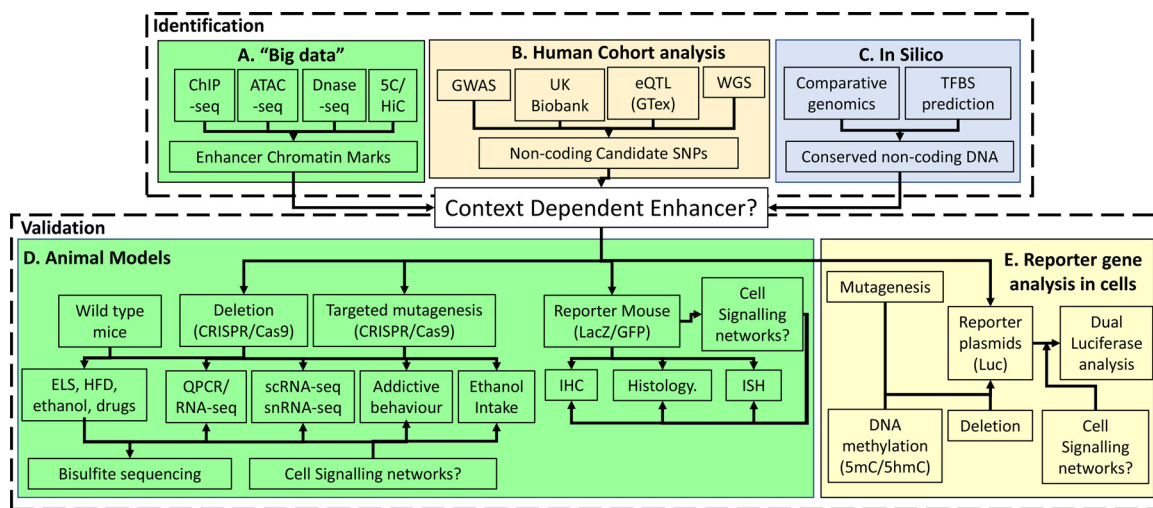


Fig. 1. Flow diagram demonstrating the relationship between different techniques designed to allow (A-C) identification of putative context dependent enhancers involved in AUD and (D and E) functional validation. ChIP-seq, chromatin immunoprecipitation sequencing; ATAC-seq, assay for transposase-accessible chromatin sequencing; 5C, carbon copy chromatin conformation capture; GWAS, genome wide association analysis; eQTL, expression quantitative trait loci; WGS, whole genome sequencing; TFBS, transcription factor binding site; LacZ/GFP, β galactosidase/green fluorescent protein, ELS, early life stress; HFD, high fat diet; QPCR, quantitative PCR; sc/snRNA-seq, single cell/single nucleus RNA-seq; IHC, immunohistochemistry; ISH, in situ hybridisation; 5mC, 5-methyl cytosine; 5hmC, 5-hydroxymethylcytosine.

two major GWAS studies identified other loci with an association to AUD and identified several genes including *KLB* (β -klotho), *GCKR* (Glucokinase regulatory protein), *CADM2* (Cell Adhesion molecule 2), *FAM69C* (Family with sequence similarity 69, member C), *STPG2* (Sperm Tail PG-Rich Repeat Containing 2) and *DNAJB14* (DnaJ Heat Shock Protein Family (Hsp40) Member B14) with a subsequent replicating study also identifying *JCAD* (Junctional Cadherin 5 Associated) *SLC39A13* (Solute Carrier Family Member 13) and *CRHR1* (Corticotropin Releasing Hormone Receptor 1) [5,6]. The most interesting of these genes is *CRHR1* whose involvement in AUD and the stress response, has been extensively explored [11]. However, the specific SNP identified in this study (rs62062288) falls within an intron of the *MAPT* gene (encodes Tau protein) and over 200 kb away from the transcriptional start site of the *CRHR1* gene, possibly reflecting the degrees of linkage disequilibrium (LD; where groups of alleles do not segregate randomly in a population) present within this region of the genome. Therefore, it is unclear how this SNP may functionally contribute to the presentation of AUD. Significantly, out of the 1292 SNPs that exceeded the $p < 1 \times 10^{-6}$ threshold required to achieve significance in GWAS studies only 25 of these SNPs fell within exonic regions [6]. These results are fairly typical of the data derived from current GWAS analyses of complex disease where the vast majority of SNPs which exceed the $p < 1 \times 10^{-6}$ threshold of significance are intronic or intergenic [10,12]. Based on these observations it is likely that the greatest burden of AUD causing SNPs do not lie within the coding regions of genes but within the unknown, and enigmatic, non-coding genome. Consequently, the main aim of the current review will not be to explore the known genetics or neuroscience of AUD in any depth, a subject which has been well reviewed in a number of other publications [8,9], but to briefly and critically appraise what we know about the information sources contained in the "non-coding" genome that are important in health, what techniques are currently used to understand the role of the non-coding genome in alcohol intake and what we need to do in the future to better understand its biology.

What information is needed to build a healthy human?

Other than the genetic information required to produce correct proteins, what other information sources are contained within the human genome that are important for normal human development health and behaviour including ethanol intake? If we consider that 10% of the hu-

man genome is under selective pressure (regions of the genome whose mutation rates are constrained due to their importance in survival) and only 1.7–1.9% of the genome encodes proteins, this suggests that at least 8% of the human genome, which does not encode proteins, is essential for health [13]. A high proportion of this genome is likely to be comprised of sequences that include promoters, enhancers, silencers and insulator regions that are critical to modulating expression levels of protein coding genes and non-coding functional RNA species, including microRNAs and long non-coding RNA, in specific cells and tissues [14,15]. A comprehensive review further describing promoter and enhancer biology, and the possible involvement of these sequences in a number of diseases, such as cancer and congenital malformations, has recently been published [14]. Intriguingly, gene regulatory sequences are lacking in unicellular organisms which suggests that these sequences may have evolved to support the development of multicellular organisms where their role is to coordinate the cell-cell interactions required for organ and body development [16]. Consequently, because we now know that most complex disease associated SNPs are found in the non-coding genome, a common feature of GWAS analyses of complex disease [10,12], it could be argued that the genetic causes of complex disease are less a function of "what genes encode" and more about ensuring that these genes are expressed within the correct cells, at the correct times in the correct amounts and in response to the correct signal, a property known as context-dependency.

Identifying regulatory elements

If we accept that the non-coding genome acts as the major reservoir of information required to build functional healthy humans (a concept not yet widely accepted), and that this information is compromised in most complex disease, how do we go about identifying the functional components of the non-coding genome? Promoter regions are the best known and understood regulatory regions in the human genome. They largely consist of sequences of DNA next to the transcriptional start sites of genes that are required to bind RNA polymerase II and other proteins that, together, comprise the transcriptional pre-initiation complex [17]. Their functions are distance and orientation dependant on the transcriptional start sites (TSS) that they control. However, although they are critically important in health and are affected by genetic and epigenetic changes, promoter sequences comprise a smaller proportion of the hu-

man genome than even coding regions. In addition, promoters, on their own, are unable to support the high levels of tissue specificity essential for the functional roles of many proteins [18].

Accordingly, other sequences are required to support the expression of genes critical to health in specific cells, in specific amounts and in response to specific cues. These sequences include enigmatic elements known as enhancers that interact with, and increase the activity of, promoter regions [15]. They are functionally distinct from promoters in that they are distance and orientation independent with respect to TSSs. Whilst our understanding of these sequences has increased enormously over the past 20 years, there is still confusion of how they can be reliably identified, what they are, their significance for health and disease and how their activity is affected by polymorphisms or environmental cues.

Attempts to identify enhancers

To date, the greatest effort to understand the non-coding genome can be represented by the 2012 ENCODE consortium release that consisted of a series of research papers which claimed to have mapped all the protein coding regions as well as all of the enhancer, silencer and insulator regions thereby representing an “Encyclopaedia of the Human Genome” [13]. These studies were based on the discovery that functional elements within the non-coding genome could be detected thanks to chromatin modification signatures and protein binding [19]. Therefore, a combination of techniques based around next generation sequencing (NGS) or “Big Data” approaches have been developed to understand the regulatory genome at a genome wide level. These include Dnase1 sensitivity-sequencing (DNase-seq) and “Assay for Transposase-Accessible Chromatin-sequencing” (ATAC-seq) that identify regions of the genome denuded of histones. Chromatin immunoprecipitation sequencing (ChIP-seq) is used to detect the interaction of different histone modifications and DNA binding proteins thought to be diagnostic of functional components. For active enhancers these include, histone 2, lysine 4 mono-methylation (H3K4me), histone 3 lysine 27 acetylation (H3K27ac) and Histone acetyltransferase p300 (p300). For active promoters these histone marks are Histone 3, lysine 4 tri-methylation (H3K4Me3) and for insulators (CCCTC-Binding factor (CTCF)). Chromatin conformation capture techniques such as 5C/Hi-seq allow for the detection of long-distance interactions within the genome which is now known to be organised into topological association domains (TAD), delimited by insulators, which have not only been conserved between tissues but also between species [21]. The influence of enhancer regions within specific regions of the genome is delimited by insulators [21]. In the ENCODE project these NGS based techniques were generally used to analyse the genomes of easily grown and accessible transformed human cell line monocultures [13,20]. Based on these analysis, it was concluded that >80% of the human genome was functional, a conclusion that was not without controversy [22]. These studies also suggested that regions displaying chromatin signatures characteristic of enhancers were not conserved, an observation subsequently supported by studies based on H3K27ac distributions in the disaggregated liver cells of several vertebrate species [23]. The inference from these studies was that most enhancer sequences evolve and de-evolved rapidly during evolution and that few enhancers are conserved during evolution.

However, earlier studies demonstrated evidence of functional enhancer heterogeneity; such that enhancers associated with distinct cellular functions may be partitioned based on specific combinations of multiple histone modifications [24]. In addition, a functional study of ENCODE predicted enhancer regions indicated that only 26% of these enhancers were active [25]. Moreover, deleting the genes encoding the Trithorax-related (Trr) proteins, responsible for adding H3K4me chromatin marks (characteristic of enhancers) from the *Drosophila* genome, did not significantly interrupt normal development [26]. Consequently, although chromatin marks will remain extremely useful in the future, there are concerns that extrapolating major conclusion about enhancer

activity, based on a limited number of chromatin marks, may not be that helpful in understanding enhancer biology [27]. Accordingly, it has been recommended that the definition of an enhancer detected using “enhancer specific” chromatin modifications should only be accepted if supported by functional data, preferably derived using *in-vivo* models [28].

Identifying enhancers through expression quantitative trait loci

Techniques such as GWAS have been essential in dissecting the genetic architecture of major diseases including AUD [5,6]. However, another method that has come to the fore in the identification of functional regulatory regions which influence health is the identification of expression quantitative trait loci (eQTLs, Fig. 1B). eQTLs are SNPs whose allelic variants are associated with differences in gene expression [29]. The Genotype–Tissue Expression (GTEx) study (<https://gtexportal.org/home/>) is the most extensive multi-tissue eQTL catalogue produced to date and is easily accessible via the internet. The GTEx consortium collected data from over 15,000 postmortem tissue specimens from 838 genotyped donors, representing 49 tissues [30]. Use of the GTEx database represents a useful way to identify functional regions within the human genome and a number of studies on primary human tissues have already made important inroads into identifying the mechanisms that contribute to the aetiology of schizophrenia [31].

Context dependant enhancers and evolutionary conservation

The evolutionary constraints placed on gene coding sequence is a function of the mostly inflexible 3-letter codon usage that determines the sequence and identity of the amino acids that make functional proteins. Thus, when added to the importance of proteins in health and species fitness, it is clear why protein coding sequences are highly conserved. Nevertheless, even though enhancer sequences do not encode proteins, is it possible that enhancer sequences can also be conserved? Can conservation be used as a method to identify context-specific enhancers? To answer these questions, we need to understand how enhancers work and in what form information is stored in their structure.

Put most simplistically, enhancers are comprised of many different transcription-factor binding sites (4–20 bp long) that are clustered together within a short section of DNA typically less than one kilobase in length. The binding site selection of many transcription factors is promiscuous such that an individual transcription factor can bind several different sequences, although with different levels of affinity [32]. However, a series of elegant experiments have shown that the process of enhancer evolution is not random and that the precise identity, order and spacing of the transcription factor binding sites that make up enhancers, known as enhancer “syntax”, is often functionally constrained and critical in defining tissue specificity [33–35]. Moreover, enhancer context-dependency is also reliant on levels of affinity of different TFs to their binding sites within enhancers whereby optimising the binding site of a given transcription factor produces ectopic activity of the enhancer in different tissues thereby reducing its specificity [33–35]. In consequence, although the binding promiscuity of TFs would suggest that enhancers evolve rapidly through evolution, the need to preserve syntax and to achieve the finely balance TF affinity needed to achieve tissue specificity would argue against this. Indeed, most studies of functionally proven enhancers argue strongly in favour of their high degrees of conservation through evolution. For example, high levels of sequence conservation are associated with functionally verified enhancers that coordinate expression of the interleukin genes [36], the Sonic hedgehog (Shh) enhancer [37] and Pierre-Robin-sequence (PRS)enhancer [38]. Another exemplary study of GWAS associated SNPs associated with neuroblastoma succeeded in identifying a SNP (rs2168101 G>T) within a highly conserved enhancer inside intron 1 of the LIM Domain Only 1 (LMO1) gene [39]. Additionally, high throughput analysis of highly conserved enhancer sequences analysed

using reporter genes in mouse embryos demonstrated that >70% of conserved non-coding regions had observable tissue specific enhancer function [40]. So, despite the widespread acceptance of chromatin markers as enhancer proxies, the case for using comparative genomics to detect functional enhancer sequences also remains strong (Fig. 1C). However, once a putative context-specific enhancer has been identified, what methods are available to allow us to deduce its function and how variables such as polymorphic variation, environmental changes and signal transduction events affect this function?

Methods of validating putative enhancer function

If identifying enhancers seems problematic, analysing their context-dependant functional activity is even more tricky and is the subject of much disagreement and debate. Enhancer activity is most often initially assayed using reporter assays where putative enhancer DNA is cloned into a reporter plasmid that also contains a promoter region driving the expression of an easily quantifiable protein product (Fig. 1D and E). Once cloned, the candidate enhancer can be cut up with enzymes or subjected to site directed mutagenesis to define their functional components (Fig. 1D and E). In the past, these reporters have included chloramphenicol acetyl-transferase (the basis of the CAT-assay), LacZ (that encodes the β galactosidase gene) and various forms of luciferase [41]. Luciferases such as firefly luciferase, are considered the most accurate as they can detect changes in gene expression over many orders of magnitude [42]. These reporter plasmids are then transfected into different cell lines and the quantities of reporter protein expressed are assayed biochemically. Reporter assays in cell lines can be carried out rapidly, at relatively little expense and can be easily scaled up [43]. However, one major disadvantage is that the cell lines often used may not provide the context appropriate for activating many enhancers. Thus, the current trend of producing “high throughput platforms” to functionally characterise enhancer regions on a whole genome level using cultured cell lines, although generating huge amounts of data, may have a limited ability to shed light on the role of context-dependant enhancers within the human brain.

The use of transgenic animals has helped to overcome many of the disadvantages of monoculture cell analysis. Reporter plasmids containing reporter genes such as LacZ or, more recently, different derivatives of the Green Fluorescent Protein (GFP), are used to make transgenic zebrafish [44] or microinjected into the pronuclei of single cell mouse embryos [45] (Fig. 1D). Although the use of these models is labour intensive and is less amenable to “scaling up”, they do provide a critical glimpse into the tissue specificity of enhancers whose activity may be undetectable in monoculture cell lines.

The benefits of CRISPR/Cas9

With the development of CAS9/CRISPR technology, it is now possible to rapidly delete enhancer regions from the mouse genome by microinjecting CAS9 mRNA or protein with single guide RNA (sgRNA) into the cytoplasm of 1-cell mouse embryos, a much easier process than pronuclear injection of reporter plasmids [46]. Although care must be taken with the possibility of generating “off-target” events, CAS9/CRISPR technology has largely superseded the previously widely used method of knocking out/in genes using embryonic stem cell targeting which was time consuming and expensive [47]. Briefly, the cytoplasm of single-cell mouse embryos are injected with either CAS9 protein (pre-incubated with sgRNA) or CAS9 mRNA and at least two sgRNA molecules, and allowed to develop to the 2-cell stage. These embryos are then oviduct transferred into a pseudopregnant host female mouse where they develop into pups.

Although more challenging, CAS9/CRISPR technology can also be used to introduce human allelic variants in the mouse genome allowing a functional comparison of the effects of allelic variants on behaviour and health *in-vivo* [48] (Fig. 1D). This approach relies on the cells own

homologous directed repair (HDR) mechanisms which are attracted to cut DNA. Hence, in the presence of a repair template, usually a 100 bp strand of DNA which is co-injected with CAS9-sgRNA complex, the cell will attempt to repair the CAS9 cut strand using a “Trojan Horse” repair template. The problem with this approach is that non-homologous end joining (NHEJ), which competes against HDR within the cell, is a much more active process in somatic cells with the result that only 10% of repairs within the cell are HDR directed [49]. Thus, although a very useful method of introducing allelic variants more development to encourage HDR over NHEJ repair pathways are required before CRISPR/CAS9 can persuasively compete against ES targeting in the short term [50].

Analysis of CRISPR/CAS9 enhancer knockouts

Whilst the main benefit of enhancer CRISPR knockouts is the ability to behaviourally test mice to assess the effects of deleting the enhancer on ethanol intake or co-morbidities such as anxiety, appetite and depression, (Fig. 1D) another benefit of these experiments is in determining the effects of these enhancer knockouts on the expression of down-stream genes which may also be involved in modulating ethanol intake. Thus, brain tissues can be recovered from these mice, subjected to RNA-seq analysis and compared to the expression of genes in wild type mice to determine which genes are modulated by the deleted enhancer. Although RNA-seq is able to determine the effects of deleting enhancers on the expression of the whole genome, the resolution of the technique at the cellular level is poor and relies on the fine dissection abilities of the operator. However, a recently developed technique called single cell RNA-seq (scRNA-seq) [51] allows the operator to define specific cell types based on the transcriptomes of individual cells. Briefly, disaggregated cells from dissected tissues are segregated either by fluorescent activated cell sorting (FACS) sorting or using a droplet-based-methods, into individual aqueous compartments in a lipid suspension. Within these compartments, cells are lysed and mRNA converted to cDNA prior to tagging with a barcode primer unique to each compartment. The cDNA is then recovered and combined for sequencing [51]. Sequences can then be desegregated by computers and the data is displayed as principal component analyses where cells are categorised based on their transcriptomes (Fig. 1D). In the case of neuronal tissues, which are comprised of heavily interdigitated cells, a refinement of this technology allows for the recovery of individual nuclei, which contain between 20 and 50% of the total cell mRNA, which can then be sorted and analysed in place of whole cells [52]. This technique can also be used to analyse frozen tissues where the integrity of the cells has been compromised by the freezing process.

Thus, in combination with CAS9/CRISPR technology, it is now possible to identify which genes are regulated by enhancers, at the cellular level, on a genome wide scale and to finally deduce the maximum distance over which enhancers can influence promoter activity.

Conserved enhancers that modulate alcohol intake

Based on previous observations that many enhancer elements have been highly conserved through evolution we tested the hypothesis that highly conserved sequences next to the coding sequences of genes known to control ethanol intake could represent context-dependant enhancers with a critical role in controlling ethanol intake [40]. We first explored the role of an enhancer sequence within the cannabinoid-1 receptor (CB1) gene (CNR1), that we called ECR1, which had been conserved since the last common ancestor of humans and fish (400 million years) [53]. This sequence was of interest as it also contained a SNP that had been associated with addictive behaviours [54] [55] and alcohol abuse [56,57]. Our initial analyses using reporter assays in primary cell lines suggested that the ECR1 sequence acted as an enhancer sequence whose presence influenced the activity of the promoter region of the CNR1 gene [53]. Initially, we were able to demonstrate that the

allelic variants of the ECR1 enhancer had a differential effect on the activity of the CNR1 promoter and responded differently to signal transduction agonists [58,59]. Based on these observations we undertook a functional analysis of the ECR1 enhancer in mice by deleting this enhancer from the mouse genome using CRISPR-Cas9 technology. Initial examination of these animals showed that, not only had the expression of the *Cnr1* gene been significantly reduced in parts of the brain that included the hippocampus, but that the hypothermia response to CB1 agonists had also been significantly reduced consistent with a reduction in *Cnr1* expression [58,60]. Subsequent analysis of these animals also demonstrated a significant reduction in ethanol intake and an altered anxiety phenotype in male and female ECR1KO mice [58,59]. To the best of our knowledge, these studies represent the first evidence that conserved enhancer regions play a role in influencing ethanol intake and anxiety.

We also explored the regulation of the *GAL* gene that encodes the galanin neuropeptide and has been associated with ethanol intake [61,62]. Genetic studies in humans had identified genotypes around the *GAL* locus that had associations with excess ethanol intake [63]. In a similar manner to ECR1, we used comparative genomics to identify a highly conserved polymorphic region of DNA that we called GAL5.1, that had also been conserved in birds and reptiles (350 million years) and lay 42 kb from the human *GAL* gene. We isolated this DNA sequence from human DNA and used it to produce transgenic reporter mice that expressed the β gal marker protein in cells of the hypothalamus and amygdala that also expressed galanin [64]. Subsequent luciferase analysis of human polymorphic variants of GAL5.1 in primary hypothalamic cell culture demonstrated a significant difference in the strength of this enhancer [64]. Based on these observations we examined the association of GAL5.1 polymorphic variants with ethanol abuse in the UK Biobank cohort ($n = 115,865$) and demonstrated a significant association between the GG genotype of this enhancer, ethanol intake and anxiety in men [65] (Fig. 1B). Intriguingly, CRISPR deletion of GAL5.1 (GAL5.1KO) almost completely ablated the expression of the *GAL* gene in all the GAL5.1KO mouse cell types analysed. Most importantly, deleting GAL5.1 produced mice that drank less ethanol whilst males suffered less anxiety mirroring our observations within the UK-Biobank [66]. Taken together, and in light of functional studies from other labs [36–40] these studies strongly suggest that enhancer regions critical to supporting tissue specific gene regulation can be highly conserved, most probably due to a need to conserve the syntax and the specific DNA binding specificities required to achieve robust levels of tissue-specificity. Critically, retrospective analysis of both ECR1 and GAL5.1 using the available ENCODE data base failed to identify chromatin marks diagnostic of active enhancers. The most likely reason for this observation was that neither ECR1 or GAL5.1 were active in the cell lines used by ENCODE. Consequently, our observations agree with previous conclusions that the enhancer status of a given sequence should only be accepted if supported by functional data, preferably derived using *in-vivo* models [28]

Signal transduction networks and enhancer polymorphisms

One of the major ambitions of medicine is to develop a personalised therapeutic approach to treating conditions such as AUD and anxiety which, in a large proportion of the population, resist current treatments [67]. The mechanisms controlling cell-cell interactions represent an important source of targets for the development of the personalised drug therapies of the future [68]. Receptor activation at the cell surface is followed by a cascading network of signal transduction interactions in the cytoplasm that terminate within the nucleus through the activation of DNA binding proteins [69]. Once activated through processes that include post-translational modifications, these proteins then assemble in a specific order on enhancer elements within the genome and recruit other factors that remodel the chromatin thereby controlling the transcriptional gene response [70]. For this cascade of events to unfold in an

appropriate manner and to generate an appropriate transcriptional response, the interaction between activated transcription factors and their target enhancers is critical [71]. Since the protein components of signal transduction mechanisms have been so strongly conserved through evolution, it is highly likely that the plasticity that generate differences in drug response resides at the level of enhancer variance [72]. It is therefore important to understand the effects of enhancer variance on enhancer response to signal transduction activation (Fig. 1D and E) [70].

To identify the cell signalling networks that control tissue specific activity of the GAL5.1 enhancer, we exposed primary hypothalamic cell cultures transfected with reporter plasmids under the control of either the GAL5.1 or ECR1 enhancers. In the case of GAL5.1, we demonstrated that neither protein kinase A or MAPkinases could significantly affect its activity in primary hypothalamic cells. However, when exposed to an agonist of protein kinase C (PKC) signalling, we observed a very significant increase in activity that was replicated in neuroblastoma cells [65]. Further dissection of the mechanisms governing the PKC demonstrated that co-transfection of cells with a plasmid expressing the EGR1 transcription factor further boosted the response of GAL5.1 to PKC agonists and Chromatin immunoprecipitation (ChIP) showed that EGR1 bound a single highly conserved consensus sequence within GAL5.1⁶⁶. Our most interesting observation came when we repeated these experiments with an allelic variant of GAL5.1. GAL5.1 hosts two common polymorphisms in perfect linkage disequilibrium (LD, rs2513280 (C/G) and rs2513281 (A/G)) to give a major allele (GG, 70–80%) and a minor allele (CA, 20–30%). Intriguingly, the CA allele demonstrated a significantly reduced response to PKC agonism and EGR1 expression [65]. Considering that the CA haplotype was also protective against anxiety and alcohol abuse in men, this difference suggests a direction for the development of anti-anxiety drugs, based on PKC antagonism, that may play a role in also reducing alcohol abuse in GG men.

Effects of environment on enhancer methylation and activity

The human genome is subject to a form of biochemical modification called DNA-methylation which is altered by environmental influence [73]. The best characterised form of methylation occurs on the 5th carbon of the cytosine ring (5mC) of the CpG dinucleotide, which has a significant influence on gene expression [73]. 5mC is initially deposited by the DNMT3A and DNMT3B proteins and is maintained by another protein called DNMT1 [73]. De-methylation can occur due to a failure of DNMT1 to continue to replicate methylation following cell division (passive de-methylation), or through active de-methylation that involved the stepwise degradation of 5mC by the Ten-eleven translocation proteins (TET1–3) to form 5-hydroxymethylation (5hmC) and a number of other forms (5fC and 5caC) [74]. Much is known about the effects of 5mC on the activity of promoter regions where high levels of methylation within the CpG island of many promoters is associated with reduced promoter activity due to binding of methyl-DNA binding proteins MBD1, MBD2, MBD4, and MeCP2 [75]. However, much less is known about the role of DNA methylation in enhancer elements [76]. Previous studies of the roles of 5mC in enhancer activity have explored the effects of environmental factors such as early life stress on 5mC levels within an enhancer region that controls the expression of the *AVP* gene that expresses the arginine vasopressin neuropeptide [77]. These elegant experiments showed that ELS induced hypomethylation of the *AVP* enhancer resulted in elevated levels of *AVP* expression in later life which could then be associated with increased depression like behaviours in ELS exposed animals [77]. These experiments drew a direct link between an environmental stimulus, changes in 5mC levels in an enhancer and changes in behaviour in later life. To determine a possible role for DNA-methylation in the activity of the GAL5.1 enhancer we exposed pregnant wild-type mice to standard low-fat diet or a choice of high-fat (60% calories from fat) and low-fat diet. We observed that levels of methylation of GAL5.1 were significantly elevated in male animals who were exposed to maternal high fat diet *in utero* [65]. In addition,

we demonstrated that methylation of GAL5.1 repressed the activity of the GAL5.1 enhancer even when stimulated by PKC agonism or transfection of active EGR1 transcription factor [65]. These studies support the hypothesis that one of the ways that environmental conditions affect health is through epigenetic modification of enhancers [76]. If we also consider that the GG haplotype of GAL5.1 contains a CpG site lacking in the CA haplotype we can also see how enhancers might serve as a nexus between genetics and environment.

Considering the known impact of DNA methylation on regulatory activity, gene expression and, hence phenotype, could the involvement of the environment in altering enhancer function through DNA methylation impact on the veracity of GWAS data? Thus, if human phenotypes can be changed through the methylation of enhancer elements, could this mask or exaggerate the presentation of disease phenotypes such as depression, anxiety or addiction?

The role of DNA-methylation in enhancer activity is further complicated by observations that a product of the active removal of 5mC, 5hmC, seems to have a different effect on enhancer activity than 5mC where 5hmC is associated with active enhancers [78-80]. These findings are further complicated by the fact that the usual method used of analysing CpG methylation in the genome is through bisulfite sequencing which is unable to differentiate between 5mC and 5hmC [81]. Our studies also show that there is considerable variation in levels of enhancer methylation between different tissues such that 5mC/5hmC levels in amygdala and hypothalamus are twice what they are in the hippocampus raising questions of the relevance of peripheral blood based DNA methylation data in understanding DNA-methylation levels in the central nervous system [65]. Clearly, a great deal more research is required to understand the interaction of DNA methylation on enhancer activity before we can truly understand the roles and influences of enhancers on health and disease.

Enhancers as the regulatory basis of co-morbidities

The relationship between enhancer and genes is not straight forward such that one enhancer may influence the expression of many genes and the influence of one gene may be under that influence of many enhancers [82]. Indeed, changes in the relationships between the more plastic regulatory genome, such as enhancer co-option and loss, and the relatively fixed coding genome in vertebrates is likely to have been the major driving force in human evolution [83]. Moreover, it is well established that the maintenance of synteny blocks (regions where the same genes are clustered in the genomes of diverse vertebrates) reflects the need for many enhancers and genes to interdigitate [83,84]. A further complication is that a single enhancer may drive the expression of individual genes in many different cell types. For example, the GAL5.1 enhancer is active in the periventricular nucleus, medial nucleus and arcuate nucleus of the hypothalamus and in the medial amygdala [66]. This may explain the fact that deletion of GAL5.1, not only reduced ethanol intake in both sexes, but also decreased fat intake in both sexes and anxiety in male animals [65,66]. Accordingly, given the ability of individual enhancers to drive expression into many tissues, and to affect many behaviours, we should not be surprised that any polymorphism or DNA methylations which affects an individual enhancer may result in co-morbidities such as obesity, alcohol abuse and chronic anxiety. Again, more analysis of enhancers, their relationships to gene expression and the effects of polymorphic and DNA-methylation on their activities, is essential to understand the basis of human disease susceptibilities.

Conclusions

Following the sequencing of the human genome, numerous promises of huge advances in our understanding of health and disease, and the subsequent production of new therapeutic technologies, raised hopes for the understanding and treatment of chronic disorders including alcohol abuse within 10 years. However, 22 years have elapsed and the

promised benefits, that largely justified the sequencing of the human genome, have yet to materialise.

Yet, it is striking to see how far our understanding of the human genome has progressed. We now know that, disappointingly for many, the majority of what we need to know about the basis of human health and disease is hidden in a portion of the genome previously dismissed as “Junk DNA”. In other words, our analysis of the human genome, to date, has given us a better perspective on what we need to know. Although, attempts to understand the non-coding genome, as typified by ENCODE, although dismissed by some as misguided and a lesson in showboating [22], succeeded in producing a great deal of very useful data that will continue to be analysed for decades to come.

However, it is clear that our biggest challenge is to design strategies that take account of the context-dependency displayed by many enhancers. Unfortunately, there is unlikely to be a high tech “quick fix” in this regard; where one or two markers of enhancer function will identify and characterise all context-dependant human enhancers using cell lines alone. Instead, we are in for a “long haul” where identifying functional enhancers, characterising the mechanisms regulating their context-dependency, and how they can be affected by genetic and environmental changes, necessitates the continued use of genetic manipulation of whole animal models such as zebrafish and mouse. But, thanks to our ability to rapidly engineer the genomes of vertebrate models such as mice and the rapid development of single cell sequencing technologies [85], we are in a much better position to develop a greater understanding of the role of context-dependant enhancers in normal development and health than we were even ten years ago.

Only by understanding the mechanisms that modulate the context-dependency of gene expression, and in determine the effects of polymorphisms and environment on these mechanisms, will we succeed in understanding the molecular basis of mental health issues such as AUD, and its co-morbidities, and to devise therapies to treat it.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

AMcE was funded by BBSRC project grant (BB/N017544/1) and EH was funded by Medical Research Scotland (PhD-719-2013).

References

- [1] N. Digital, *Statistics on Alcohol, England* (2020).
- [2] A. Marcos, et al., *Moderate Consumption of Beer and Its Effects on Cardiovascular and Metabolic Health: an Updated Review of Recent Scientific Evidence*, *Nutrients* 13 (2021).
- [3] W.H. Organisation, *Global Status Report On Alcohol and Health*, World Health Organisation, 2018.
- [4] B. Verhulst, M.C. Neale, K.S. Kendler, *The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies*, *Psychol. Med.* 45 (2015) 1061–1072.
- [5] T.K. Clarke, 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 (2017) 1376–1384.
- [6] S. Sanchez-Roige, et al., *Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts*, *Am. J. Psychiatry* 176 (2019) 107–118.
- [7] H.J. Edenberg, J.N. McClintick, *Alcohol Dehydrogenases, Aldehyde Dehydrogenases, and Alcohol Use Disorders: a Critical Review*, *Alcohol Clin. Exp. Res.* 42 (2018) 2281–2297.
- [8] C.R. Coker, B.N. Keller, A.C. Arnold, Y. Silberman, *Impact of High Fat Diet and Ethanol Consumption on Neurocircuitry Regulating Emotional Processing and Metabolic Function*, *Front. Behav. Neurosci.* 14 (2020) 601111.
- [9] F.M. Wurst, et al., *Alcoholism, craving, and hormones: the role of leptin, ghrelin, prolactin, and the pro-opiomelanocortin system in modulating ethanol intake*, *Alcohol Clin. Exp. Res.* 31 (2007) 1963–1967.
- [10] E.A. Boyle, Y.I. Li, J.K. Pritchard, *An Expanded View of Complex Traits: from Polygenic to Omnigenic*, *Cell* 169 (2017) 1177–1186.

- [11] D.E. Curley, A.E. Webb, D.J. Sheffler, C.L. Haass-Koffler, Corticotropin releasing factor binding protein as a novel target to restore brain homeostasis: lessons learned from alcohol use disorder research, *Front. Behav. Neurosci.* 15 (2021) 786855.
- [12] L.A. Hindorff, et al., Potential etiologic and functional implications of genome-wide association loci for human diseases and traits, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 9362–9367.
- [13] I. Dunham, et al., An integrated encyclopedia of DNA elements in the human genome, *Nature* 489 (2012) 57–74.
- [14] A. Claringbould, J.B. Zaugg, Enhancers in disease: molecular basis and emerging treatment strategies, *Trends Mol. Med.* 27 (2021) 1060–1073.
- [15] C. Benoist, P. Chambon, In vivo sequence requirements of the SV40 early promoter region, *Nature* 290 (1981) 304–310.
- [16] A. Sebe-Pedros, et al., The Dynamic Regulatory Genome of *Capsaspora* and the Origin of Animal Multicellularity, *Cell* 165 (2016) 1224–1237.
- [17] S. Davidson, N. Macpherson, J.A. Mitchell, Nuclear organization of RNA polymerase II transcription, *Biochem. Cell. Biol.* 91 (2013) 22–30.
- [18] I. Chepelev, G. Wei, D. Wangsa, Q. Tang, K. Zhao, Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization, *Cell Res.* 22 (2012) 490–503.
- [19] M.P. Creighton, et al., Histone H3K27ac separates active from poised enhancers and predicts developmental state, *Proc. Natl. Acad. Sci. U. S. A.* 107 (2020) 21931–21936.
- [20] P.H. Holmqvist, M. Mannervik, Genomic occupancy of the transcriptional co-activators p300 and CBP, *Transcription* 4 (2013) 18–23.
- [21] R.D. Acemil, I. Maeso, J.L. Gomez-Skarmeta, Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals, *Wiley Interdiscip. Rev. Dev. Biol.* 6 (2017).
- [22] D. Graur, et al., On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE, *Genome Biol. Evol.* 5 (2013) 578–590.
- [23] D. Villar, et al., Enhancer evolution across 20 mammalian species, *Cell* 160 (2015) 554–566.
- [24] G.E. Zentner, P.J. Tesar, P.C. Scacheri, Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions, *Genome Res.* 21 (2011) 1273–1283.
- [25] J.C. Kwasniewski, C. Fiore, H.G. Chaudhari, B.A. Cohen, High-throughput functional testing of ENCODE segmentation predictions, *Genome Res.* 24 (2014) 1595–1602.
- [26] R. Rickels, et al., Histone H3K4 monomethylation catalyzed by Trx and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability, *Nat. Genet.* 49 (2017) 1647–1653.
- [27] M.S. Halfon, Studying Transcriptional Enhancers: the Founder Fallacy, Validation Creep, and Other Biases, *Trends Genet.* 35 (2019) 93–103.
- [28] T.J. Cunningham, J.J. Lancman, M. Berenguer, P.D.S. Dong, G. Duyster, Genomic Knockout of Two Presumed Forelimb Tbx5 Enhancers Reveals They Are Nonessential for Limb Development, *Cell Rep.* 23 (2018) 3146–3151.
- [29] Y. Gilad, S.A. Rifkin, J.K. Pritchard, Revealing the architecture of gene regulation: the promise of eQTL studies, *Trends Genet.* 24 (2008) 408–415.
- [30] B.D. Umans, A. Battle, Y. Gilad, Where Are the Disease-Associated eQTLs? *Trends Genet.* 37 (2021) 109–124.
- [31] L.S. Hall, et al., Cis-effects on gene expression in the human prenatal brain associated with genetic risk for neuropsychiatric disorders, *Mol. Psychiatry* 26 (2021) 2082–2088.
- [32] S. Brodsky, et al., Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity, *Mol. Cell* 79 (2020) 459–471 e4.
- [33] E.K. Farley, K.M. Olson, W. Zhang, D.S. Rokhsar, M.S. Levine, Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 6508–6513.
- [34] E.K. Farley, et al., Suboptimization of developmental enhancers, *Science* 350 (2015) 325–328.
- [35] E.K. Farley, K.M. Olson, M.S. Levine, Regulatory Principles Governing Tissue Specificity of Developmental Enhancers, *Cold Spring Harb. Symp. Quant. Biol.* 80 (2015) 27–32.
- [36] G.G. Loots, et al., Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons, *Science* 288 (2000) 136–140.
- [37] L.A. Lettice, P. Devenney, C. De Angelis, R.E. Hill, The Conserved Sonic Hedgehog Limb Enhancer Consists of Discrete Functional Elements that Regulate Precise Spatial Expression, *Cell Rep.* 20 (2017) 1396–1408.
- [38] H.K. Long, et al., Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder, *Cell Stem Cell* (2020).
- [39] D.A. Oldridge, et al., Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism, *Nature* 528 (2015) 418–421.
- [40] A. Visel, J. Bristow, L.A. Pennacchio, Enhancer identification through comparative genomics, *Semin. Cell Dev. Biol.* 18 (2007) 140–152.
- [41] J. Mihalý, et al., Chromatin domain boundaries in the Bithorax complex, *Cell. Mol. Life Sci.* 54 (1998) 60–70.
- [42] J. Alam, J.L. Cook, Reporter genes: application to the study of mammalian gene transcription, *Anal. Biochem.* 188 (1990) 245–254.
- [43] F. Inoue, N. Ahituv, Decoding enhancers using massively parallel reporter assays, *Genomics* 106 (2015) 159–164.
- [44] H. Baier, E.K. Scott, Genetic and optical targeting of neural circuits and behavior—zebrafish in the spotlight, *Curr. Opin. Neurobiol.* 19 (2009) 553–560.
- [45] L. Montoliu, J.A. Blendy, T.J. Cole, G. Schutz, Analysis of the cAMP response on liver-specific gene expression in transgenic mice, *Fundam. Clin. Pharmacol.* 8 (1994) 138–146.
- [46] E.A. Hay, et al., An analysis of possible off target effects following CAS9/CRISPR targeted deletions of neuropeptide gene enhancers from the mouse genome, *Neuropeptides* 64 (2017) 101–107.
- [47] P. Singh, J.C. Schimenti, E. Bolcun-Filas, A mouse geneticist's practical guide to CRISPR applications, *Genetics* 199 (2015) 1–15.
- [48] A. Baud, J. Flint, Identifying genes for neurobehavioural traits in rodents: progress and pitfalls, *Dis. Model. Mech.* 10 (2017) 373–383.
- [49] G.J. Scott, A. Gruzdev, Genome Editing in Mouse Embryos with CRISPR/Cas9, *Methods Mol. Biol.* (2019) 23–40 1960.
- [50] C.D. Yeh, C.D. Richardson, J.E. Corn, Advances in genome editing through control of DNA repair pathways, *Nat. Cell Biol.* 21 (2019) 1468–1478.
- [51] E.J. Armand, J. Li, F. Xie, C. Luo, E.A. Mukamel, Single-Cell Sequencing of Brain Cell Transcriptomes and Epigenomes, *Neuron* 109 (2021) 11–26.
- [52] T.E. Bakken, et al., Single-nucleus and single-cell transcriptomes compared in matched cortical cell types, *PLoS One* 13 (2018) e0209648.
- [53] G. Nicoll, et al., Allele-specific differences in activity of a novel cannabinoid receptor 1 (CNR1) gene intronic enhancer in hypothalamus, dorsal root ganglia, and hippocampus, *J. Biol. Chem.* 287 (2012) 12828–12834.
- [54] A. Ketcherside, L.J. Noble, C.K. McIntyre, F.M. Filbey, Cannabinoid Receptor 1 Gene by Cannabis Use Interaction on CB1 Receptor Density, *Cannabis Cannabinoid Res.* 2 (2017) 202–209.
- [55] X. Chen, et al., Cannabinoid receptor 1 gene association with nicotine dependence, *Arch. Gen. Psychiatry* 65 (2008) 816–824.
- [56] M.J. Pava, J.J. Woodward, A review of the interactions between alcohol and the endocannabinoid system: implications for alcohol dependence and future directions for research, *Alcohol* 46 (2012) 185–204.
- [57] K.E. Hutchison, et al., The incentive salience of alcohol: translating the effects of genetic variant in CNR1, *Arch. Gen. Psychiatry* 65 (2008) 841–850.
- [58] E.A. Hay, et al., Disease-associated polymorphisms within the conserved ECR1 enhancer differentially regulate the tissue-specific activity of the cannabinoid-1 receptor gene promoter; implications for cannabinoid pharmacogenetics, *Hum. Mutat.* 41 (2020) 291–298.
- [59] E.A. Hay, et al., Genetic and epigenetic influences modulating the tissue specific regulation of the cannabinoid receptor -1 gene (CB1); implications for cannabinoid pharmacogenetics, *bioRxiv* (2019).
- [60] E.A. Hay, et al., Disruption of an enhancer associated with addictive behaviour within the cannabinoid receptor-1 gene suggests a possible role in alcohol intake, cannabinoid response and anxiety-related behaviour, *Psychoneuroendocrinology* 109 (2019) 104407.
- [61] P. Rada, N.M. Avena, S.F. Leibowitz, B.G. Hoebel, Ethanol intake is increased by injection of galanin in the paraventricular nucleus and reduced by a galanin antagonist, *Alcohol* 33 (2004) 91–97.
- [62] J.R. Barson, I. Morganstern, S.F. Leibowitz, Galanin and consummatory behavior: special relationship with dietary fat, alcohol and circulating lipids, *EXS* 102 (2011) 87–111.
- [63] I. Belfer, et al., Association of galanin haplotypes with alcoholism and anxiety in two ethnically distinct populations, *Mol. Psychiatry* 11 (2006) 301–311.
- [64] S. Davidson, et al., Differential activity by polymorphic variants of a remote enhancer that supports galanin expression in the hypothalamus and amygdala: implications for obesity, depression and alcoholism, *Neuropsychopharmacology* 36 (2011) 2211–2221.
- [65] A. McEwan, et al., The anxiety and ethanol intake controlling GAL5.1 enhancer is epigenetically modulated by, and controls preference for, high-fat diet, *Cell. Mol. Life Sci.* 78 (2021) 3045–3055.
- [66] A.R. McEwan, et al., CRISPR disruption and UK Biobank analysis of a highly conserved polymorphic enhancer suggests a role in male anxiety and ethanol intake, *Mol. Psychiatry* 26 (2021) 2263–2276.
- [67] R.G. Dos Santos, J.C. Bousso, M.A. Alcazar-Corcoles, J.E.C. Hallak, Efficacy, tolerability, and safety of serotonergic psychedelics for the management of mood, anxiety, and substance-use disorders: a systematic review of systematic reviews, *Expert Rev. Clin. Pharmacol.* 11 (2018) 889–902.
- [68] D. Ron, A. Berger, Targeting the intracellular signaling "STOP" and "GO" pathways for the treatment of alcohol use disorders, *Psychopharmacology (Berl.)* 235 (2018) 1727–1743.
- [69] E. McQueen, M. Rebeiz, On the specificity of gene regulatory networks: how does network co-option affect subsequent evolution? *Curr. Top. Dev. Biol.* 139 (2020) 375–405.
- [70] A. MacKenzie, B. Hing, S. Davidson, Exploring the effects of polymorphisms on cis-regulatory signal transduction response, *Trends Mol. Med.* 19 (2012) 99–107.
- [71] C.M. Vockley, A. Barrera, T.E. Reddy, Decoding the role of regulatory element polymorphisms in complex disease, *Curr. Opin. Genet. Dev.* 43 (2017) 38–45.
- [72] R. Nunez-Torres, et al., Association Between ABCB1 genetic variants and persistent chemotherapy-induced alopecia in women with breast cancer, *JAMA Dermatol.* 156 (2020) 987–991.
- [73] T.H. Bestor, The DNA methyltransferases of mammals, *Hum. Mol. Genet.* 9 (2000) 2395–2402.
- [74] R.M. Kohli, Y. Zhang, TET enzymes, TDG and the dynamics of DNA demethylation, *Nature* 502 (2013) 472–479.
- [75] T. Baubec, R. Ivanek, F. Lienert, D. Schubeler, Methylation-dependent and -independent genomic targeting principles of the MBD protein family, *Cell* 153 (2013) 480–492.
- [76] A. Angeloni, O. Bogdanovic, Enhancer DNA methylation: implications for gene regulation, *Essays Biochem.* 63 (2019) 707–715.

- [77] C. Murgatroyd, et al., Dynamic DNA methylation programs persistent adverse effects of early-life stress, *Nat. Neurosci.* 12 (2009) 1559–1566.
- [78] H. Stroud, S. Feng, S. Morey Kinney, S. Pradhan, S.E. Jacobsen, 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells, *Genome Biol.* 12 (2011) R54.
- [79] G.C. Hon, et al., 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation, *Mol. Cell* 56 (2014) 286–297.
- [80] H. Hashimoto, et al., Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation, *Nucleic. Acids. Res.* 40 (2012) 4841–4849.
- [81] K. Skvortsova, C. Stirzaker, P. Taberlay, The DNA methylation landscape in cancer, *Essays Biochem.* 63 (2019) 797–811.
- [82] C.H. Spurrell, D.E. Dickel, A. Visel, The Ties That Bind: mapping the Dynamic Enhancer-Promoter Interactome, *Cell* 167 (2016) 1163–1166.
- [83] A. Mackenzie, K.A. Miller, J.M. Collinson, Is there a functional link between gene interdigitation and multi-species conservation of synteny blocks? *Bioessays* 26 (2004) 1217–1224.
- [84] H. Kikuta, D. Fredman, S. Rinkwitz, B. Lenhard, T.S. Becker, Retroviral enhancer detection insertions in zebrafish combined with comparative genomics reveal genomic regulatory blocks - a fundamental feature of vertebrate genomes, *Genome Biol.* 8 (2007) S4 Suppl 1.
- [85] A. Hoffmann, D. Spengler, Single-Cell Transcriptomics Supports a Role of CHD8 in Autism, *Int. J. Mol. Sci.* 22 (2021).