TRANSCRIPTION REGULATION OF THE CLASS IV ALCOHOL DEHYDROGENASE 7 (*ADH7)*

Sowmya Jairam

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Biochemistry and Molecular Biology,
Indiana University

May 2014

Accepted by the Graduate Faculty, of Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Howard J. Edenberg, Ph.D., Chair

Brian P. Herring, Ph.D.

Doctoral Committee

David G. Skalnik, Ph.D.

January 21, 2014

Ronald C. Wek, Ph.D.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Howard Edenberg, for his guidance and support throughout my time in his lab. He fostered the ability to think independently by asking questions and letting me find the answers, even those I didn't have readily, while making sure I was on the right track. I know the lessons I have learnt and the training I have received will stand me in good stead in my career. I would also like to thank all the current and former lab members at both the MS building and BRTC, particularly Jun Wang, Dr. Jeanette McClintick, Ron Jerome and Dr. Sirisha Pochareddy for their mentoring and friendship. They have been a great help and an immense support whenever I needed to get a second opinion on a technique or a research idea, and I have enjoyed discussing both scientific and non-scientific topics with them.

I would like to thank my committee members Dr. Paul Herring, Dr. David Skalnik and Dr. Ron Wek for their time, effort and guidance, and for helping me focus on the right questions. They have not only provided suggestions that greatly helped with my projects, but also freely shared their resources for which I am grateful. I thank Dr. Paul Herring for not only letting me use his resources (CTCF and control shRNA) but also guiding me through the knockdown experiments.

I would like to thank my friends, Lakshmi Reddy and Souvik, who have been with me since the start of my PhD and with whom I have a treasure trove of great memories. Thanks as well to my other friends —Latha, Jyothi and Abhirami, fellow PhDs with

whom I've spent fun times discussing ups and downs with the program and who helped me when I needed the most. Thank you all for your friendship and company- it definitely brightened up my days. I thank my best friends Priyanka and Roopa, who have no connection with my PhD and with whom I can discuss any and all things freely.

Finally, I want to thank my beloved family whose love and support has been a very large part of my life, and a great motivation even at the bleakest times. Thank you Amma and Appa for trusting me with my choices, for instilling confidence in me and for encouraging me and loving me always. Thank you Rami, my dearest sister for just being your sweet self, for being non-judgemental and always, always loving. Thank you Uday for your love, support and your patience- I am excited and looking forward to our life together. Thank you Santhosh for your unique brand of humor that has never yet failed in bringing me out of one of my ruts. Thank you Paati and all the rest of my family who have always kept me in their prayers and their thoughts, for their love and never-ending support.

Sowmya Jairam

TRANSCRIPTION REGULATION OF THE CLASS IV ALCOHOL DEHYDROGENASE 7 (*ADH7)*

The class IV alcohol dehydrogenase (ADH7, μ-ADH, σ-ADH) efficiently metabolizes ethanol and retinol. *ADH7* is expressed mainly in the upper gastrointestinal tract with no expression in the liver unlike the other *ADHs,* and is implicated in various diseases including alcoholism, cancer and fetal alcohol syndrome. Genome wide studies have identified significant associations between *ADH7* variants and alcoholism and cancer, but the causative variants have not been identified. Due to its association with two important metabolic pathways and various diseases, this dissertation is focused on studying *ADH7* regulation and the effects of variants on this regulation using cell systems that replicate endogenous *ADH7* expression. We identified elements regulating *ADH7* transcription and observed differences in the effects of variants on gene expression. A7P-G and A7P-A, two promoter haplotypes differing in a single nucleotide at rs2851028, had different transcriptional activities and interacted with variants further upstream. A sequence located 12.5 kb upstream (7P10) can function as an enhancer. These complex interactions indicate that the effects of variants in the *ADH7* regulatory elements depend on both sequence and cellular context, and should be considered in interpretation of the association of variants with alcoholism and cancer.

The mechanisms governing the tissue-specific expression of *ADH7* remain unexplained however. We identified an intergenic region (iA1C), located between *ADH7* and *ADH1C,*

having enhancer blocking activity in liver-derived HepG2 cells. This enhancer blocking function was cell- and position- dependent with no activity seen in CP-A esophageal cells. iA1C had a similar effect on the ectopic SV40 enhancer. The CCCTC-binding factor (CTCF) bound iA1C in HepG2 cells but not in CP-A cells. Our results suggest that in liver-derived cells, iA1C blocks the effects of downstream *ADH* enhancers and thereby contributes to the cell specificity of *ADH7* expression. Thus, while genetic factors determine level of *ADH7* transcriptional activity, iA1C helps determine the cell specificity of transcription.

Howard J. Edenberg, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ADH	Alcohol dehydrogenase
AF	Activation function
ALDH	Aldehyde dehydrogenase
AP-1	Activator protein-1
APL	Acute promyelocytic leukemia
AUD	alcohol use disorder
bp	base pair
C/EBP	CCAAT/enhancer-binding protein
cDNA	complementary DNA
CDS	Coding sequence
ChIP	Chromatin immunoprecipitation
Crispr	Clustered regularly interspaced short palindromic repeats
cSNP	coding SNP
CTCF	CCCTC binding factor
DBD	DNA binding domain
dbSNP	database of short genetic variations, (http://www.ncbi.nlm.nih.gov/SNP/)
DNase I	Deoxyribonuclease I
DR	Direct repeat
DSM	Diagnostic and Statistical Manual of Mental Disorders
DTT	Dithiothreitol
EMSA	Electrophoretic mobility shift assay
ENCODE	Encyclopedia of DNA elements

ENH	Class I *ADH* enhancer
FAM	6-carboxyfluorescein
FASD	Fetal Alcohol Spectrum Disorders
FOX	Forkhead box
GI	gastrointestinal
GR	glucocorticoid receptor
GWAS	Genome wide association studies
HNF1	Hepatocyte nuclear factor 1
hTERT	Human telomerase reverse transcriptase
Kb	Kilobase
LBD	Ligand binding domain
LD	Linkage Disequilibrium
MEF	Mouse embryonic fibroblasts
MEM	minimum essential medium
Min	minutes
mRNA	messenger RNA
NAD	Nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information, NIH
NFκB	Nuclear factor kappa B
PCR	Polymerase chain reaction
qPCR	quantitative polymerase chain reaction
RA	Retinoic acid
RAR	Retinoic acid receptor

RARE	Retinoic acid response element
RPL30	Ribosomal protein L30
RT	Room temperature
RXR	Retinoid X receptor
shCTCF	short hairpin CTCF RNA
shctrl	short hairpin control RNA
shRNA	short hairpin RNA
SNP	Single nucleotide polymorphism
TSS	Translation start site
UADT	Upper aero-digestive tract

I. INTRODUCTION

1. Alcohol dehydrogenases

The medium chain dehydrogenases/reductases (MDR) constitute a large superfamily of close to 1000 enzymes that are involved in the oxidation of various alcohols, detoxification of alcohols/aldehydes and metabolism of bile acids (Persson et al., 1994). The alcohol dehydrogenases (ADHs) are the best studied among the MDR superfamily. Alcohol dehydrogenases (ADHs) are zinc-containing metalloenzymes that catalyze the reversible oxidation of various primary and secondary alcohols to their aldehydes (Hurley and Edenberg, 2012, Edenberg, 2012, Ehrig et al., 1990, Bosron and Li, 1986a). They are cytosolic enzymes that form homo- and heterodimers, each subunit approximately 40 kDa. Ethanol, one of their most important substrates, is first converted to acetaldehyde and further oxidized by aldehyde dehydrogenases (ALDHs) to acetate. They require NAD^+ as a coenzyme to function (Brändén and Eklund, 2008, Plapp, 2010). ADHs are divided into 8 different classes in vertebrates based on sequence and kinetic properties. Humans express classes I to V. Class VI is present in rat and deer mouse (Höög and Brandt, 1995); class VII and VIII in the chicken (Kedishvili et al., 1997) and zebrafish (Reimers et al., 2004) respectively. There is less than 70% sequence identity between the different classes. Dimerization occurs between members of the same class.

Humans encode seven alcohol dehydrogenases, ADH1A, ADH1B, ADH1C, ADH4, ADH5, ADH6 and ADH7, divided into five classes (Höög and Östberg, 2011). The three class I enzymes ADH1A (α), ADH1B (β) and ADH1C (γ) share more than 90% identity.

An endogenous ADH6 enzyme encoded by Class V *ADH* has not been reported. ADH4 or π (class II), ADH5 or χ (class III) and class IV enzyme (ADH7, μ-ADH, σ-ADH) are the other human ADHs. Class III ADH, ADH5, is the ancestral form and the only ADH found in invertebrates (Danielsson and Jörnvall, 1992, Julian-Sanchez et al., 2006).

Class	Gene	Protein	Substrates	Expression
I	*ADH1A*	α, ADH1A	Ethanol, retinol	Fetal and adult liver[1,2], adult kidney[3], adrenal gland[5]
I	*ADH1B*	β, ADH1B	Ethanol, retinol	Fetal and adult liver[1,2], adult kidney[1,4], adrenal gland[5], lung[1,4], skin[6], blood vessels[7]
I	*ADH1C*	γ, ADH1C	Ethanol, retinol	Adult liver[2], kidney[1], adrenal gland[5]
II	*ADH4*	π. ADH4	Ethanol, retinol	Fetal and adult liver[1,5], lower GI tract[5], spleen[8]
III	*ADH5*	χ, AH5	HMGSH, GSNO	Ubiquitous[4,5]
IV	*ADH7*	μ, σ, ADH7	Retinol, ethanol	Stomach[9,10], Upper GI tract [11,12]
V	*ADH6*	-	Ethanol	mRNA seen in liver[5]

Table 1: **Tissue specific expression and substrate specificity of human ADHs**. HMGSH stands for S-(hydroxymethyl) glutathione and GSNO for S-nitrosoglutathione. [1](Smith et al., 1971), [2](Smith et al., 1972), [3](Smith, 1986), [4](Duley et al., 1985), [5](Estonius et al., 1996), [6](Cheung et al., 1999), [7](Allali-Hassani et al., 1997), [8](Edenberg, 2000), [9](Yin et al., 1990), [10](Kedishvili et al., 1995), [11](Yin et al., 1993), [12](Westerlund et al., 2007)

2. Expression of the *ADHs*

The genes encoding the human enzymes are present in a cluster spanning approximately 365 kb on chromosome 4q23, all in the same genomic orientation **(Figure 1)** (Edenberg, 2007b). The primary site of expression for all *ADHs* except *ADH7* is the liver **(Table 1)** where they account for most of the elimination of ingested ethanol and metabolism of other alcohols. They are also expressed in other tissues at lower levels in a tissue-specific manner (Agarwal and Seitz, 2001, Hurley et al., 2003, Edenberg, 2000). The class I *ADHs* are expressed in lungs (Duley et al., 1985, Smith et al., 1971), kidneys (Smith, 1986, Duley et al., 1985, Smith et al., 1971), adrenal glands (Estonius et al., 1996) and skin (Cheung et al., 1999) among other tissues. *ADH1B* is also expressed in blood vessels (Allali-Hassani et al., 1997). The class I *ADHs* exhibit temporal expression patterns with *ADH1A* and *ADH1B* expressed in early and late fetal liver (Estonius et al., 1996, Smith et al., 1972, Smith et al., 1971) respectively and *ADH1C* expressed postnatal (Smith et al., 1972). *ADH4,* the class II isozyme is expressed in both fetal and adult liver (Smith et al., 1971, Estonius et al., 1996), and at lower levels in the lower GI tract and spleen (Edenberg, 2000). *ADH6* transcripts are found in adult and fetal liver and in the stomach, however a functional protein hasn't been reported.

Figure 1: The human ADH gene cluster. The gene cluster of approximately 365 kb is drawn to scale. The cluster is located on the long arm of chromosome 4 (4q23). Their genomic orientation is shown, ADH5 closest to the centromere. The genes are transcribed

in the direction shown. The three class I ADHs are depicted as plain arrows with the other genes as gray arrows.

Interestingly, the two *ADHs* at either end of the cluster **(Figure 1)** have expression patterns most different from the other *ADHs.* While *ADH5* is a housekeeping gene expressed ubiquitously in all tissues including brain, class IV *ADH7* is unique as the only *ADH* not expressed in the liver. Instead, it is mainly expressed in the upper gastrointestinal tract down to the stomach (Yin et al., 1990, Kedishvili et al., 1995, Yin et al., 1997, Moreno and Parés, 1991) and in cornea and other epithelial tissues (Edenberg, 2000, Agarwal and Seitz, 2001).

Studies in rodents show *Adh* (Footnote 1) tissue specific expression pattern similar to their human counterparts *(Galter et al., 2003).* Rodents have 5 *Adh* classes similar to humans, but class I has only one isoform, *Adh1*. Interestingly, the expression of *Adh1* and class IV *Adh7* genes seemed to follow a complementary pattern with only one or the other expressed in any given space (Vaglenova et al., 2003, Westerlund et al., 2007). In the stomach where both are expressed, *Adh1* is expressed in the gastric mucosa and *Adh7* in the gastric pits. In general, *Adh1* is prevalent in the inner cell layers whereas *Adh7* is expressed in the outer layers (Vaglenova et al., 2003, Westerlund et al., 2007) that have high cell turnover and require retinoic acid for epithelial maintenance. *Adh3* is the ancestral ubiquitous class III alcohol dehydrogenase.

[1] The rodent alcohol dehydrogenases are represented as Adh for the protein and *Adh* for the gene. The rodent class II Adh4, class III Adh5 and class IV Adh7 were earlier referred to as Adh2, Adh3 and Adh4 respectively.

3. ADH substrates

The alcohol dehydrogenases metabolize a wide range of primary and secondary alcohols/aldehydes (Holmes, 1994, Boleda et al., 1993) including formaldehyde, 3β-hydroxy-5β-steroids, ω-hydroxy fatty acids, with ethanol and retinol remaining the best studied.

A. Ethanol

All ADHs can metabolize ethanol but with different affinities and catalytic efficiencies **(Table 2)**. The class I enzymes have the lowest Km for ethanol, and account for greater than 70% of ethanol metabolism in the liver (Hurley et al., 2003). ADH4 contributes towards most of the remaining 30% alcohol metabolism in the liver with a Km of 34 mM (Hurley et al., 2003, Li et al., 1977). ADH7 has a lower affinity (higher Km) and highest capacity (larger Vmax) for oxidizing ethanol (Kedishvili et al., 1995). Although ethanol is mostly metabolized in the liver by the hepatic class I enzymes, gastric ADH7 can limit the amount of ethanol entering the bloodstream (Han et al., 1998, Vaglenova et al., 2003). This first-pass metabolism of alcohol by ADH7 can be significant if large amounts of alcohol are consumed in short time, as in episodes of binge drinking. ADH5 is a glutathione dependent enzyme with an important role in elimination of formaldehyde but little role in ethanol metabolism (Kaiser et al., 1993, Li et al., 1977).

Class	Enzyme/ Kinetic constants	**Ethanol**		**Retinol**	
		Affinity	Turnover	Affinity	Efficiency
I	ADH1A	high	Low	moderate	Low
	ADH1B	highest	Moderate	moderate	Low
	ADH1C	high	Low	low	Low
II	ADH4	low	Low	high	Low
III	ADH5	negligible	Low	negligible	
IV	ADH7	low	Highest	high	High

Table 2: Affinity and efficiency for ethanol (Edenberg, 2007a, Hurley and Edenberg, 2012) and retinol (Han et al., 1998, Yang et al., 1994, Parés et al., 2008) metabolism by the ADHs. Km represents the binding affinity for the substrate and indicates the substrate concentration at which the enzyme works at 50% capacity. Vmax or kcat is the turnover number and is defined as the number of substrate molecules converted per minute at saturating substrate concentrations. The class I enzymes with the lowest Km values are the most efficient at low ethanol concentrations, whereas ADH7 has lower binding affinity and high efficiency at saturating ethanol concentrations.

I. Alcohol related diseases

Alcoholism is a polygenic, multi-factorial disease influenced by the additive effects of genes at several loci, and also influenced by social and environmental factors. Alcohol related deaths are the third preventable cause of preventable deaths in the United States (Mokdad et al., 2004, Danaei et al., 2009). The Diagnostic and Statistical Manual of

Mental Disorders (DSM), published by the American Psychiatric Association, define the criteria for diagnosis of alcoholism. In 2007, lifetime alcohol dependence prevalence was estimated to be 12.5% of the US population whereas 3.8% were currently alcohol dependent (Hasin et al., 2007). This study was based on DSM-4 criteria, which defined alcohol abuse as harmful or unhealthy drinking habits despite recurrent social, legal, interpersonal and work-related problems. A result of continual alcohol abuse leading to physical or mental addiction, alcohol dependence or alcoholism was defined as a maladaptive pattern of drinking with three or more of the following symptoms occurring in a one year period: alcohol tolerance, withdrawal, drinking more than intended and unsuccessful attempts to control, excessive time spent in alcohol related activities, impaired social or work related activities and continued used despite physical or psychological consequences.

According to the current DSM (DSM-5) (American Psychiatry Association, 2013), a diagnosis of alcohol use disorder (AUD) requires at least 2 of the following criteria: tolerance, withdrawal, impaired control over alcohol use, unsuccessful efforts to control drinking, alcohol use interfering with important social, occupational or recreational activities and excessive time spent in alcohol related activities, continued consumption despite knowledge of the problem, alcohol craving, recurrent drinking in hazardous situations or resulting in failure of fulfilling obligations. DSM-5 removed the distinction between alcohol abuse and dependence and instead uses alcohol use disorder where the severity of the AUD is dependent on the number of criteria (Dawson et al., 2013). The

threshold for diagnosis of moderate AUD requires 2-3 criteria and severe AUD requires 4+ criteria.

Alcohol consumption can disrupt normal bodily functions including those of the brain, liver, heart, pancreas and the immune system. It is associated with several diseases including liver cirrhosis, steatosis, cardiovascular disease and increases risk for cancer (Cargiulo, 2007, Parry et al., 2011). Alcohol can also increase risk of other health problems like nerve damage, insomnia, erectile dysfunction, dementia and depression. Consumption of alcohol by a woman during any stage of pregnancy can lead to developmental defects in the fetus, which are collectively called Fetal Alcohol Spectrum Disorders (FASD) (Sokol et al., 2003, Zelner and Koren, 2013). FASD includes Fetal Alcohol Syndrome, partial Fetal Alcohol Syndrome, Alcohol-related Neurodevelopmental Disorder, and Alcohol-related birth defects. Along with the direct effects of alcohol, indirect effects resulting from a poor general health status can also cause or exacerbate existing problems. For example, deficiency of the essential nutrient thiamine or vitamin B1 is common in up to 80% of alcoholics, which can lead to serious brain disorders like the Wernicke-Korsakoff syndrome (Martin et al., 2003). Wernicke-Korsakoff syndrome is a severe, debilitating condition whose symptoms include mental confusion, muscle coordination problems and difficulty with learning and memory. In addition, alcohol misuse has undesirable socio-economic consequences including poverty, professional instability, domestic violence, traffic accidents and fractured family units.

II. Genetics of alcoholism

Studies have shown that genetics play an important role in the risk for alcoholism (McGue, 1999, Mayfield et al., 2008, Kendler et al., 1997, Goodwin et al., 1974, Goodwin et al., 1973, Grant, 1998, Munn-Chernoff et al., 2013, Sørensen et al., 2011). Compared to offspring of parents with no alcohol use disorder (AUD, defined previously), offspring with one AUD parent have a 2.5-fold increased risk for developing AUD, with the risk increasing to 4.4-fold for those with 2 AUD parents (Yoon et al., 2013). Children adopted away from alcoholic parents have the same risk for AUD as their siblings raised by the biological parents (Goodwin et al., 1974, Goodwin et al., 1973). These studies imply an important role for genetics in the development of AUD.

Allelic variations in several genes have been reported to be associated with the risk for alcoholism (Treutlein and Rietschel, 2011, Marcella and Jens, 2012, Edenberg and Foroud, 2013, Rietschel and Treutlein, 2013). Apart from the *ADHs* and *ALDHs,* these include the gamma-aminobutyric acid A receptor, alpha 2 (*GABRA2*) (Edenberg et al., 2004, Lappalainen et al., 2005, Covault et al., 2004, Fehr et al., 2006) and the adjacent *GABRG1* (Mary-Anne et al., 2008, Jonathan et al., 2007, Ittiwut et al., 2012), cholinergic receptor, muscarinic 2 (*CHRM2*) (Luo et al., 2005, Dick et al., 2007, Wang et al., 2004), cholinergic receptor, nicotinic, alpha 5 (*CHRNA5*) (Wang et al., 2008), *CHRNB3* and *CHRNA3* (Haller et al., 2013), opioid receptor, kappa 1 (*OPRK1*) (Xuei et al., 2007, Edenberg et al., 2008a), nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (*NFκB1*) (Edenberg et al., 2008b), solute carrier family 6 (neurotransmitter reporter), membrane 4 (*SLC6A4*) (Jian et al., 2013), peroxisomal trans-2-enoyl-coA

reductase (*PECR*) (Treutlein et al., 2009), autism susceptibility candidate 2 (*AUTS2*) (Schumann et al., 2011), importin 11-5-hydroxytryptamine (serotonin) receptor 1A region *(IPO11-HTR1A)* (Zuo et al., 2013b) and ankyrin repeat and kinase domain containing 1 (*ANKK1)* (Kendler et al., 2011).

III. Role of the Alcohol dehydrogenases

The rate of ethanol metabolism determines the concentration of ethanol and its metabolite acetaldehyde in the different tissues, which in turn influences the effects of ethanol consumption on liver and the other organs. There exist variants of ADH isoforms that alter enzyme activities and influence a person's ability to metabolize ethanol and consequently, the risk for developing alcoholism (Thomasson et al., 1993b, Goate and Edenberg, 1998, Osier et al., 1999, Hoog et al., 1986, Mizoi et al., 1983, Bosron and Li, 1986a, Rietschel and Treutlein, 2013, Edenberg and Foroud, 2013, Hurley and Edenberg, 2012). Apart from coding variants causing functional differences in enzyme activity, noncoding variants can alter the level of gene expression, resulting in differential enzyme levels that can also influence the rate of alcohol metabolism (Guindalini et al., 2005, Edenberg et al., 1999, Birley et al., 2008, Birley et al., 2009, Han et al., 2005) and thereby the susceptibility of individuals to alcoholism and other diseases (Edenberg, 2007b).

Functional variants of *ADH1B* and *ADH1C* have been studied extensively. The three most commonly studied *ADH1B* isoforms (Bosron et al., 1985, Bosron and Li, 1987, Hurley and Edenberg, 2012) differ in two coding SNPs, rs1229984 and rs2066702 at

amino acid positions 48 and 370 respectively, with the reference *ADH1B*1* encoding the β_1 isoform of the enzyme with Arginine (Arg) residues at both positions. *ADH1B*2* encodes the β_2 isoform which has histidine (His) at position 48 (ADH1B-Arg48His) (Jörnvall et al., 1984) and the *ADH1B*3* encoded β_3 isoform has the amino acid cysteine (Cys) instead of arginine at position 370 (ADH1B-Arg370Cys) (Burnell et al., 1987). The β_2 and β_3 isoforms of ADH1B have much faster ethanol turnover (80-90 fold higher Vmax) than the reference β_1 isoform (Hurley and Edenberg, 2012, Edenberg et al., 2006).

The *ADH1B*2* allele is very common in East Asian populations with an allele frequency of 73%, and is also present in European and American populations but at a much lower frequency (The 1000 genomes project, 2012). In East Asian populations where *ADH1B*2* is the major allele, several reports strongly indicate it protects against alcohol dependence in an additive manner (Chen et al., 1999, Whitfield, 1997, Whitfield, 2002, Thomasson et al., 1993b, Edenberg, 2011, Li et al., 2011, Thomasson et al., 1993a). The protective effect of *ADH1B*2* was seen in other populations as well (Li et al., 2011, Bierut et al., 2012, Carr et al., 2002). In both European American and African American populations, *ADH1B*2* is associated with reduced risk for AUD, and also with a lower maximum number of drinks consumed in a 24 hour period (Bierut et al., 2012). It reduces binge drinking (Luczak et al., 2002) and the risk for alcoholism in Jews (Hasin et al., 2002), and is associated with unpleasant reactions following alcohol consumption (Carr et al., 2002). It is proposed that the rapid ethanol turnover causes at least a transient elevation in acetaldehyde levels in one or more tissues (Israel et al., 2011, Bosron and Li, 1986b). Since acetaldehyde causes unpleasant reactions that include nausea and facial flushing, it

can lead to an aversion to further alcohol consumption. The *ADH1B*3* allele is seen only in individuals of African descent, and was reported to protect against alcohol dependence (Gizer et al., 2011, Edenberg, 2007b, Edenberg et al., 2006) and the teratogenic effects of alcohol (Scott and Tayor, 2007), with the protective effect of the maternal allele continuing in the children (Dodge et al., 2014).

The two most well studied *ADH1C* coding single nucleotide polymorphisms (SNPs) result in amino acid changes in two positions, 272 and 350 (Hoog et al., 1986). The γ_1 isoform encoded by *ADH1C*1* has Arg and isoleucine (Ile) at positions 272 and 350 respectively, whereas the γ_2 enzyme encoded by *ADH1C*2* has glutamine (Gln) and valine (Val) residues. The differences between the kinetic properties of the two ADH1C isoforms are small compared to those between the ADH1B enzymes (Hurley and Edenberg, 2012, Edenberg, 2007a). Though *ADH1C* has several reports showing association with alcohol dependence (Montane-Jaime et al., 2006, Zintzaras et al., 2006, Chen et al., 1996, Konishi et al., 2003, Konishi et al., 2004, Mulligan et al., 2003, Thomasson et al., 1993b, Day et al., 1991, Thomasson et al., 1993a), the significant and independent association of *ADH1C* variants with alcoholism have been inconclusive because of the strong Linkage Disequilibrium (LD) with the *ADH1B* alleles (Osier et al., 1999, Wall, 2005, Chen et al., 1999, Choi et al., 2005). However, a recent meta-analysis showed that *ADH1C*-Ile350Val is associated with alcohol dependence and abuse with the strongest effects seen in Asian populations (Li et al., 2012).

ADH noncoding variants have also been associated with alcohol dependence. A study of 110 SNPs across the *ADH* cluster showed that variations in *ADH1A, ADH1B* and *ADH4* were associated with alcoholism in European American families, with the associations strongest for SNPs located between the intron closest to the CDS and a region approximately 20 kb downstream of *ADH4* (Edenberg et al., 2006). A SNP in the *ADH4* promoter was found to be associated with alcohol dependence in a Brazilian population (Guindalini et al., 2005) and with lower blood alcohol levels in individuals carrying the *ALDH2**487Glu/Glu isoform (Kimura et al., 2009). This *ADH4* promoter SNP, which is located at position -136 relative to the translational start site, was reported to have a two-fold difference in activity between the two alleles *in vitro* (Edenberg et al., 1999). Variants in *ADH1C* are associated with risk in Native Americans (Mulligan et al., 2003)

IV. ADHs, alcohol and cancer

Chronic alcohol consumption is associated with an increased risk for various kinds of cancer (Seitz et al., 2012, Seitz et al., 2001, Haas et al., 2012, Friedenson, 2012). A meta-analysis found that alcohol most strongly influenced risk for upper aerodigestive tract cancers, with significant increase also seen for cancers of the stomach, colon, rectum, liver, breast and ovaries (Bagnardi et al., 2001). A threshold level of consumption at which alcohol did not pose an increased risk for cancer was not identified. The product of ethanol metabolism, acetaldehyde, probably plays a major role in alcohol related carcinogenesis **(Figure 2)**. Acetaldehyde is a toxic mutagen and carcinogen and can modify DNA by forming DNA adducts (Salaspuro, 2009). Genetic linkage studies have reported that individuals with polymorphisms in the *ADH* genes that generate

acetaldehyde or the *ALDHs* that further oxidize it show an increased risk for cancer (Yokoyama and Omori, 2003, Druesne-Pecollo et al., 2009, Visapää et al., 2004, Yokoyama et al., 1998). *ADH1C*1* is associated with increased risk for cancer in heavy drinkers (Seitz and Meier, 2007) whereas *ADH1B*1* is associated with increased risk for esophageal cancer even in the never/rare and moderate drinkers with the risk increasing with increasing alcohol consumption (Yang et al., 2010). *ADH1B*2* protects against cancers of the upper aerodigestive tract (McKay et al., 2011, Hashibe et al., 2008), possibly indirectly by limiting drinking. It has also been proposed that, in contrast to the damaging effects of prolonged exposure to acetaldehyde, limited exposure can activate DNA repair mechanisms providing protection against deleterious mutations and thus, cancer (Israel et al., 2011).

Several other mechanisms for alcohol related carcinogenesis have been postulated (Purohit et al., 2005, Boffetta and Hashibe, 2006, Pöschl and Seitz, 2004). Alcohol can act as a solvent for carcinogenic compounds and can have a local toxic effect resulting in cell injury and hyper proliferation, the first step to malignant transformation. Induction of the cytochrome CYP-450 2E1, associated with the generation of reactive oxygen species (ROS) and activation of procarcinogens to carcinogens, can also affect cancer risk. Other potential mechanisms include changes in global DNA methylation, increase in estrogen concentration, particularly in the context of breast cancer and nutritional deficiencies, and folate metabolism, among others. Finally, one of the most significant is the impairment of the retinoid metabolism/retinoic acid biosynthesis pathway (Yokoyama et al., 2012, Han et al., 1998, Duester, 1991, Zachman and Grummer, 1998) by ADHs. Retinoic acid is an

important signaling molecule (Rhinn and Dollé, 2012) and anti-carcinogen (Uzzaman et al., 2011, Garattini et al., 2007, Bushue and Wan, 2010) that is required for cell differentiation and maintenance of epithelial tissues.

B. Retinol

Retinol, which is oxidized reversibly to retinaldehyde (RALD) on the pathway to retinoic acid (RA), is another important ADH substrate. Retinoic acid is an important signaling molecule (Rhinn and Dollé, 2012) and anti-carcinogen (Siddikuzzaman et al., 2011) **(Figure 2).** The ADHs are part of the cytosolic medium-chain dehydrogenase (MDR) superfamily, and along with the short-chain dehydrogenases (SDR), constitute the enzymes involved in retinol metabolism: the retinol dehydrogenases (Persson et al., 2008). As in ethanol metabolism, the various enzymes have different activities for retinol metabolism **(Table 2),** with the NAD^+ dependent enzymes (including ADHs) proposed to be best suited for retinol oxidation and $NADP^+$ dependent enzymes for retinaldehyde reduction based on kinetic constants (Parés et al., 2008).

Among the ADHs, the class I enzymes have similar Km values for retinol metabolism, between 30 to 76 μM (Kedishvili et al., 1995, Parés et al., 2008, Han et al., 1998, Yang et al., 1994). *Adh1*$^{-/-}$ null mice are phenotypically similar to wild type mice when maintained on a vitamin A sufficient diet (Molotkov et al., 2002b, Molotkov et al., 2002a). On a vitamin A deficient diet, both the *Adh1* double null mutants and wild type mice show similar growth deficiencies that could be rescued on vitamin A supplementation. However, the *Adh1*$^{-/-}$ mice show decreased metabolism of excess retinol

leading to RA toxicity (Molotkov et al., 2002b, Molotkov et al., 2002a) compared to wild type mice. ADH4, with low Km and low kcat values, is more active for retinol metabolism (Kedishvili et al., 1995, Parés et al., 2008, Han et al., 1998, Yang et al., 1994). Both the class I and class II enzymes are important in elimination of hepatic retinol excess.

ADH5 has little activity with retinoids but *Adh5*$^{-/-}$ mutant mice showed postnatal lethality and growth deficiency, correlating with decreased retinoic acid generation, which could be rescued with supplement vitamin A (Molotkov et al., 2002b). The low but ubiquitous RA generation by ADH5 may be important in tissues that lack the more active enzymes and only require low levels of RA. ADH5 also has a role in elimination of excess retinol.

ADH7 has the highest activity for both retinol oxidation and retinaldehyde reduction and the highest efficiency for retinol metabolism with low Km and high kcat values **(Table 2)** (Kedishvili et al., 1995, Parés et al., 2008, Han et al., 1998, Yang et al., 1994, Gallego et al., 2006) among all retinol dehydrogenases. *Adh5*$^{-/-}$ mice and *Adh7*$^{-/-}$ (null) mice are phenotypically similar to wild type mice on vitamin A sufficient diet but exhibit severe postnatal lethality and growth deficiencies on a vitamin A deficient diet (Deltour et al., 1999, Molotkov et al., 2002b, Molotkov et al., 2002a). Thus, ADH7 is important in RALD generation and RA biosynthesis in the tissues where it is expressed but not in the elimination of excess retinol, indicating the ADHs have non-overlapping functions in retinol metabolism. The ubiquitous ADH5 is involved in low-level RA generation in all tissues. The class I and class IV ADHs exhibit much greater activity towards retinol and

are necessary for RA generation and signaling, mediated through the retinoid receptors, in tissues where they are expressed, particularly on a vitamin A deficient diet.

4. Retinoid receptors

The biological responses to retinoids, which include both natural ligands (retinol/vitamin A metabolites) and synthetic analogs, are mediated by intracellular retinoic acid receptors (RARs) and retinoid X receptors (RXR) (Bushue and Wan, 2010, Germain et al., 2006, Connolly et al., 2013, Chambon, 1996, Aranda and Pascual, 2001). Retinoid receptors belong to the larger superfamily of steroid/thyroid hormone nuclear receptors that share a common modular structure consisting of a variable amino-terminal domain (A/B), a highly conserved central DNA binding domain (C or DBD), a hinge region (D) and a well conserved carboxy terminal ligand binding domain (E or LBD). The DBD contains two highly conserved zinc fingers important for DNA binding specificity, whereas the LBD has a common structure formed by 12 α-helices that undergo a conformational change on ligand binding creating binding surfaces for transcription coregulators.

The receptors are present in both the nucleus and the cytoplasm, with heterodimerization with RXRs and ligand binding shifting the subcellular localization to the nucleus (Kumar et al., 2006, Kawata, 2001, Kesler and Paschal, 2009). Nuclear localization is mediated by nuclear localization signals, typically located in the DBD-hinge region (C/D domain) of the receptors, by interacting with nuclear import receptors at the cytoplasmic side. The receptor-importin complex is transported through the nuclear pore complex to the nucleus, and the receptor released from the complex in a RAN-GTP dependent manner.

The nuclear receptors are typically exported to the cytoplasm by export receptors that interact with nuclear export signals within the nuclear receptors and function in a manner opposite to the import receptors, or by piggybacking on binding protein partners such as transcription coregulators. Nuclear receptors can undergo multiple rounds of nucleocytoplasmic shuttling in their lifecycle. While nuclear localization is essential for the classical or genomic pathway of regulating gene transcription by binding target gene promoters, cytoplasmic localization may be important for efficient receptor turnover and for nongenomic functions such as interactions with various signaling pathways occurring in the cytoplasm (Kesler and Paschal, 2009, Germain et al., 2006, Connolly et al., 2013).

The RAR subfamily binds two naturally occurring ligands, all-*trans*-retinoic-acid and 9-*cis*-retinoic-acid, whereas RXRs bind 9-*cis*-retinoic-acids alone. Each subfamily has multiple isotypes (α, β and γ), which in turn have multiple isoforms due to alternative splicing and differential usage of two promoters (Chambon, 1996, Germain et al., 2006, Robinson-Rechavi et al., 2003, Aranda and Pascual, 2001). There are two major isoforms of RARα (α1 and α2) and RARγ (γ1 and γ2), and four major isoforms of RARβ (β1-β4). The different isoforms of each retinoid receptor subfamily have highly conserved DNA- and ligand binding domains, but differ in their NH_2-terminal regions (A/B domain). The retinoid receptors contain two transcription activation functions that can act synergistically: a constitutive ligand-independent function AF-1 in the A/B region that is specific to each isoform and a ligand-dependent AF-2 function in the LBD that is common to all isoforms of a specific receptor isotype. The AF-2 function contains an activation domain core motif that is highly conserved in all transcriptionally active

nuclear hormone receptors and is essential for ligand-dependent interactions of the receptor, but not for ligand or DNA binding.

The retinoid receptors bind as RAR/RXR heterodimers to retinoic acid response elements (RAREs) to mediate the appropriate cellular response to ligand binding (Chambon, 1996, Germain et al., 2006, Aranda and Pascual, 2001). RAREs are direct repeats of the core recognition motif PuGG/TTCA separated by a 5 bp (DR5) or 2 bp (DR2) or 1 bp (DR1) spacer. The RAR/RXR heterodimer is nonpermissive and does not lead to activation on binding by an RXR ligand alone. However, binding of the RAR ligand allows activation with the subsequent binding of the RXR ligand resulting in synergism and increase in the transcriptional response to the RAR ligand. In the absence of retinoic acid, RAR/RXR heterodimers repress transcription by interacting with the corepressors NCoR, SMRT, Sin3A and Sin3B and recruiting histone deacetylases (HDACs) or DNA methyltransferases (DNMTs) that leads to a condensed chromatin structure. Retinoic acid binding causes a conformational change in the RARs leading to the release of corepressors and recruitment of transcriptional coactivators and subsequent activation of gene transcription. Ligand bound nuclear receptors are found in a complex with multiple proteins of diverse functions, including the coactivators with histone acetyltransferase (HAT) activity, ATP-driven chromatin remodeling activity and finally proteins that interact with and recruit the basal transcription machinery. This is the classical mechanism of regulation by retinoic acid receptors which leads to cell differentiation, cell arrest and apoptosis.

RARs/RXRs can also heterodimerize with other receptors, including the estrogen receptor (ER), peroxisome proliferator-activated receptor (PPAR) and vitamin D receptor (VDR). In this nonclassical or nongenomic pathway where they form unconventional heterodimers, RARs and RXRs regulate their partner's signaling pathways, often having functions opposite to the classical pathway (Balmer and Blomhoff, 2002, Connolly et al., 2013, Aranda and Pascual, 2001). These contrary effects include inhibition of apoptosis and upregulation of pro-survival genes in cancer cells in response to RA binding instead of promoting differentiation, which may allow cancer cells to develop resistance to RA treatment. The AF-1 transactivation and ligand binding domains of RARs are phosphorylated by multiple kinases in response to various signals, which can lead to ligand independent activation of target genes. RARs also interfere with AP-1, NF-κB, and VEGF mediated regulation and chromatin remodeling activities thus integrating diverse signaling pathways.

Disruption in retinoid signaling pathways are linked to cancer. Retinoids have been successfully used as therapeutic agents for cancer treatment, primarily because of their ability to arrest proliferation and induce differentiation (Connolly et al., 2013, Bushue and Wan, 2010, Evans and Kaye, 1999, Germain et al., 2006). Exogenous expression of the tumor suppressor RARβ, which is epigenetically silenced in many tumors, causes RA-dependent and RA-independent apoptosis and growth arrest. The best-studied example of RA as a cancer therapeutic tool is acute promyelocytic leukemia (APL). APL is caused by chromosomal translocation between RARα and promyelocyte leukemia (PML) genes resulting in aberrant signaling of both proteins. The fusion PML- RARα

fusion protein binds the corepressors NCoR and SMRT more efficiently than wild type RARα and causes repression of RARα target genes, which in turn prevents differentiation of APL cells. Treatment with supraphysiological doses of all *trans* retinoic acid (ATRA) leads to remission in APL patients; patients that develop resistance to ATRA and relapse can be effectively treated with other RARα specific agonists for complete remission.

However, retinoids have not been as successful in the treatment of solid tumors despite their success with hematological malignancies (Connolly et al., 2013, Bushue and Wan, 2010, Evans and Kaye, 1999, Germain et al., 2006). A possible mechanism of resistance to RA therapy is the epigenetic silencing of RARβ early in carcinogenesis, which can potentially be overcome by combining RA therapy with DNMT and HDAC inhibitors. Other possible mechanisms include the presence of cancer stem cells that resist differentiation, loss of RARβ expression by mechanisms other than epigenetic silencing and crosstalk between retinoid and other signaling pathways allowing cancer cells to bypass specific targeted therapies. Additionally, the therapeutic potential of retinoids is limited by the associated toxic side affects caused by their pleiotropic functions. The side effects of varying severity include teratogenicity, headache, bone toxicity and mucocutaneous cytotoxicity. Thus, further studies are focused on both overcoming the resistance to retinoids and decreasing the adverse side effects to them.

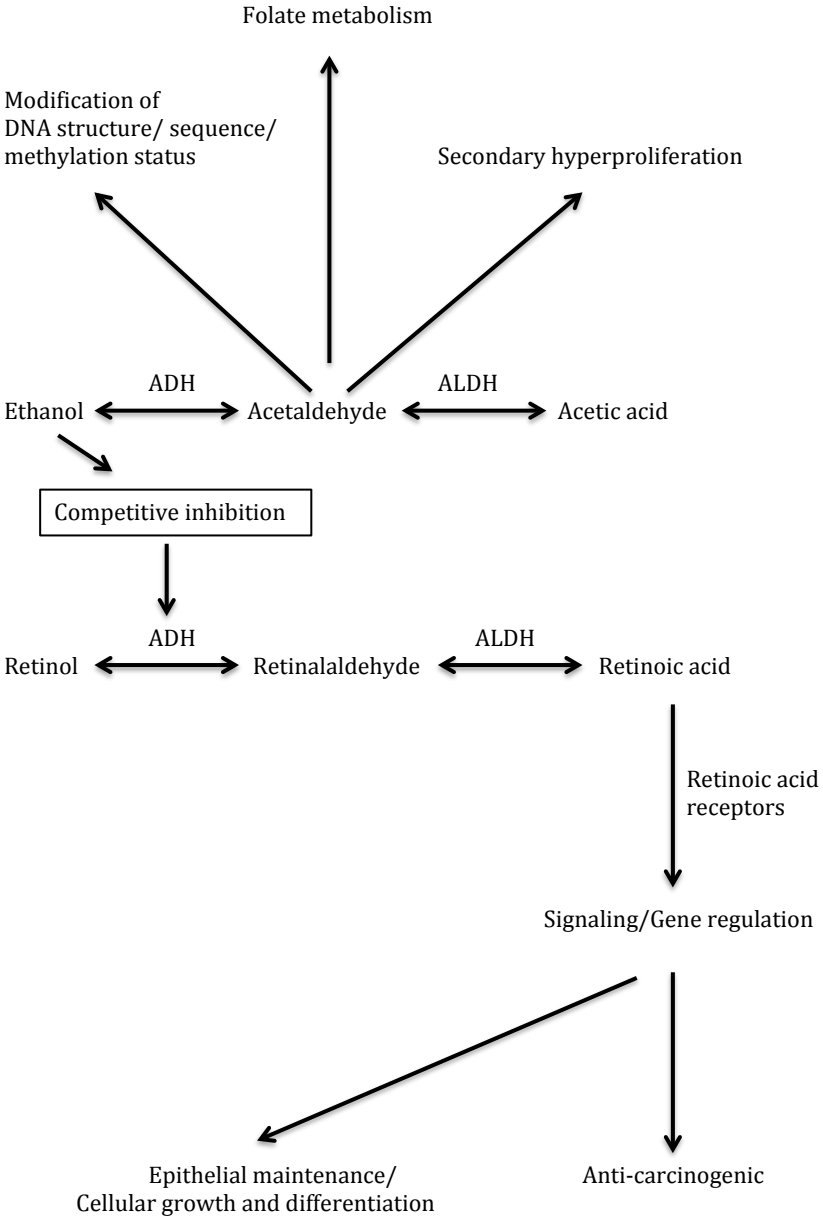


Figure 2: ADH substrates and their biological significance. Ethanol and retinol compete for the ADH catalytic site. Ethanol can competitively inhibit retinol metabolism with the inhibitory constants of each ADH (Ki) similar to the corresponding affinity constant (Km). Ethanol metabolite acetaldehyde is a toxic mutagen and carcinogen while retinoic acid has important roles in signaling and directing cellular growth and differentiation with anti-carcinogen properties.

Since ethanol and retinol bind the same active site, ethanol competitively inhibits retinol metabolism (Yin et al., 2003, Duester, 1991) **(Figure 2)** by class I, class II and class IV ADHs with inhibition constants ranging from 0.037 to 11 mM (Han et al., 1998). Retinol metabolism by the ADH linked pathway can be significantly disrupted by physiologically attainable concentrations of ethanol from social/heavy drinking. This inhibition of the retinoic acid biosynthesis pathway is implicated in fetal alcohol syndrome (Zachman and Grummer, 1998, Duester, 1991) and cancer (Purohit et al., 2005, Boffetta and Hashibe, 2006).

5. *ADH7* and disease associations

Genome wide association studies (GWAS) have identified associations of *ADH7* with alcoholism, drug dependence and cancer **(Table 5)**. Association studies are useful in identifying variants contributing to diseases like alcoholism, which is multifactorial and influenced by several different genes (Murray and Gurling, 1982). Alleles identified by GWAS are typically not the causal variants, but are in linkage disequilibrium (LD) with them, particularly where the association with disease is not apparent (Marian, 2012, Clark, 2003). LD is when alleles at two or more loci associate in a non-random fashion (are inherited together more often than would be expected based on their individual frequencies) (Reich et al., 2001). LOD, D' and r^2 are three measures for LD with a D' or r^2 score of 1 indicating the alleles have not been separated by recombination. The identification of the significant association of any SNP with a disease phenotype from a GWAS study represents the finding that might represent any of the SNPs that are in a high degree of LD with the originally identified SNP.

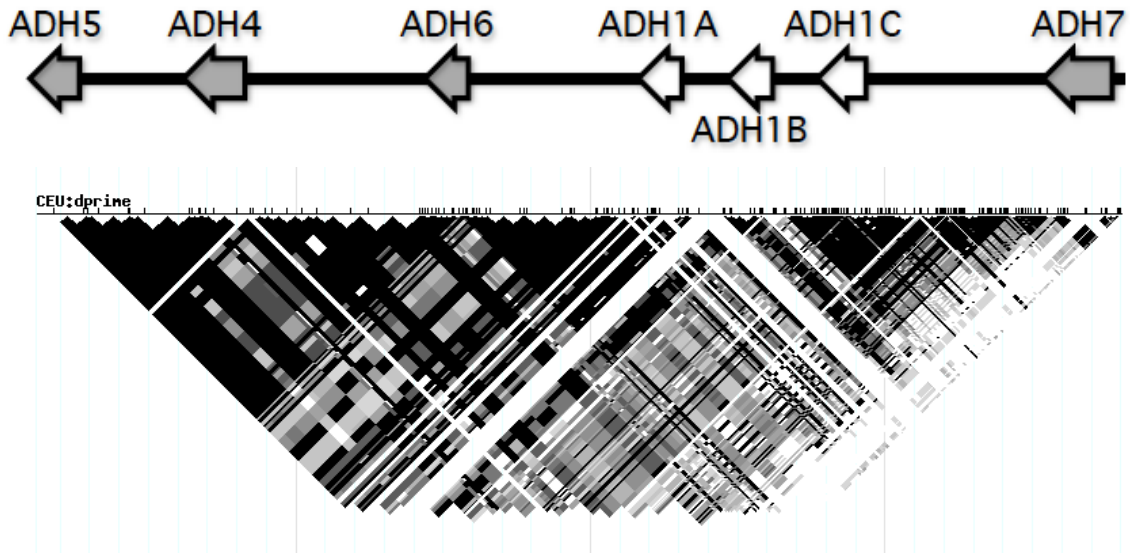


Figure 3: Linkage disequilibrium in the *ADH* region. LD among single-nucleotide polymorphisms (SNPs) in the alcohol dehydrogenase (*ADH*) genes represented by the D' values in CEU population (International HapMap Consortium, 2010, The International HapMap, 2005) (Hapmap Data Rel 27 Phase II+III, Feb09, on NCBI B36 assembly, dbSNP b126). D' is a measure of LD between two genetic markers with a value of D'= 1 meaning no recombination has occurred and the markers are in complete LD. The intensity of the shading is proportional to the extent of LD between the markers with darker shades representing greater LD and lighter shades representing lower LD. SNPs within genes generally are in high LD with each other, whereas SNPs in different genes typically have a lower LD between them. Positions of the *ADHs* are indicated on the top. Most of the *ADH* genes except *ADH7* are in moderately strong LD (indicated by the mostly clear areas in the vicinity of *ADH7*).

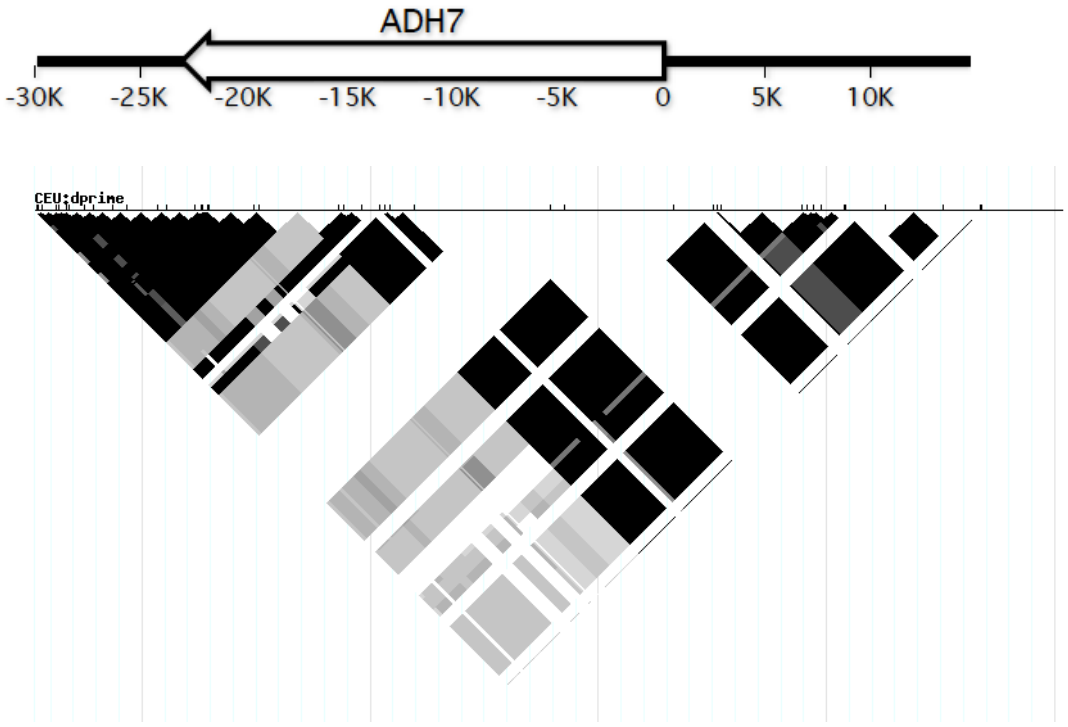


Figure 4: LD plot of the *ADH7* gene. LD is represented by D' values in CEU population (International HapMap Consortium, 2010, The International HapMap, 2005) (Hapmap Data Rel 27 Phase II+III, Feb09, on NCBI B36 assembly, dbSNP b126). The intensity of the box color is proportional to the strength of LD between two genetic markers. The *ADH7* gene and upstream and downstream regions are depicted in the picture at the top; +1 represents the ADH7 translational start site. *ADH7* is divided into two LD blocks with a recombination hotspot within the gene; each block is in moderately strong LD within itself but not with the other block.

LD analysis in the *ADH* region reveals an interesting pattern. There is little LD between *ADH7* and the rest of the *ADH* cluster, with a break in the intergenic region between *ADH7* and *ADH1C* **(Figure 3).** *ADH7* itself is divided into two haplotype blocks with a recombination hotspot within intron 7 of the gene **(Figure 4)** (Han et al., 2005, Birley et al., 2008, Edenberg et al., 2006). SNPs located in the 5' haplotype block are associated with both blood alcohol and breath alcohol concentrations, and account for approximately 18% of a major quantitative trait locus within the *ADH* region influencing alcohol metabolism, or 11% of the total genetic variance (Birley et al., 2009, Birley et al., 2008). The C allele of rs1154458, located in intron 6 of *ADH7,* is associated with protection against alcoholism (Osier et al., 2004, Han et al., 2005). Among individuals carrying the incontrovertibly protective *ADH1B*2* polymorphism, the allele frequency of rs1154458-C (0.135) was reported to be significantly lower in alcoholics compared to rs1154458-G (0.345), whereas the allele frequencies in nonalcoholic controls were similar (0.331 for rs1154458-C vs 0.324 for rs1154458-G). This indicates that the protection cannot be attributed to *ADH1B*2* alone (Osier et al., 2004). The absence of significant linkage disequilibirum between rs1154458 or any other known *ADH7* SNP and the *ADH1B*2* allele indicates that the genetic association of *ADH7* with alcoholism is an independent association (Han et al., 2005).

Two *ADH7* non-synonymous coding variants are known. rs1573496 causes an Alanine (Ala) to Glycine (Gly) substitution at position 92, whereas rs59534319 (which is rare to uncommon in most populations) causes Lysine (Lys) to Glutamic acid (Glu) substitution at position 238. The GG and CG genotypes at rs1573496 confer a protective effect

against cancers of the upper aero-digestive tract (McKay et al., 2011, Hashibe et al., 2008, Wei et al., 2010), where ADH7 is expressed. There is no reported association of rs59534319 with disease, although it is only 30-bp away from the synonymous SNP rs971074, which is associated with a higher risk for drug dependence (Levran et al., 2009, Luo et al., 2007). Since neither of the non-synonymous coding variants rs1573496 explain all *ADH7* associations, it is likely that many of these associations reflect differences in regulation of expression, as has been found for other complex diseases (Ward and Kellis, 2012, Pennisi, 2011, Gregory and Jay, 2011). It is therefore important to study *ADH7* transcription regulation.

6. Elements regulating gene transcription

A) Promoters

Transcriptional regulation involves complex interaction of different kinds of *cis*-acting regulatory elements, proteins that bind these elements, and chromatin structure (Maston, 2006). The gene promoter may be the most important regulatory sequence because all the events preceding transcriptional activation ultimately lead to the basal transcription machinery at the gene promoter. The ubiquitous and highly conserved RNA polymerase II (RNAPII) core promoter drives transcription in all eukaryotic cells, and is the minimum sequence required to initiate gene transcription containing the transcription start site. (Juven-Gershon et al., 2008, Juven-Gershon and Kadonaga, 2010). Core promoter recognition is the first step in the mechanism of transcription initiation (Thomas and Chiang, 2006). In combination with non-prototypical binding partners, the core

promoter can also influence regulation of cell specific transcription (Goodrich and Tjian, 2010). While the core promoter typically has binding sites for general transcription factors, the proximal promoter immediately upstream of the transcription start site contains primary regulatory elements and specific transcription binding sites. The region upstream to the gene promoter typically contains transcriptional regulatory elements marked by characteristics such as conservation across species, DNaseI hypersensitivity, histone modifications and binding sites for transcription factors (Maston, 2006). Upstream regulatory elements can affect gene expression by affecting promoter activity.

B) Enhancers and silencers

Enhancers, silencers and Locus Control Regions (LCRs) can control gene expression in an orientation-independent and position-independent manner, and from locations as remote as 100 kb from the gene (Bondarenko et al., 2003, Dean, 2006, Ong and Corces, 2011, Jin et al., 2011). Enhancers bind activator proteins that activate transcription either by directly recruiting general transcription factors and RNA polymerase II and/or by recruiting chromatin remodeling complexes that make the chromatin accessible to general transcription factors and RNA polymerase II (Szutorisz et al., 2005, Arnosti and Kulkarni, 2005, Chin-Tong and Victor, 2011). Though enhancers can activate promoters over long distance, they generally activate the nearest promoter preferentially. On the other hand, silencers repress transcription either by interfering with general transcription factors assembly or by binding repressors that in turn inhibit assembly of general transcription factors (Ogbourne and Antalis, 1998). LCRs enhance the expression of

linked genes in a tissue specific and copy number dependent manner (Dean, 2006, Li et al., 2002).

C) Insulators

Insulators or Boundary elements restrict the effects of long-range chromatin interactions. The eukaryotic genome is organized into domains comprised of individual genes or clusters of genes with distinct patterns of expression (Lunyak, 2008, Kadauke and Blobel, 2009). Active and inactive chromatin domains are often in close proximity to one another, and enhancer and silencer elements operate over large distances to regulate the genes in these domains. Boundary elements or insulators composed of specific sequences function to prevent regulatory elements present within a domain from promiscuously activating or suppressing the expression of genes located within adjacent domains (Bushey et al., 2008, Barkess and West, 2012, Moltó et al., 2009).

Chromatin insulators have emerged as important factors in the spatial and topological organization of higher order chromatin structures and functional transcriptional domains (Phillips-Cremins and Corces, 2013, Van Bortle and Corces, 2012). Insulators typically exhibit either one or both of two characteristics 1) "enhancer blocking" function, i.e., when placed between the enhancer and promoter they block enhancers from activating the promoter and 2) "barrier" function of protecting transgenes from position effects. Insulators can interact with each other, with other regulatory elements and with the nuclear architecture forming loops that encourage specific enhancer-promoter interactions while restricting others (Murrell, 2011, Valenzuela and Kamakaka, 2006,

Wallace and Felsenfeld, 2007) depending on various factors including cell type and target enhancers and promoters. All vertebrate insulators identified so far bind CTCF, a zinc finger protein that can recognize diverse DNA elements (Phillips and Corces, 2009, Holwerda and De Laat, 2013).

D) CTCF

CTCF is a highly conserved ubiquitous zinc finger protein and can function as both a transcriptional activator (Vostrov and Quitschke, 1997) and repressor (Filippova et al., 1996), and is involved in insulation, genomic imprinting and X-chromosome inactivation (Filippova, 2008). CTCF is known to interact with various transcription factors, chromatin modifying enzymes and other regulatory proteins, with the end result of activation, repression or insulation dependent on the combination of factors present at the site in a particular cell or tissue (Merkenschlager and Odom, 2013, Lee and Iyer, 2012, Barkess and West, 2012). CTCF can also form homodimers, and can interact with the nuclear lamina causing DNA looping and formation of *cis*- and *trans*- chromatin domains and influencing cross talk between gene promoters and regulatory elements (Ren et al., 2012, Phillips and Corces, 2009, Williams and Flavell, 2008). Cell specific interactions between CTCF bound sequences are known to regulate the cell specific expression of linked genes such as the beta-globin cluster (Junier et al., 2012, Ren et al., 2012). CTCF binding is dependent on the DNA sequence methylation status since methylation of the CpG nucleotides within the CTCF binding site can disrupt binding (Wang et al., 2012). CTCF binding and function can also be influenced by the neighboring DNA binding

factors (Weth and Renkawitz, 2011). CTCF binding is, therefore, a marker for potential regulatory function.

7. Transcription regulation of the *ADHs*

Proximal promoter elements are important in regulating local gene expression and those pertaining to ADHs have been studied extensively to understand the regulation of ADH expression (Edenberg, 2000). The proximal promoters of different ADH classes are packed with binding sites for multiple proteins, particularly liver specific transcription factors considering that the liver is the primary site of expression for 6 of the seven human *ADHs*. All *ADHs* except *ADH5* contain a TATA box, whereas *ADH5* contains a GC-rich sequence instead. The CCAAT/enhancer-binding protein (C/EBP) and activator protein-1 (AP-1) family of transcription factors, specificity protein 1 (Sp1), nuclear factor-1 (NF-1), upstream stimulatory factor (USF), hepatocyte nuclear factor 1 (HNF1), glucocorticoid receptors (GR) and retinoic acid receptors (RAR, RXR) (Edenberg, 2000) are implicated in *ADH* regulation.

CCAAT/enhancer-binding proteins (C/EBP) are liver enriched transcription factors and play an important role in the regulation of all the *ADH* genes (Stewart et al., 1991, Brown et al., 1996, Edenberg, 2000). They are a highly conserved family of proteins that have basic leucine zipper domain, which is involved in the dimerization and DNA binding (Tsukada et al., 2011). The AP-1 family of basic-region leucine zipper (bZIP) transcription factors include JUN, FOS, ATF and MAF sub-family proteins that can homo- and hetero-dimerize. AP-1 proteins contribute both to basal gene expression, as

well as in response to various stimuli including serum, growth factors and oncoproteins, and a variety of environmental stresses, such as UV radiation (Karin et al., 1997). They can activate or repress transcription, and have both oncogenic and anti-oncogenic properties (Robert and Erwin, 2003) depending on the context. The HNF1 (or Transcription factor 1 (TCF1)) transcription factor is essential for the activation of several liver specific genes though it is also expressed in the epithelial cells of endoderm derived organs (Tronche and Yaniv, 1992). HNF1-α binds as a homo- or heterodimer with its paralog HNF1-β, a transcription factor with the same DNA sequence specificity (Mendel et al., 1991). Mutations in HNF1-α and HNF1-β are causal factors for maturity-onset diabetes of the young (MODY) types 3 and 5 respectively (Raile et al., 2008, Gardner and Tai, 2012).

Known *cis*-regulatory elements in the *ADH* proximal promoter regions, however, do not fully explain the tissue-specific expression of the *ADHs*. The human class I *ADH* genes share 80-90% similarity up to 300 bp upstream of the transcription start site, but show different spatio-temporal expression (Edenberg, 2000). Regions extending to 110 kb upstream and 104 kb downstream could induce *ADH* class I expression in liver at levels comparable to endogenous levels in mice (Szalai et al., 2002); 12 kb upstream and 23 kb downstream regions were not sufficient to activate the class I *ADH* promoter. This suggests that distal elements are involved in *ADH* regulation. A HNF1-bound enhancer located in the intergenic region between *ADH1C* and *ADH7* was found to be both necessary and sufficient to induce expression of all 3 human *ADH* class I genes in liver (Su et al., 2006). Distal elements have also been identified upstream of *ADH1A*, one of

them regulating tissue-specificity (Dannenberg et al., 2005). A potent FOXA-dependent enhancer, which is liver specific and is located upstream of *ADH4,* increased *ADH4* and *ADH1B* promoter activities 50-fold in hepatoma cells (Pochareddy and Edenberg, 2010). However, distant regulatory mechanisms of the other classes of *ADH* genes have not been addressed yet, especially those that regulate the differential pattern of *ADH7* and *ADH* class I tissue specific expression.

8. *ADH7* transcription regulation

A 232 bp proximal promoter of *ADH7* is functional in HeLa, CV-1 (monkey kidney) and H4IIE-C3 (rat liver) cell lines (Kotagiri and Edenberg, 1998). Four transcription factor binding sites were identified within the proximal promoter. Site A contains a consensus AP-1 binding site and was bound by transcription factors including c-Jun, a component of AP-1, in the three cell lines tested. Mutation of the AP-1 transcription factor binding site preventing c-Jun binding decreased promoter activity to 8-12% of the wild type promoter, indicating a critical role for AP-1 in *ADH7* gene regulation (Kotagiri and Edenberg, 1998). Multiple binding sites for C/EBP transcription factors were identified. Site-directed mutagenesis of site B that could be bound by C/EBP reduced promoter activity by 75% indicating the importance of this site in *ADH7* regulation. Site C could be bound by both c-Jun and C/EBPα, and the mutated sequence doubled activity in CV-1 cells while having no significant effect in H4IIE-C3 cells making it the only cell-specific regulatory element in this proximal promoter. The DNA footprint for the fourth site D was observed only in CV-1 cells, suggesting it may be cell specific. Sites A, C and D are

conserved in the orthologous mouse gene *Adh3* indicating they may be functional in the mouse (Kotagiri and Edenberg, 1998).

Co-transfection of the *ADH7* promoter with C/EBP-α or C/EBP-β expression plasmids led to decreases in promoter activity in CV-1 cells that do not express their endogenous C/EBP-α (Kotagiri and Edenberg, 1998). Since these C/EBP transcription factors especially C/EBP-α are prominently expressed in the liver, the negative effect of C/EBP co-expression suggests one mechanism that contributes to *ADH7* not being expressed there. However, these elements do not fully explain the tissue specific transcriptional regulation of *ADH7*.

Other transcription factors, including the retinoic acid receptor family, STAT family and HNF family proteins, are predicted by PROMO (Messeguer et al., 2002) to bind in the *ADH7* proximal region. The Transfac Matrix database (Matys et al., 2003) predicted binding sites for the transcription factors FOXO3 and FOXJ2 in the 1.5 kb region immediately upstream of *ADH7,* but ENCODE (Rosenbloom et al., 2013) data do not show binding of any transcription factors in that region in the cell lines tested. In a genome-wide study of alcohol dependence in Han Chinese, *ADH7* was reported to be hypomethylated in alcoholics, although the sites were not shown (Zuo et al., 2013a). The UCSC Genome Browser shows two sites of methylation in the proximal promoter region (at bp -148 and -787) and one site further upstream (at -12059); these are partially methylated in many cell types, including HepG2. None are at polymorphic sites, and their effect on gene expression is not known. Binding of NFκB, Maf (v-maf avian

Musculoaponeurotic fibrosarcoma oncogene homolog) and IRF (interferon regulatory factor) proteins have been shown to bind sequences further upstream in some cell lines.

9. Research Objectives

The overarching objective of this dissertation is the study of *ADH7* transcription regulation. *ADH7* is involved in several important metabolic pathways and is unique among the *ADHs* in its expression, affinity and efficiency for its substrates and in having little to no LD with the rest of the cluster. Genome wide association studies (GWAS) have shown that *ADH7* is independently and significantly associated with several diseases including widely prevalent alcoholism and cancer, but its regulation is not fully explained. This dissertation is a study on the proximal and more distal elements involved in cell-specific *ADH7* regulation, and how genetic variants affect function. Studies were done in cell systems with contrasting *ADH7* expression, using bioinformatics and experimental approaches.

The first main research objective is the identification of elements regulating *ADH7* transcription and the effect of variants on this regulation. Since all except one of the reported disease associations are noncoding and remain unexplained, at least some are likely from regulatory element variants affecting transcriptional activity that are in LD with the GWAS identified genetic markers. An understanding of the *ADH7* regulatory regions and the effects of variants can help interpret the association data and point toward likely causal variants (Gregory and Jay, 2011). Since most of the disease associations belong to the *ADH7* 5' LD block and regulatory elements typically reside upstream of the

coding region and are conserved across species, the 12.5 kb conserved region upstream of *ADH7* was the focus of this study.

The second objective was to gain an understanding of the tissue-specific expression of *ADH7,* specifically the lack of expression in the liver despite the proximity of the class I *ADH* enhancer located in the intergenic region between *ADH1C* and *ADH7* and other liver functional *ADH* enhancers. *In silico* prediction tools were first used to identify potential chromatin insulators based on the hypothesis that such elements can block liver enhancers from wrongly activating *ADH7* in the liver. Then experimental studies were done to study the functional role of the test element in relevant cell systems. This study investigated one mechanism by which *ADH7* expression in the liver is regulated.

10. Cell systems

The cell systems used in this dissertation replicate *ADH7* tissue specific expression and therefore make great *in vitro* models for studying cell specific expression. CP-A esophageal cells express *ADH7* endogenously while HepG2 hepatoma cells do not express *ADH7,* similar to the pattern seen in the respective tissues. MEF mouse embryonic fibroblasts also express *Adh7.*

A. HepG2 cells

HepG2 human hepatoma cells (ATCC, Manassas, VA; HB-8065) are adherent cells with an epithelial morphology. The cells grow in monolayers forming small aggregates. The cell line was derived from a fifteen year old Caucasian male with a well-differentiated

hepatocellular carcinoma (Knowles et al., 1980). It is non-tumorigenic in immunosuppressed mice. Because of the absence of viral infection, HepG2 cells are a popular model for studying hepatocellular carcinoma and other liver diseases (Xia et al., 2013, Costantini et al., 2013). They express both intact and inducible phase I (including several *ADHs* **(Table 3)**) and phase II enzymes required for activation and detoxification of xenobiotics, and therefore are a good *in vitro* model for studying drug metabolism (Mersch-Sundermann et al., 2004, Diamond et al., 1980).

B. CP-A cells

CP-A cells (ATCC, CRL-4027) (Palanca-Wessels et al., 2003, Palanca-Wessels et al., 1998) are an hTERT immortalized cell line obtained from a non-dysplastic Barrett's Esophagus tissue of an adult male. They have an epithelial like morphology and grow as adherent cells in monolayers. CP-A cells represent a great model to study esophageal cancers particularly Barrett's esophagus(Kosoff et al., 2012) and *ADH7* regulation because it is the only human cell line identified so far that expresses *ADH7* endogenously.

C. MEF cells

MEF (mouse embryonic fibroblast) cells (Xu, 2001) immortalized by infection with a recombinant retrovirus expressing SV40 large T antigen are frequently used as feeder cells to support the growth of embryonic stem (ES) cells and to maintain them in an undifferentiated state (Villa-Diaz et al., 2013). They express *Adh7* endogenously and can be used as a second point for comparison on *ADH7* regulation.

II. MATERIALS AND METHODS

1. Bioinformatics

The mammalian conservation track in the UCSC genome browser was used to identify the regions of conservation between human and other vertebrates in the sequence upstream of *ADH7*. Published literature and online databases including ALFRED (Allele frequency database) (Rajeevan et al., 2012) and HGMD (Human gene mutation database) (Stenson et al., 2009) were used to identify disease associated SNPs. Linkage Disequilibrium (LD) data from the NCBI Hapmap database (Hapmap Data Rel 27, Feb 09) (International HapMap Consortium, 2010) and the 1000 genome database (The 1000 genomes project, 2012) (based on Ensembl v69, Oct 2012) were used to identify LD between SNPs in the vicinity of *ADH7* and disease-associated SNPs. The Hapmap and dbSNP databases were used to identify naturally occurring haplotypes and the samples to use as templates for obtaining them. The insulatordb database (Bao et al., 2008) was used to identify potential CTCF binding sites in the *ADH* region, particularly in the vicinity of *ADH7*.

2. Cloning of test fragments and their variants

Two naturally occurring haplotypes of an 841-bp *ADH7* proximal promoter (A7P-G/A; -19 to -859-bp relative to *ADH7* translation start site) were generated by PCR amplification from an anonymous human DNA sample using Invitrogen Platinum Pfx polymerase (Life Technologies, Grand Island, NY; catalog 11708-013). The two promoter haplotypes, A7P-A and A7P-G were cloned into HindIII and BglII sites in the multiple cloning site of the pXP2 luciferase vector (SK, 1988), oriented so that the

promoters drive luciferase expression. Restriction sites for NcoI, Acc65I, NotI and XhoI were included in the forward primer of A7P to facilitate further subcloning. Fragments of approximately 1 kb each extending up to -12.5 kb upstream of the *ADH7* gene were amplified from human DNA using Invitrogen Platinum Pfx polymerase. DNA samples used as templates in PCR amplification of naturally occurring haplotypes of 7P5, 7P6, 7P8 and 7P10 **(Table 4)** were obtained from the Coriell Institute for Medical Research (Camden, New Jersey, USA). Fragments 7P2 to 7P10 (except 7P6) were cloned into Acc65I and XhoI sites of pXP2 upstream of promoters A7P-G and A7P-A, and tested for effect on each promoter activity. 7P6 was cloned into NcoI and XhoI sites of pXP2 upstream of A7P-G and A7P-A because Acc65I cut within the 7P6 sequence. 7P10 sub-fragments were cloned into HindIII and XhoI sites upstream of A7P-A. 7P10 sub-fragment E2 was cloned into HindIII and XhoI sites upstream of A7P-A in the reverse orientation to make E2flip, and into ApaI and SwaI sites to make E2far. E2 was cloned into HindIII and XhoI sites upstream of luciferase gene to make E2prom. All primers are listed in Appendix A **(Table 8)**.

The A7P-A promoter haplotype cloned into pXP2 was used as the parent construct for insulator assays. The approximately 240 bp fragment corresponding to class I specific HNF1 bound enhancer (ENH) (Su et al., 2006) was cloned into PciI and NdeI sites approximately 1 kb upstream of the A7P promoter fragment in pXP2. ENH forward primer contained restriction sites for MluI, BssHII, NruI and the reverse primer had AsiSI, AfeI, NgoMIV restriction sites to facilitate further sub-cloning. The known insulator element, XL9 (Majumder et al., 2008) and the 965 bp test fragment located 2 kb

upstream of HNF1 bound enhancer identified as a potential CTCF binding site *in silico* were cloned into the AsiSI and NgoMIV sites, and/or the MluI and NruI sites for the enhancer blocking assays. DNA samples used as templates in PCR amplification of the three naturally occurring haplotypes of iA1C were obtained from the Coriell Institute for Medical Research (Camden, New Jersey, USA). All primers are listed in Appendix B **(Table 9)**.

For tests with non-homologous elements, the 965 bp iA1C fragment was cloned into the pGL3 control vector (Groskreutz et al., 1995) containing the luciferase reporter gene driven by the SV40 promoter, with the SV40 enhancer more than 2 kb upstream of the SV40 promoter. The test and control elements were placed between the promoter and enhancer between the MluI and BglII sites in the multiple cloning site (MCS) immediately upstream of the SV40 promoter.

3. Site directed mutagenesis

The QuikChange Lightning Multi Site-Directed Mutagenesis kit (Agilent Technologies, Santa Clara, CA; 210516) and unlabeled oligos (iA1C-Osh) used for EMSA assays were used to make full length shuffled iA1C plasmid (with only the specific 5-6 nt different from the WT sequence). 1.5 µl of 100 ng/µl DNA template and 1.5 µl of 5 µM mutagenic primer were added to relevant amounts of dNTP mix, 10X QuikChange Lightning multi reaction buffer, QuikChange Lightning Multi enzyme blend and double-distilled water to make up 25 µl reaction volume as per manufacturer's instructions. The PCR reaction was set up as follows: 95C – 2 min; <95C – 20 s, 55C – 30 s, 65C – 4.5 min> 30 cycles; 65C

– 5 min; 4C. 1 μl of DpnI restriction enzyme was added directly to the PCR product and vortexed gently to mix, then incubating the reaction for 5 min at 37C to digest parental strand. 1.5 μl of DpnI treated mutated PCR product was transformed into XL10-Gold ultracompetent cells according to manufacturer's instructions.

4. Cell culture and Transient transfections

HepG2 human hepatoma cells (ATCC, Manassas, VA; HB-8065) were cultured in MEM (ATCC) with 10% fetal bovine serum (Invitrogen, Carlsbad, CA), 4 mM glutamine (Thermo Scientific Hyclone, Waltham, MA) and 1X Penicillin and Streptomycin (Thermo Scientific Hyclone) on cell bind surface plates (Corning Inc., Tewksbury, MA; CLS3296) at 37°C. Transient transfections in HepG2 cells were done by seeding 0.8 x 10^5 cells per well in Corning cell bind surface 12-well plates (Corning Inc., Tewksbury, MA; CLS3336). 24 h after seeding, medium was changed and cells were transfected with 1.52 pmoles of test DNA, 15 ng of pCMV β-galactosidase plasmid (Clontech, Mountain View, CA) and enough pUC19 DNA to get a total DNA amount of 1 μg per well. 2 μl per well Fugene HD (Roche, Indianapolis, IN) was used as the transfection reagent.

CP-A cells (ATCC, CRL-4027) are an hTERT immortalized cell line obtained from a non-dysplastic Barrett's Esophagus tissue. CP-A cells were cultured in Keratinocyte-SFM (Invitrogen, Carlsbad, CA; 17005-042, with each 500 ml supplemented with 25 mg Bovine pituitary extract and 2.5 μg human recombinant epidermal growth factor supplied with the medium) plus 10% FBS and 1% Penicillin-Streptomycin (ATCC, 30-2300) at 37 °C. For transient transfections**,** CP-A cells were seeded at 3.75 X 10^5 cells per well in

12-well cell bind plates (Corning Inc., Tewksbury, MA; CLS3336). Transfections were done 24 h after seeding as described previously.

MEF (mouse embryonic fibroblast) cells immortalized by infection with a recombinant retrovirus expressing SV40 large T antigen were cultured in DMEM supplemented with 10% fetal bovine serum (Invitrogen, Carlsbad, CA), and 1X Penicillin and Streptomycin (Thermo Scientific Hyclone, Waltham, MA) at 37°C. MEF cells were seeded at 0.8 X 10^5 cells per well in 12-well cell bind plates (Corning Inc., Tewksbury, MA; CLS3336) for reporter gene assays. Transfections were done 24 h after seeding as described previously.

5. Reporter gene assays

HepG2 cells were collected in ice cold 1X PBS made from 10 X PBS solution (Fisher Bioreagents, Hampton, New Hampshire; BP3994) 30 h after transfection. Cells were collected by centrifugation and the resulting pellet was resuspended in 100 μl of 1X Reporter Lysis buffer (Promega, Madison, WI; E3971). Cell lysates were prepared by repeated freeze-thaw cycles in dry ice and water. Lysates were centrifuged to pellet cell debris and the supernatants transferred to new tubes. Luciferase activity was measured on a Spectramax LS (Molecular Devices, Sunnyvale, CA) using the Luciferase assay system (Promega, Madison, WI; E1501), with 20 μl of the extract. β-galactosidase assays were carried out with 5 μl extract using the Galacto-Light System (Tropix, Bedford, MA).

CP-A cells were harvested 24 h after transfections by washing once with 1X PBS, and scraping into 250 μl 1X Reporter lysis buffer. Lysates were then freeze-thawed

repeatedly and centrifuged to pellet down cell debris, as described earlier. 20 μl and 10 μl of CP-A cell supernatant was used for luciferase and β-galactosidase assays respectively. MEF cells were collected 24 h after transfection in 250 μl 1X Reporter lysis buffer, and assayed for luciferase and β-galactosidase activity as earlier.

All test fragments were transfected in at least 3 independent experiments, with each experiment having at least 3 replicates. Relative activity of a test construct was calculated by normalizing each luciferase activity to the internal control β-galactosidase activity, to correct for the transfection efficiency, and then calculating the ratio of normalized luciferase activity of test fragment to that of the promoter haplotype driving the luciferase expression in the construct. P-values were calculated using two-tailed t-tests of the normalized values.

6. Nuclear and cytoplasmic extraction

Nuclear extracts were prepared from HepG2 cells using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermoscientific Pierce, Waltham, MA; 78835), following the manufacturer's protocol. Protein concentrations were measured by Bio-Rad protein assay (Bio-Rad, Hercules, CA; 500-0002).

7. Electrophoretic Mobility Shift Assay (EMSA)

EMSAs were carried out with 40 bp double-strand oligonucleotides with 10 bp on either side of the 20 bp predicted CTCF binding site. Oligonucleotides were synthesized (Integrated DNA Technologies) with a 5′ 6-FAM label on one of the strands and then

annealed to complementary unlabeled oligonucleotides in 10 mM Tris (pH 8.0), 1 mM EDTA (pH 8.0) and 100 mM sodium chloride. Protein binding reactions were carried out with 0.25 μM of the annealed oligonucleotides and 6 or 8 μg of the nuclear extracts in 10 mM HEPES (pH 7.9), 60 mM potassium chloride, 5 mM magnesium chloride, 1 mM EDTA, 1 mM DTT, 7% glycerol, 1 μg of poly (dIdC), 1X protease inhibitor and 0.1 mM ZnSO4. Nuclear extracts were first incubated with premix containing all components except oligonucleotides for 5-10 min on ice.

Labeled oligonucleotides (0.5 μM per reaction) were then incubated with the nuclear extract for 16 min at 25°C. In competitor assays, unlabeled competitor oligonucleotides in molar excess to the labeled oligonucleotides were added to the reaction and incubated at RT for 10 min before addition of the probe. For supershift assays, 2 μg of the antibody was added to each reaction. CTCF (Millipore 07-729), and CTCF (Santa Cruz., sc-5916X) antibodies were tested. DNA–protein complexes were electrophoresed on 6% polyacrylamide Novex DNA Retardation Gels (Invitrogen; EC63652BOX) for 45 min in ice-cold 0.5X TBE running buffer (45 mM Tris-borate and 1 mM EDTA, pH 8.3) and then scanned with fluorescent image analyzer FLA-5100 (Fujifilm, Valhalla, NY) at 473 nm with LPB filter. The gel was run on ice to avoid overheating.

8. Chromatin immunoprecipitation (ChIP)

ChIP assays were done in HepG2 and CP-A cells with Anti-CTCF (Millipore; 07-729), positive control H3 and negative control IgG antibodies supplied with the SimpleChIP Enzymatic Chromatin IP kit (Cell Signaling Technology, Danvers, MA; 9003). Assays

were done following manufacturers protocol, with sonication conditions optimized for the two cell types. CP-A cells were sonicated 3 times for 7 s each and HepG2 cells were sonicated 5 times for 15 s each, with a 60 second interval on ice between sonication bursts. The immunoprecipitations were done overnight at 4C with gentle rotation, followed by incubation with magnetic beads for 3 h at 4C. Following DNA elution, quantification was done by qPCR and standard PCR. Eleven primer pairs amplifying overlapping iA1C sub-fragments approximately 100 bp in size were tested for enrichment in the Anti-CTCF IP sample with yields calculated as % input. Human RPL30 primers (Cell Signaling Technology, Danvers, MA; 7014S) specific to exon 3 of RPL30 gene bound by Histone H3 served as a random non-specific control for CTCF binding; and as positive control for H3 binding.

9. Adenovirus infection of cells

HepG2 cells were plated in two 12 well plates with each plate having alternate columns seeded at 60-70% confluence (24 wells in total- 12 each for shCTCF and shControl). 24 h after seeding, first column in each plate was infected with adenovirus encoding CTCF shRNA (shCTCF) and third column infected with adenovirus encoding control shRNA (shctrl). Both shCTCF and shctrl adenovirus were generously provided by Dr. Paul Herring (Rodenberg et al., 2010, Ishihara et al., 2006). The adenovirus was prepared for infection by diluting in EMEM Complete medium (used for culturing HepG2 cells) by adding 20 μl/well of each virus to 400 μl/well of EMEM Complete medium (700 μl/well complete medium for 6 well plates). After removing medium from cells, they were washed once with ice cold 1X PBS; after which 420 μl of diluted shCTCF or shctrl was

added directly to the cells. Cells were returned to incubator for 4 h and an additional 1ml complete medium added at this point. Cells were then harvested 72 h after infection. Media was changed daily until harvest. 6 wells of each plate (shCTCF or shctrl infections) were used for RNA and protein extraction respectively. For RNA extraction, 400 μl Trizol was added directly to cells, incubated at RT for 3 min and then pipetted up and down to break up DNA. The cells lysed in Trizol were stored at -80C until lysates were ready to be used, after which the regular protocol for Trizol extraction was followed. cDNA synthesis was done from both Trizol extracted RNA and column purified RNA. For protein extraction, cells were washed with ice cold 1X PBS and then lysed in 150 μl RIPA buffer +inhibitors placed on ice and let sit at RT for 10-15 min. Cell lysates were then scraped and collected into prechilled 1.5 ml Eppendorf tubes, spun 2 min at 4C and supernatant transferred to fresh tubes (2 tubes---one for most of the sample and one for protein assay ~ 20 μl). The crude protein extracts were then frozen at -80C.

10. RNA extraction and purification, Real time PCR

HepG2 and CP-A RNA were extracted using Trizol (Life Technologies, Carlsbad) and purified using the Qiagen RNeasy mini kit (Qiagen, Germantown, MD; 74106) following manufacturer's protocol. DNA concentrations were determined by absorbance at 260 nm using a Nanodrop spectrophotometer (Thermo Scientific, Waltham, MA). Superscript III First-strand synthesis system (Invitrogen, Carlsbad, CA; 18080-051) was used to synthesize cDNA from 1 μg RNA following the manufacturer's protocol. qPCR assays were performed using 10 μl of 2X Power SYBR Green mastermix (Life technologies, Carlsbad, CA; 4367659), 2 μl of 3 mM primers, 3 μl of 2X diluted cDNA and water in

the StepOnePlus Real-time PCR system (Life Technologies, Carlsbad, CA; 4376598). No RT and no DNA controls were included in each experiment.

III. CHAPTER I [2]

SINGLE NUCLEOTIDE POLYMORPHISMS INTERACT TO AFFECT *ADH7* TRANSCRIPTION

[2] Large portions of the text in this chapter are reproduced from the published article: Jairam, S and Edenberg, HJ. Single Nucleotide Polymorphisms interact to affect ADH7 transcription JAIRAM, S. & EDENBERG, H. J. 2014. Single-Nucleotide Polymorphisms Interact to Affect ADH7 Transcription. *Alcoholism: Clinical and Experimental Research***,** n/a-n/a.

1. INTRODUCTION

Genome wide studies have identified associations of *ADH7* with alcoholism, drug dependence and cancer. SNPs located in the 5' haplotype block are associated with both blood alcohol and breath alcohol concentrations **(Table 5)**. The intronic SNP rs1154458 is linked to protection against alcoholism. Two *ADH7* non-synonymous coding variants are known, rs1573496 (Ala92Gly) and rs59534319 (Lys238Glu; rare to uncommon in most populations). Rs1573496-G confers a protective effect against cancers of the upper aero-digestive tract where *ADH7* is expressed. There is no reported association of rs59534319 with disease, although it is only 30 bp away from the synonymous SNP rs971074, which is associated with a higher risk for drug dependence. It is likely that many of these associations reflect differences in regulation of expression, as has been found for other complex diseases (Ward and Kellis, 2012, Pennisi, 2011, Gregory and Jay, 2011). An understanding of the *ADH7* regulatory regions and the effects of variants on regulation can help interpret the association data and point toward likely causal variants.

The goal of this study is to identify elements regulating *ADH7* transcription and examine how genetic variants affect regulation. Two human cell lines and a mouse cell line with contrasting *ADH7* expression were chosen to study *ADH7* gene regulation. CP-A epithelial cells, derived from a non-dysplastic Barrett's esophageal sample immortalized by hTERT (Palanca-Wessels et al., 1998), and MEF (mouse embryonic fibroblast) cells immortalized with the SV40 T antigen express *ADH7* endogenously. HepG2 human

hepatoma cells (Knowles et al., 1980) do not express *ADH7.* We identified regulatory elements, including some with cell-specificity, and detected interactions among them.

2. RESULTS

A. *ADH7* proximal promoter variant rs2851028 and its effect on activity

There are two haplotypes of the 822 bp fragment that extends from -1 bp to -823 bp with respect to the *ADH7* translation start site, A7P-A and A7P-G, that differ at a single SNP: rs2851028, at -630 bp **(Figure 5)**. Both haplotypes were tested for promoter activity by transient transfections in CP-A esophageal cells and MEF mouse embryonic fibroblasts that express *ADH7* and in HepG2 hepatoma cells that do not **(Table 3)**. These cells represent the complementary pattern of *ADH7* expression seen in humans (Westerlund et al., 2007, Vaglenova et al., 2003). Both haplotypes of the promoter fragment were active in all tested cell types **(Figure 6)**. The promoter construct with the minor G allele was 1.6- to 2-fold more active than the promoter with the A allele independent of cell type. The empty vector pXP2 had no activity in either cell type.

Gene	CP-A	HepG2	MEF	Gene
ADH1A	-	-	**N/A**	*Adh1*
ADH1B	-	-		
ADH1C	+	-		
ADH4	-	+	-	*Adh4*
ADH5	+	+	+	*Adh5*
ADH6	+	+	**N/A**	*Adh6*
ADH7	+	-	+	*Adh7*

Table 3: *ADH* gene expression in CP-A human esophageal, HepG2 human hepatoma and MEF mouse embryonic fibroblast cells. Expression was tested using RT-PCR and the results are depicted as a "+" for gene expressed or "-" for no gene expression. The human *ADH* genes are listed on the left and mouse *Adhs* on the right. N/A indicates that the only mouse class I *Adh* (*Adh1*) and *Adh6* were not tested.

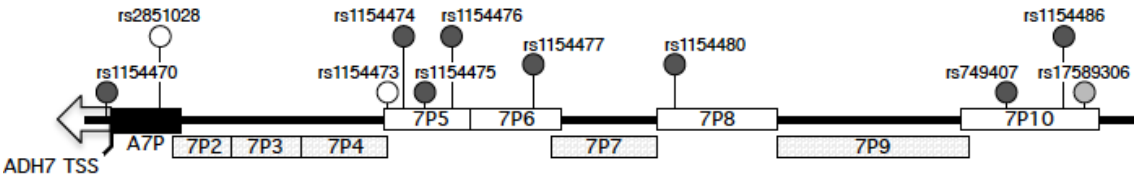


Figure 5: Map of the *ADH7* promoter and upstream fragments. Map is drawn to scale in chromosomal orientation with promoter (black solid), monomorphic fragments (below the line) and polymorphic fragments (on the line). Transcription occurs right to left (arrow) from the translation start site (TSS). SNPs are shown, with the three LD blocks represented as solid black circles, plain circles or grey circle (rs17589306).

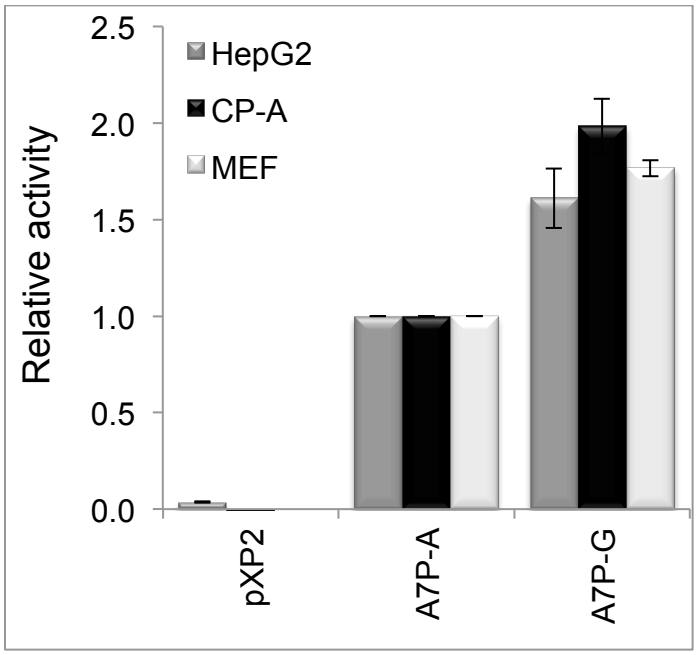


Figure 6: *ADH7* promoter haplotypes are differently active in HepG2, CP-A and MEF cells. The relative activity is the ratio of promoter activity of the G construct (A7P-G) to the A construct (A7P-A) within each cell line. Standard errors of mean are shown ($n \geq 12$). P-value for the difference between the A and G constructs was 3×10^{-5} (HepG2 cells), 1×10^{-6} (CP-A cells) and 9×10^{-8} (MEF cells).

B. Upstream regulatory regions

Conservation extended approximately 12.5 kb upstream of *ADH7*, with only patches of conservation seen in the rest of the intergenic sequence between *ADH7* and *C4orf17,* which flanks the ADH gene cluster on one end. We therefore focused our studies on this 12.5 kb region. Fragments of approximately 1 kb **(Figure 5)** were cloned upstream of the A7P-G and A7P-A promoters and tested for effects on promoter activity by transient transfections. Opposite haplotypes of fragments having multiple known haplotypes were tested **(Table 4).**

Test fragment (Location)	dbSNP	DNA Source	Haplotypes tested	Haplotype frequency		
				CEU	YRI	CHB+ JPT
A7P (-19 to -859)	rs2851028	AN	G A	0.18 0.83	0 1	0.58 0.42
7P5 (-3475 to -4667)	rs1154473, rs1154474, rs1154475, rs1154476	NA12057 NA12751	ATAC GCGT	0.41 0.30	0.51 0	0.57 0.24
7P6 (-4571 to -5741)	rs1154477 rs1442477	NA12057 NA12751	GG AG	0.69 0.31	0.92 0.08	0.76 0.24
7P8 (-6954 to -8492)	rs1154480, rs1154481	NA07000	GC AG	0.05 0.36	- -	0.23 0
7P10 (-10815 to -12578)	rs749407, rs1154486 rs17589306	AN	CTT TTT	0.48 0.36	0.78 0.08	0.78 0.21

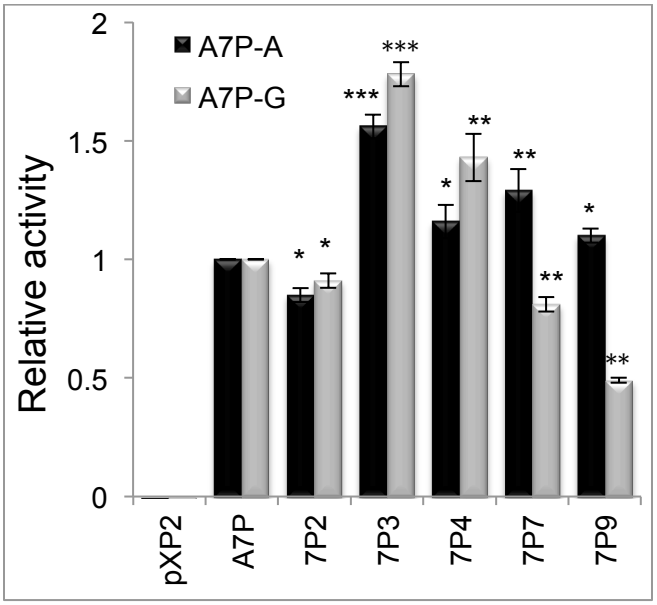
Table 4: Upstream fragments and their variants tested. Test fragment location is with respect to the ADH7 TSS. DNA source: Samples from Coriell Institute (NA number) or anonymous DNA used to amplify test fragments of different haplotypes. Haplotypes tested are listed in order of the SNPs. Haplotype frequencies in three populations obtained from the Hapmap (The International HapMap, 2005) and 1000 genome (THE 1000 GENOMES PROJECT, 2012) databases; CEU, Northern and Western Europeans from Utah; CHB+JPT, East Asian Han Chinese and Japanese and YRI, Yoruba from Nigeria).

C. Monomorphic fragments

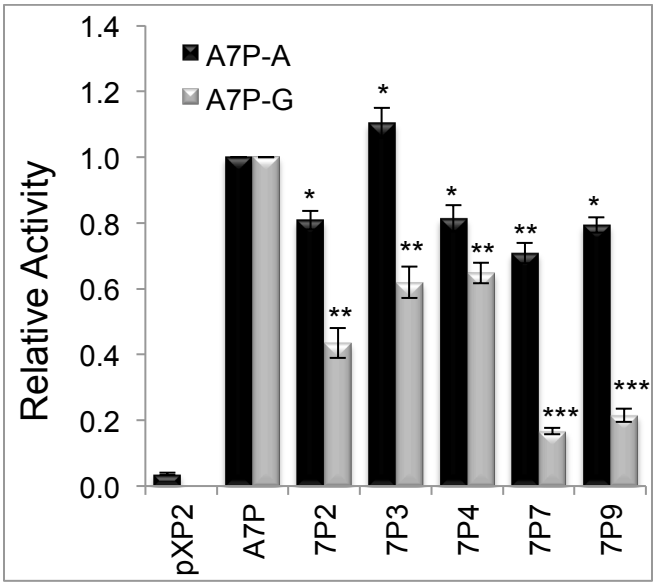
Fragments 7P2, 7P3, and 7P4, together spanning from -762 bp to -3479 bp, did not have any reported SNPs, nor did fragments 7P7 and 7P9 **(Figure 5)**. These monomorphic fragments showed different activity depending on which promoter haplotype they were combined with and which cells they were tested in. Fragment 7P2 reduced activity of both promoters by 10-15% in CP-A cells **(Figure 7A)** and reduced the activity of the A7P-A promoter by a similar 20% in HepG2 cells, but had a significantly larger effect on A7P-G activity in HepG2 cells (55% reduction) **(Figure 7B).** In MEF cells, 7P2 considerably decreased both promoter activities but had a stronger effect on A7P-G (51%) **(Figure 7C).** 7P3 significantly increased both promoter activities in CP-A cells but had little effect on the activity of A7P-A in HepG2 cells, and significantly reduced activity of A7P-G (by 38%) in HepG2 cells. On the other hand, it had no effect on A7P-G in MEF cells, but increased A7P-A activity by 30%. 7P4 had stronger effects on the A7P-G promoter in CP-A and HepG2 cells, but in different directions: it increased A7P-A and A7P-G promoter activities in CP-A cells (by 15% and 40% respectively) **(Figure 7A)**, while decreasing both activities in HepG2 cells (by 20% and 35% respectively). 7P4 had an opposite effect on the two promoters in MEF cells: increasing A7P-A activity by 20% and reducing A7P-G activity by 12% **(Figure 7C)**. In CP-A cells, 7P7 increased A7P-A activity but decreased A7P-G activity to a similar extent, whereas in HepG2 cells it decreased both promoters, with a much stronger effect on A7P-G **(Figure 7B)**. 7P7 showed a more pronounced effect in MEF cells by increasing A7P-A activity 100% and decreasing A7P-G by 25% **(Figure 7C).** 7P9 had little effect on A7P-A activity in CP-A

cells, but strongly decreased A7P-G activity. In HepG2 cells and MEF cells, 7P9 decreased activity of both promoters, with a much stronger effect on A7P-G.

7A. **CP-A cells**



7B. **HepG2 cells**



7*C.* MEF cells

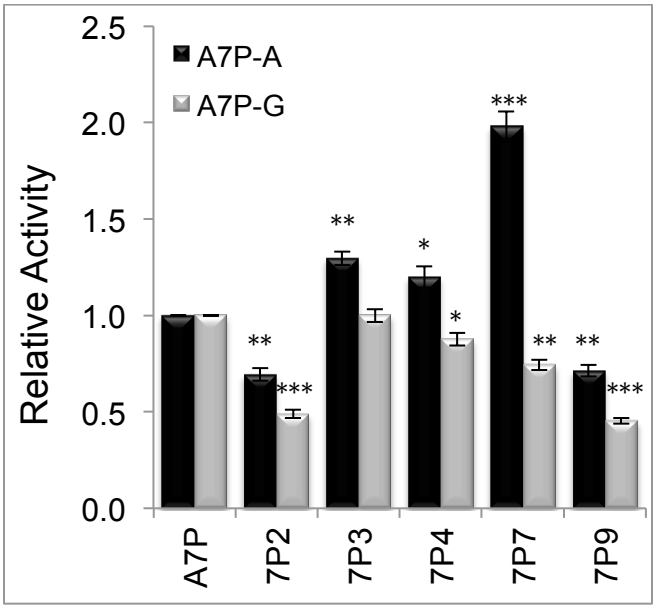


Figure 7: Activity of monomorphic fragments on the two promoter haplotypes. Transient transfections of the monomorphic upstream sequences 7P2, 7P3, 7P4, 7P7 and 7P9 were done in **A)** CP-A cells, **B)** HepG2 cells and **C)** MEF cells. Relative activity of each construct represents the ratio of normalized luciferase activity of the test fragment to the A7P-A (black) promoter or A7P-G (grey shaded), whichever the fragment was tested in. Scales of vertical axes are different. Error bars represent standard errors of mean. T-tests were done for each construct and the corresponding promoter construct in each cell type. Significant differences between transcriptional activities (calculated as normalized luciferase activities) are represented by stars: * indicate p-values ≤ 0.03; ** indicate p-values ≤ 0.001; *** indicate p-values $\leq 1 \times 10^{-8}$.

D. Polymorphic fragments

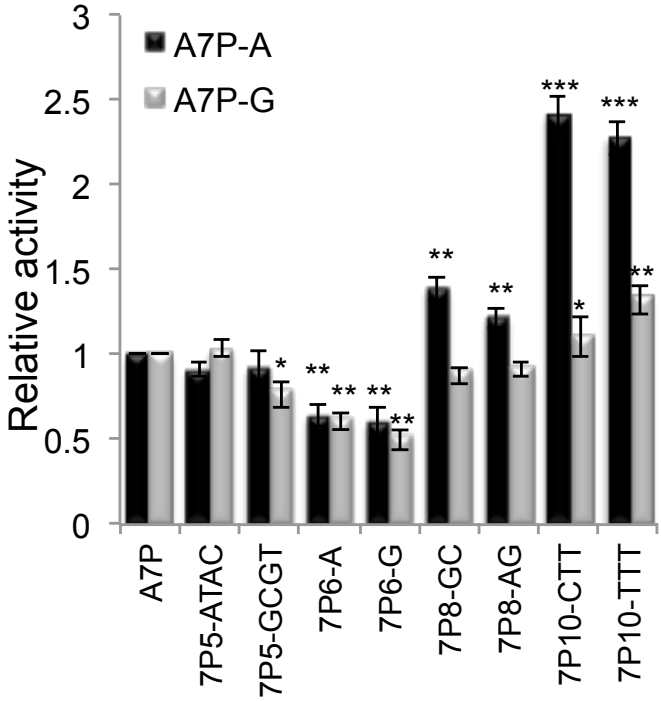
Fragments 7P5, 7P6, 7P8 and 7P10 contain SNPs **(Figure 5)** in linkage disequilibrium with variants having significant disease associations **(Table 5)** (Osier et al., 2004, Han et al., 2005, Birley et al., 2008, Birley et al., 2009, Hashibe et al., 2008, McKay et al., 2011, Cadoni et al., 2012, Levran et al., 2009, Wei et al., 2010). Transient transfections were done with two naturally occurring haplotypes of each of these fragments **(Table 4)** cloned upstream of both A7P-A and A7P-G promoters. 7P5-ATAC had no effect on either promoter activity in CP-A cells **(Figure 8A)** while strongly reducing both promoter activities (about 50%) in HepG2 cells **(Figure 8B**). 7P5-GCGT showed both promoter-specific and cell-specific function: it had no effect on A7P-A but decreased A7P-G activity by 20% in CP-A cells; in HepG2 cells, it increased A7P-A activity by 20% but decreased A7P-G activity by 40%. In MEF cells, 7P5-ATAC decreased both promoter activities whereas 7P5-GCGT showed promoter specific effects by increasing A7P-A by 55% and decreasing A7P-G by 30% **(Figure 8C)**.

Both 7P6 haplotypes decreased both promoter activities by 36-47% in CP-A cells **(Figure 8A)**. In HepG2 cells **(Figure 8B**), 7P6-A reduced A7P-A activity by 20% with no significant effect of 7P6-G. Both 7P6 haplotypes showed a dramatic 85% reduction of A7P-G activity. Both 7P6 variants decreased both promoter activities in MEF cells. The 7P8 haplotypes 7P8-GC and 7P8-AG increased A7P-A activity in CP-A cells by 40% and 20% respectively but had nearly no effect on A7P-G **(Figure 8A)**. In HepG2 cells **(Figure 8B**), both 7P8 variants decreased A7P-A activity by 20% and A7P-G activity by approximately 60%. 7P8-AG had opposite effects in MEF cells increasing A7P-A by

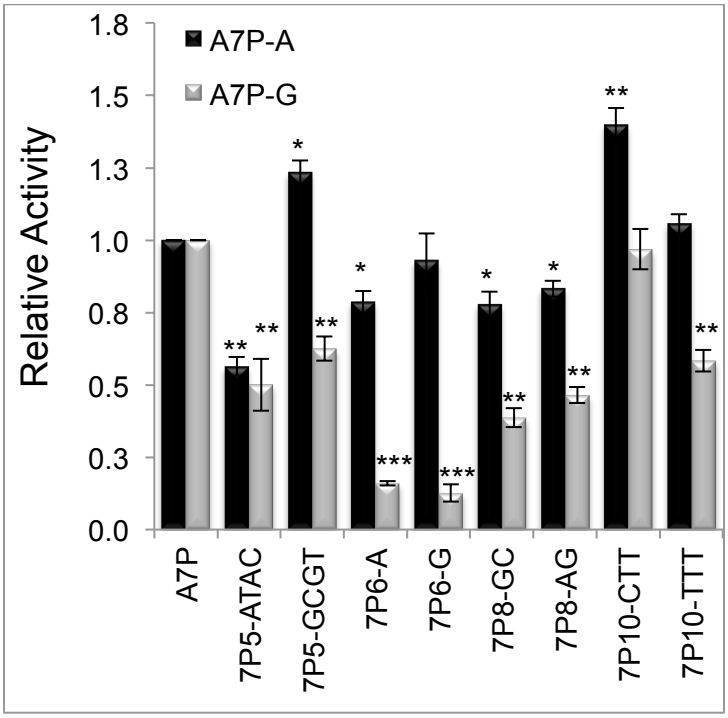
60% and decreasing A7P-G activity by a similar amount, while 7P8-GC decreased both promoter activities **(Figure 8C)**.

The 7P10 variants had the greatest positive impact on promoter activity. In CP-A cells, both 7P10 haplotypes increased the A7P-A promoter activity 2.3-2.4 fold while having significantly smaller effects on A7P-G activity (1.1-1.3 fold) **(Figure 8A)**. In HepG2 cells, 7P10-CTT increased A7P-A activity by 40% with no effect on A7P-G, and 7P10-TTT had no effect on A7P-A but decreased A7P-G activity by 40% **(Figure 8B)**. In MEF cells, similar interactions between variants were seen with 7P10 but with a more pronounced effect **(Figure 8C)**. 7P10-CTT decreased both promoter activities, but the TTT variant had a strong promoter-specific effect. 7P10-TTT increased A7P-A activity 2-fold while decreasing A7P-G activity by half.

8*A*. CP-A cells



***8B*. HepG2 cells**



***8C.* MEF cells**:

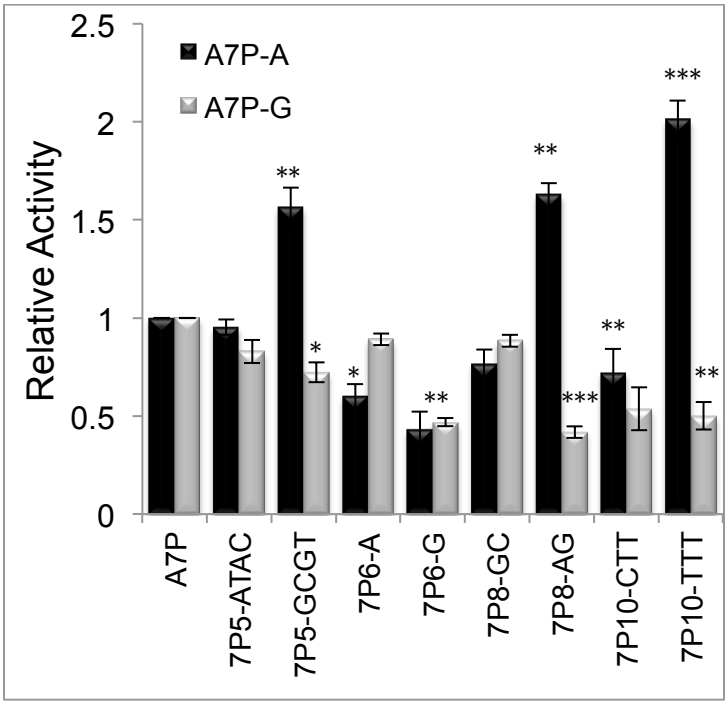
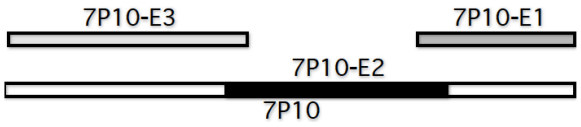


Figure 8: Activity of polymorphic fragments on the two promoter haplotypes. Transient transfections of two naturally occurring haplotypes of each polymorphic fragment (7P5, 7P6, 7P8 and 7P10) were done in **A)** CP-A cells, **B)** HepG2 cells and **C)** MEF cells. Each sequence was tested for effect on both promoters A7P-G and A7P-A. Relative activity of each construct represents the ratio of normalized luciferase activity to the promoter on which it was tested, A7P-G (grey shaded) or A7P-A (black). Error bars represent standard errors of mean. Scales of vertical axes are different. P-values reflect the differences between the normalized luciferase activities of each construct and the corresponding promoter construct. Relative activities are shown with * indicating p-values ≤ 0.008; ** indicate p-values $\leq 8 \times 10^{-5}$; *** indicate p-values $\leq 2 \times 10^{-11}$

7P10, which had the greatest effect on activity of the A7P-A promoter, was fragmented into three partially overlapping sub-sequences to better localize the regulatory element(s) **(Figure 9A)**. Transient transfections cells showed that one of the sub-fragments, 7P10-E2 had the strongest effect on promoter activity in both CP-A and MEF cells **(Figure 9B)**, which was more than twice that of the whole fragment. The other sub-fragments 7P10-E1 and 7P10-E3 had some activity, but less than the whole fragment. The effect of each sub-fragment was very similar in the two cell types tested.

Given its location and relatively strong effect on promoter activity, 7P10-E2 was tested to determine if it functioned as an enhancer in CP-A cells **(Figure 10A)**. It did not function alone as a promoter (7P10-E2prom; **(Figure 10B)**). It did function nearly as well in either orientation (E2flip vs E2), and when moved further away from the promoter, fulfilling characteristics of an enhancer **(Figure 10B).**

9A.



9B.

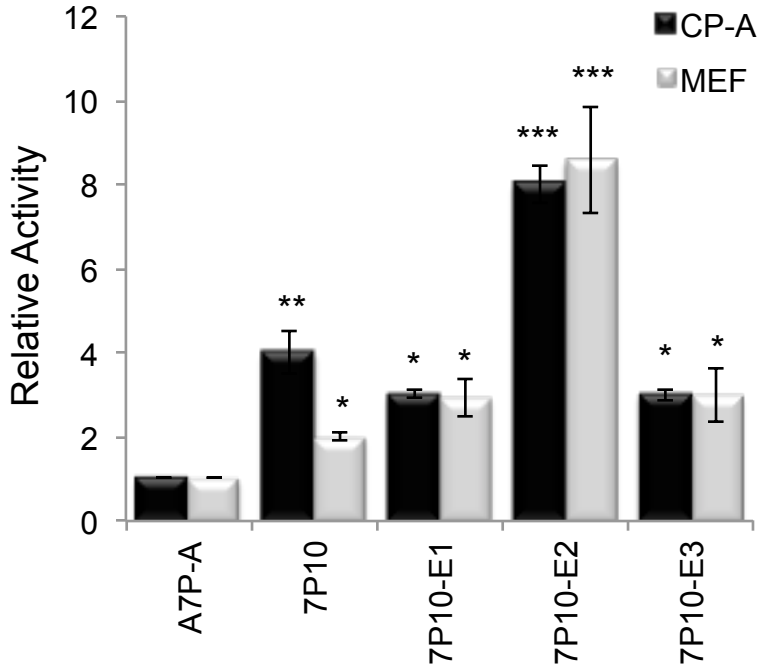
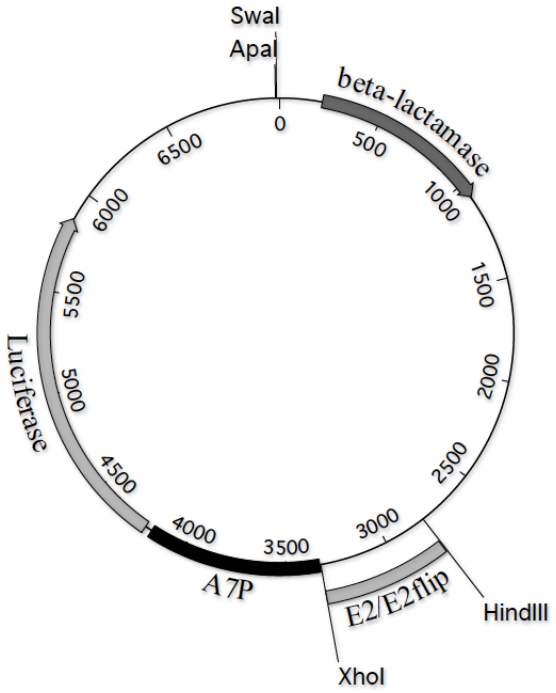


Figure 9: Localization of regulatory elements in fragment 7P10. A) 7P10 sub-fragments approximately 500 bp in size are shown as oriented on the chromosome, with E1 being the farthest upstream from ADH7. **B)** Activity of sub-fragments of 7P10 in CP-A (black solid) and MEF cells (grey solid). P-values reflect the differences between the normalized luciferase activities of each construct and the promoter construct in corresponding cell type. * indicate p-values $\leq 7 \times 10^{-4}$; ** indicate p-values $\leq 1 \times 10^{-6}$; *** indicate p-values $\leq 4 \times 10^{-12}$

10A.



10B.

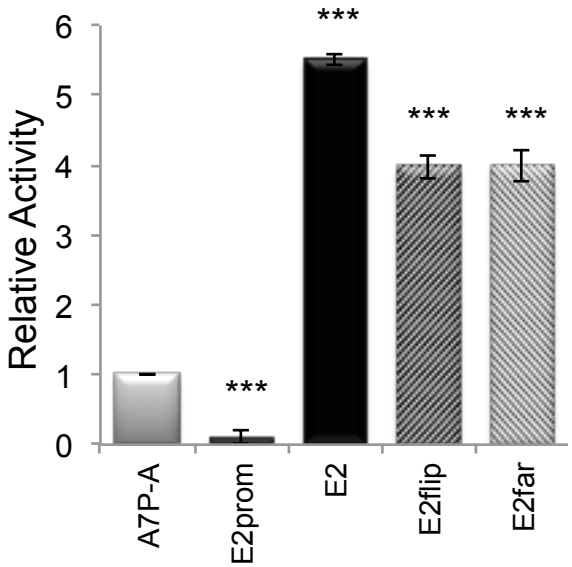


Figure 10: Upstream fragment 7P10-E2 acts as an enhancer. A) Map of the pXP2 vector construct with E2 or E2flip (E2 in reverse orientation) cloned into the HindIII and XhoI sites immediately upstream of A7P drawn to scale. E2far contains E2 cloned into

the ApaI and SwaI sites 3.5 kb away from the promoter to test for position effects. **B)** 7P10-E2 tested for promoter activity (E2prom) and enhancer properties, orientation independence (E2flip) and position independence (E2far), by transient transfections in CP-A cells. Activity was measured relative to A7P-A. Error bars represent standard error of means. P-values reflect the differences between the normalized luciferase activity of each construct and A7P-A, and are all $\leq 1 \times 10^{-9}$ indicated by ***

3. DISCUSSION

We have identified elements regulating *ADH7* promoter activity in physiologically relevant cell systems, and observed functional differences in the activities of naturally occurring haplotypes of both the promoter and upstream regulatory elements. There are cell-specific differences, and also differences in how the upstream sequences affect the promoter with the A allele at rs2851028 vs. the G allele. In both CP-A and MEF cells, which express *ADH7* endogenously, some fragments increase activity of the promoter with the A allele more than they do the promoter with the G allele, and others increase the A-promoter while decreasing activity of the G-promoter with the general trend favoring the former pattern. Most of the upstream fragments reduce promoter activity in HepG2 cells, which do not express endogenous *ADH7*, with a stronger effect on the promoter with the G allele. In general, regulatory elements have similar effects in CP-A and MEF cells, being particularly apparent in the effects of 7P10 sub-fragments in the two cell types, indicating that regulation is dependent on the status of *ADH7* expression. Our study of polymorphic fragments carrying opposite haplotypes of upstream sequences shows that variants have significant effects on regulatory function depending on cell type and DNA sequence.

The pattern of linkage disequilibrium among the upstream variants and those associated with diseases shows three distinct LD blocks **(Table 5)**. Variants in block 1, which contains the A7P promoter with rs2851028 and the upstream 7P5 fragment with rs1154473, are in strong LD ($r^2 \geq 0.8$, $D' \geq 0.9$) with variant rs1154458, associated with alcoholism (Osier et al., 2004, Han et al., 2005), rs2654849, associated with age of onset

of regular alcohol use (van Beek JH, 2010) and rs1154460, associated with upper aero-digestive tract (UAT) cancers (Hakenewerth et al., 2011, Oze et al., 2009). The more active allele of the promoter, rs2851028-G, is in LD with the C allele of rs1154458, which protects against alcoholism (Osier et al., 2004, Han et al., 2005). This is analogous to the faster-metabolizing isoforms of *ADH1B* and *ADH1C*, which have protective effects on risk for alcoholism and excessive drinking (Chen et al., 1999, Thomasson et al., 1993, Edenberg and Foroud, 2013, Hurley and Edenberg, 2012, Bierut et al., 2012). Though the 7P5 haplotypes have similar effects relative to each promoter in CP-A cells, there is a potential synergistic interaction between the two 7P5-GCGT haplotype and rs2851028-A in HepG2 cells. The minor G upstream allele of rs1154473 located in 7P5 and belonging to this LD block gains binding sites for several transcription factors including STAT4, NFκB, YY1 and RelA proteins.

rs1154460, also in block 1 **(Table 5)**, is associated with upper aerodigestive cancers, the risk increasing with increasing alcohol consumption (Hakenewerth et al., 2011). In this case, it is the less active A allele of rs2851028 that is in LD with the risk allele for cancer, rs1154460-A; the cancer risk increased with increasing alcohol consumption (Hakenewerth et al., 2011, Oze et al., 2009). The direction of affect is similar to the association of the less active *ADH1C*2* with cancer, with the risk increasing additively with alcohol consumption (Peters et al., 2005, Bongaerts et al., 2011, Xue et al., 2012). Retinoic acid and its precursor retinol are anti-carcinogens (Siddikuzzaman et al., 2011) and signaling molecules (Rhinn and Dollé, 2012) that are important for the maintenance of epithelial tissues (Osanai et al., 2007, Everts, 2012) including those expressing *ADH7*.

Although there are other retinol dehydrogenases, ADH7 has the greatest efficiency for oxidation of retinol to retinaldehyde (Parés et al., 2008) and an important, non-redundant role to play in retinoic acid synthesis. The tissue-specific expression of *ADH7,* the role of RA as an anti-carcinogen and in maintenance of these tissues, and the role of ADH7 in the retinol metabolism suggest that the inhibition of this RA biosynthesis pathway can increase risk for cancer in the tissues that express *ADH7.* This is highlighted both by morphological changes including inflammation and intestinal metaplasia in the gastric mucosa when ADH7 activity is reduced (Matsumoto et al., 2005) and by the severe growth and survival defects when $Adh7^{-/-}$ mice are raised on a vitamin-A deficient diet (Deltour et al., 1999). By influencing ADH7 protein levels, the less active A7P-A promoter can slow retinol metabolism relative to the more active A7P-G promoter in the presence of ethanol due to the competition for the catalytic site between the two substrates, particularly at high ethanol concentrations. This can in turn disrupt RA functions including cell differentiation and growth arrest leading to increased risk for cancer **(Figure 2)**. The risk for cancer increasing with increasing alcohol consumption supports this possibility. The promoter A7P can bind transcription factors C/EBP, CREB and AP-1 (Kotagiri and Edenberg, 1998) and has consensus binding sites for others including HNF1, Oct1, FOXP3 and GRalpha transcription factors (Myers et al., Messeguer et al., 2002, Farré et al., 2003). A7P-G loses the HNF1 binding sites, but gains IRF-2 transcription factor binding site supporting the possibility of allele-specific regulatory protein binding that may be responsible for difference in activity.

LD block	SNPs in tested fragments	Tested Fragments	Disease phenotype and associated SNPs
1	rs2851028 rs1154473	A7P 7P5	**rs1154458**- Protection against alcoholism (Han et al., 2005, Osier et al., 2004) **rs2654849**- Age at onset of regular alcohol use (van Beek JH, 2010) **rs1154460**- UAT cancers (Oze et al., 2009, Hakenewerth et al., 2011)
2	rs1154474 rs1154475 rs1154476 rs1154477 rs1154480 rs749407	7P5 7P5 7P5 7P6 7P8 7P10	**rs1154461**- Alcohol metabolism (Birley et al., 2009, Birley et al., 2008) **rs1154468**- Alcohol metabolism **rs894363**- Alcohol metabolism **rs1154470**- Alcohol metabolism; Extraversion and conscientiousness in Substance-dependent subjects (Luo et al., 2008)
3	rs17589306	7P10	**rs971074**- Drug dependence & heroin addiction (Luo et al., 2007, Levran et al., 2009), **rs1573496**- UAT cancer (Hashibe et al., 2008, McKay et al., 2011, Wei et al., 2010, Cadoni and Pandolfini, 2012)

Table 5: Upstream SNPs in LD with disease associated SNPs. Upstream SNPs in LD with disease associated SNPs and the test fragments containing them are listed**.** SNPs with $r^2 \geq 0.8$, $D' \geq 0.9$ are defined as within an LD block. The disease phenotypes associated with each SNP are listed, along with the references.

Block 2 contains 6 upstream SNPs in fragments 7P5, 7P6, 7P8 and 7P10, located between -4667 bp to –12578 bp **(Table 5)**. These upstream fragments affect regulatory function in a cell- and promoter haplotype-dependent manner. The minor alleles of these SNPs are in LD with the minor alleles of SNPs associated with the early stages of alcohol metabolism (Birley et al., 2008, Birley et al., 2009) but have no association with alcoholism or cancer. This suggests that different *cis*-sequences and possibly different mechanisms underlie the various disease phenotypes.

The third LD block comprises the 7P10 enhancer SNP rs17589306, in LD with two SNPs associated with cancer **(Table 5)** (Hashibe et al., 2008, Wei et al., 2010, McKay et al., 2011, Cadoni and Pandolfini, 2012) and drug dependence (Levran et al., 2009, Luo et al., 2007) but not, thus far, with alcoholism or alcohol metabolism (Birley et al., 2008). One of these, rs1573496, is a coding SNP associated with upper aerodigestive cancer (Hashibe et al., 2008, McKay et al., 2011, Wei et al., 2010, Cadoni and Pandolfini, 2012); it is in complete LD with the enhancer SNP rs17589306. rs971074, a synonymous SNP, is associated with drug dependence (Luo et al., 2007, Levran et al., 2009). The other non-synonymous coding polymorphism rs59534319, 30 bp away from rs971074, is uncommon to rare in most populations, and no associations have yet been reported; it is, however, possible that it contributes to the reported associations in this linkage block. Thus the effects of a particular regulatory polymorphism are complex, and depend upon the sequence context of other polymorphisms and on cell type. This can complicate interpretation of association data.

7P10-E2 is DNaseI hypersensitive in many cell lines (Sandstrom et al.). It was shown to bind various transcription factors (Myers et al.), including the NFkB and AP-1 superfamily of transcription factors, in cells having no endogenous ADH7. The AP-1 family related Maf (v-maf avian Musculoaponeurotic fibrosarcoma oncogene homolog) proteins, that bind 7P10-E2 in some cell lines, are frequently associated with enhancer elements and are expressed in the esophagus (Motohashi et al., 2004, Chen et al., 2012) (where *ADH7* is expressed). The peaks of these binding sites coincide with a CpG element that is methylated in several cell lines, including HepG2 (Myers and Absher, Costello et al.), suggesting that 7P10-E2 CpG methylation may influence regulatory protein binding. The histone mark H3K4me1 is enriched in this region in many ENCODE cell lines (Bernstein, The Encode Project, 2011), but not H3K27ac (Bernstein, The Encode Project, 2011). H3K27ac distinguishes active enhancers from poised or inactive enhancers that contain the H3K4me1 mark alone (Creyghton et al., 2010), indicating a poised enhancer at 7P10-E2 that is activated in cells expressing *ADH7* including CP-A cells. 7P5 can bind CTCF in some cell lines (Myers et al., The Encode Project, 2011) as well as NF-κB and JunD transcription factors. Since the cells tested by ENCODE do not express *ADH7* endogenously, the data may represent a poised enhancer in 7P10-E2.

Our study suggests that variants affecting *ADH7* gene regulation show combinatorial and cell-specific interactions. These findings highlight the complexity of interpreting the effects of individual SNPs, because they are dependent upon other SNPs that are in *cis*. Detailed analyses of haplotypes that are associated with alcohol-related traits will be

needed to identify which SNPs are functional and the consequences of different combinations of SNPs.

IV. CHAPTER II

AN ENHANCER- BLOCKING ELEMENT REGULATES THE CELL-SPECIFIC EXPRESSION OF ALCOHOL DEHYDROGENASE 7

1. INTRODUCTION

ADH7 is unique among the *ADHs* in being expressed mainly in the esophagus and gastric mucosa but not in the liver, the primary site of expression of the other 6 ADHs (Engeland and Maret, 1993, Kedishvili et al., 1995). Regulatory elements extending up to -799 bp are active in HeLa, CV-1 monkey kidney, and H4IIE3 rat liver cells (Kotagiri and Edenberg, 1998). The promoter and other regulatory elements in the *ADH7* proximal region extending 12.5 kb upstream from the translation start site are also active in *ADH7* expressing CP-A, MEF and HepG2 cells, although there are cell-specific differences in the level of transcriptional activity (Chapter I) (Jairam and Edenberg, 2014). This suggests that other, more distant factors contribute to the cell-specific expression of *ADH7*.

Enhancers can activate promoters over long distances (Bulger and Groudine, 2011, Chin-Tong and Victor, 2011, Arnosti and Kulkarni, 2005). Several *ADH* enhancers have been identified that activate more than one *ADH* promoter including the potent FOXA-dependent liver-specific enhancer located upstream of *ADH4* (Pochareddy and Edenberg, 2010) and the class I *ADH* HNF1-bound enhancer, located 10 kb downstream of *ADH7* in the 60 kb intergenic region between *ADH1C* and *ADH7* (Su et al., 2006). Yet, despite the presence and proximity of functional *ADH* enhancers, *ADH7* is not expressed in the liver,

suggesting the possibility of an element that blocks the effect of these enhancers on *ADH7.* Chromatin insulators are known to exhibit this enhancer blocking activity, and all known vertebrate insulators bind the CCCTC binding factor CTCF. CTCF binding is, therefore, a candidate for enhancer blocking insulator function in the *ADH* locus.

Bioinformatics analyses suggested that a region between the *ADH* class I enhancer and *ADH7* is potentially an insulator element that can prevent the HNF1-bound enhancer from activating *ADH7*. We have analyzed the function of this region in cell systems that replicate the cell-specific pattern of endogenous *ADH7* expression observed *in vivo:* CP-A human esophageal cells and MEF (mouse embryonic fibroblast) cells that express *ADH7* and HepG2 human hepatoma cells that do not. We have demonstrated that this region can function as an insulator element that can contribute to cell-specific *ADH7* expression.

2. RESULTS

A. Identification of iA1C, an enhancer blocker

Since vertebrate insulator elements containing enhancer-blocking activity are typically associated with the CCCTC binding factor (CTCF), we searched the insulatordb (Bao et al., 2008) database for potential CTCF binding sites in the vicinity of *ADH7* and found a 965 bp sequence we called iA1C in the intergenic region between the enhancer and *ADH7* **(Figure 11)**

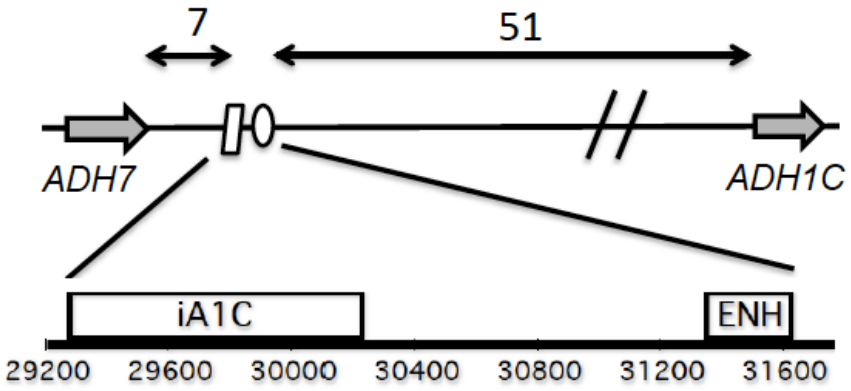


Figure 11: Location of the *ADH* class I enhancer (ENH) and iA1C. Top line is the part of chromosome 4 containing ADH7 and ADH1C; distances are in kb. Below, iA1C and ENH are expanded; distances (bp) are measured from the ADH7 translation start site (TSS, +1). iA1C is the sequence from 29277:30222.

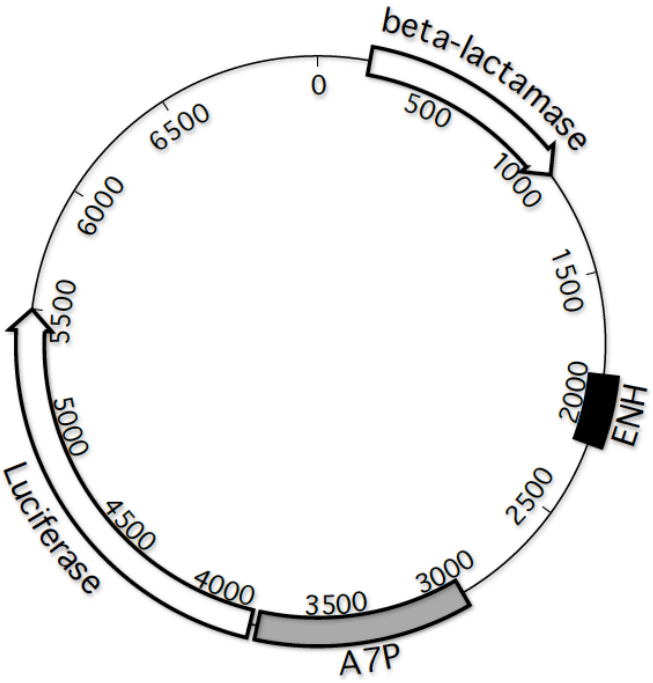


Figure 12: **Map of the pXP2 plasmid construct used for insulator assays.** The enhancer was cloned 1 kb upstream of the *ADH7* promoter A7P. Test and control fragments were cloned into sites on either side of ENH. Map is drawn to scale, distances in bp.

B. iA1C is a cell- and position-dependent insulator

To test the potential enhancer blocking activity of the iA1C fragment, we cloned it upstream of the *ADH7* promoter (A7P, with allele A at the functional SNP rs2851028 (Jairam and Edenberg, 2014) on either or both sides of the enhancer ENH **(Figure 12).** Constructs were tested for effect on enhancer function in MEF and CP-A cells that express endogenous *ADH7* (Chapter I*)* and in HepG2 cells that do not. ENH increased A7P promoter activity 2.2 fold in MEF cells, 2.4 fold in CP-A cells and 3.4 fold in HepG2 cells. iA1C significantly increased transcriptional activity of the enhancer-

containing plasmids independent of its placement with respect to ENH in both CP-A and MEF cells though the effects were more pronounced in MEF cells **(Figure 13).**

In HepG2 cells, iA1C caused a 60% reduction in enhancer activity when placed between ENH and A7P. When placed outside the enhancer, iA1C had a significant but much smaller effect on activity. Flanking ENH by iA1C on both sides had a dramatic and nearly complete block on ENH function, with the luciferase expression of the flanking construct only slightly higher (1.17 fold) than that of the promoter construct alone **(Figure 13).**

When tested in combination with the known insulator element XL9 (that binds CTCF and regulates expression of the *HLA-DRB1* and *HLA-DQA1* genes (Majumder et al., 2008)), iA1C showed a similar cell- and position dependent behavior with no enhancer blocking function in CP-A and MEF cells (that express *ADH7)* **(Figure 14)**. In HepG2 cells, iA1C placed between ENH and the promoter blocked enhancer activity by 50% when combined with XL9 on the other side of ENH. The construct with the reverse placement of iA1C and XL9 (XL9-ENH-iA1C-A7P) affected A7P activity by 2.4 fold with a 40% reduction in ENH activity, similar to the effect seen with XL9 flanking ENH on both sides **(Figure 14)**. Thus, iA1C had enhancer-blocking activity in HepG2 cells, and this activity was cell-specific and dependent on the placement of iA1C with respect to the enhancer.

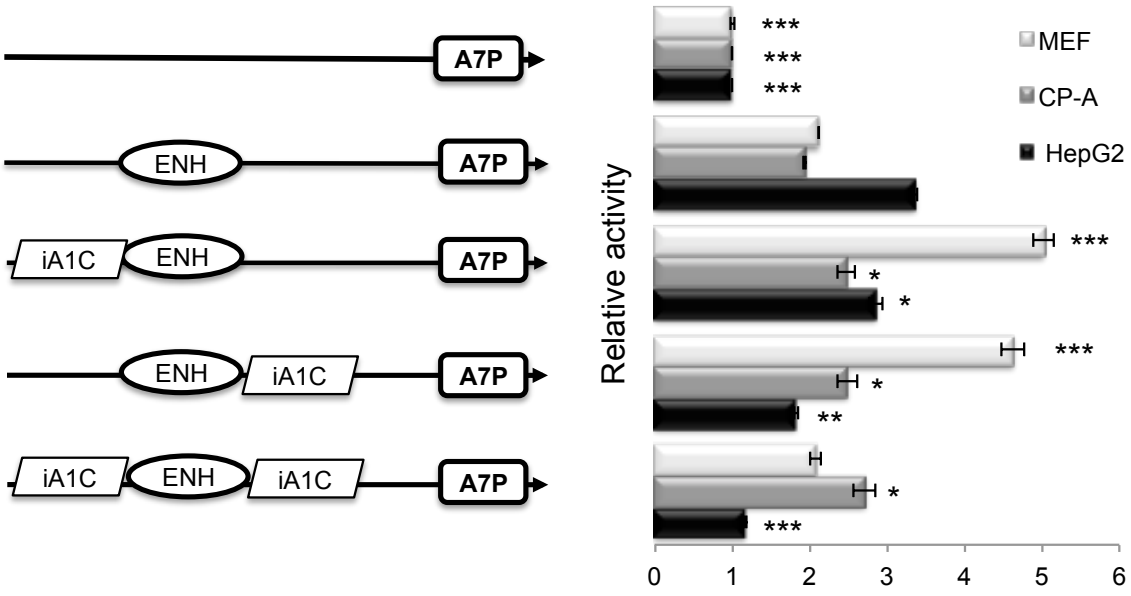


Figure 13: iA1C is an insulator with cell-specific enhancer blocking activity. Effects of iA1C in MEF, CP-A and HepG2 cells tested using constructs with iA1C cloned on either or both sides of ENH as depicted were transiently transfected into the cells. Relative activities were the ratios of normalized luciferase activity of each construct to that of the promoter (A7P) construct shown at top in the same cell line. Error bars indicate standard errors of the mean. P-values reflect the differences between the normalized luciferase activity of each construct and the enhancer construct in corresponding cell type. * indicate p-values ≤ 0.006; ** indicate p-values $\leq 5 \times 10^{-6}$; *** indicate p-values $\leq 8 \times 10^{-8}$

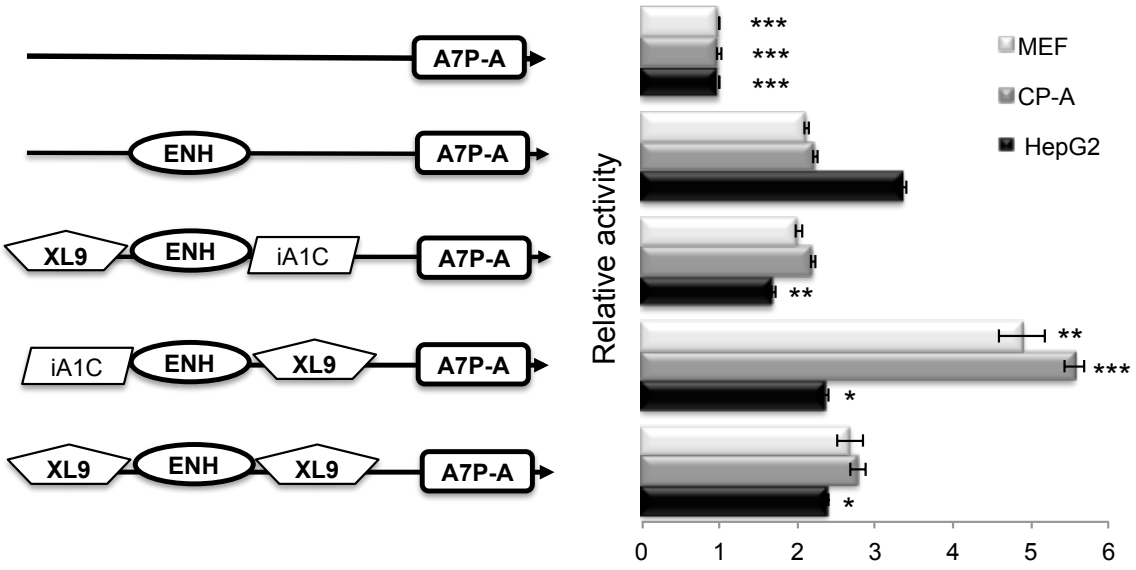


Figure 14: iA1C function with XL9 insulator. Constructs with test iA1C and control XL9 fragments cloned on either or both sides of ENH as depicted were transiently transfected in MEF, CP-A and HepG2 cells. Luciferase activities were calculated by normalizing to the internal control β -galactosidase. Relative activities were determined by calculating the ratios of normalized luciferase activity of each construct to that of the promoter A7P construct. Error bars indicate standard errors of the mean. P-values reflect the differences between the normalized luciferase activity of each construct and the enhancer construct in corresponding cell type. * indicate p-values $\leq 5 \times 10^{-4}$; ** indicate p-values $\leq 1 \times 10^{-6}$; *** indicate p-values $\leq 4 \times 10^{-8}$

C. Sub-fragmentation of iA1C to localize insulator function

To localize the enhancer blocking activity, iA1C was fragmented into two overlapping shorter fragments of 630 bp and 395 bp, and tested for function. Transient transfections in HepG2 cells showed that enhancer activity was reduced by about 75% when flanked by either sub-fragment **(Figure 15)**. The full- length iA1C was a more potent enhancer blocker than its sub-fragments, nearly completely eliminating enhancer effect.

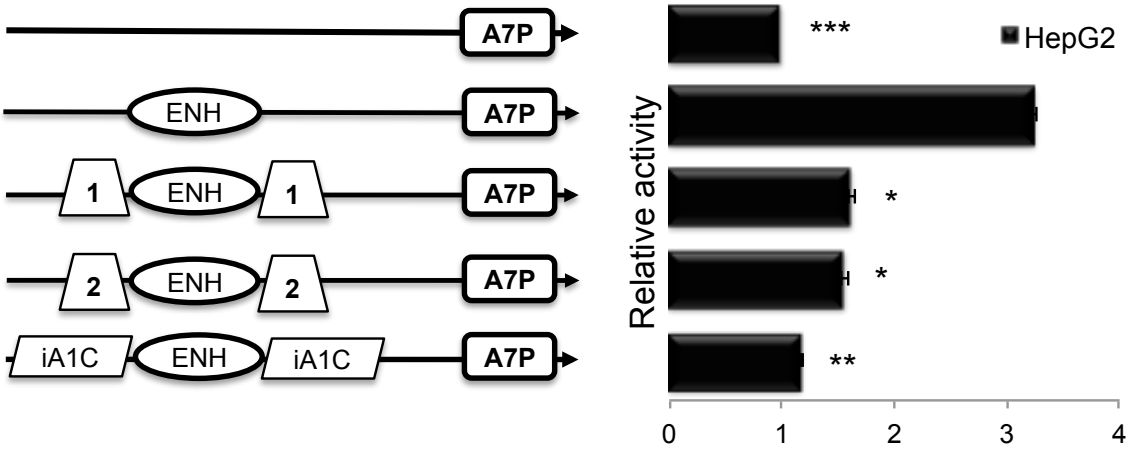
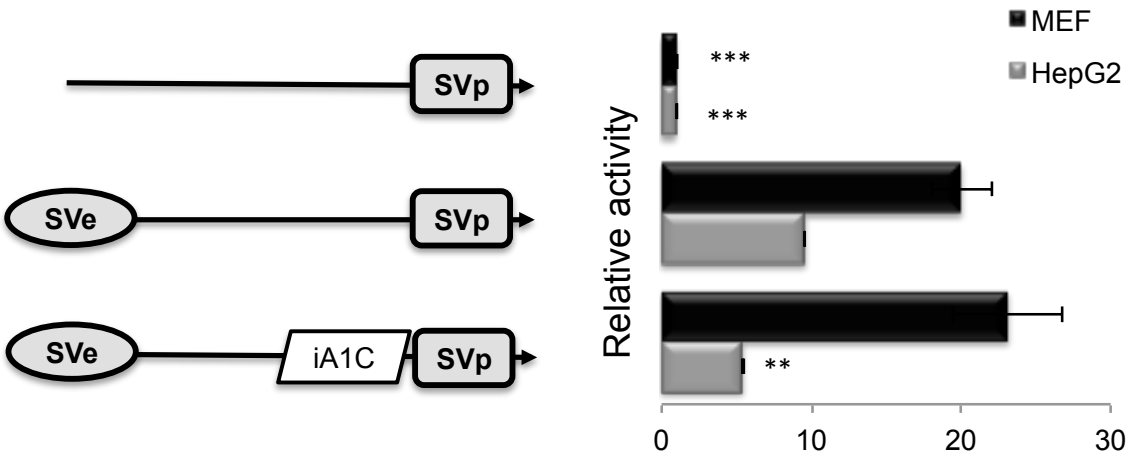


Figure 15: **Effects of iA1C sub-fragments iA1C-1 and iA1C-2 in HepG2 cells relative to full length iA1C.** The sub-fragments iA1C-1 and iA1C-2 contain the sequences from 29277: 29917 and 29918:30222 with respect to ADH7 TSS. Relative activities in HepG2 cells were determined as the ratio of normalized luciferase activity of each construct to that of the promoter vector. Standard errors of mean are shown. . P-values reflect the differences between the normalized luciferase activity of each construct and the enhancer construct. . * indicate p-values $\leq 1 \times 10^{-6}$; ** indicate p-values $\leq 8 \times 10^{-9}$; *** indicate p-values $\leq 3 \times 10^{-10}$

D. iA1C function with heterologous enhancer and promoter

The pGL3 control vector containing the SV40 enhancer and SV40 promoter driving luciferase expression (Groskreutz et al., 1995) was used to study the function of iA1C with heterologous elements. Since the SV40 promoter and enhancer typically work well in nearly all cell types, they represent a great heterologous system, particularly to study cell-specificity. The SV40 enhancer increased SV40 promoter activity by 180 fold in CP-A cells and approximately 10 fold in MEF and HepG2 cells, indicating that both the SV40 enhancer and promoter were functional in our cell systems **(Figure 16)**. In both CP-A and MEF cells, the transcriptional activity of the construct with iA1C cloned immediately upstream of the SV40 promoter was similar to the transcriptional activity of pGL3 control vector, indicating no effect of iA1C on enhancer function. In HepG2 cells however, iA1C decreased SV40 enhancer activity by 40% **(Figure 16).** Thus, iA1C exhibits similar behavior independent of the enhancer and promoter elements, but dependent on the cellular context and status of endogenous *ADH7* expression.

16A)



16B)

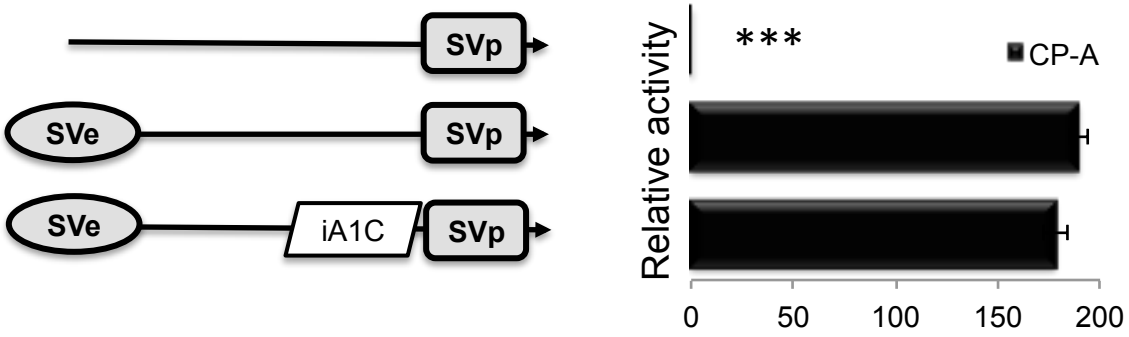


Figure 16: iA1C functions as a cell specific insulator with heterologous enhancer and promoter. Transient transfections in A) MEF and HepG2 cells and B) CP-A cells, with pGL3 promoter vector (containing the SV40 promoter SVp), pGL3 control vector (containing SVp and the SV40 enhancer SVe), and iA1C cloned between SVp and SVe in the pGL3 control vector. Relative activities represent the ratio of normalized luciferase activity of each construct to that of the pGL3 promoter vector. P-values are the differences between the normalized luciferase activity of each construct and the SVe construct in the corresponding cell type. p-values $\leq 2.5 \times 10^{-5}$ are indicated by **; p-values $\leq 1 \times 10^{-9}$ are indicated by ***

E. iA1C variants affect function

Variants of regulatory elements can affect function. To study the effects of iA1C variants on its enhancer blocking activity, three constructs each containing a naturally occurring iA1C haplotype **(Table 6)** positioned between SV40 enhancer and promoter were tested by transient transfections in HepG2 cells. The enhancer increased SV40 promoter activity 10-fold in HepG2 cells; the AC haplotype of iA1C used in earlier experiments reduced enhancer activity by 44% **(Figure 17)**. We tested two additional haplotypes of iA1C **(Table 6)** in this system. The iA1C-GC haplotype had a similar effect while iA1C-GT had a slightly (but significantly) greater effect, reducing enhancer activity by 50% **(Figure 17).**

rs1442490	**rs1442489**	**Haplotype frequency**	**DNA source**
A	**C**	30.8	NA07000
G	T	63.4	NA12248
G	C	5.8	NA12248

Table 6: Variants of iA1C tested for effects on function. The alleles of the two SNPs making up the naturally occurring haplotypes of iA1C, their haplotype frequencies in Northern and Western European populations from Utah (CEU) and the DNA templates from Coriell Institute for Medical Research used for obtaining the sequences are listed.

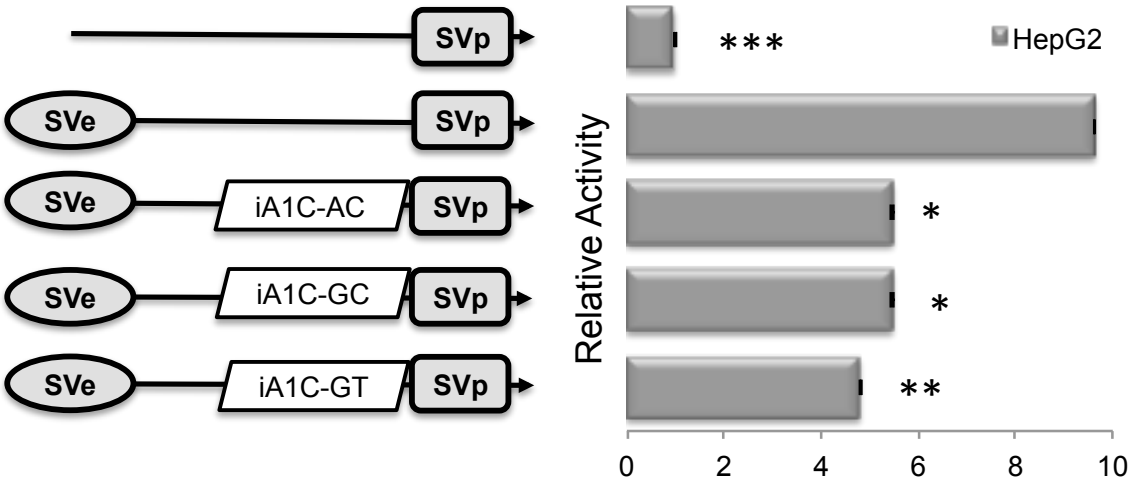


Figure 17: iA1C variants have a significant but slight effect on its enhancer blocking activity. Effects of different naturally occurring haplotypes of iA1C tested with the heterologous SV40 elements in HepG2 cells. iA1C-AC is the haplotype used in earlier assays. Standard errors of mean are shown. P-values represent the differences between the normalized luciferase activity of each construct and SVe construct containing SV40 enhancer and are indicated by * for values$\leq 5 \times 10^{-5}$; ** for values$\leq 1.75 \times 10^{-6}$; *** for values$= 1 \times 10^{-9}$

F. iA1C binding to CTCF *in vitro*

To test binding of the vertebrate insulator binding protein CTCF to iA1C in HepG2 cells *in vitro*, Electrophoretic Mobility Shift Assays (EMSA) were used. EMSA assays with iA1C-O, a 40 bp FAM-labeled oligonucleotide within iA1C **(Figure 18)** identified from the insulatordb database as a potential CTCF binding site, showed shifts in the binding when incubated with HepG2 nuclear extract **(Figure 19; lane 2)**.

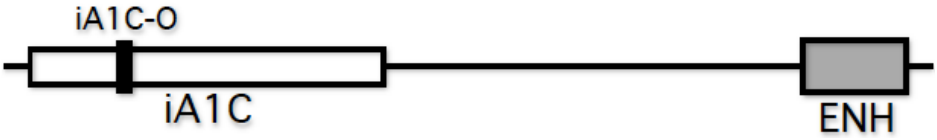


Figure 18: A depiction of iA1C-O (Table 9), the 40 bp binding site predicted to bind CTCF. The location relative to full length iA1C and the enhancer ENH is shown. The map is drawn to scale with iA1C-O represented as a black box within iA1C.

Competition assays with unlabeled oligonucleotides showed that loss of shift occurred with addition of increasing amount (5X and 10X) of specific competitor oligos (cold iA1C-O; lanes 8-9) or CTCF consensus oligos (CTCF-CON; lanes 10-11), but not with the non-specific competitor (NS; lane 3) indicating the binding was specific. Competition with oligonucleotides containing the iA1C sequence shuffled to disrupt the consensus CTCF binding site (iA1C-Osh; lanes 4-7) resulted in loss of both specific and nonspecific bands **(Figure 19)**. However, addition of the Anti-CTCF antibody did not result in a supershift **(Figure 20; lane 3)** because the antibody directly bound the target oligo **(Figure 20; lane 11)**. Similar direct interaction between CTCF antibody and positive control β-globin oligonucleotide that is known to bind CTCF was observed resulting in inconclusive evidence of whether or not CTCF binds to iA1C-O **(Figure 20; lane 12)**.

Labeled iA1C-O	+	+	+	+	+	+	+	+	+	+	+
HepG2 NE		+	+	+	+	+	+	+	+	+	+
NS			25X								
iA1C-Osh				5X	10X	25X	40X				
Cold iA1C-O								5X	10X		
CTCF-CON										5X	10X

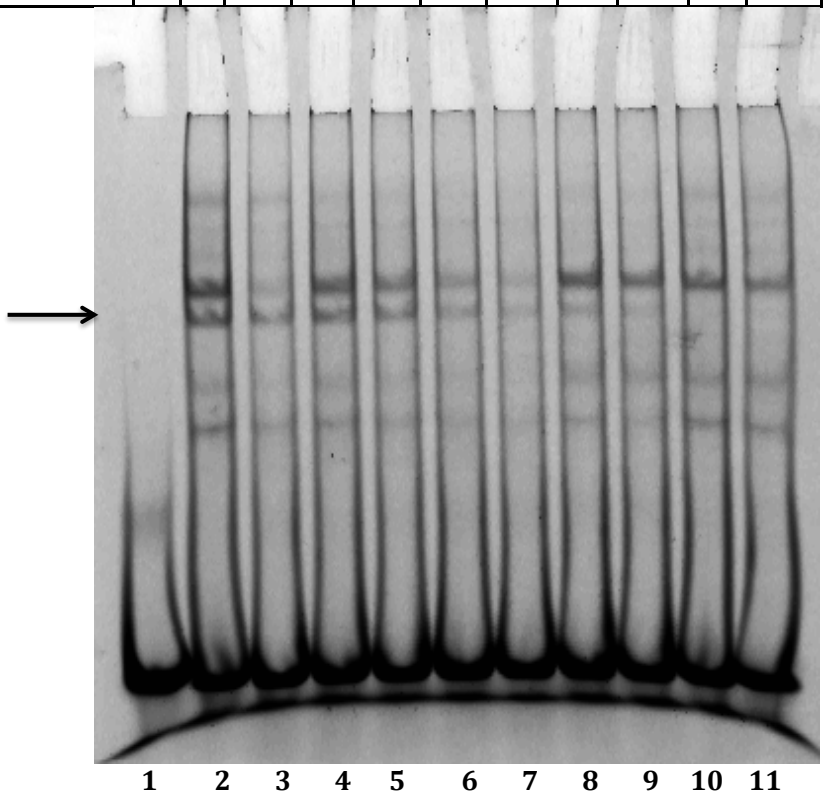


Figure 19: Electrophoretic mobility shift assays (EMSA) to study protein binding to iA1C-O *in vitro*. Assays were done using labeled iA1C 40 bp oligonucleotide probe (iA1C-O) which was predicted to bind CTCF and HepG2 nuclear extract (HepG2 NE). Competition assays were done with nonspecific control (NS), iA1C-O shuffled to disrupt the consensus CTCF binding site (iA1C-Osh), cold iA1C-O and CTCF consensus oligo (CTCF-CON). An arrow points toward the specific band corresponding to iA1C binding to the HepG2 nuclear extract. Lanes are listed at the bottom.

Labeled iA1C-O	+	+	+	+	+	+					+	
Labeled β-g							+	+	+	+		+
HepG2 NE		+	+	+	+	+	+	+	+	+		
Cold iA1C-O (10X)				+								
iA1C-Osh (10X)					+							
CTCF-CON (10X)						+				+		
β-globin (10X)									+			
Anti-CTCF			+					+			+	+

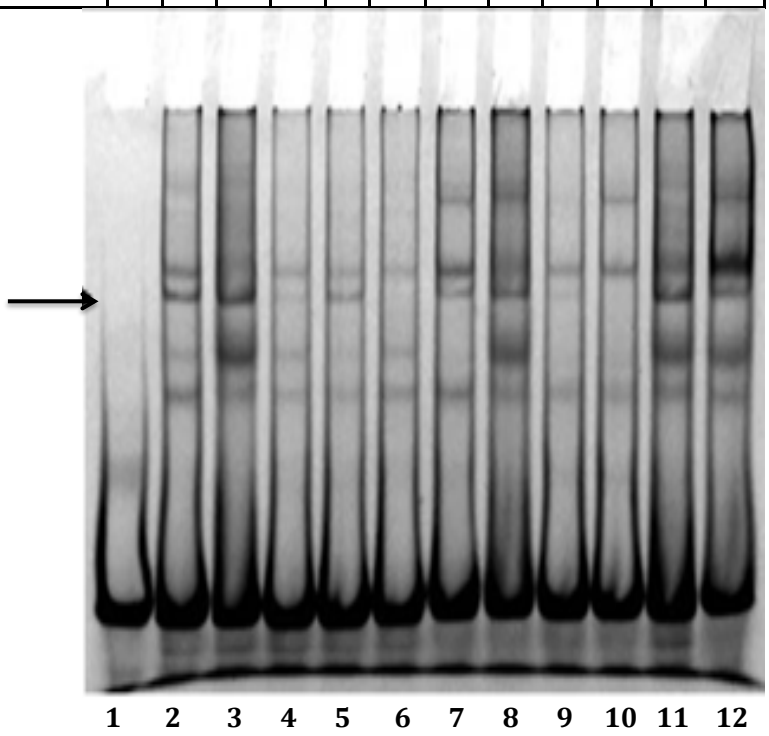


Figure 20: Supershift assay with Anti-CTCF. To confirm binding, supershift assays were done using CTCF antibody and the iA1C-O. Controls included β-globin labeled probe and unlabeled cold probes (iA1C-O; iA1C-Osh=shuffled iA1C; CTCF-CON=CTCF consensus oligo and β-globin). iA1C and β-globin labeled probes were incubated with Anti-CTCF alone (for control) or along with the HepG2 nuclear extracts. The components of each lane are shown. An arrow points toward the specific band corresponding to iA1C and β-globin binding to the HepG2 nuclear extract.

G. iA1C-O, the 40 bp sequence identified by insulatordb is not required for insulator activity

To test if the 40 bp sequence iA1C-O contained insulator activity, transient transfections were done with the wild type and mutated oligonucleotides (used in the EMSA assays) cloned on either or both sides of the enhancer ENH into A7P promoter construct. The mutations were done so that the sequence no longer had the CTCF consensus-binding site. Transient transfections showed that neither the wild type nor the mutated oligo sequences had insulator activity, instead up-regulating ENH activity **(Figure 21).** Similar experiments with the full-length iA1C sequence mutated via site-directed mutagenesis (iA1C-SH) **(Figure 22)** to no longer contain the predicted CTCF consensus-binding site showed no difference in the enhancer blocking activity when compared to wild type full length iA1C.

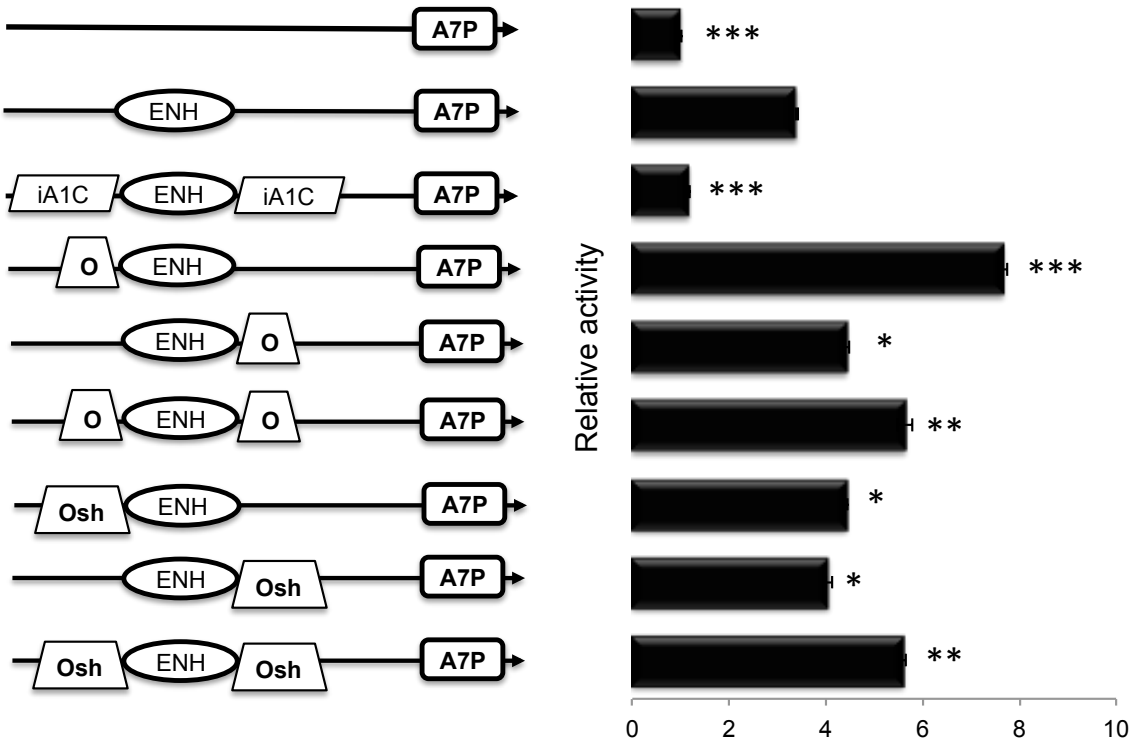


Figure 21: iA1C-O does not contain enhancer blocking activity in HepG2 cells.

Transient transfections in HepG2 cells with wild type iA1C oligonucleotide and mutated oligonucleotide sequences to test enhancer blocking function of the sequence predicted to bind CTCF *in silico*. Constructs with the wild type (O) or shuffled oligos (Osh) used in the EMSA assays cloned on either or both sides of ENH were tested, and none of them show enhancer blocking activity. Relative activities represent the ratio of normalized luciferase activity of each construct to that of the A7P promoter vector. T-tests for significance of the differences between the normalized luciferase activities of each construct and the enhancer construct ENH were carried out: * indicate p-values ≤ 0.08; ** indicate p-values $\leq 9 \times 10^{-4}$; *** indicate p-values $\leq 1 \times 10^{-8}$

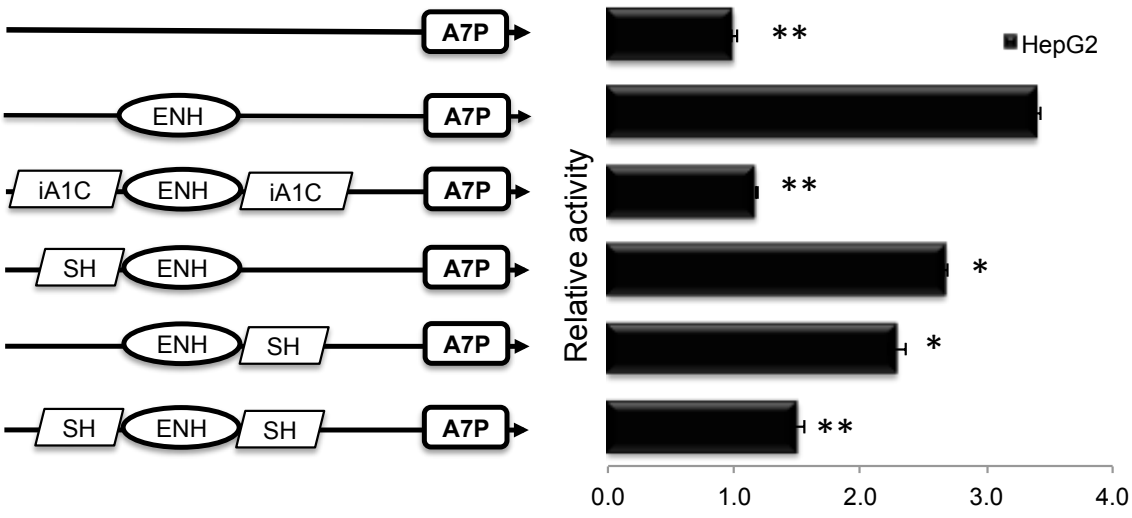


Figure 22: Full length iA1C shuffled to disrupt the predicted CTCF binding site does not affect insulator function. Transient transfections in HepG2 cells with full length iA1C mutated to disrupt the consensus CTCF binding site cloned on either or both sides of the enhancer ENH. In this depiction, SH represents the full length sequence mutated to no longer contain the CTCF consensus site predicted *in silico*, instead carrying the shuffled 40 bp sequence used in EMSA assays. Relative activities represent the ratio of normalized luciferase activity of each construct to that of the A7P promoter vector. T-tests were done for significant differences between the normalized luciferase activity of each construct and the enhancer construct ENH. * indicate p-values $\leq 8 \times 10^{-4}$; ** indicate p-values $\leq 1 \times 10^{-8}$

H. CTCF binds multiple sites within iA1C *in vivo*

Since chromatin immunoprecipitation assays can detect *in vivo* binding and do not have some of the limitations of EMSA assays, ChIP assays were performed to identify any CTCF binding sites within iA1C. CTCF binding and enrichment of the target sequences in the immunoprecipitated samples are represented as yield, which are the ratio of signals obtained from ChIP to signals from a 4% input sample (the non-immunoprecipitated chromatin sample and serves as a control and as a measure of starting DNA material relative to which the immunoprecipitated chromatin can be measured)**.** Of the eleven iA1C sub-fragments **(Figure 24A)** tested for CTCF binding by ChIP, the greatest binding was seen for iA1C-D, with the adjacent sub-fragment iA1C-C also showing high yield **(Figure 24B)**. Another spike in yield is seen for iA1C-J, indicating multiple binding sites at or near the ends of iA1C in HepG2 cells. A random sequence from exon 3 of RPL30 housekeeping gene does not bind CTCF in HepG2 cells, indicating the binding is specific. RPL30 binds the positive control histone H3 antibody, but not IgG confirming the validity of RPL30 as a control **(Figure 23).**

In CP-A cells however, iA1C did not bind CTCF **(Figure 24B).** ChIP assays with positive control H3 and negative control IgG and the target sequence RPL30 indicate the ChIP assays worked as expected. Thus, there are significantly different binding patterns of CTCF to iA1C sub-sequences in HepG2 and CP-A cells.

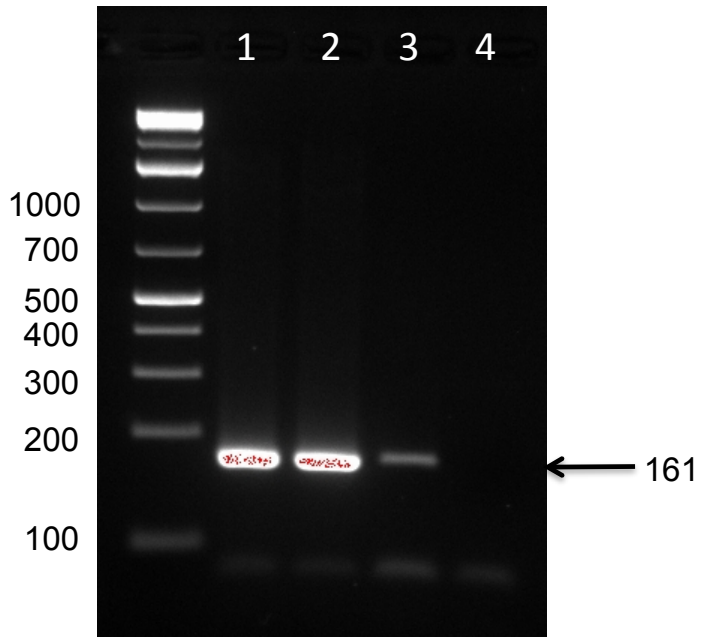
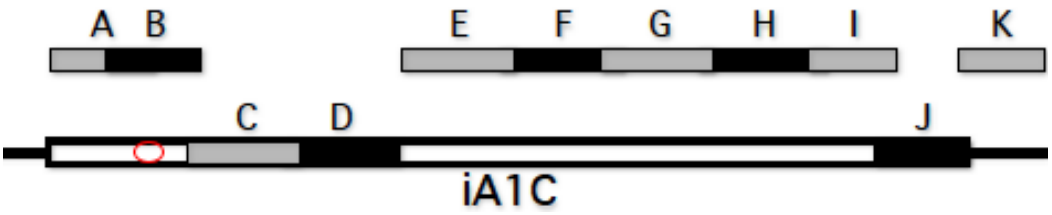


Figure 23: Semi-quantitative ChIP analysis of CTCF binding to controls using a 1% agarose gel run. Each of the four lanes shows PCR amplification of the control human exon 3 RPL30 sequence from different templates. Product from Lane 1) contains the 2% input sample as template, 2) positive control H3 and 3) negative control rabbit. Lane 4) shows the no DNA control for PCR. The molecular weights of bands up to 1000 bp are listed in bp on left side of the panel. On the right side, an arrow points toward the band corresponding to RPL30 amplicon of size 161 bp.

24A)



24B)

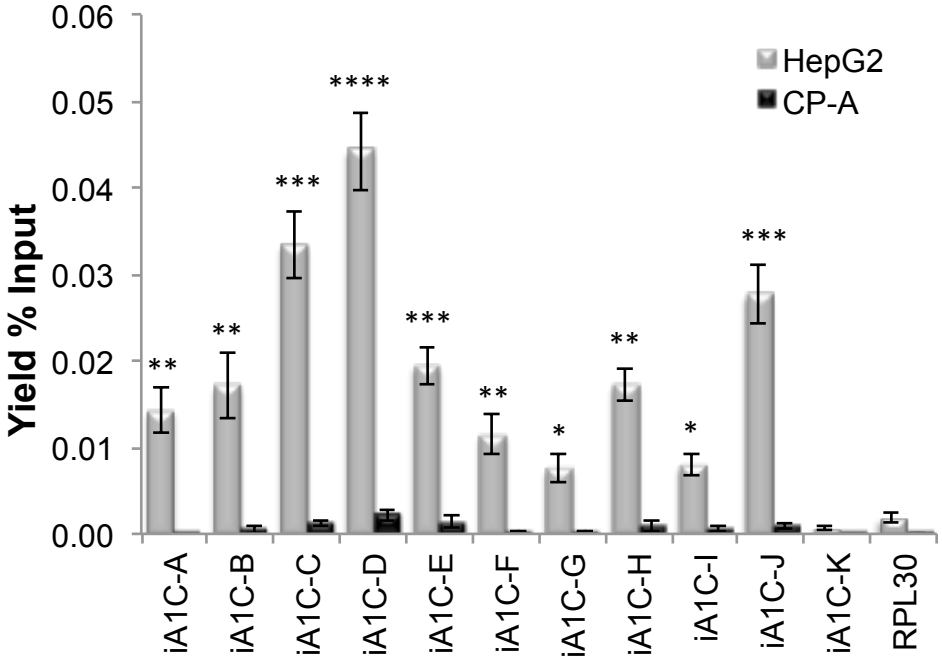


Figure 24: CTCF binds multiple sites within iA1C *in vivo*, but only in HepG2 cells.

A) iA1C subsequences used in Chromatin immunoprecipitation (ChIP) assays. iA1C- A to J are 100-150 bp sequences encompassing the full length of iA1C. K is immediately downstream of iA1C and serves as a control, along with human RPL30. The targets C, D and J with the greatest enrichment for CTCF in HepG2 cells are depicted on sequence, with other targets above sequence. iA1C-O, the 40 bp oligonucleotide sequence tested by

EMSA assays is shown as an oval within iA1C. **B)** CTCF binding to iA1C subsequences in HepG2 and CP-A cells, determined by ChIP. A random sequence in the housekeeping gene RPL30 serves as a control. Yields represent the ratio of signals obtained from ChIP to signals from a 4% input sample, which is the non-immunoprecipitated chromatin sample and serves as a control and as a measure of starting DNA material relative to which the immunoprecipitated chromatin can be measured. Means and standard errors from two or three biological replicates and at least 12 technical replicates in total are shown. Statistical significance between yields (2^ (Input Ct-IP Ct)) of each target and RPL30 control were calculated by t-tests. * indicate p-values ≤ 0.015; ** indicate p-values ≤ 0.002; *** indicate p-values $\leq 1 \times 10^{-5}$; **** indicate p-values $\leq 4 \times 10^{-7}$

I. CTCF knockdown is insufficient to induce *ADH7* in liver cells

Since iA1C is an insulator that cell specifically blocks activation of the *ADH7* promoter and binds CTCF with the same specificity, it is likely that CTCF plays a role in iA1C enhancer blocking activity and regulation of *ADH7* expression. To study the effect of CTCF knockdown on *ADH7* expression, HepG2 cells were infected with adenovirus encoding CTCF shRNA (shCTCF). Cells infected simultaneously with control shRNA (shctrl) served as reference. Samples from untreated HepG2 and CP-A cells served as controls. qPCR assays with cDNA amplified from RNA samples (extracted from the shCTCF or shctrl infected cells) showed that CTCF was highly expressed in cells treated with either shRNA **(Table 7)**, with a 2-fold reduction in CTCF expression in the shCTCF samples relative to shctrl samples. CTCF was still highly expressed even with the 50% decrease achieved by this CTCF shRNA **(Table 7)** and was insufficient to induce *ADH7* expression in HepG2 cells to a level that was detectable by qPCR assay. The control genes *H19* and *PUMA*, which are moderately expressed in HepG2 cells and are affected by changes in CTCF levels according to published literature (Gomes and Espinosa, 2010), showed 1-1.5 fold induction with CTCF knockdown.

cDNA sample	Target	Ct Mean	Stdev	2^ (shCTCF Ct-shctrl Ct)
shctrl	*CTCF*	25.54	0.48	
shCTCF	*CTCF*	26.72	0.19	2.26
shctrl	*ADH7*	39.36		
shCTCF	*ADH7*	-		-
shctrl	*H19*	32.90	0.20	
shCTCF	*H19*	33.47	0.31	1.48
shctrl	*PUMA*	28.71	0.20	
shCTCF	*PUMA*	28.78	0.07	1.05

Table 7: Changes in the levels of gene transcripts in HepG2 cells on CTCF knockdown assayed by qPCR. The knockdown achieved a 2 fold reduction in expression between the CTCF shRNA and control shRNA samples. ADH7 mRNA in control samples is virtually undetectable, whereas there is a very modest induction in H19 and PUMA mRNA levels with CTCF knockdown. shctrl and shCTCF represent cDNA from the control shRNA and CTCF shRNA infected samples respectively. Differential gene expression is measured by the value of 2^ (shCTCF Ct-shctrl Ct) and mean Ct values for each target with both cDNA samples is shown. Biological replicates=6. Technical replicates=24 for each target mRNA.

3. DISCUSSION

In this chapter, we identified an intergenic element, designated iA1C, that is located between the *ADH7* gene and the class I *ADH* specific enhancer ENH (Su et al., 2006) and blocks ENH activity in the *ADH7* promoter in HepG2 cells. The enhancer blocking activity was dependent on the placement of iA1C with respect to the enhancer and promoter, acting only when iA1C was placed between the two elements. This insulator function was specific to cell-type, with no enhancer blocking activity in *ADH7*-expressing CP-A cells or MEF cells. CTCF bound multiple sites within iA1C in HepG2 cells but no binding was seen in CP-A cells, suggesting CTCF binding may be required for iA1C insulator function. Since CTCF is highly expressed in both HepG2 and CP-A cells, the cell specific binding is not a function of CTCF availability.

iA1C showed cell-specificity even on a widely expressed heterologous promoter-enhancer pair from SV40. Thus, the cell-specificity resides within iA1C itself. Interestingly, the effects of the various regulatory elements were very similar in some aspects while different in others in the two cell types that express *ADH7,* CP-A cells and MEF cells. However, the general trend remained similar with the upstream elements showing promoter specific effects: greater increase or smaller decrease when combined with A7P-A and the opposite with A7P-G. Similarly, iA1C had no enhancer blocking activity with either the *ADH* HNF1-bound enhancer (ENH) or the much more potent SV40 enhancer in both *ADH7* expressing cell types. iA1C insulator function and cell-specificity with the potent SV40 enhancer suggests that it protects the *ADH7* promoter from not just the proximal class I *ADH* enhancer, but from other distal enhancers as well.

Chromatin insulators have emerged as important factors in the spatial and topological organization of higher order chromatin structures and functional transcriptional domains (Phillips-Cremins and Corces, 2013, Van Bortle and Corces, 2012). Insulators can employ any one or a combination of mechanisms depending on cell type and target enhancers and promoters (Valenzuela and Kamakaka, 2006, Wallace and Felsenfeld, 2007, Bushey et al., 2008, Zhu et al., 2007). All vertebrate insulators identified so far bind CTCF. CTCF can interact with itself forming homodimers, with other regulatory proteins (Merkenschlager and Odom, 2013, Weth and Renkawitz, 2011), and also with the nuclear lamina causing DNA looping and formation of *cis*- and *trans*- chromatin domains and influencing cross talk between gene promoters and regulatory elements (Phillips and Corces, 2009, Williams and Flavell, 2008). Cell specific interactions between CTCF bound sequences are known to regulate the cell specific expression of linked genes such as the b-globin cluster (Junier et al., 2012, Ren et al., 2012).

Since our results show that unbound iA1C correlates with a lack of enhancer blocking activity, it is probable that CTCF binding is a necessary event for iA1C function as an insulator on this element. However, the test for an induction of *ADH7* expression in HepG2 cells by knockdown of CTCF mediated by adenoviral shRNA proved ineffectual. Because of the complete absence of *ADH7* in HepG2 cells, a high level of induction would be needed to register in a qPCR assay. CTCF is also very highly expressed in HepG2 cells so that even with a 50% knockdown (or a ΔΔCt value of 2), CTCF mRNA was present at high levels indicating insufficient CTCF mRNA degradation. However, while mRNA degradation is a primary mechanism of shRNA mediated gene silencing,

other mechanisms include translation repression. To test for gene silencing by translation repression, CTCF protein levels in the CTCF shRNA and control shRNA treated cells should be compared to correctly ascertain knockdown efficiency. Irrespective of the level of CTCF knockdown, because of the complete absence of *ADH7* in untreated HepG2 cells, a high level of induction would be needed to register in a qPCR assay. Such induction was not achieved in this experiment. Therefore, evidence of the fate of *ADH7* expression on 2-fold CTCF mRNA knockdown is inconclusive.

CTCF levels in untreated HepG2 cells (Ct=23.5) is 10 fold lower than GAPDH levels (Ct=18.7); relatively, CTCF levels in CP-A cells (Ct=25.5) were 4 fold lower than those in HepG2 cells starting with similar RNA and cDNA amounts and handled similarly. The high levels of CTCF in both HepG2 and CP-A cells indicate mechanisms other than CTCF availability influence binding to iA1C. Some of these potential mechanisms that can play a role in cell specific CTCF binding include the differential methylation status of the CTCF binding sites within iA1C, differently compacted chromatin structures (open vs closed chromatin) and competition/ differential occupation of regulatory protein binding sites (including CTCF binding sites) in the two cell types.

iA1C is DNaseI hypersensitive in nearly all cell lines tested by ENCODE (Bernstein et al.), which does not include CP-A or MEF cells, and is also predicted to bind the typical CTCF binding partner, Rad21 (Myers et al., The Encode Project, 2011) which is a component of the cohesin complex. ENH is located less than 2 kb downstream of iA1C, and the intervening sequence is a DNaseI hypersensitive region bound by several

different transcription factors including AP-1 and NF-κB family (Myers et al., The Encode Project, 2011). The MafK and MafF transcription factors belonging to the AP-1 family can bind this region as well as 7P10-E2, the enhancer located 12 kb upstream of *ADH7* gene. The Maf (v-maf avian Musculoaponeurotic fibrosarcoma oncogene homolog) proteins can dimerize by physically interacting with each other (Motohashi et al., 2004). This 2 kb region encompassing iA1C and ENH may be a potential locus control region, regulating the tissue-specific expression of multiple *ADH* genes.

The ENCODE project has identified potential CTCF binding sites (Myers et al., Bernstein et al., The Encode Project, 2011) in the intergenic regions between *ADH5-ADH4, ADH1C-ADH7* and *ADH7-C4orf17* including a site corresponding to our element iA1C. *ADH7,* which is not expressed in the liver at all unlike the adjacent *ADHs,* may be part of a chromatin domain formed by the interactions between iA1C and the CTCF binding site between *ADH7* and C4orf17. The necessity of such a structure is emphasized by fact that the genes further upstream of C4orf17 are either housekeeping genes like *TRMT10A* (tRNA Methyl Transferase Homolog 10A) or differently expressed like *MTTP* (Microsomal Triglyceride Transfer Protein). Indeed, the ENCODE ChIP-PET Interactions track identifies potential interactions between iA1C and the CTCF site between *ADH7-C4orf17* supporting our model. Our results show that iA1C protects the promoter from the non-homologous SV40 enhancer indicating that it may be involved in blocking other *ADH7* non-relevant regulatory elements like the much stronger *ADH4* enhancer 4E3.

We have shown that iA1C insulates the *ADH7* gene from the class I *ADH* enhancer ENH, and also the heterologous SV40 enhancer, in liver derived cells. We demonstrate that the function of iA1C is cell-specific, as is the binding of CTCF to iA1C. Thus we conclude that the insulator iA1C helps determine the tissue specificity of *ADH7* expression. There is a small but significant difference in function due to genetic variation at rs1442489. Chapter I demonstrated cell specificity and significant effects of genetic variations in several more proximal elements, including an enhancer 7P10 and the *ADH7* promoter itself (Jairam and Edenberg, 2014). Thus, a combination of factors is important for the overall and precise regulation of *ADH7* transcription.

V. CONCLUSIONS and FUTURE DIRECTIONS

The main objective of my dissertation was the identification of elements regulating *ADH7* transcription and variants having significant effects on regulation. Since most of the reported disease associations belong to the 5' LD block of *ADH7,* the first research objective focused on the 12.5 kb conserved region upstream of *ADH7* and its variants. This was based on the hypothesis that the causal variants for at least some of the reported associations affect *ADH7* regulation and expression levels, since the only known non-synonymous cSNP was associated with cancer but not with alcoholism or alcohol metabolism. An alteration of gene expression can influence enzyme levels and the contribution of ADH7 to substrate metabolism, and ultimately risk for alcoholism and cancer.

An analysis of the LD pattern between the SNPs in the upstream region and disease associated SNPs showed that each major disease phenotype was associated with a distinct LD block comprising SNPs with little overlap between the 3 blocks. Upstream variants were identified through their effects on regulatory function and *ADH7* promoter activity. The promoter SNP, that is in LD with SNPs associated with alcoholism and alcohol related cancer and causes a 2-fold difference in activity, is particularly important. The causal variant for a particular disease phenotype is not always the SNP identified by an association study; instead, any SNP having significant association with a particular disease phenotype actually represents the association of at least one SNP that is in LD with the original SNP. In cases such as this where the originally identified SNP has no apparent effect on protein levels or gene expression, it is likely the causal variant is a

SNP that is in LD with the original SNP and affects either transcript or protein levels. Thus, the promoter SNP is a potential causal variant for protection against alcoholism and risk for UADT cancers. Another important finding is the identification of the enhancer 7P10 and its variants that significantly affect function, and are in LD with the only reported non-synonymous coding SNP as well as SNPs associated with alcohol metabolism. The promoter and upstream variants interact in a cell-specific and allele-specific manner complicating the interpretation of effects of individual SNPs. These regulatory elements are functional in HepG2 cells and do not explain the tissue-specific expression of *ADH7*.

My second objective was to better understand *ADH7* tissue specific expression, particularly the lack of *ADH7* expression in the liver in the presence of proximal functional enhancers. This was accomplished by the identification of iA1C, a cell-specific and position-dependent insulator that blocked the class I *ADH* HNF1 bound enhancer (ENH) from activating the *ADH7* promoter in hepatoma cells; it behaved similarly with the more potent heterologous SV40 enhancer. iA1C also bound CTCF cell-specifically, validating the original hypothesis and experimental approach to the project which was based on the *in silico* identification of iA1C as a potential CTCF binding site on the premise that a chromatin insulator was involved. Based on its location (between *ADH7* and the rest of the *ADH* cluster) and properties, iA1C likely protects *ADH7* from being expressed in the liver by blocking ectopic *ADH* enhancers (mainly liver-specific). It could also work the other way as a barrier element to block *ADH7* enhancers (7P10-E2) from activating other ADHs in ectopic tissues **(Figure 25)**.

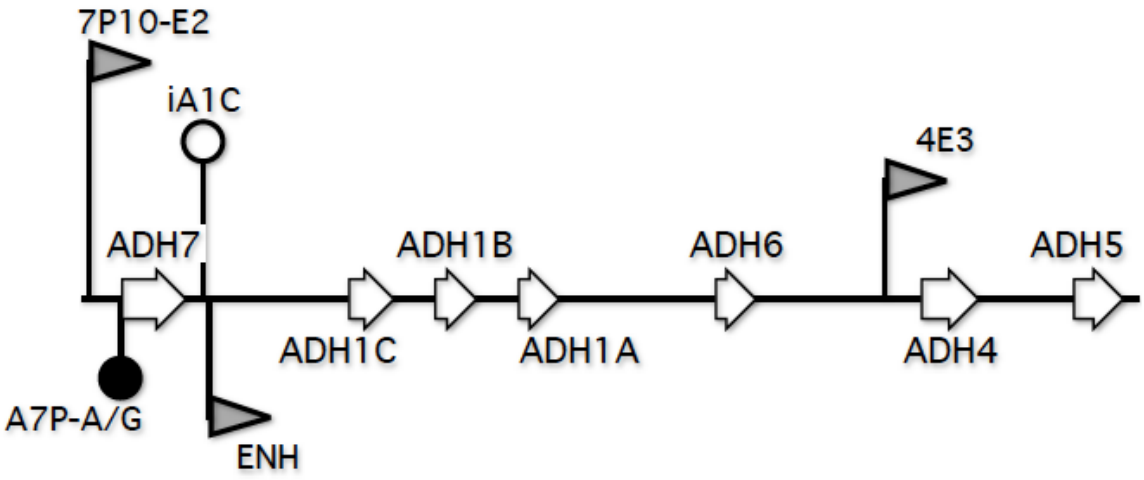


Figure 25: Regulatory elements in the *ADH* region. *ADH* regulatory elements with the strongest effects and that are relevant to this thesis mapped in a diagram drawn to scale. The *ADHs* are listed on sequence in the direction of transcription orientation. Known enhancers are listed as line arrows, iA1C as plain circle and the *ADH7* promoter as solid circle. The insulator iA1C blocks the *ADH7* promoter A7P-A/G from activation in the liver by other *ADH* enhancers, ENH and likely 4E3 (Pochareddy and Edenberg, 2010). It could also block the *ADH7* enhancer from activating the other *ADH* promoters in ectopic tissues like the esophagus.

Together, this dissertation identified elements that regulate *ADH7* tissue specific expression and variants affecting this regulation in physiologically relevant cell systems. While the upstream regulatory elements have smaller but still significant effects in the cell-specific expression of *ADH7*, iA1C has a more important role. On the other hand, the influence of the upstream variants on regulation is more potent compared to the effect of the smaller, but still significant effect of the iA1C variants. This is consistent with the

fact that most reported disease associations of *ADH7* lie within the 5' LD block. It is likely that the smaller genetic influence on iA1C function is related to the need to preserve the integrity of *ADH7* (and possibly other *ADHs)* tissue-specific expression. In conclusion, genetic factors determine the level of *ADH7* transcriptional activity while iA1C helps determine the fate of transcription, thus providing an overall picture of *ADH7* regulation. Human and mouse cell lines replicating *ADH7* tissue specific expression were used as models in this study and while they cannot duplicate the full complexity of *in vivo* systems, they worked well in the context of this dissertation.

Future work could continue and build on this knowledge using a combination of bioinformatics and experimental approaches. A more detailed analysis of the regulatory element haplotypes associated with diseases could be done to identify additional functional SNPs, particularly rare alleles. My work focused mainly on common haplotypes occurring naturally in the Western and Northern European populations (CEU). Future work on rarer haplotypes needs to be undertaken since it has been shown that the protective *ADH1B*2* allele that is common mainly in the East Asians has a significant protective effect in other populations as well. Opposite haplotypes of important regulatory regions can be tested first to gain an initial understanding of the potential effect of the variants in the haplotype. For example, 7P5-ATAC and 7P5-GCGT had a significant difference in function in HepG2 cells suggesting that a more detailed study of the other haplotypes is needed to help decipher the likely causal variant. The differences in activities of promoter and regulatory element haplotypes could be due to differential binding of regulatory proteins. *In silico* prediction tools like PROMO and

Transfac can be used to help prioritize potential candidates. The list can be further narrowed by looking for transcription factors important for expression of other genes in specific epithelial tissues like the esophagus, which can then be tested for binding target sequences by *in vivo* and *in vitro* binding assays. Haplotype specific binding can be studied by searching for regulatory proteins that only bind one haplotype of the target sequence, losing their consensus binding sites on alteration of the DNA sequence. The enhancer 7P10-E2 can be characterized further. The genome wide ENCODE consortium and 1000 Genome data available publicly can also be used to identify other elements with regulatory potential by looking for characteristics like histone modifications, DNaseI hypersensitivity and transcription factor binding.

The iA1C insulator could also be characterized further. The potential barrier function of iA1C could be studied by stably transfecting a reporter gene sequence flanked by iA1C into HepG2 cells, followed by testing for protection against position effects by assaying reporter activity. Epigenetic studies like DNA methylation and histone modifications can be done to decipher iA1C properties; comparison studies in HepG2 and CP-A cells could explain the cell specific CTCF binding. CTCF bound insulators can form domains by interacting with each other and with the nuclear architecture. iA1C binding to the nuclear architecture can be studied by the nuclear matrix attachment assay or DamID technique which is based on the targeted adenine methylation (in the sequence GATC) of sequences interacting *in vivo* with the HepG2 nuclear lamina integrated lamin-DNA adenine methyltransferase fusion protein (Dam).

The ENCODE consortium and insulatordb database predict other potential CTCF binding sites in the *ADH* cluster. Chromosome conformation capture (3C) assays can be used to study interactions between iA1C and the other potential CTCF bound sites. The potential formation of domains by the interactions between the CTCF bound sites and the nuclear lamina could play a vital role in the tissue specific expression of all the *ADHs.* The approximately 1 kb region between iA1C and the class I enhancer ENH is DNaseI hypersensitive in many cell lines (Bernstein et al., The Encode Project, 2011) and is bound by many different transcription factors (Myers et al., The Encode Project, 2011) indicating high regulatory potential. However, since none of the cell lines tested by ENCODE express *ADH7,* they are not the best systems to study *ADH* expression. Therefore, while the transcription factors identified by ENCODE do not affirmatively predict binding or function in relevant cell systems, they can be used as a guide for further studies in this region. Knockdown assays resulting in greater (nearly complete) knockdown of CTCF can be attempted to test *ADH7* induction. Similar knockdown experiments to specifically test the reversal of enhancer blocking by iA1C can be done by co-transfection of iA1C plasmids and CTCF shRNA adenovirus. CTCF heterozygous mice are viable (Heath et al., 2008, Moore et al., 2012) and can be used to study the profile of *ADH* expression in the liver and other tissues of the heterozygotes.

Genome editing technology can be used to manipulate the different regulatory elements *in vivo* for more detailed studies of function. For example, iA1C sequence can be directly deleted in HepG2 cells and effect on *ADH* expression tested. Genome editing technology uses artificially engineered nucleases to insert, delete or replace DNA by creating specific

double strand breaks in the genome, which are then repaired by the cell's endogenous mechanisms of homology directed repair and nonhomologous end-joining. The recently discovered Crispr technology, which was first identified as a natural defense system used by bacteria against invading viruses, can be used to target any region of the genome with very high accuracy with the help of the DNA endonuclease Cas9 (Crispr associated nuclease-9) (Richter et al., 2013). Since the only requirement for Crispr mediated genome editing is the Cas9 nuclease and a small guide RNA that can base pair with the target sequence and guide the endonuclease to generate double strand breaks in the sequence, it is a simple, straightforward tool for site-specific engineering that is easy to implement and can be used to generate a wealth of physiologically relevant information *in vivo.* These studies will help further the understanding of regulation of the alcohol dehydrogenases and the correlation between expression and functional significance and diseases.

APPENDIX A

Description	Primer name	Sequence
A7P	HE3879	CGCAAGCTTCCATGGTACCGCGGCCGCTCGAGGTCTGTAATGGTTAAAC
A7P	HE1048	CACTCGAGCTGTATTTCTGCAAACATAGAC
7P2	HE1051	GAGGTACCTTTATGCTGTCACCACCC
7P2	HE1052	CACTCGAGGTTTTCATCTACTTGGG
7P3	HE1053	GAGGTACCGTATGGACAGGGTAACCG
7P3	HE1054	CACTCGAGCTTACTCACAAGCCAAAT
7P4	HE1086	GAGGTACCTGCTGTGTTTTGATGCC
7P4	HE1087	CACTCGAGTTGAACCTGGAAGTGCTAC
7P5	HE1084	GAGGTACCTCTCTTTCACCTTCCTCG
7P5	HE1085	CACTCGAGAAACTCTCCATTTCACTTAGC
7P6	HE3881	CATGCCATGGCGCTCTCATCTCATCTTGAC
7P6	HE1083	CACTCGAGTCCTTGGAGGTCTAAGTCTC
7P7	HE1080	CAGGTACCAGGTCTTTGAGGAATCACC
7P7	HE1081	CACTCGAGAACCCCATCACTTTAGTCC
7P8	HE1078	CAGGTACCAAAAAGTGATGAAGGATGC
7P8	HE1079	CACTCGAGAAGACATAACTGCCATTCG
7P9	HE1076	CAGGTACCTGAGGTTCTGGCATACG
7P9	HE1077	GACTCGAGTTTGTGTCATTACCGAGG
7P10	HE3781	GACGGTACCTGAGTTGCTGGTCTGTCCC
7P10	HE3782	CCGCTCGAGGCTTGAATGATACCTTGTGCTC
7P10-E1	HE4620	CCCAAGCTTGAGTTGCTGGTCTGTCCCACA
7P10-E1	HE4621	CCGCTCGAGAGTTACAAATGTCACAAGTGTGAGTG
7P10-E2	HE4622	CCCAAGCTTCTCATTGAATGAATAAAAGATTTATTGCC
7P10-E2	HE4623	CCGCTCGAGGACAACACTGATACCTAAATGTAAACTAC
7P10-E3	HE4624	CCCAAGCTTCTGGCTGGTAGAACTTTCATTTACTAAG
7P10-E3	HE4625	CCGCTCGAGCTTTGATTTTGGACTTCTGGTGttc
7P10-E2flip	HE4800	CCGCTCGAGCTCATTGAATGAATAAAAGATTTATTGCC
7P10-E2flip	HE4801	CCCAAGCTTGACAACACTGATACCTAAATGTAAACTAC
7P10-E2far	HE4802	CGCAATGGGCCCCTCATTGAATGAATAAAAGATTTATTGCC
7P10-E2far	HE4803	CCCATTTAAATGACAACACTGATACCTAAATGTAAACTAC

Table 8: Primers used for generating test and control fragments for CHAPTER I.

Primer names, the complete sequence including restriction sites and the test fragments amplified from the primer pairs are listed.

APPENDIX B

ENH	HE3532	CACCATATGGCCGGCATCAGCGCTGCGATCGCTCTGATTCTGAAAGCTGAG
ENH	HE3533	CTCACATGTACGCGTATCGCGCGCTCGCGAGAGTATGAGTTTTGAGG
iA1C	HE3552	ATGCGATCGCATCTTGGAGCACCCATCACTTGAGCG
iA1C	HE3553	ATTGCCGGCCCTTCAGGTTTTCCACCAGACAAATC
iA1C	HE3546	CGGACGCGTATCTTGGAGCACCCATCACTTGAGCG
iA1C	HE3547	ACTTCGCGACCTTCAGGTTTTCCACCAGACAAATC
XL9	HE3542	ATGCGATCGCTTCGGTGTCCAAGTGTCAAAAAGG
XL9	HE3543	ATTGCCGGCTGTGGTATGAAAGTGAGGGTAGGTG
XL9	HE3538	GGGACGCGTTTCGGTGTCCAAGTGTCAAAAAGG
XL9	HE3539	ACTTCGCGATGTGGTATGAAAGTGAGGGTAGGTG
iA1C-O (LAI	HE4551	5'FAM-CAGTACTTATGAGGGCAGGAAGGTGGCAGCATTTAGCTATC
iA1C-O	HE4553	CAGTACTTATGAGGGCAGGAAGGTGGCAGCATTTAGCTATC
iA1C-O	HE4552	GATAGCTAAATGCTGCCACCTTCCTGCCCTCATAAGTACTG
iA1C-Osh	HE4574	CAGTACTTATGAGGGTAGGACCGTACGAGCATTTAGCTATC
iA1C-Osh	HE4575	GATAGCTAAATGCTCGTACGGTCCTACCCTCATAAGTACTG
NONSPECIF	HE4468	AGTGCAGACACAGAAAGTTTCACTTAACTTTCTACACCTAA
NONSPECIF	HE4469	TTAGGTGTAGAAAGTTAAGTGAAACTTTCTGTGTCTGCACT
CTCF-CON	HE4568	CAATGACATGATGGCCAGCAGAGGGCGCATGGCCCTGGGG
CTCF-CON	HE4569	CCCCAGGGCCATGCGCCCTCTGCTGGCCATCATGTCATTG
iA1C-SH	HE4618	CAGTACTTATGAGGGTAGGACCGTACGAGCATTTAGCTATC
iA1C-SH	HE4619	GATAGCTAAATGCTCGTACGGTCCTACCCTCATAAGTACTG
iA1C-A	HE4833	TTGGAGCACCCATCACTTG
iA1C-A	HE4834	CAAATGAGGACGCAAAAGC
iA1C-B	HE4835	CCTGAGTCCCCAAAGTTCAC
iA1C-B	HE4836	GGAATGGAAAACCAAACACTGTATG
iA1C-C	HE4837	GTTTGGTTTTCCATTCCTGAG
iA1C-C	HE4838	TGCTGCCACCTTCCTGC
iA1C-D	HE4839	TGAGGGCAGGAAGGTGG
iA1C-D	HE4840	GGAAAAGAGAAGCAATCAAGG
iA1C-E	HE4841	CCAGCAGAACTGAGAAAATAAGG
iA1C-E	HE4842	TGACAAAAGTGGGAGGATGAG
iA1C-F	HE4843	CATCCTCCCACTTTTGTCATG
iA1C-F	HE4844	CCCAATAATTCCCTTAGTATATGAAAC
iA1C-G	HE4845	GTTTCATATACTAAGGGAATTATTGGG
iA1C-G	HE4846	ATGGGATCATTCTGTGGGATT
iA1C-H	HE4847	TCCCACAGAATGATCCCAT
iA1C-H	HE4848	CCACTGAATGCTTAGGGTTG
iA1C-I	HE4849	TCAACCCTAAGCATTCAGTGG
iA1C-I	HE4850	TTCTCAGCCTAATCAGTAAGACG
iA1C-J	HE4851	CGTCTTACTGATTAGGCTGAGAA
iA1C-J	HE4852	CCTTCAGGTTTTCCACCAGA
iA1C-K	HE4853	GGAAAACCTGAAGGTAGGACC
iA1C-K	HE4854	TTCAACATCCCTTATTGGAGTG

Table 9: Primers and oligonucleotide probes used in CHAPTER II. Primer/oligo names, the complete sequence including restriction sites and the test fragments amplified from the primer pairs are listed.

REFERENCES

AGARWAL, D. P. & SEITZ, H. K. 2001. *Alcohol in Health and Disease*, CRC Press.

ALLALI-HASSANI, A., MARTÍNEZ, S. E., PERALBA, J. M., VAGLENOVA, J., VIDAL, F., RICHART, C., FARRÉS, J. & PARÉS, X. 1997. Alcohol dehydrogenase of human and rat blood vessels: Role in ethanol metabolism. *FEBS Letters,* 405**,** 26-30.

AMERICAN PSYCHIATRY ASSOCIATION, T. 2013. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition,* Arlington, VA, American Psychiatric Association.

ARANDA, A. & PASCUAL, A. 2001. Nuclear Hormone Receptors and Gene Expression. *Physiological Reviews,* 81**,** 1269-1304.

ARNOSTI, D. N. & KULKARNI, M. M. 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry,* 94**,** 890-898.

BAGNARDI, V., BLANGIARDO, M., LA VECCHIA, C. & CORRAO, G. 2001. Alcohol consumption and the risk of cancer: a meta-analysis. *Br J Cancer,* 85**,** 1700-1705.

BALMER, J. E. & BLOMHOFF, R. 2002. Gene expression regulation by retinoic acid. *Journal of Lipid Research,* 43**,** 1773-1808.

BAO, L., ZHOU, M. & CUI, Y. 2008. CTCFBSDB: a CTCF-binding site database for characterization of vertebrate genomic insulators. *Nucleic Acids Research,* 36**,** D83-D87.

BARKESS, G. & WEST, A. G. 2012. Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. *Epigenomics,* 4**,** 67-80.

BERNSTEIN, B. H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE.

BERNSTEIN, B. H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE.

BERNSTEIN, B., MYERS, R. M., SNYDER, M., CRAWFORD, G. & STAMATOYANNOPOULOS, J. Integrated Regulation from ENCODE tracks.

BIERUT, L. J., GOATE, A. M., BRESLAU, N., JOHNSON, E. O., BERTELSEN, S., FOX, L., AGRAWAL, A., BUCHOLZ, K. K., GRUCZA, R., HESSELBROCK, V., KRAMER, J., KUPERMAN, S., NURNBERGER, J., PORJESZ, B., SACCONE, N. L., SCHUCKIT, M., TISCHFIELD, J., WANG, J. C., FOROUD, T., RICE, J. P. & EDENBERG, H. J. 2012. ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Mol Psychiatry,* 17**,** 445-450.

BIRLEY, A. J., JAMES, M. R., DICKSON, P. A., MONTGOMERY, G. W., HEATH, A. C., MARTIN, N. G. & WHITFIELD, J. B. 2009. ADH single nucleotide polymorphism associations with alcohol metabolism in vivo. *Human Molecular Genetics,* 18**,** 1533-1542.

BIRLEY, A. J., JAMES, M. R., DICKSON, P. A., MONTGOMERY, G. W., HEATH, A. C., WHITFIELD, J. B. & MARTIN, N. G. 2008. Association of the gastric alcohol dehydrogenase gene ADH7 with variation in alcohol metabolism. *Human Molecular Genetics,* 17**,** 179-189.

BOFFETTA, P. & HASHIBE, M. 2006. Alcohol and cancer. *The Lancet Oncology,* 7**,** 149-156.

BOLEDA, M. D., SAUBI, N., FARRES, J. & PARES, X. 1993. Physiological Substrates for Rat Alcohol Dehydrogenase Classes: Aldehydes of Lipid Peroxidation, ω-Hydroxyfatty Acids, and Retinoids. *Archives of Biochemistry and Biophysics,* 307**,** 85-90.

BONDARENKO, V. A., LIU, Y. V., JIANG, Y. I. & STUDITSKY, V. M. 2003. Communication over a large distance: enhancers and insulators. *Biochemistry and Cell Biology,* 81**,** 241-251.

BOSRON, W. F. & LI, T.-K. 1986a. Genetic polymorphism of human liver alcohol and aldehyde dehydrogenases, and their relationship to alcohol metabolism and alcoholism. *Hepatology,* 6**,** 502-510.

BOSRON, W. F. & LI, T. K. 1986b. Genetic polymorphism of human liver alcohol and aldehyde dehydrogenases, and their relationship to alcohol metabolism and alcoholism. *Hepatology,* 6**,** 502-510.

BOSRON, W. F. & LI, T. K. 1987. Catalytic properties of human liver alcohol dehydrogenase isoenzymes. *Enzyme,* 37**,** 19-28.

BOSRON, W. F., YIN, S.-J. & LI, T. K. 1985. Purification and characterization of human liver beta 1 beta 1, beta 2 beta 2 and beta Ind beta Ind alcohol dehydrogenase isoenzymes. *Prog Clin Biol Res,* 174**,** 193-206.

BRÄNDÉN, C.-I. & EKLUND, H. 2008. Coenzyme-Induced Conformational Changes and Substrate Binding in Liver Alcohol Dehydrogenase. *Ciba Foundation*

Symposium 60 - Molecular Interactions and Activity in Proteins. John Wiley & Sons, Ltd.

BROWN, C. J., ZHANG, L. & EDENBERG HOWARD, J. 1996. Gene expression in a young multigene family: tissue-specific differences in the expression of the human alcohol dehydrogenase genes ADH1, ADH2, and ADH3. *DNA Cell Biol,* 15**,** 187-196.

BURNELL, J. C., CARR, L. G., DWULET, F. E., EDENBERG, H. J., LI, T. K. & BOSRON, W. F. 1987. The human beta 3 alcohol dehydrogenase subunit differs from beta 1 by a Cys for Arg-369 substitution which decreases NAD(H) binding. *Biochem Biophys Res Commun.,* 146**,** 1127-1133.

BUSHEY, A. M., DORMAN, E. R. & CORCES, V. G. 2008. Chromatin Insulators: Regulatory Mechanisms and Epigenetic Inheritance. *Molecular cell,* 32**,** 1-9.

BUSHUE, N. & WAN, Y.-J. Y. 2010. Retinoid pathway and cancer therapeutics. *Advanced Drug Delivery Reviews,* 62**,** 1285-1298.

CADONI, G., BOCCIA, S., PETRELLI, L., DI GIANNANTONIO, P., ARZANI, D., GIORGIO, A., DE FEO, E., PANDOLFINI, M., GALLÌ, P., PALUDETTI, G. & RICCIARDI, G. 2012. A review of genetic epidemiology of head and neck cancer related to polymorphisms in metabolic genes, cell cycle control and alcohol metabolism. *ACTA otorhinolaryngologica italica,* 32**,** 1-11.

CARGIULO, T. 2007. Understanding the health impact of alcohol dependence. *American Journal of Health-System Pharmacy,* 64**,** S5-S11.

CARR, L. G., FOROUD, T., STEWART, T., CASTELLUCCIO, P., EDENBERG, H. J. & LI, T.-K. 2002. Influence of ADH1B polymorphism on alcohol use and its

subjective effects in a Jewish population. *American Journal of Medical Genetics,* 112**,** 138-143.

CHAMBON, P. 1996. A decade of molecular biology of retinoic acid receptors. *The FASEB Journal,* 10**,** 940-54.

CHEN, C.-C., LU, R.-B., CHEN, Y.-C., WANG, M.-F., CHANG, Y.-C., LI, T.-K. & YIN, S.-J. 1999. Interaction between the Functional Polymorphisms of the Alcohol-Metabolism Genes in Protection against Alcoholism. *American journal of human genetics,* 65**,** 795-807.

CHEN, H., LI, J., LI, H., HU, Y., TEVEBAUGH, W., YAMAMOTO, M., QUE, J. & CHEN, X. 2012. Transcript Profiling Identifies Dynamic Gene Expression Patterns and an Important Role for Nrf2/Keap1 Pathway in the Developing Mouse Esophagus. *PLoS ONE,* 7**,** e36504.

CHEN, W. J., LOH, E. W., HSU, Y. P., CHEN, C. C., YU, J. M. & CHENG, A. T. 1996. Alcohol-metabolising genes and alcoholism among Taiwanese Han men: independent effect of ADH2, ADH3 and ALDH2. *Br J Psychiatry,* 168**,** 762-767.

CHEUNG, C., SMITH, C. K., HOOG, J.-O. & HOTCHKISS, S. A. M. 1999. Expression and Localization of Human Alcohol and Aldehyde Dehydrogenase Enzymes in Skin. *Biochemical and Biophysical Research Communications,* 261**,** 100-107.

CHIN-TONG, O. & VICTOR, G. C. 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics,* 12**,** 283-293.

CHOI, I.-G., SON, H.-G., YANG, B.-H., KIM, S. H., LEE, J.-S., CHAI, Y.-G., SON, B. K., KEE, B. S., PARK, B. L., KIM, L. H., CHOI, Y. H. & SHIN, H. D. 2005.

Scanning of genetic effects of alcohol metabolism gene (ADH1B and ADH1C) polymorphisms on the risk of alcoholism. *Human Mutation,* 26**,** 224-234.

CLARK, A. G. 2003. Finding genes underlying risk of complex disease by linkage disequilibrium mapping. *Current Opinion in Genetics & Development,* 13**,** 296-302.

CONNOLLY, R. M., NGUYEN, N. K. & SUKUMAR, S. 2013. Molecular Pathways: Current Role and Future Directions of the Retinoic Acid Pathway in Cancer Prevention and Treatment. *Clinical Cancer Research,* 19**,** 1651-1659.

COSTANTINI, S., DI BERNARDO, G., CAMMAROTA, M., CASTELLO, G. & COLONNA, G. 2013. Gene expression signature of human HepG2 cell line. *Gene,* 518**,** 335-345.

COSTELLO, J., NAGARANJAN, R., FOUSE, S., JOHNSON, B., HONG, C., SHCHORS, K., HEINE, V. M., ROWITCH, D. H., BILENKY, M., D'SOUZA, C., NIELSEN, C., ZHAO, Y., DELANEY, A., VARHOL, R., THIESSEN, N., JONES, S. S. J., MARRA, M. A., HIRST, M., WANG, T., XING, X., FIORE, C., SCHILLEBEECKX, M., BALLINGER, T. J., HAUSSLER, D. & TURECKI, G. UCSF Brain DNA Methylation.

COVAULT, J., GELERNTER, J., HESSELBROCK, V., NELLISSERY, M. & KRANZLER, H. R. 2004. Allelic and haplotypic association of GABRA2 with alcohol dependence. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* 129B**,** 104-109.

CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M.,

SHARP, P. A., BOYER, L. A., YOUNG, R. A. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*.

DANAEI, G., DING, E. L., MOZAFFARIAN, D., TAYLOR, B., REHM, J., MURRAY, C. J. L. & EZZATI, M. 2009. The Preventable Causes of Death in the United States: Comparative Risk Assessment of Dietary, Lifestyle, and Metabolic Risk Factors. *PLoS Med,* 6**,** e1000058.

DANIELSSON, O. & JÖRNVALL, H. 1992. "Enzymogenesis": classical liver alcohol dehydrogenase origin from the glutathione-dependent formaldehyde dehydrogenase line. *Proceedings of the National Academy of Sciences,* 89**,** 9247-9251.

DANNENBERG, L. O., CHEN, H.-J. & EDENBERG, H. J. 2005. GATA-2 and HNF-3β Regulate the Human Alcohol Dehydrogenase 1A (ADH1A) Gene. *DNA Cell Biol,* 24**,** 543-552.

DAWSON, D. A., GOLDSTEIN, R. B. & GRANT, B. F. 2013. Differences in the Profiles of DSM-IV and DSM-5 Alcohol Use Disorders: Implications for Clinicians. *Alcoholism: Clinical and Experimental Research,* 37**,** E305-E313.

DAY, C. P., BASHIR, R., JAMES, O. F., BASSENDINE, M. F., CRABB, D. W., THOMASSON, H. R., LI, T. K. & EDENBERG, H. J. 1991. Investigation of the role of polymorphisms at the alcohol and aldehyde dehydrogenase loci in genetic predisposition to alcohol-related end-organ damage. *Hepatology,* 14**,** 798-801.

DEAN, A. 2006. On a chromosome far, far away: LCRs and gene expression. *Trends in genetics : TIG,* 22**,** 38-45.

DELTOUR, L., FOGLIO, M. H. & DUESTER, G. 1999. Impaired retinol utilization in Adh4 alcohol dehydrogenase mutant mice. *Developmental Genetics,* 25**,** 1-10.

DIAMOND, L., KRUSZEWSKI, F., ADEN, D. P., KNOWLES, B. B. & BAIRD, W. M. 1980. Metabolic activation of benzo[a]pyrene by a human hepatoma cell line. *Carcinogenesis,* 1**,** 871-875.

DICK, D. M., AGRAWAL, A., WANG, J. C., HINRICHS, A., BERTELSEN, S., BUCHOLZ, K. K., SCHUCKIT, M., KRAMER, J., NURNBERGER, J., TISCHFIELD, J., EDENBERG, H. J., GOATE, A. & BIERUT, L. J. 2007. Alcohol dependence with comorbid drug dependence: genetic and phenotypic associations suggest a more severe form of the disorder with stronger genetic contribution to risk*. *Addiction,* 102**,** 1131-1139.

DODGE, N. C., JACOBSON, J. L. & JACOBSON, S. W. 2014. Protective effects of the alcohol dehydrogenase-ADH1B*3 allele on attention and behavior problems in adolescents exposed to alcohol during pregnancy. *Neurotoxicology and Teratology,* 41**,** 43-50.

DRUESNE-PECOLLO, N., TEHARD, B., MALLET, Y., GERBER, M., NORAT, T., HERCBERG, S. & LATINO-MARTEL, P. 2009. Alcohol and genetic polymorphisms: effect on risk of alcohol-related cancer. *The Lancet Oncology,* 10**,** 173-180.

DUESTER, G. 1991. A Hypothetical Mechanism for Fetal Alcohol Syndrome Involving Ethanol Inhibition of Retinoic Acid Synthesis at the Alcohol Dehydrogenase Step. *Alcoholism: Clinical and Experimental Research,* 15**,** 568-572.

DULEY, J. A., HARRIS, O. & HOLMES, R. S. 1985. Analysis of Human Alcohol- and Aldehyde-Metabolizing Isozymes by Electrophoresis and Isoelectric Focusing. *Alcoholism: Clinical and Experimental Research,* 9**,** 263-271.

EDENBERG, H. J. 2000. Regulation of the mammalian alcohol dehydrogenase genes. *Prog Nucleic Acid Res Mol Biol.,* 64**,** 295-341.

EDENBERG, H. J. 2007a. The genetics of alcohol metabolism: Role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alc Res Health,* 30**,** 5-13.

EDENBERG, H. J. 2007b. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res Health,* 30**,** 5-13.

EDENBERG, H. J. 2011. Common and Rare Variants in Alcohol Dependence. *Biological Psychiatry,* 70**,** 498-499.

EDENBERG, H. J. 2012. Genes contributing to the development of alcoholism: an overview. *Alc Health Res,* 34**,** 336-338.

EDENBERG, H. J., DICK, D. M., XUEI, X., TIAN, H., ALMASY, L., BAUER, L. O., CROWE, R. R., GOATE, A., HESSELBROCK, V., JONES, K., KWON, J., LI, T.-K., NURNBERGER, J. I., O CONNOR, S. J., REICH, T., RICE, J., SCHUCKIT, M. A., PORJESZ, B., FOROUD, T. & BEGLEITER, H. 2004. Variations in GABRA2, Encoding the ±2 Subunit of the GABAA Receptor, Are Associated with Alcohol Dependence and with Brain Oscillations. *American journal of human genetics,* 74**,** 705-714.

EDENBERG, H. J. & FOROUD, T. 2013. Genetics and alcoholism. *Nat Rev Gastroenterol Hepatol,* 10**,** 487-494.

EDENBERG, H. J., JEROME, R. E. & LI, M. 1999. Polymorphism of the human alcohol dehydrogenase 4 (ADH4) promoter affects gene expression. *Pharmacogenetics and Genomics,* 9**,** 25-30.

EDENBERG, H. J., WANG, J., TIAN, H., POCHAREDDY, S., XUEI, X., WETHERILL, L., GOATE, A., HINRICHS, T., KUPERMAN, S., NURNBERGER, J. I., SCHUCKIT, M., TISCHFIELD, J. A. & FOROUD, T. 2008a. A regulatory variation in OPRK1, the gene encoding the κ-opioid receptor, is associated with alcohol dependence. *Human Molecular Genetics,* 17**,** 1783-1789.

EDENBERG, H. J., XUEI, X., CHEN, H.-J., TIAN, H., WETHERILL, L. F., DICK, D. M., ALMASY, L., BIERUT, L., BUCHOLZ, K. K., GOATE, A., HESSELBROCK, V., KUPERMAN, S., NURNBERGER, J., PORJESZ, B., RICE, J., SCHUCKIT, M., TISCHFIELD, J., BEGLEITER, H. & FOROUD, T. 2006. Association of alcohol dehydrogenase genes with alcohol dependence: a comprehensive analysis. *Human Molecular Genetics,* 15**,** 1539-1549.

EDENBERG, H. J., XUEI, X., WETHERILL, L. F., BIERUT, L., BUCHOLZ, K., DICK, D. M., HESSELBROCK, V., KUPERMAN, S., PORJESZ, B., SCHUCKIT, M. A., TISCHFIELD, J. A., ALMASY, L. A., NURNBERGER, J. I. & FOROUD, T. 2008b. Association of NFKB1, which encodes a subunit of the transcription factor NF-κB, with alcohol dependence. *Human Molecular Genetics,* 17**,** 963-970.

EHRIG, T., BOSRON, W. F. & LI, T.-K. 1990. Alcohol and aldehyde dehydrogenase. *Alcohol and Alcoholism,* 25**,** 105-116.

ESTONIUS, M., SVENSSON, S. & HÖÖG, J.-O. 1996. Alcohol dehydrogenase in human tissues: localisation of transcripts coding for five classes of the enzyme. *FEBS Letters,* 397**,** 338-342.

EVANS, T. R. J. & KAYE, S. B. 1999. Retinoids: present role and future potential. *Br J Cancer,* 80**,** 1-8.

EVERTS, H. B. 2012. Endogenous retinoids in the hair follicle and sebaceous gland. *Biochimica et biophysica acta,* 1821**,** 222-229.

FARRÉ, D., ROSET, R., HUERTA, M., ADSUARA, J. E., ROSELLÓ, L., ALBÀ, M. M. & MESSEGUER, X. 2003. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Research,* 31**,** 3651-3653.

FEHR, C., SANDER, T., TADIC, A., LENZEN, K. P., ANGHELESCU, I., KLAWE, C., DAHMEN, N., SCHMIDT, L. G. & SZEGEDI, A. 2006. Confirmation of association of the GABRA2 gene with alcohol dependence by subtype-specific analysis. *Psychiatric Genetics,* 16**,** 9-17.

FILIPPOVA, G. N. 2008. Genetics and epigenetics of the multifunctional protein CTCF. *Current topics in developmental biology,* 80**,** 337-360.

FILIPPOVA, G. N., FAGERLIE, S., KLENOVA, E. M., MYERS, C., DEHNER, Y., GOODWIN, G., NEIMAN, P. E., COLLINS, S. J. & LOBANENKOV, V. V. 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Molecular and Cellular Biology,* 16**,** 2802-13.

FRIEDENSON, B. 2012. Alcohol, acetaldehyde and breast cancer risk. *The Breast,* 21**,** 612.

GALLEGO, O., BELYAEVA, O. V., PORTE, S., RUIZ, X. F., STETSENKO, A. V., SHABROVA, E. V., KOSTEREVA, N. V., FARRES, J., PARÉS, X. & KEDISHVILI, N. Y. 2006. Comparative functional analysis of human medium-chain dehydrogenases, short-chain dehydrogenases/reductases and aldo-keto reductases with retinoids. *Biochemical Journal,* 399**,** 101-109.

GALTER, D., CARMINE, A., BUERVENICH, S., DUESTER, G. & OLSON, L. 2003. Distribution of class I, III and IV alcohol dehydrogenase mRNAs in the adult rat, mouse and human brain. *European Journal of Biochemistry,* 270**,** 1316-1326.

GARATTINI, E., GIANNI, M. & TERAO, M. 2007. Cytodifferentiation by Retinoids, a Novel Therapeutic Option in Oncology: Rational Combinations with Other Therapeutic Agents. *In:* GERALD, L. (ed.) *Vitamins & Hormones.* Academic Press.

GARDNER, D. S. & TAI, S. E. 2012. Clinical features and treatment of maturity onset diabetes of the young (MODY). *Diabetes Metab Syndr Obes,* 5**,** 101-108.

GERMAIN, P., CHAMBON, P., EICHELE, G., EVANS, R. M., LAZAR, M. A., LEID, M., DE LERA, A. R., LOTAN, R., MANGELSDORF, D. J. & GRONEMEYER, H. 2006. International Union of Pharmacology. LX. Retinoic Acid Receptors. *Pharmacological Reviews,* 58**,** 712-725.

GIZER, I. R., EDENBERG, H. J., GILDER, D. A., WILHELMSEN, K. C. & EHLERS, C. L. 2011. Association of Alcohol Dehydrogenase Genes with Alcohol-Related

Phenotypes in a Native American Community Sample. *Alcoholism: Clinical and Experimental Research,* 35**,** 2008-2018.

GOATE, A. M. & EDENBERG, H. J. 1998. The genetics of alcoholism. *Current Opinion in Genetics & Development,* 8**,** 282-286.

GOMES, N. P. & ESPINOSA, J. M. 2010. Gene-specific repression of the p53 target gene PUMA via intragenic CTCF–Cohesin binding. *Genes & Development,* 24**,** 1022-1034.

GOODRICH, J. A. & TJIAN, R. 2010. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet,* 11**,** 549-558.

GOODWIN, D. W., SCHULSINGER, F., HERMANSEN, L., GUZE, S. B. & WINOKUR, G. 1973. Alcohol problems in adoptees raised apart from alcoholic biological parents. *Archives of General Psychiatry,* 28**,** 238-243.

GOODWIN, D. W., SCHULSINGER, F., MØLLER, N., HERMANSEN, L., WINOKUR, G. & GUZE, S. B. 1974. Drinking problems in adopted and nonadopted sons of alcoholics. *Archives of General Psychiatry,* 31**,** 164-169.

GRANT, B. F. 1998. The impact of a family history of alcoholism on the relationship between age at onset of alcohol use and DSM-IV alcohol dependence: results from the National Longitudinal Alcohol Epidemiologic Survey. *Alcohol Health Res World,* 22**,** 144-147.

GROSKREUTZ, D. J., SHERF, B. A., WOOD, K. V. & SCHENBORN, E. T. 1995. Increased expression and convenience with the new pGL3 Luciferase Reporter Vectors. *Promega Notes***,** 2-8.

GUINDALINI, C., SCIVOLETTO, S., FERREIRA, R. G. M., BREEN, G., ZILBERMAN, M., PELUSO, M. A. & ZATZ, M. 2005. Association of Genetic Variants in Alcohol Dehydrogenase 4 with alcohol dependence in Brazilian patients *Am J Psychiatry,* 162**,** 1005-1007.

HAAS, S. L., YE, W. & LÖHR, J.-M. 2012. Alcohol consumption and digestive tract cancer. *Current Opinion in Clinical Nutrition & Metabolic Care,* Publish Ahead of Print**,** 10.1097/MCO.0b013e3283566699.

HAKENEWERTH, A. M., MILLIKAN, R. C., RUSYN, I., HERRING, A. H., NORTH, K. E., BARNHOLTZ-SLOAN, J. S., FUNKHOUSER, W. F., WEISSLER, M. C. & OLSHAN, A. F. 2011. Joint Effects of Alcohol Consumption and Polymorphisms in Alcohol and Oxidative Stress Metabolism Genes on Risk of Head and Neck Cancer. *Cancer Epidemiology Biomarkers & Prevention,* 20**,** 2438-2449.

HALLER, G., KAPOOR, M., BUDDE, J., XUEI, X., EDENBERG, H., NURNBERGER, J., KRAMER, J., BROOKS, A., TISCHFIELD, J., ALMASY, L., AGRAWAL, A., BUCHOLZ, K., RICE, J., SACCONE, N., BIERUT, L. & GOATE, A. 2013. Rare missense variants in CHRNB3 and CHRNA3 are associated with risk of alcohol and cocaine dependence. *Human Molecular Genetics*.

HAN, C.-L., LIAO, C.-S., WU, C.-W., HWONG, C.-L., LEE, A.-R. & YIN, S.-J. 1998. Contribution to first-pass metabolism of ethanol and inhibition by ethanol for retinol oxidation in human alcohol dehydrogenase family. *European Journal of Biochemistry,* 254**,** 25-31.

HAN, Y., OOTA, H., OSIER, M. V., PAKSTIS, A. J., SPEED, W. C., ODUNSI, A., OKONOFUA, F., KAJUNA, S. L. B., KAROMA, N. J., KUNGULILO, S., GRIGORENKO, E., ZHUKOVA, O. V., BONNE-TAMIR, B., LU, R.-B., PARNAS, J., SCHULZ, L. O., KIDD, J. R. & KIDD, K. K. 2005. Considerable Haplotype Diversity within the 23kb Encompassing the ADH7 Gene. *Alcoholism: Clinical and Experimental Research,* 29**,** 2091-2100.

HASHIBE, M., MCKAY, J. D., CURADO, M. P., OLIVEIRA, J. C., KOIFMAN, S., KOIFMAN, R., ZARIDZE, D., SHANGINA, O., WUNSCH-FILHO, V., ELUF-NETO, J., LEVI, J. E., MATOS, E., LAGIOU, P., LAGIOU, A., BENHAMOU, S., BOUCHARDY, C., SZESZENIA-DABROWSKA, N., MENEZES, A., DALL'AGNOL, M. M., MERLETTI, F., RICHIARDI, L., FERNANDEZ, L., LENCE, J., TALAMINI, R., BARZAN, L., MATES, D., MATES, I. N., KJAERHEIM, K., MACFARLANE, G. J., MACFARLANE, T. V., SIMONATO, L., CANOVA, C., HOLCATOVA, I., AGUDO, A., CASTELLSAGUE, X., LOWRY, R., JANOUT, V., KOLLAROVA, H., CONWAY, D. I., MCKINNEY, P. A., ZNAOR, A., FABIANOVA, E., BENCKO, V., LISSOWSKA, J., CHABRIER, A., HUNG, R. J., GABORIEAU, V., BOFFETTA, P. & BRENNAN, P. 2008. Multiple ADH genes are associated with upper aerodigestive cancers. *Nat Genet,* 40**,** 707-709.

HASIN, D. S., AHARONOVICH, E., LIU, X., MAMMAN, Z., MATSEOANE, K., CARR, L. G. & LI, T.-K. 2002. Alcohol and ADH2 in Israel: Ashkenazis, Sephardics, and recent Russian immigrants. *Am J Psychiatry,* 159**,** 1432-1434.

HASIN, D. S., STINSON, F. S., OGBURN, E. & GRANT, B. F. 2007. Prevalence, correlates, disability, and comorbidity of dsm-iv alcohol abuse and dependence in the united states: Results from the national epidemiologic survey on alcohol and related conditions. *Archives of General Psychiatry,* 64**,** 830-842.

HEATH, H., DE ALMEIDA, C. R., SLEUTELS, F., DINGJAN, G., VAN DE NOBELEN, S., JONKERS, I., LING, K. W., GRIBNAU, J., RENKAWITZ, R., GROSVELD, F., HENDRIKS, R. W. & GALJART, N. 2008. CTCF regulates cell cycle progression of αβ T cells in the thymus. *The EMBO Journal,* 27**,** 2839-2850.

HOLMES, R. S. 1994. Alcohol dehydrogenases: a family of isozymes with differential functions. *Alcohol Alcohol Suppl,* 2**,** 127-130.

HOLWERDA, S. J. B. & DE LAAT, W. 2013. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 368.

HÖÖG, J.-O. & BRANDT, M. 1995. Mammalian Class VI Alcohol Dehydrogenase. *In:* WEINER, H., HOLMES, R. & WERMUTH, B. (eds.) *Enzymology and Molecular Biology of Carbonyl Metabolism 5.* Springer US.

HOOG, J.-O., HEDÉN, L.-O., LARSSON, K., JORNVALL, H. & VON BAHR-LINDSTROM, H. 1986. The γ1 and γ2 subunits of human liver alcohol dehydrogenase. *European Journal of Biochemistry,* 159**,** 215-218.

HÖÖG, J.-O. & ÖSTBERG, L. J. 2011. Mammalian alcohol dehydrogenases – A comparative investigation at gene and protein levels. *Chemico-Biological Interactions,* 191**,** 2-7.

HURLEY, T. D. & EDENBERG, H. J. 2012. Genes Encoding Enzymes Involved in Ethanol Metabolism. *Alc Res Health,* 34**,** 339-44.

HURLEY, T. D., EDENBERG, H. J. & LI, T.-K. 2003. Pharmacogenomics of Alcoholism. *Pharmacogenomics.* Wiley-VCH Verlag GmbH & Co. KGaA.

INTERNATIONAL HAPMAP CONSORTIUM, T. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature,* 467**,** 52-58.

ISHIHARA, K., OSHIMURA, M. & NAKAO, M. 2006. CTCF-Dependent Chromatin Insulator Is Linked to Epigenetic Remodeling. *Molecular cell,* 23**,** 733-742.

ISRAEL, Y., RIVERA-MEZA, M., QUINTANILLA, M. E., SAPAG, A. & TAMPIER, L. 2011. Acetaldehyde Burst Protection of ADH1B*2 Against Alcoholism: An Additional Hormesis Protection Against Esophageal Cancers Following Alcohol Consumption? *Alcoholism: Clinical and Experimental Research,* 35**,** 806-810.

ITTIWUT, C., YANG, B.-Z., KRANZLER, H. R., ANTON, R. F., HIRUNSATIT, R., WEISS, R. D., COVAULT, J., FARRER, L. A. & GELERNTER, J. 2012. GABRG1 and GABRA2 Variation Associated with Alcohol Dependence in African Americans. *Alcoholism: Clinical and Experimental Research,* 36**,** 588-593.

JAIRAM, S. & EDENBERG, H. J. 2014. Single-Nucleotide Polymorphisms Interact to Affect ADH7 Transcription. *Alcoholism: Clinical and Experimental Research***,** n/a-n/a.

JIAN, C., JAMES, J. H. & DAWEI, L. 2013. Multi-Cultural Association of the Serotonin Transporter Gene (SLC6A4) with Substance Use Disorder. *Neuropsychopharmacology,* 38**,** 1737-1747.

JIN, F., LI, Y., REN, B. & NATARAJAN, R. 2011. Enhancers: Multi-dimensional signal integrators. *Transcription,* 2**,** 226-230.

JONATHAN, C., JOEL, G., KEVIN, J., RAYMOND, A. & HENRY, R. K. 2007. Markers in the 5′-Region of GABRG1 Associate to Alcohol Dependence and are in Linkage Disequilibrium with Markers in the Adjacent GABRA2 Gene. *Neuropsychopharmacology,* 33**,** 837-848.

JÖRNVALL, H., HEMPEL, J., VALLEE, B. L., BOSRON, W. F. & LI, T. K. 1984. Human liver alcohol dehydrogenase: amino acid substitution in the beta 2 beta 2 Oriental isozyme explains functional properties, establishes an active site structure, and parallels mutational exchanges in the yeast enzyme. *Proceedings of the National Academy of Sciences,* 81**,** 3024-3028.

JULIAN-SANCHEZ, A., RIVEROS-ROSAS, H. & PINA, E. 2006. Are class III alcohol dehydrogenases the ancestral group inside the medium-chain dehydrogenase/reductase superfamily? *The FASEB Journal,* 20.

JUNIER, I., DALE, R. K., HOU, C., KÉPÈS, F. & DEAN, A. 2012. CTCF-mediated transcriptional regulation through cell type-specific chromosome organization in the β-globin locus. *Nucleic Acids Research,* 40**,** 7718-7727.

JUVEN-GERSHON, T., HSU, J.-Y., THEISEN, J. W. M. & KADONAGA, J. T. 2008. The RNA polymerase II core promoter — the gateway to transcription. *Current Opinion in Cell Biology,* 20**,** 253-259.

JUVEN-GERSHON, T. & KADONAGA, J. T. 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology,* 339**,** 225-229.

KADAUKE, S. & BLOBEL, G. A. 2009. Chromatin loops in gene regulation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* 1789**,** 17-25.

KAISER, R., FERNÁNDEZ, M. R., PARÉS, X. & JÖRNVALL, H. 1993. Origin of the human alcohol dehydrogenase system: implications from the structure and properties of the octopus protein. *Proceedings of the National Academy of Sciences,* 90**,** 11222-11226.

KARIN, M., LIU, Z.-G. & ZANDI, E. 1997. AP-1 function and regulation. *Current Opinion in Cell Biology,* 9**,** 240-246.

KAWATA, M. 2001. Subcellular Steroid/Nuclear Receptor Dynamics. *Archives of Histology and Cytology,* 64**,** 353-368.

KEDISHVILI, N. Y., BOSRON, W. F., STONE, C. L., HURLEY, T. D., PEGGS, C. F., THOMASSON, H. R., POPOV, K. M., CARR, L. G., EDENBERG, H. J. & LI, T.-K. 1995. Expression and Kinetic Characterization of Recombinant Human Stomach Alcohol Dehydrogenase: ACTIVE-SITE AMINO ACID SEQUENCE EXPLAINS SUBSTRATE SPECIFICITY COMPARED WITH LIVER ISOZYMES. *Journal of Biological Chemistry,* 270**,** 3625-3630.

KEDISHVILI, N. Y., GOUGH, W. H., CHERNOFF, E. A. G., HURLEY, T. D., STONE, C. L., BOWMAN, K. D., POPOV, K. M., BOSRON, W. F. & LI, T.-K. 1997. cDNA Sequence and Catalytic Properties of a Chick Embryo Alcohol Dehydrogenase That Oxidizes Retinol and 3β,5α-Hydroxysteroids. *Journal of Biological Chemistry,* 272**,** 7494-7500.

KENDLER, K. S., KALSI, G., HOLMANS, P. A., SANDERS, A. R., AGGEN, S. H., DICK, D. M., ALIEV, F., SHI, J., LEVINSON, D. F. & GEJMAN, P. V. 2011. Genomewide Association Analysis of Symptoms of Alcohol Dependence in the Molecular Genetics of Schizophrenia (MGS2) Control Sample. *Alcoholism: Clinical and Experimental Research,* 35**,** 963-975.

KENDLER, K. S., PRESCOTT, C. A., NEALE, M. C. & PEDERSEN, N. L. 1997. Temperance board registration for alcohol abuse in a national sample of swedish male twins, born 1902 to 1949. *Archives of General Psychiatry,* 54**,** 178-184.

KESLER, C. T. & PASCHAL, B. M. 2009. Nucleocytoplasmic Transport of Steroid Hormone Receptors. *In:* KEHLENBACH, R. H. (ed.) *Nuclear Transport.* Landes Bioscience.

KIMURA, Y., NISHIMURA, F. T., ABE, S., FUKUNAGA, T., TANII, H. & SAIJOH, K. 2009. Polymorphisms in the promoter region of the human class II alcohol dehydrogenase (ADH4) gene affect both transcriptional activity and ethanol metabolism in Japanese subjects. *The Journal of Toxicological Sciences,* 34**,** 89-97.

KNOWLES, B. B., HOWE, C. & ADEN, D. P. 1980. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. *Science,* 209**,** 497-499.

KONISHI, T., CALVILLO, M., LENG, A. I.-S., FENG, J., LEE, T., LEE, H., SMITH, J. L., SIAL, S. H., BERMAN, N., FRENCH, S., EYSSELEIN, V., LIN, K.-M. & YVONNE WAN, Y. U.-J. 2003. The ADH3*2 and CYP2E1 c2 alleles increase

the risk of alcoholism in Mexican American men. *Experimental and Molecular Pathology,* 74**,** 183-189.

KONISHI, T., LUO, H.-R., CALVILLO, M., MAYO, M. S., LIN, K.-M. & WAN, Y.-J. Y. 2004. ADH1B*1, ADH1C*2, DRD2 (−141C Ins), and 5-HTTLPR Are Associated With Alcoholism in Mexican American Men Living in Los Angeles. *Alcoholism: Clinical and Experimental Research,* 28**,** 1145-1152.

KOSOFF, R. E., GARDINER, K. L., MERLO, L. M. F., PAVLOV, K., RUSTGI, A. K. & MALEY, C. C. 2012. Development and characterization of an organotypic model of Barrett's esophagus. *Journal of Cellular Physiology,* 227**,** 2654-2659.

KOTAGIRI, S. & EDENBERG, H. J. 1998. Regulation of Human Alcohol Dehydrogenase Gene ADH7: Importance of an AP-1 Site. *DNA Cell Biol.,* 17**,** 583-590.

KUMAR, S., SARADHI, M., CHATURVEDI, N. K. & TYAGI, R. K. 2006. Intracellular localization and nucleocytoplasmic trafficking of steroid receptors: An overview. *Molecular and Cellular Endocrinology,* 246**,** 147-156.

LAPPALAINEN, J., KRUPITSKY, E., REMIZOV, M., PCHELINA, S., TARASKINA, A., ZVARTAU, E., SOMBERG, L. K., COVAULT, J., KRANZLER, H. R., KRYSTAL, J. H. & GELERNTER, J. 2005. Association Between Alcoholism and γ-Amino Butyric Acid α2 Receptor Subtype in a Russian Population. *Alcoholism: Clinical and Experimental Research,* 29**,** 493-498.

LEE, B.-K. & IYER, V. R. 2012. Genome-wide Studies of CCCTC-binding Factor (CTCF) and Cohesin Provide Insight into Chromatin Structure and Regulation. *Journal of Biological Chemistry,* 287**,** 30906-30913.

LEVRAN, O., LONDONO, D., O'HARA, K., RANDESI, M., ROTROSEN, J., CASADONTE, P., LINZY, S., OTT, J., ADELSON, M. & KREEK, M. J. 2009. Heroin addiction in African Americans: a hypothesis-driven association study. *Genes, Brain and Behavior,* 8**,** 531-540.

LI, D., ZHAO, H. & GELERNTER, J. 2011. Strong Association of the Alcohol Dehydrogenase 1B Gene (ADH1B) with Alcohol Dependence and Alcohol-Induced Medical Diseases. *Biological Psychiatry,* 70**,** 504-512.

LI, D., ZHAO, H. & GELERNTER, J. 2012. Further clarification of the contribution of the ADH1C gene to vulnerability of alcoholism and selected liver diseases. *Human Genetics,* 131**,** 1361-1374.

LI, Q., PETERSON, K. R., FANG, X. & STAMATOYANNOPOULOS, G. 2002. Locus control regions. *Blood,* 100**,** 3077-3086.

LI, T.-K., BOSRON, W. F., DAFELDECKER, W. P. & LANGE, L., G. 1977. Isolation of II-alcohol dehydrogenase of human liver: Is it a determinant of alcoholism? *Proceedings of the National Academy of Sciences,* 74**,** 4378-4381.

LUCZAK, S. E., SHEA, S. H., CARR, L. G., LI, T.-K. & WALL, T. L. 2002. Binge Drinking in Jewish and Non-Jewish White College Students. *Alcoholism: Clinical and Experimental Research,* 26**,** 1773-1778.

LUNYAK, V. V. 2008. Boundaries. Boundaries…Boundaries??? *Current Opinion in Cell Biology,* 20**,** 281-287.

LUO, X., KRANZLER, H. R., ZUO, L., WANG, S., BLUMBERG, H. P. & GELERNTER, J. 2005. CHRM2 gene predisposes to alcohol dependence, drug

dependence and affective disorders: results from an extended case–control structured association study. *Human Molecular Genetics,* 14**,** 2421-2434.

MAJUMDER, P., GOMEZ, J. A., CHADWICK, B. P. & BOSS, J. M. 2008. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *The Journal of Experimental Medicine,* 205**,** 785-798.

MARCELLA, R. & JENS, T. 2012. The genetics of alcohol dependence. *Annals of the New York Academy of Sciences*.

MARIAN, A. J. 2012. Molecular genetic studies of complex phenotypes. *Translational research : the journal of laboratory and clinical medicine,* 159**,** 64-79.

MARTIN, P. R., SINGLETON, C. K. & HILLER-STURMHOFEL, S. 2003. The role of thiamine deficiency in alcoholic brain disease. *Alcohol Res Health,* 27**,** 134-143.

MARY-ANNE, E., COLIN, A. H., QIAOPING, Y., BERNARD, A., MATTI, V. & DAVID, G. 2008. GABRG1 and GABRA2 as Independent Predictors for Alcoholism in Two Populations. *Neuropsychopharmacology,* 34**,** 1245-1254.

MASTON, G. A., EVANS, SARA K, GREEN, MICHAEL R. 2006. Transcriptional Regulatory Elements in the Human Genome. *Ann Review Genomics Hum Genet,* 7**,** 29-59.

MATSUMOTO, M., YOKOYAMA, H., SUZUKI, H., SHIRAISHI-YOKOYAMA, H. & HIBI, T. 2005. Retinoic acid formation from retinol in the human gastric mucosa: role of class IV alcohol dehydrogenase and its relevance to morphological changes. *American Journal of Physiology - Gastrointestinal and Liver Physiology,* 289**,** G429-G433.

MATYS, V., FRICKE, E., GEFFERS, R., GÖSSLIN, E., HAUBROCK, M., HEHL, R., HORNISCHER, K., KARAS, D., KEL, A. E., KEL-MARGOULIS, O. V., KLOOS, D.-U., LAND, S., LEWICKI-POTAPOV, B., MICHAEL, H., MÜNCH, R., REUTER, I., ROTERT, S., SAXEL, H., SCHEER, M., THIELE, S. & WINGENDER, E. 2003. TRANSFAC®: transcriptional regulation, from patterns to profiles. *Nucleic Acids Research,* 31**,** 374-378.

MAYFIELD, R. D., HARRIS, R. A. & SCHUCKIT, M. A. 2008. Genetic factors influencing alcohol dependence. *British Journal of Pharmacology,* 154**,** 275-287.

MCGUE, M. 1999. The Behavioral Genetics of Alcoholism. *Current Directions in Psychological Science,* 8**,** 109-115.

MCKAY, J. D., TRUONG, T., GABORIEAU, V., CHABRIER, A., CHUANG, S.-C., BYRNES, G., ZARIDZE, D., SHANGINA, O., SZESZENIA-DABROWSKA, N., LISSOWSKA, J., RUDNAI, P., FABIANOVA, E., BUCUR, A., BENCKO, V., HOLCATOVA, I., JANOUT, V., FORETOVA, L., LAGIOU, P., TRICHOPOULOS, D., BENHAMOU, S., BOUCHARDY, C., AHRENS, W., MERLETTI, F., RICHIARDI, L., TALAMINI, R., BARZAN, L., KJAERHEIM, K., MACFARLANE, G. J., MACFARLANE, T. V., SIMONATO, L., CANOVA, C., AGUDO, A., CASTELLSAGUÉ, X., LOWRY, R., CONWAY, D. I., MCKINNEY, P. A., HEALY, C. M., TONER, M. E., ZNAOR, A., CURADO, M. P., KOIFMAN, S., MENEZES, A., WÜNSCH-FILHO, V., NETO, J. E., GARROTE, L. F., BOCCIA, S., CADONI, G., ARZANI, D., OLSHAN, A. F., WEISSLER, M. C., FUNKHOUSER, W. K., LUO, J., LUBIŃSKI, J., TRUBICKA, J., LENER, M., OSZUTOWSKA, D., SCHWARTZ, S. M., CHEN,

C., FISH, S., DOODY, D. R., MUSCAT, J. E., LAZARUS, P., GALLAGHER, C. J., CHANG, S.-C., ZHANG, Z.-F., WEI, Q., STURGIS, E. M., WANG, L.-E., FRANCESCHI, S., HERRERO, R., KELSEY, K. T., MCCLEAN, M. D., MARSIT, C. J., NELSON, H. H., ROMKES, M., BUCH, S., NUKUI, T., ZHONG, S., LACKO, M., MANNI, J. J., PETERS, W. H. M., HUNG, R. J., MCLAUGHLIN, J., VATTEN, L., NJØLSTAD, I., GOODMAN, G. E., FIELD, J. K., LILOGLOU, T., VINEIS, P., CLAVEL-CHAPELON, F., PALLI, D., TUMINO, R., KROGH, V., PANICO, S., GONZÁLEZ, C. A., QUIRÓS, J. R., MARTÍNEZ, C., NAVARRO, C., ARDANAZ, E., LARRAÑAGA, N., et al. 2011. A Genome-Wide Association Study of Upper Aerodigestive Tract Cancers Conducted within the INHANCE Consortium. *PLoS Genet,* 7**,** e1001333.

MENDEL, D. B., HANSEN, L. P., GRAVES, M. K., CONLEY, P. B. & CRABTREE, G. R. 1991. HNF-1 alpha and HNF-1 beta (vHNF-1) share dimerization and homeo domains, but not activation domains, and form heterodimers in vitro. *Genes & Development,* 5**,** 1042-1056.

MERKENSCHLAGER, M. & ODOM, D. T. 2013. CTCF and Cohesin: Linking Gene Regulatory Elements with Their Targets. *Cell,* 152**,** 1285-1297.

MERSCH-SUNDERMANN, V., KNASMÜLLER, S., WU, X.-J., DARROUDI, F. & KASSIE, F. 2004. Use of a human-derived liver cell line for the detection of cytoprotective, antigenotoxic and cogenotoxic agents. *Toxicology,* 198**,** 329-340.

MESSEGUER, X., ESCUDERO, R., FARRÉ, D., NÚÑEZ, O., MARTÍNEZ, J. & ALBÀ, M. M. 2002. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics,* 18**,** 333-334.

MIZOI, Y., TATSUNO, Y., ADACHI, J., KOGAME, M., FUKUNAGA, T., FUJIWARA, S., HISHIDA, S. & IJIRI, I. 1983. Alcohol sensitivity related to polymorphism of alcohol-metabolizing enzymes in Japanese. *Pharmacology Biochemistry and Behavior,* 18, Supplement 1**,** 127-133.

MOKDAD, A. H., MARKS, J. S., STROUP, D. F. & GERBERDING, J. L. 2004. Actual causes of death in the united states, 2000. *JAMA,* 291**,** 1238-1245.

MOLOTKOV, A., DELTOUR, L., FOGLIO, M. H., CUENCA, A. E. & DUESTER, G. 2002a. Distinct Retinoid Metabolic Functions for Alcohol Dehydrogenase Genes Adh1 and Adh4 in Protection against Vitamin A Toxicity or Deficiency Revealed in Double Null Mutant Mice. *Journal of Biological Chemistry,* 277**,** 13804-13811.

MOLOTKOV, A., FAN, X., DELTOUR, L., FOGLIO, M. H., MARTRAS, S., FARRÉS, J., PARÉS, X. & DUESTER, G. 2002b. Stimulation of retinoic acid production and growth by ubiquitously expressed alcohol dehydrogenase Adh3. *Proceedings of the National Academy of Sciences,* 99**,** 5337-5342.

MOLTÓ, E., FERNÁNDEZ, A. & MONTOLIU, L. 2009. Boundaries in vertebrate genomes: different solutions to adequately insulate gene expression domains. *Briefings in Functional Genomics & Proteomics,* 8**,** 283-296.

MONTANE-JAIME, K., MOORE, S., SHAFE, S., JOSEPH, R., CROOKS, H., CARR, L. & EHLERS, C. L. 2006. ADH1C*2 allele is associated with alcohol dependence and elevated liver enzymes in Trinidad and Tobago. *Alcohol (Fayetteville, N.Y.),* 39**,** 81-86.

MOORE, J. M., RABAIA, N. A., SMITH, L. E., FAGERLIE, S., GURLEY, K., LOUKINOV, D., DISTECHE, C. M., COLLINS, S. J., KEMP, C. J.,

LOBANENKOV, V. V. & FILIPPOVA, G. N. 2012. Loss of Maternal CTCF Is Associated with Peri-Implantation Lethality of <italic>Ctcf</italic> Null Embryos. *PLoS ONE,* 7**,** e34915.

MORENO, A. & PARÉS, X. 1991. Purification and characterization of a new alcohol dehydrogenase from human stomach. *The Journal of biological chemistry,* 266**,** 1128-1133.

MOTOHASHI, H., KATSUOKA, F., ENGEL, J. D. & YAMAMOTO, M. 2004. Small Maf proteins serve as transcriptional cofactors for keratinocyte differentiation in the Keap1–Nrf2 regulatory pathway. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 6379-6384.

MULLIGAN, C., ROBIN, R., OSIER, M., SAMBUUGHIN, N., GOLDFARB, L., KITTLES, R., HESSELBROCK, D., GOLDMAN, D. & LONG, J. 2003. Allelic variation at alcohol metabolism genes (ADH1B, ADH1C, ALDH2) and alcohol dependence in an American Indian population. *Human Genetics,* 113**,** 325-336.

MUNN-CHERNOFF, M. A., DUNCAN, A. E., GRANT, J. D., WADE, T. D., AGRAWAL, A., BUCHOLZ, K. K., MADDEN, P. A. F., MARTIN, N. G. & HEATH, A. C. 2013. A Twin Study of Alcohol Dependence, Binge Eating, and Compensatory Behaviors. *J Stud Alcohol Drugs,* 74**,** 664-673.

MURRAY, R. M. & GURLING, H. M. 1982. Alcoholism: polygenic influence on a multifactorial disorder. *British journal of hospital medicine,* 27**,** 328-334.

MURRELL, A. 2011. Setting up and maintaining differential insulators and boundaries for genomic imprinting. *Biochemistry and Cell Biology,* 89**,** 469-478.

MYERS, R. M. & ABSHER, D. CpG Methylation by Methyl 450K Bead Arrays from ENCODE/HAIB

MYERS, R. M., SNYDER, M., GERSTEIN, M., WEISSMAN, S., FARNHAM, P. & STRUHL, K. ENCODE Transcription Factor ChIP-Seq.

OGBOURNE, S. & ANTALIS, T. M. 1998. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem. J.,* 331**,** 1-14.

ONG, C.-T. & CORCES, V. G. 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet,* 12**,** 283-293.

OSANAI, M., NISHIKIORI, N., MURATA, M., CHIBA, H., KOJIMA, T. & SAWADA, N. 2007. Cellular Retinoic Acid Bioavailability Determines Epithelial Integrity: Role of Retinoic Acid Receptor α Agonists in Colitis. *Molecular Pharmacology,* 71**,** 250-258.

OSIER, M., PAKSTIS, A. J., KIDD, J. R., LEE, J.-F., YIN, S.-J., KO, H.-C., EDENBERG, H. J., LU, R.-B. & KIDD, K. K. 1999. Linkage Disequilibrium at the ADH2 and ADH3 Loci and Risk of Alcoholism. *American journal of human genetics,* 64**,** 1147-1157.

OSIER, M. V., LU, R. B., PAKSTIS, A. J., KIDD, J. R., HUANG, S. Y. & KIDD, K. K. 2004. Possible epistatic role of ADH7 in the protection against alcoholism. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* 126B**,** 19-22.

PALANCA-WESSELS, M. C., BARRETT, M. T., GALIPEAU, P. C., ROHRER, K. L., REID, B. J. & RABINOVITCH, P. S. 1998. Genetic analysis of long-term

Barrett's esophagus epithelial cultures exhibiting cytogenetic and ploidy abnormalities. *Gastroenterology,* 114**,** 295-304.

PALANCA-WESSELS, M. C. A., KLINGELHUTZ, A., REID, B. J., NORWOOD, T. H., OPHEIM, K. E., PAULSON, T. G., FENG, Z. & RABINOVITCH, P. S. 2003. Extended lifespan of Barrett's esophagus epithelium transduced with the human telomerase catalytic subunit: a useful in vitro model. *Carcinogenesis,* 24**,** 1183-1190.

PARÉS, X., FARRÉS, J., KEDISHVILI, N. & DUESTER, G. 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families. *Cellular and Molecular Life Sciences,* 65**,** 3936-3949.

PARRY, C. D., PATRA, J. & REHM, J. 2011. Alcohol consumption and non-communicable diseases: epidemiology and policy implications. *Addiction,* 106**,** 1718-1724.

PERSSON, B., HEDLUND, J. & JÖRNVALL, H. 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families. *Cellular and Molecular Life Sciences,* 65**,** 3879-3894.

PERSSON, B., ZIGLER, J. S. & JÖRNVALL, H. 1994. A Super-Family of Medium-Chain Dehydrogenases/Reductases (MDR). *European Journal of Biochemistry,* 226**,** 15-22.

PHILLIPS, J. E. & CORCES, V. G. 2009. CTCF: Master Weaver of the Genome. *Cell,* 137**,** 1194-1211.

PHILLIPS-CREMINS, J. E. & CORCES, V. G. 2013. Chromatin Insulators: Linking Genome Organization to Cellular Function. *Molecular cell,* 50**,** 461-474.

PLAPP, B. V. 2010. Conformational changes and catalysis by alcohol dehydrogenase. *Archives of biochemistry and biophysics,* 493**,** 3-12.

POCHAREDDY, S. & EDENBERG, H. J. 2010. Identification of a FOXA-dependent enhancer of human alcohol dehydrogenase 4 (ADH4). *Gene,* 460**,** 1-7.

PÖSCHL, G. & SEITZ, H. K. 2004. Alcohol and cancer. *Alcohol and Alcoholism,* 39**,** 155-165.

PUROHIT, V., KHALSA, J. & SERRANO, J. 2005. Mechanisms of alcohol-associated cancers: introduction and summary of the symposium. *Alcohol (Fayetteville, N.Y.),* 35**,** 155-160.

RAILE, K., KLOPOCKI, E., WESSEL, T., DEISS, D., HORN, D., MÜLLER, D., ULLMANN, R. & GRÜTERS, A. 2008. HNF1B Abnormality (Mature-Onset Diabetes of the Young 5) in Children and Adolescents: High prevalence in autoantibody-negative type 1 diabetes with kidney defects. *Diabetes Care,* 31**,** e83.

RAJEEVAN, H., SOUNDARARAJAN, U., KIDD, J. R., PAKSTIS, A. J. & KIDD, K. K. 2012. ALFRED: an allele frequency resource for research and teaching. *Nucleic Acids Research,* 40**,** D1010-D1015.

REICH, D. E., CARGILL, M., BOLK, S., IRELAND, J., SABETI, P. C., RICHTER, D. J., LAVERY, T., KOUYOUMJIAN, R., FARHADIAN, S. F., WARD, R. & LANDER, E. S. 2001. Linkage disequilibrium in the human genome. *Nature,* 411**,** 199-204.

REIMERS, M. J., HAHN, M. E. & TANGUAY, R. L. 2004. Two Zebrafish Alcohol Dehydrogenases Share Common Ancestry with Mammalian Class I, II, IV, and V

Alcohol Dehydrogenase Genes but Have Distinct Functional Characteristics. *Journal of Biological Chemistry,* 279**,** 38303-38312.

REN, L., WANG, Y., SHI, M., WANG, X., YANG, Z. & ZHAO, Z. 2012. CTCF Mediates the Cell-Type Specific Spatial Organization of the <italic>Kcnq5</italic> Locus and the Local Gene Regulation. *PLoS ONE,* 7**,** e31416.

RHINN, M. & DOLLÉ, P. 2012. Retinoic acid signalling during development. *Development,* 139**,** 843-858.

RICHTER, H., RANDAU, L. & PLAGENS, A. 2013. Exploiting CRISPR/Cas: Interference Mechanisms and Applications. *International Journal of Molecular Sciences,* 14**,** 14518-14531.

RIETSCHEL, M. & TREUTLEIN, J. 2013. The genetics of alcohol dependence. *Annals of the New York Academy of Sciences,* 1282**,** 39-70.

ROBERT, E. & ERWIN, F. W. 2003. AP-1: a double-edged sword in tumorigenesis. *Nature Reviews Cancer,* 3**,** 859-868.

ROBINSON-RECHAVI, M., GARCIA, H. E. & LAUDET, V. 2003. The nuclear receptor superfamily. *Journal of Cell Science,* 116**,** 585-586.

RODENBERG, J. M., HOGGATT, A. M., CHEN, M., TOUW, K., JONES, R. & HERRING, B. P. 2010. Regulation of serum response factor activity and smooth muscle cell apoptosis by chromodomain helicase DNA-binding protein 8. *American Journal of Physiology - Cell Physiology,* 299**,** C1058-C1067.

ROSENBLOOM, K. R., SLOAN, C. A., MALLADI, V. S., DRESZER, T. R., LEARNED, K., KIRKUP, V. M., WONG, M. C., MADDREN, M., FANG, R.,

HEITNER, S. G., LEE, B. T., BARBER, G. P., HARTE, R. A., DIEKHANS, M., LONG, J. C., WILDER, S. P., ZWEIG, A. S., KAROLCHIK, D., KUHN, R. M., HAUSSLER, D. & KENT, W. J. 2013. ENCODE Data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Research,* 41**,** D56-D63.

SALASPURO, M. 2009. Acetaldehyde as a common denominator and cumulative carcinogen in digestive tract cancers. *Scandinavian Journal of Gastroenterology,* 44**,** 912-925.

SANDSTROM, R., BOYLE, A., SONG, L., CRAWFORD, G., GIRESI, P., LIEB, J., FUREY, T., LIU, Z., MCDANIELL, R., LEE, B.-K., IYER, V., FLICEK, P., KEEFE, D., BIRNEY, E. & GRAF, S. ENCODE group. Digital DNaseI Hypersensitivity clusters.

SCHUMANN, G., COIN, L. J., LOURDUSAMY, A., CHAROEN, P., BERGER, K. H., STACEY, D., DESRIVIÈRES, S., ALIEV, F. A., KHAN, A. A., AMIN, N., AULCHENKO, Y. S., BAKALKIN, G., BAKKER, S. J., BALKAU, B., BEULENS, J. W., BILBAO, A., DE BOER, R. A., BEURY, D., BOTS, M. L., BREETVELT, E. J., CAUCHI, S., CAVALCANTI-PROENÇA, C., CHAMBERS, J. C., CLARKE, T.-K., DAHMEN, N., DE GEUS, E. J., DICK, D., DUCCI, F., EASTON, A., EDENBERG, H. J., ESKO, T., FERNÁNDEZ-MEDARDE, A., FOROUD, T., FREIMER, N. B., GIRAULT, J.-A., GROBBEE, D. E., GUARRERA, S., GUDBJARTSSON, D. F., HARTIKAINEN, A.-L., HEATH, A. C., HESSELBROCK, V., HOFMAN, A., HOTTENGA, J.-J., ISOHANNI, M. K., KAPRIO, J., KHAW, K.-T., KUEHNEL, B., LAITINEN, J., LOBBENS, S., LUAN, J. A., MANGINO, M., MAROTEAUX, M., MATULLO,

G., MCCARTHY, M. I., MUELLER, C., NAVIS, G., NUMANS, M. E., NÚÑEZ, A., NYHOLT, D. R., ONLAND-MORET, C. N., OOSTRA, B. A., O'REILLY, P. F., PALKOVITS, M., PENNINX, B. W., POLIDORO, S., POUTA, A., PROKOPENKO, I., RICCERI, F., SANTOS, E., SMIT, J. H., SORANZO, N., SONG, K., SOVIO, U., STUMVOLL, M., SURAKK, I., THORGEIRSSON, T. E., THORSTEINSDOTTIR, U., TROAKES, C., TYRFINGSSON, T., TÖNJES, A., UITERWAAL, C. S., UITTERLINDEN, A. G., VAN DER HARST, P., VAN DER SCHOUW, Y. T., STAEHLIN, O., VOGELZANGS, N., VOLLENWEIDER, P., WAEBER, G., WAREHAM, N. J., WATERWORTH, D. M., WHITFIELD, J. B., WICHMANN, E. H., WILLEMSEN, G., WITTEMAN, J. C., YUAN, X., ZHAI, G., ZHAO, J. H., ZHANG, W., MARTIN, N. G., METSPALU, A., et al. 2011. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proceedings of the National Academy of Sciences,* 108**,** 7119-7124.

SCOTT, D. M. & TAYOR, R. E. 2007. Health-related effects of genetic variations of alcohol-metabolizing enzymes in African Americans. *Alcohol Res Health,* 30**,** 18-21.

SEITZ, H. K., MATSUZAKI, S., YOKOYAMA, A., HOMANN, N., VÄKEVÄINEN, S. & WANG, X. D. 2001. Alcohol and Cancer. *Alcoholism: Clinical and Experimental Research,* 25**,** 137S-143S.

SEITZ, H. K. & MEIER, P. 2007. The role of acetaldehyde in upper digestive tract cancer in alcoholics. *Translational research : the journal of laboratory and clinical medicine,* 149**,** 293-297.

SEITZ, H. K., PELUCCHI, C., BAGNARDI, V. & VECCHIA, C. L. 2012. Epidemiology and Pathophysiology of Alcohol and Breast Cancer: Update 2012. *Alcohol and Alcoholism,* 47**,** 204-212.

SK, N. 1988. Luciferase reporter gene vectors for analysis of promoters and enhancers. *Biotechniques,* 6**,** 454-458.

SMITH, M. 1986. Genetics of Human Alcohol and Aldehyde Dehydrogenases. *In:* HARRIS, H. & HIRSCHHORN, K. (eds.) *Advances in Human Genetics 15.* Springer US.

SMITH, M., HOPKINSON, D. A. & HARRIS, H. 1971. Developmental changes and polymorphism in human alcohol dehydrogenase. *Annals of Human Genetics,* 34**,** 251-271.

SMITH, M., HOPKINSON, D. A. & HARRIS, H. 1972. Alcohol dehydrogenase isozymes in adult human stomach and liver: evidence for activity of the ADH3 locus. *Annals of Human Genetics,* 35**,** 243-253.

SOKOL, R. J., DELANEY-BLACK, V. & NORDSTROM, B. 2003. Fetal alcohol spectrum disorder. *JAMA,* 290**,** 2996-2999.

SØRENSEN, H. J., MANZARDO, A. M., KNOP, J., PENICK, E. C., MADARASZ, W., NICKEL, E. J., BECKER, U. & MORTENSEN, E. L. 2011. The Contribution of Parental Alcohol Use Disorders and Other Psychiatric Illness to the Risk of

Alcohol Use Disorders in the Offspring. *Alcoholism: Clinical and Experimental Research,* 35**,** 1315-1320.

STENSON, P., MORT, M., BALL, E., HOWELLS, K., PHILLIPS, A., THOMAS, N. & COOPER, D. 2009. The Human Gene Mutation Database: 2008 update. *Genome Med,* 1**,** 13.

STEWART, M. J., SHEAN, M. L., PAEPER, B. W. & DUESTER, G. 1991. The role of CCAAT/enhancer-binding protein in the differential transcriptional regulation of a family of human liver alcohol dehydrogenase genes. *Journal of Biological Chemistry,* 266**,** 11594-11603.

SU, J.-S., TSAI, T.-F., CHANG, H.-M., CHAO, K.-M., SU, T.-S. & TSAI, S.-F. 2006. Distant HNF1 Site as a Master Control for the Human Class I Alcohol Dehydrogenase Gene Expression. *Journal of Biological Chemistry,* 281**,** 19809-19821.

SZALAI, G., XIE, D., WASSENICH, M., VERES, M., CECI, J. D., DEWEY, M. J., MOLOTKOV, A., DUESTER, G. & FELDER, M. R. 2002. Distal and proximal cis-linked sequences are needed for the total expression phenotype of the mouse alcohol dehydrogenase 1 (Adh1) gene. *Gene,* 291**,** 259-270.

SZUTORISZ, H., DILLON, N. & TORA, L. S. 2005. The role of enhancers as centres for general transcription factor recruitment. *Trends in biochemical sciences,* 30**,** 593-599.

THE 1000 GENOMES PROJECT, C. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature,* 491**,** 56-65.

THE ENCODE PROJECT, C. 2011. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol,* 9**,** e1001046.

THE INTERNATIONAL HAPMAP, C. 2005. A haplotype map of the human genome. *Nature,* 437**,** 1299-1320.

THOMAS, M. C. & CHIANG, C.-M. 2006. The General Transcription Machinery and General Cofactors. *Critical Reviews in Biochemistry and Molecular Biology,* 41**,** 105-178.

THOMASSON, H., CRABB, D., EDENBERG, H. & LI, T.-K. 1993a. Alcohol and aldehyde dehydrogenase polymorphisms and alcoholism. *Behavior Genetics,* 23**,** 131-136.

THOMASSON, H. R., CRABB, D. W., EDENBERG, H. J. & LI, T.-K. 1993b. Alcohol and aldehyde dehydrogenase polymorphisms and alcoholism. *Behavior Genetics,* 23**,** 131-136.

TREUTLEIN, J., CICHON, S., RIDINGER, M. & ET AL. 2009. GEnome-wide association study of alcohol dependence. *Archives of General Psychiatry,* 66**,** 773-784.

TREUTLEIN, J. & RIETSCHEL, M. 2011. Genome-Wide Association Studies of Alcohol Dependence and Substance Use Disorders. *Current Psychiatry Reports,* 13**,** 147-155.

TRONCHE, F. & YANIV, M. 1992. HNF1, a homeoprotein member of the hepatic transcription regulatory network. *BioEssays,* 14**,** 579-587.

TSUKADA, J., YOSHIDA, Y., KOMINATO, Y. & AURON, P. E. 2011. The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription

factors is a multifaceted highly-regulated system for gene regulation. *Cytokine,* 54**,** 6-19.

UZZAMAN, S., GURUVAYOORAPPAN, C. & BERLIN GRACE, V. M. 2011. All Trans Retinoic Acid and Cancer. *Immunopharmacology and Immunotoxicology,* 33**,** 241-249.

VAGLENOVA, J., MARTÍNEZ, S. E., PORTÉ, S., DUESTER, G., FARRÉS, J. & PARÉS, X. 2003. Expression, localization and potential physiological significance of alcohol dehydrogenase in the gastrointestinal tract. *European Journal of Biochemistry,* 270**,** 2652-2662.

VALENZUELA, L. & KAMAKAKA, R. T. 2006. Chromatin Insulators*. *Annual Review of Genetics,* 40**,** 107-138.

VAN BORTLE, K. & CORCES, V. G. 2012. Nuclear Organization and Genome Function. *Annual Review of Cell and Developmental Biology,* 28**,** 163-187.

VILLA-DIAZ, L. G., ROSS, A. M., LAHANN, J. & KREBSBACH, P. H. 2013. Concise Review: The Evolution of human pluripotent stem cell culture: From feeder cells to synthetic coatings. *STEM CELLS,* 31**,** 1-7.

VISAPÄÄ, J. P., GÖTTE, K., BENESOVA, M., LI, J., HOMANN, N., CONRADT, C., INOUE, H., TISCH, M., HÖRRMANN, K., VÄKEVÄINEN, S., SALASPURO, M. & SEITZ, H. K. 2004. Increased cancer risk in heavy drinkers with the alcohol dehydrogenase 1C*1 allele, possibly due to salivary acetaldehyde. *Gut,* 53**,** 871-876.

VOSTROV, A. A. & QUITSCHKE, W. W. 1997. The Zinc Finger Protein CTCF Binds to the APBβ Domain of the Amyloid β-Protein Precursor Promoter: EVIDENCE

FOR A ROLE IN TRANSCRIPTIONAL ACTIVATION. *Journal of Biological Chemistry,* 272**,** 33353-33359.

WALL, T. L. 2005. Genetic Associations of Alcohol and Aldehyde Dehydrogenase With Alcohol Dependence and Their Mechanisms of Action. *Therapeutic Drug Monitoring,* 27**,** 700-703.

WALLACE, J. A. & FELSENFELD, G. 2007. We gather together: insulators and genome organization. *Current opinion in genetics & development,* 17**,** 400-407.

WANG, H., MAURANO, M. T., QU, H., VARLEY, K. E., GERTZ, J., PAULI, F., LEE, K., CANFIELD, T., WEAVER, M., SANDSTROM, R., THURMAN, R. E., KAUL, R., MYERS, R. M. & STAMATOYANNOPOULOS, J. A. 2012. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Research,* 22**,** 1680-1688.

WANG, J. C., GRUCZA, R., CRUCHAGA, C., HINRICHS, A. L., BERTELSEN, S., BUDDE, J. P., FOX, L., GOLDSTEIN, E., REYES, O., SACCONE, N., SACCONE, S., XUEI, X., BUCHOLZ, K., KUPERMAN, S., NURNBERGER, J., RICE, J. P., SCHUCKIT, M., TISCHFIELD, J., HESSELBROCK, V., PORJESZ, B., EDENBERG, H. J., BIERUT, L. J. & GOATE, A. M. 2008. Genetic variation in the CHRNA5 gene affects mRNA levels and is associated with risk for alcohol dependence. *Molecular Psychiatry,* 14**,** 501-510.

WANG, J. C., HINRICHS, A. L., STOCK, H., BUDDE, J., ALLEN, R., BERTELSEN, S., KWON, J. M., WU, W., DICK, D. M., RICE, J., JONES, K., NURNBERGER, J. I., TISCHFIELD, J., PORJESZ, B., EDENBERG, H. J., HESSELBROCK, V., CROWE, R., SCHUCKIT, M., BEGLEITER, H., REICH,

T., GOATE, A. M. & BIERUT, L. J. 2004. Evidence of common and specific genetic effects: association of the muscarinic acetylcholine receptor M2 (CHRM2) gene with alcohol dependence and major depressive syndrome. *Human Molecular Genetics,* 13**,** 1903-1911.

WEI, S., LIU, Z., ZHAO, H., NIU, J., WANG, L.-E., EL-NAGGAR, A. K., STURGIS, E. M. & WEI, Q. 2010. A single nucleotide polymorphism in the alcohol dehydrogenase 7 gene (alanine to glycine substitution at amino acid 92) is associated with the risk of squamous cell carcinoma of the head and neck. *Cancer,* 116**,** 2984-2992.

WESTERLUND, M., BELIN, A. C., FELDER, M. R., OLSON, L. & GALTER, D. 2007. High and complementary expression patterns of alcohol and aldehyde dehydrogenases in the gastrointestinal tract. *FEBS Journal,* 274**,** 1212-1223.

WETH, O. & RENKAWITZ, R. 2011. CTCF function is modulated by neighboring DNA binding factors. *Biochemistry and Cell Biology,* 89**,** 459-468.

WHITFIELD, J. B. 1997. Meta-analysis of the effects of alcohol dehydrogenase genotype on alcohol dependence and alcoholic liver disease. *Alcohol and Alcoholism,* 32**,** 613-619.

WHITFIELD, J. B. 2002. Alcohol Dehydrogenase and Alcohol Dependence: Variation in Genotype-Associated Risk between Populations. *American journal of human genetics,* 71**,** 1247-1250.

WILLIAMS, A. & FLAVELL, R. A. 2008. The role of CTCF in regulating nuclear organization. *The Journal of Experimental Medicine,* 205**,** 747-750.

XIA, J., GAO, J., INAGAKI, Y., KOKUDO, N., NAKATA, M. & TANG, W. 2013. Flavonoids as potential anti-hepatocellular carcinoma agents: Recent approaches using HepG2 cell line. . *Drug Discov Ther. ,* 7**,** 1-8.

XU, J. 2001. Preparation, Culture, and Immortalization of Mouse Embryonic Fibroblasts. *Current Protocols in Molecular Biology.* John Wiley & Sons, Inc.

XUEI, X., FLURY-WETHERILL, L., BIERUT, L., DICK, D., NURNBERGER, J., FOROUD, T. & EDENBERG, H. J. 2007. The opioid system in alcohol and drug dependence: Family-based association study. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* 144B**,** 877-884.

YANG, S.-J., YOKOYAMA, A., YOKOYAMA, T., HUANG, Y.-C., WU, S.-Y., SHAO, Y., NIU, J., WANG, J., LIU, Y., ZHOU, X.-Q. & YANG, C.-X. 2010. Relationship between genetic polymorphisms of ALDH2 and ADH1B and esophageal cancer risk: A meta-analysis. *World J Gastroenterol,* 16**,** 4210-4220.

YANG, Z.-N., DAVIS, G. J., HURLEY, T. D., STONE, C. L., LI, T.-K. & BOSRON, W. F. 1994. Catalytic Efficiency of Human Alcohol Dehydrogenases for Retinol Oxidation and Retinal Reduction. *Alcoholism: Clinical and Experimental Research,* 18**,** 587-591.

YIN, S.-J., CHOU, C.-F., LAI, C.-L., LEE, S.-L. & HAN, C.-L. 2003. Human class IV alcohol dehydrogenase: kinetic mechanism, functional roles and medical relevance. *Chemico-Biological Interactions,* 143–144**,** 219-227.

YIN, S.-J., CHOU, F.-J., CHAO, S.-F., TSAI, S.-F., LIAO, C.-S., WANG, S.-L., WU, C.-W. & LEE, S.-C. 1993. Alcohol and Aldehyde Dehydrogenases in Human

Esophagus: Comparison With the Stomach Enzyme Activities. *Alcoholism: Clinical and Experimental Research,* 17**,** 376-381.

YIN, S.-J., HAN, C.-L., LIAO, C.-S. & WU, C.-W. 1997. Expression, Activities, and Kinetic Mechanism of Human Stomach Alcohol Dehydrogenase. Inference for first-pass metabolism of ethanol in mammals. *Advances in Experimental Medicine and Biology,* 414**,** 347-355.

YIN, S.-J., WANG, M.-F., LIAO, C.-S., CHEN, C. M. & WU, C.-W. 1990. Identification of a human stomach alcohol dehydrogenase with distinctive kinetic properties. *Biochem Int.,* 22**,** 829-35.

YOKOYAMA, A., MURAMATSU, T., OHMORI, T., YOKOYAMA, T., OKUYAMA, K., TAKAHASHI, H., HASEGAWA, Y., HIGUCHI, S., MARUYAMA, K., SHIRAKURA, K. & ISHII, H. 1998. Alcohol-related cancers and aldehyde dehydrogenase-2 in Japanese alcoholics. *Carcinogenesis,* 19**,** 1383-1387.

YOKOYAMA, A. & OMORI, T. 2003. Genetic Polymorphisms of Alcohol and Aldehyde Dehydrogenases and Risk for Esophageal and Head and Neck Cancers. *Japanese Journal of Clinical Oncology,* 33**,** 111-121.

YOKOYAMA, H., YOKOYAMA, H.-S. & HIBI, T. 2012. The Interaction Between the Metabolism of Retinol and Ethanol in Esophageal Mucosa – A Possible Mechanism of Esophageal Cancer in Alcoholics,. *In:* JAZII, P. F. R. (ed.) *Esophageal Cancer - Cell and Molecular Biology, Biomarkers, Nutrition and Treatment.*

YOON, G., WESTERMEYER, J., KUSKOWSKI, M. A. & NESHEIM, L. 2013. Impact of the number of parents with alcohol use disorder on alcohol use disorder in offspring: a population-based study. *J Clin Psychiatry.,* 74**,** 795-801.

ZACHMAN, R. D. & GRUMMER, M. A. 1998. The Interaction of Ethanol and Vitamin A as a Potential Mechanism for the Pathogenesis of Fetal Alcohol Syndrome. *Alcoholism: Clinical and Experimental Research,* 22**,** 1544-1556.

ZELNER, I. & KOREN, G. 2013. Alcohol Consumption Among Women. *J Popul Ther Clin Pharmacol* 20**,** e201-e206.

ZINTZARAS, E., STEFANIDIS, I., SANTOS, M. & VIDAL, F. 2006. Do alcohol-metabolizing enzyme gene polymorphisms increase the risk of alcoholism and alcoholic liver disease? *Hepatology,* 43**,** 352-361.

ZUO, L., ZHANG, H., MALISON, R. T., LI, C.-S. R., ZHANG, X.-Y., WANG, F., LU, L., LU, L., WANG, X., KRYSTAL, J. H., ZHANG, F., DENG, H.-W. & LUO, X. 2013a. Rare ADH Variant Constellations are Specific for Alcohol Dependence. *Alcohol and Alcoholism,* 48**,** 9-14.

ZUO, L., ZHANG, X.-Y., WANG, F., LI, C.-S. R., LU, L., YE, L., ZHANG, H., KRYSTAL, J. H., DENG, H.-W. & LUO, X. 2013b. Genome-Wide Significant Association Signals in IPO11-HTR1A Region Specific for Alcohol and Nicotine Codependence. *Alcoholism: Clinical and Experimental Research,* 37**,** 730-739.

CURRICULUM VITAE

Sowmya Jairam

EDUCATION AND EXPERIENCE

INDIANA UNIVERSITY, Indianapolis, IN (2006-2014)
Major: PhD in Biochemistry and Molecular Biology
***Minor**: Cancer Biology/Oncology*

Dissertation: Identified complex interactions between single nucleotide polymorphisms (SNPs) that determine level of transcriptional activity of alcohol dehydrogenase 7 gene (*ADH7)*, and a chromatin insulator that determines fate of *ADH7* transcription – thus providing, for the first time, an overview of *ADH7* gene regulation.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, Hyderabad, India **(**2002-2006)
Bachelors of technology in Biotechnology
GPA (%): 78% (First class with distinction) ● Ranked 1st in national level science symposium ● Student representative to school management ● School Liaison officer

PUBLICATIONS

- **Jairam S** and Edenberg HJ. Single Nucleotide Polymorphisms interact to affect *ADH7* transcription *[in press]*.
- **Jairam S** and Edenberg HJ. The enhancer blocker iA1C is involved in the cell-specific expression of alcohol dehydrogenase *ADH7 [submitted for review]*.

ABSTRACTS AND PRESENTATIONS

- **Jairam S,** and Edenberg H. "*ADH7 gene regulation"*, Dept of Biochemistry and Molecular biology, Indiana University School of Medicine, 2012.
- **Jairam S,** and Edenberg H. "*ADH7 gene regulation by proximal and distal regulatory elements*", Dept of Biochemistry and Molecular biology, Indiana University School of Medicine, 2010.
- **Jairam S**. and Gayatri, V. "*Mutations and Mutagenesis",* National symposium at Jawaharlal Nehru technological University, 2005.