Addiction, Mental Health, and Infectious Disease: A complex web of genetic interactions

A Thesis

Submitted to the Faculty

of

Drexel University

by

Latifa Fatima Jackson

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

October 2014

DEDICATIONS

This work is dedicated to my parents Robert and Fatimah and my siblings: Malik, Sekou, Aisha, Hasan and Salim, with innumerable thanks for your lifetime of support.

My husband Dr. Raouf Ghomrasni whose encouragement meant everything.

And Allah (SWT), who makes all things possible.

ACKNOWLEDGMENTS

1. Dr. Aydin Tozeren- whose advice and support has been invaluable.
2. My committee Members:
    a. Dr. Uri Hershberg
    b. Dr. Ahmet Sacan
    c. Dr. Donald McEachron
    d. Dr. Rene Doursat
3. The Tozeren Lab especially to Ceylan Tanes and Max Shestov. I have benefitted greatly from our collaborations, your helpful discussion, support and generosity.

Contents

List of Tables 7
List of Figures 9
ABSTRACT 13
Introduction 16
Opioid, Dopamine and GABA Addiction 23
Addiction Intersections with Schizophrenia, Depression and Bipolar Disorder 24
Infectious disease and Immunity 25
Specific Aim I: Identification of Addiction Hotspots 27
Introduction 27
METHODS 30
Addiction linked genes and genome hotspots 30
Functional Annotation of Hotspot Genes 31
Genetic variation in hotspots in population subtypes 31
RESULTS 33
Molecular functions and biological processes of addiction genes 33
Identification and annotation of addiction hotspots on the human genome 38
Addiction hotspots exhibit population subtype dependent polymorphisms 45
Distant connections of hotspot polymorphisms 52
DISCUSSION 54
Specific Aim II: Addiction and Mental Health 59
Introduction 59
Methods 60
Addiction linked genes and genome hotspots 60
Functional Annotation of Addiction and Mental Health Hotspot Genes 61
Results 62
Overlap in Gene Lists for Addiction, Schizophrenia, Bipolar Disorder and Depression 62
Functional annotation of Schizophrenia, bipolar, depression and addiction genes 65
Numerous targets found in Hotspot windows 66
Discussion 69
Specific Aim III: Addiction and Immunity 73
Introduction 73

Methods 75
Identification of Immune and addiction gene clusters 75
Annotation of Genes in the Addiction and Immunity Windows 76
Populations Studied and Determination of Environmental Context 76
Geographical mapping of correlated variants 79
Results 80
Addiction and Immunity Hotspots 80
Addiction and Immunity Hotspot Annotation 80
Annotation of Tropical identified SNPs 83
Acute Inflammatory Response and Hepatitis B 89
Discussion 93
Conclusions 97
Summary of Findings 97
Appendices 116
Appendix A: KEGG Pathways of Morphine Addiction in humans. 116
Appendix B: KEGG Pathway of Dopaminergic Synapse in Humans. 117
Appendix C: KEGG Pathway of Systematic Lupus Erythematosus in humans. 118
Appendix D: Final SNP Table- All SNPs generated for the all allelic polymorphisms found in the genome. 119
Appendix E: GWAS3D- Identification of the European long distance interactions 120
Curriculum Vitae 121

.

List of Tables

Table 1: Population subtypes considered at Addiction Hotspots. The eleven HapMap populations grouped into geographical regions. Populations are described by ethnic identity and by their sample collection location.

Table2: Hypergeometric test

Table 3: Addiction genes within hotspots show distinct functional classification. The table shows the gene ontology molecular functions and biological processes statistically enriched in addiction hotspots (Benjamini coefficient < 0.03). Annotations in bold correspond to the addiction genes identified within a cluster and not the entire genome region.

Table 4: SNPs identified as significant in comparisons between Africans, Asians and Europeans for the 6p21.2 addiction hotspot. All SNPs were tested for significant differences in allele frequencies using a chi square test (cutoff < 1e-5) in pairwise regional comparisons: Africans (ASW, YRI, LWK, and MKK), Europeans (CEU and TSI), and Asians (CHB, CHD, JPT). The Allele frequencies of variants in populations are given with the p values of the grouped regional ethnic populations.

Table 5: GWAS derived Addiction Alleles showed significant differences in allele frequencies between Asians and Non-Asian populations. Alleles identified as significant in GWAS studies were typed in the 11 HapMap populations.

Table 6: Addiction, bipolar disorder, depression, and schizophrenia form eight genomic hotspots. When genes were mapped to the genome, we found eight regions smaller than 1.5 Mb each with 11- 38 genes identified from NCBI Gene curated gene lists. Additionally we consider the genes interspersed with our set of addiction and mental health genes.

Table 7: Drug binding sites were annotated for the genes in the hotspots. When genes identified in the hotspots were annotated for drug interactions, we found 16 drug binding sites for both commercially developed and illicit substances.

Table 8: Genes identified at the intersection of Addiction and Immunity.

Table 9: Functional Annotation of Addiction and Immunity hotspot crosstalk regions.

Table 10: Top five SNPs have shared patterns of allele frequency differences in HapMap populations as in HGDP populations. SNPs are identified with their location type and gene.

Table 9: GWAS3D analysis of the tropical segregating polymorphisms showed interacting SNPs, transcriptions factors and ACP2, a gene involved in

de-phosphorylation. We characterized the set of SNPs that were frequency divergent in tropical and non-tropical living populations. The set had interacting SNPs that were locally located, often in the same gene with the exception of rs901746 which interacted with rs2167079 at ACP2, an adjacent gene. Three of the interacting SNPs sat within transcription factors, with the ACP2 gene SNP (rs2167079) sat at the intersection of 62 different transcription factors.

List of Figures

Figure 1: Intersection of the Addiction gene sets identified in KEGG, NCBI OMIM and NCBI Gene. Genes involved in dopamine, opiate and GABA addiction were surveyed in each data set including: the Kyoto Encyclopedia of Genes and Genomes (N=139 genes), The Online Mendelian Inheritance Map (N= 42 genes) and the National Center for Bioinformatics Information Gene (N=587 genes).

Figure 2: Flowchart of dissertation inputs and outputs. NCBI Gene searches will be used to identify complex disease hotspots, and then examining the variation found in human populations derived from the HapMap SNP dataset, the Human Genome Diversity Panel, and the 1000 Genomes SNP dataset.

Figure 3: Flowchart for identifying biologically relevant addiction genes. The search terms used at NCBI Gene to populate a list of unique addiction genes are shown. The heat maps indicate intersection of gene sets in three classes of addiction: dopamine, opiate, and GABA. Addiction hotspots were defined at a genomic region with six or more addiction genes within a 1-1.5Mb genomic window. A Venn diagram shows the comparison with the 387 genes identified through an alternate addiction study(1).

Figure 4: Gene ontology molecular function annotations for dopamine, GABA and opiate addictions. The top 75 significant functional annotations were obtained (Benjamini > 0.01) for gene ontology molecular function for each addiction gene class. The vertical axis of each graph shows the $-\log_{10}$ (p) while the horizontal axis indicates the functional annotations. The gene lists are color coded for opiates (green), GABA (red), and dopamine (blue).

Figure 5: Genes shared between dopamine, opiate and GABA addiction sets. The figure provides the gene symbols for intersections of gene sets corresponding to dopamine, opiate, and GABA addictions.

Figure 6: Seven addiction hotspots identified on the human genome. Each pie chart represents a hotspot with the scaled size in proportion to the total number of genes within the 1- 1.5Mb window. Three of the hotspots were functionally related exclusively to GABA function while four hotspots were mixed in addiction activity. Genes that lie within each hotspot were identified through USCS Genome Browser.

Figure 7: Hotspots participating in acute and chronic alcoholism pathways. The alcohol addiction pathway contains genes located in all seven of the addiction hotspots, each are colored according to the legend. They participate in post-synaptic processes in both acute and chronic alcohol signaling. Additionally each component in the alcoholism pathway has a bar above the gene/component indicating their involvement in dopamine, GABA or opiate addiction.

Figure 8: Allelic distributions at addiction hotspots show population and regional variation. Each graphic considers polymorphism distributions of addiction hotspots

against the neutral distribution (in red). Hotspot distributions were tested against the neutral loci using the non-parametric Kolmogorov-Smirnov test. Distribution graphs with grey-scaled lines are those not significantly different from the neutral distribution.

Figure 9: Heat maps of SNP distribution on hotspots show population and geographical Patterning. Comparisons of the average reference allele frequencies are shown for each addiction hotspot across populations using -log10 of the p values obtained from the Kolmogorov-Smirnov test. The variation in neutral regions shows very little demographic patterning while the hotspot located at Chr6: 25.7-26.4 shows considerably more regional blocks.

Figure 10: Hotspot SNPs with significantly altered expression in regional population subtypes. Significant polymorphisms were found for regional population comparisons between African-Asian (N=112), African-European (N=126), and European-Asian (N=122) using chi squared test with a p value cutoff of 1e-5 as shown by the blue line in the figure. The value for the red line cut off is 5e-8.

Figure 11: Significant SNPs characterized in the Yoruba population show local and long range interactions. Polymorphisms identified as significant in African to European comparison were projected onto the GWAS3D platform. At least six long-range trans-chromosomal interactions were identified and numerous local interactions were observed.

Figure12: Genes shared between dopamine, opiate and GABA addiction sets with common mental health conditions. The Venn diagram provides the gene symbols (N=1968 genes) for intersections of gene sets corresponding to dopamine, opiate, and GABA addictions with bipolar disorder, depression, and schizophrenia. The gray boxed genes are those genes common to all disorders.

Figure 13: Eight bipolar, depression, schizophrenia and addiction hotspots identified on the human genome. Each karyotype of a chromosome shows the location of the genes that are represents as hotspots. Genes that lie within each hotspot were identified through USCS Genome Browser. The insets for each hotspot show genes identifies curated genes (green) and interspersed genes (black) with a red band showing the entire hotspot region.

Figure 14: Gene ontology molecular function annotations for addiction, bipolar disorder, depression and schizophrenia. The top 25 significant functional annotations were obtained (Benjamini > 0.01) for gene ontology molecular function for each addiction, bipolar, depression, schizophrenia gene class. The vertical axis of each graph shows the $-\log_{10}$ (p) while the horizontal axis indicates the functional annotations. The gene lists are color coded for depression (green), bipolar disorder (red), schizophrenia (blue). Addiction genes are a composite of opiate, GABA, and dopamine addiction genes (purple). Shown are those 13 functions that are shared between the gene sets.

Figure 15: Ethnic populations surveyed for immunity and addiction crosstalk from the HapMap and Human Genome Diversity Project. The HapMap populations (populations described in Table 1) and the HGDP populations: HGDP Africans (green)- Moazibite, Mandenka, Yoruba, Biaka, Mbuti, San, NE Bantu, and SAf Bantu. HGDP Asians (purple); N. Asia: Oroquen, Daur, N. Han, Hezhen, Japanese, Uygur, Xibo); C. China: Han, Yi, She, Tu; S.E. Asia: Naxi, Lahu, Dai, Miao, Cambodian), S. Asian (burgundy): HGDP Europeans (blue): Adygei, Basque, French, North Italian, Orcadian, Russian, Sardinian, and Tuscan ; and HGDP Oceanians (orange): Melanesia, Papua

Figure 16: Five SNPs underlying the Chr 11 hotspot vary between high tropical versus low tropical living HGDP populations. Allele frequencies were plotted in the tropical living climates (blue) versus those living in temperate climates (red). The inset shows the hotspot locations.

Figure 17: GWAS3D analysis of the five single nucleotide polymorphisms found to differ in tropical versus non-tropical living populations show that SNPs have local and long range interactions. AMBRA1 showed 2 long range interactions on chromosomes 3 and 8 along with a local interaction on chromosome 11. The ACP2 gene (rs2167079), an interacting partner of rs901746, had three interactions: at Xq26.1, 1q21.1, and 7p15.2

Fig 18: Neutral SNPs in global human populations and the Acute Inflammatory response gene set surveyed in a diverse set of African and East Asian populations. Neutral SNPs only show significant differences between Africans and non-African populations. A Pearson correlation of Hepatitis B prevalence and mean acute inflammatory response allele frequency shows a trend toward correlation ($p=0.08$).

ABSTRACT

Opiate, dopamine and GABA addictions are complex diseases with strong genetic components. These three substance disorders represent significant costs to the global judicial and healthcare systems. The treatment of addiction is further confounded by the co-occurrence of other pathologies that complicate treatment regimes. For example, addiction and mental health are well-characterized co-morbidities. Mental health conditions such as depression, bipolar disorder and schizophrenia have clear genetic synergies between the prevalence of one mental health condition and addiction. This dissertation focuses on the characterization of addiction hotspots in the genome, their interplay with mental health genetics and then examines how infectious disease burden is correlated to the rise of immune and addiction variants.

Molecular genetics, metabolism analyses, epigenetic and association studies have contributed to current understandings of the genetic components of addiction disorder phenotypes. The resulting literature curated gene sets can be used to identify the modules and pathways mediating shared addiction, mental health and immune disorders. Studying addiction, mental health and immune genes in a geographically diverse sample of human populations is critical to understanding the role that evolutionary factors play in the rise and maintenance of variation potentially underlying addiction phenotypes. These human population comparisons are possible due to the recent expansion of human polymorphism databases, such as the HapMap Project, the Human Genome Diversity Panel and the 1000 genomes datasets. Careful comparisons of allele frequencies in human populations can point to those polymorphisms for which

both functional and evolutionary histories converge to either promote or inhibit addiction, mental health, and immune susceptibility.

We can project curated addiction genes onto gene ontology categories and cellular pathways to draw a bioinformatics portrayal of addiction and its interplay with mental health and immunity. These addiction genes lists as well as schizophrenia, depression, and bipolar disorder gene sets can be further projected onto the genome to portray the overlap between addiction and mental health disorders. This can also serve as a tool to discover additional genes that play a candidate role in mental illness and addiction. Functionally annotating these regions using existing databases such as the Kyoto Encyclopedia of Genes and Genomes allows for robust characterization of the roles that genes and genomic regions play in modulating addiction phenotypes. This approach enables the identification of candidate genes sitting adjacent to known addiction hotspot genes and the subsequent identification of the candidate polymorphisms in a diverse array of human populations. Finally the addition of new databases of genome wide association studies can inform candidate polymorphisms for addiction, mental health and immune response to infectious disease.

Introduction

A central tenet of evolutionary analysis is the idea that natural selection influences allele frequencies in populations affected by an environmental selective agent. The mechanism of natural selection is governed by four principles: the occurrence of trait or phenotypic variation in populations, the fact that trait variation is heritable through Mendelian genetic mechanisms, that this genetic variation leads to differential fitness and survivability, there is competition among individuals in the population with those of highest fitness contributing the most to reproduction (2, 3). These principles were further supported by the evolutionary modern synthesis in the 1930s, which merged Darwinian natural selection with Mendelian mechanisms of genetic inheritance. With both the mechanisms and the mode of inheritance clarified, many turned their attention to the selective agents that could create differences in allele frequencies in human populations. In 1949, Haldane predicted that infectious disease what a primary driver of human immune variation (4). No selective agents have been characterized as more important in human history in shaping allele frequency differences in and between human populations than infectious disease (2, 5-17). This view has been largely adopted to become prevalent in human evolutionary biology, human genetics and has served as the foundational underpinnings of the movement towards personalized genetics.

This view has been borne out by numerous studies of human immunity related allozymes conducted on infectious diseases such as the host-pathogen-vector triangle that characterizes Malaria (*P. falciparum* in Africa and *P. vivax* in Asia) (18). Malaria is a widely prevalent disease currently affecting as many as 2.7 billion people, and killing

approximately 1 million people a year, most of whom are young and living in Africa (19-22). As such, malarial disease is one of the most significantly impactful infectious diseases studied in recorded human history. It is known to have shaped human history and survival in equatorial regions, and has been documented to have had extended geographical ranges. Recent literature has identified that *Plasmodium falciparum* has been thought to have spread approximately 13,000 years ago in concordance with Neolithic human expansion (23). This corresponds to estimates of dating for malarial disease in human populations (24, 25). The severity of infectious diseases such as malaria infection in current human populations and the supposition that malaria has exerted similar types of selection pressures throughout its shared history with *Homo sapiens* makes it and other infectious disease strong candidates for disease- driven selection in humans (19, 20, 26, 27).

Gene studies associated with infectious diseases have fallen in three classes: erythrocyte, cell receptor, and immune genes. We previously lacked the necessary insights into the full scope of human loci impacted by interactions with infectious diseases in human hosts. This predicament stemmed largely from technological limitations to large scale molecular locus characterization. This has been largely rectified with pyrosequencing techniques and high throughput sequencing to discern both genomic and polymorphic regions of interest (28-31) . Recent advances in sequencing and functional genomic typing of genic regions has greatly broadened the extent and depth of our understanding of molecular targets of human evolution (29).

Infectious disease continues to be viewed as the major cause of the differential survival that is known to shape allele frequency differences. We have noted that the study of selection in human populations rarely makes the leap to what fitness or evolutionary consequences that human migration (largely to the United States) via the trans-Atlantic slave trade and American immigration has caused. In my dissertation work, I will look for links between the immune driven selection on infectious disease traits and common chronic diseases for which American are co-susceptible- addiction disorders and mental health conditions (32, 33). These two broad disorders are often co-morbidities with immune deterioration in chronically mental ill (34, 35) or with chronic substance abusers (36-39). We chose these two broad classes of disorders based on their co-morbidity in American populations (32, 36, 40, 41), and the shared clinical manifestations among chronic dopamine, opiate and GABA addiction sufferers (32, 39, 42).

Given the confluence of these two chronic conditions in human populations and the observation that they both share strong immunity overlaps in the clinical literature, it would be particularly interesting to characterize how much genetic overlap exists in these seemingly etiologically different disorders. We choose to study the most common illicit addictive substances- cocaine, methamphetamine, heroin, morphine, alcohol, and GHB. These six addictive substances fall into three drug classes: dopamine, opiate and GABA receptor based addictions they were also chosen based on their clinical similarities in chronic abusers.

Currently we have identified the genetic inputs to our analyses (immune, addiction and mental health genes), and the evolutionary framework within which we are interested in contextualizing our work. When we considered how to best exploit the publicly available datasets, we considered those curated by the three datasets: Online Mendelian Inheritance Map (OMIM), the National Center for Bioinformatics Information Gene (NCBI Gene), and The Kyoto Encyclopedia of Genes and Genomes (KEGG). Each of these datasets contains genes identified through experimental or clinical studies to be associated with a specified addiction, mental health or immunity disorder. OMIM is a compendium of human genes and genetic phenotypes (43-45). OMIM is curated by Johns Hopkins University and housed at both omim.org and at NCBI's portal. The NCBI GENE dataset is curated database of genes derived from fully sequenced genomes housed on the NCBI web portal. NCBI Gene integrates information from OMIM, and creates links to OMIM, at both the gene and the phenotype levels (Gene Help). KEGG is an interactive database that seeks to unify molecular processes, genetics and chemical interactions together to form a cohesive picture of how genes products interact with each other and their environment (46-50). KEGG can be queried to address both how genes of interest interact in genome and to identify which genes act in particular phenotypes or disorders(47, 48). KEGG is curated by the Kanehisa Labs jointly at the Kyoto University and the University of Tokyo. Each of these datasets provides some benefits for our analyses input files.

In order to assess which dataset is most appropriate the mine for the initial list of genes that will serve as the input upon which we will perform computational analyses, we

surveyed the opiate, dopamine, and GABA addiction in all three datasets and determined which one had the most robust gene list with which to begin analyses. In Figure 1, we demonstrate that OMIM contained 42 gene results, NCBI GENE contained 587 gene results and KEGG contained 139 gene results. Additionally NCBI GENE contained all but 50 genes of the OMIM and KEGG gene sets. Based on these analyses and the previously stated linkages between OMIM and NCBI GENE, we have determined that NCBI GENE is the most robust dataset with respect to diversity of the genes represented and inclusion of the genes captured by other publicly available gene datasets.

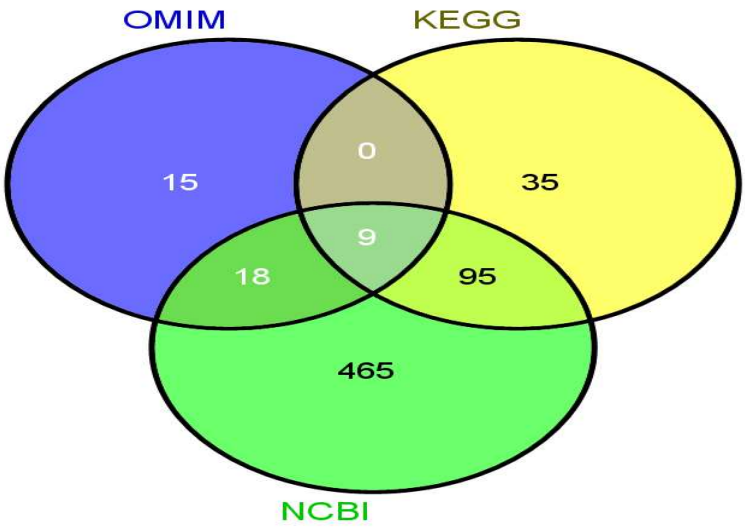


Figure 1: Intersection of the Addiction gene sets identified in KEGG, NCBI OMIM and NCBI Gene. Genes involved in dopamine, opiate and GABA addiction were surveyed in each data set including: the Kyoto Encyclopedia of Genes and Genomes (N=139 genes), The Online Mendelian Inheritance Map (N= 42 genes) and the National Center for Bioinformatics Information Gene (N=587 genes).

After determining that we could use the NCBI dataset as a strong input set, we wanted to outline the workflow that we intend to employ during the execution of this dissertation.

In Figure 2 we outline the work flowchart for this dissertation. This dissertation seeks to identify significant human SNPs of interest in chronic disorders for which we have limited understanding of the biomarkers for disease. I propose to use gene sets identified through the NCBI Gene lists which are by and large the most complete data sets of their kind. These gene lists for selected types of addiction, mental health and immune disorders will be mined for each disease constellation of interest.

We will use NCBI gene sets as inputs covering addiction (specific aim 1), addiction with mental health (specific aim 2), and addiction with immunity (specific aim 3). We will then conduct computational analyses to determine whether genes involved in different disorders shared genomic locality. If this was the case then we could use these locations to characterize the human variation present and make inference about the relationship between this variation and geographical/environmental variation.

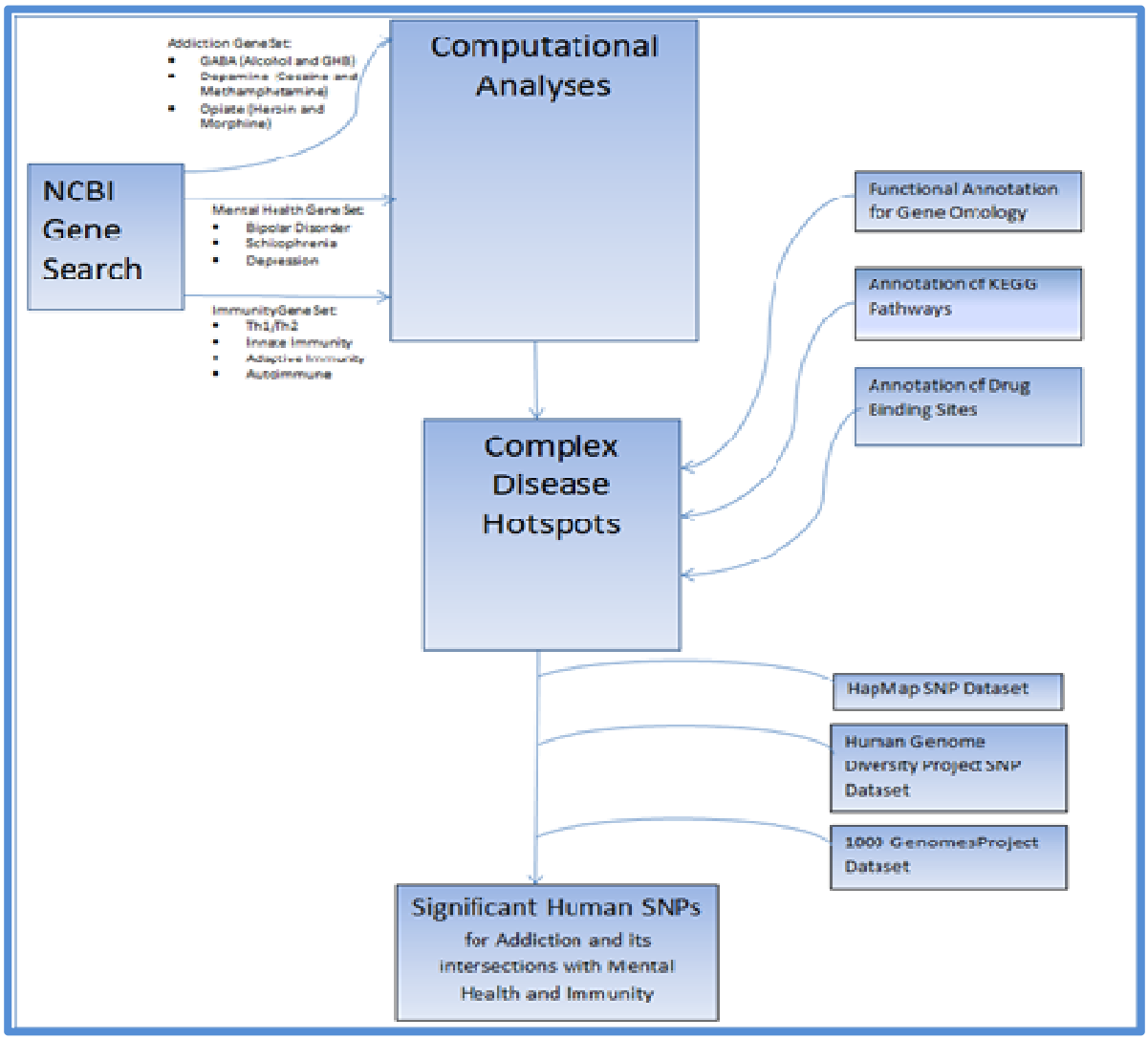


Figure 2: Flowchart of dissertation work inputs and outputs. NCBI Gene searches will be used to conduct computational analyses, identify complex disease hotspots, and then examining the variation found in human populations derived from the HapMap SNP dataset, the Human Genome Diversity Panel, and the 1000 Genomes SNP dataset.

Opioid, Dopamine and GABA Addiction

Addiction disorders represent a major economic and health cost to human populations (2). Addiction is loosely defined as a chronic relapsing spectrum disorder characterized by loss of control over substance taking (3-5). It is a behavior-based phenomenon

representing a diverse array of psychological, biological, and genetic attributes and environmental and cultural factors (6-8). While the genetic components of addiction is not fully characterized, clear evidence in the form of single gene association studies, and genome wide association studies suggest a large role for heritable variation in shaping the phenotypes observed in dopamine, opiate and GABA based addictions. These studies provide a robust substrate on which to consider how addiction genes might be working together to create the functional characteristics of addiction seen in clinical populations.

Traditional approaches to studying multiple gene interactions have focused on identifying how genes form gene networks and thereby contribute to functional pathways. These two approaches have provided significant insights into the functional role that genes play in relationship their network neighbor genes. These types of studies have led to the curation of a large list of genetic contributors to complex disease phenotypes such as substance addictions, mental health conditions, and infectious diseases. Currently, substance addiction contains a large set of genes identified as participating to the generation of its phenotypes. Researchers of addiction have also mapped many of these addiction related genes onto gene ontology pathways to further elucidate their role in addiction. This wealth of genetic targets for addiction has not lead to an understanding of which genes or polymorphic variants are the most important in identifying substance addiction susceptibility or addiction disorder progression. Faced with a wealth of genetic targets for analysis, we try to identify those genes that appear to have genetic relationships beyond their roles as functional agents in pathways.

Addiction Intersections with Schizophrenia, Depression and Bipolar Disorder

As with addiction, neurological disorders such as depression, bipolar disorder, and schizophrenia represent a significant strain on health and judicial entities. Both addiction and mental health disorder classes have long been identified as co-morbid conditions (9). While we can make clear inferences about the role that genes play in Mendelian genetic disorders, characterizing the genetic underpinnings of complex disease has been significantly more challenging. The intersection of two complex diseases such as addiction and mental provides an additional level of genetic complexity. Both Substance addiction and mental health are behavior-based phenomenon representing a diverse array of psychological, biological, and genetic attributes and environmental and cultural factors (6-8, 10, 11).

Schizophrenia is a mental illness that affects 1% of the global human population (12). It is identified as a disorder that disrupts brain neural networks and is characterized by hallucinations, delusions, lack of willpower, and cognitive deficits (13). Genealogy studies have shown that there is a strong genetic component to schizophrenia with genetic components accounting for as much as 80% of the risk variance (14). Linkage studies and candidate gene approaches have identified over 1000 candidate genes associated with schizophrenia.

Bipolar disorder is complex and severe mental disorder found in approximately 2.6 percent of American adults (15). It is a disorder with genetic concordance rates for bipolar disorder in twin studies has been estimated to be between 60% and 80% (16).

Along with schizophrenia, there are a number of clinical symptoms overlap with bipolar disorder that may lead to these two disorders being consolidated into one mental health classification (17) . Finally, depression is a mood disorder experienced by 1 in 6 Americans (18). These three mental health conditions taken together represent one of the top 10 diseases in terms of life years lost (19). Examining the intersection of genes involved in these complex disorders allows us to potentially determine the role that each play and how they contribute to trait sharing.

Infectious disease and Immunity

Human immunological interactions with their environment are the substrate for natural selection. One approach to studying natural selection in humans has been to examine single genes in a population to directly assess selection caused by some environmental effect (i.e. HBB and malaria, SLC24A5 and UV exposure)(20, 21). An analysis conducted on global populations has demonstrated success in identifying variation in allele frequencies between populations taking into account diet, subsistence strategy and ecoregions (22, 23).

This approach allows for finer scale patterns of molecular evolution to be observed within a Bayesian framework. This method also harnesses the strengths of a candidate gene approach by selecting metabolic pathways involved in specific traits of interest (24). Hancock et al. find that signatures of selection are seen in multiple pathways in human populations (22, 23). Indeed this approach can be seen as a major innovation in multivariate analysis for factors affecting gene frequencies in human populations. Applying this approach to a chronic disease like addiction will be a novel use of a

relatively underutilized approach. The interplay of addiction, location, and immunity invoking exposure to pathogens are compelling external forces that impact individual survival and therefor gene patterns.

By examining correlations between environmental, addiction, and disease conditions with allele frequencies, I will be able to search for allele frequencies consistent with signatures of selection on multi-locus traits. This method is useful in identifying adaptations (whether tolerance or resistance focused) for complex infectious diseases. Tropical populations provide an opportunity to examine how environmental effects affect complex traits because they exhibit a variety of subsistence strategies, population histories and exposures to infectious disease. They also live in a variety of ecological regions, nutritional contexts and latitudes. These various conditions make them a living laboratory for studies in natural selection. Until recently there was not sufficient sampling coverage in Asian populations to make inference about populations. The confluence of novel population datasets spanning the world and a new methodology that can exploit a candidate gene-like approach to identifying selection in human populations makes tropical dwelling populations a compelling system to assess natural selection events.

Specific Aim I: Identification of Addiction Hotspots

Introduction

Addiction disorders represent a major economic and health cost to human populations (51). Addiction is loosely defined as a chronic relapsing spectrum disorder characterized by loss of control over substance taking (52-54). It is a behavior-based phenomenon representing a diverse array of psychological, biological, and genetic attributes and environmental and cultural factors(55-57). Additionally, individuals addicted to illicit substances are stigmatized, with widespread marginalization of rehabilitation and recovery services, perhaps facilitating return to recidivism (58-60). Current methods in overcoming addiction induced destructive behavior include those emphasizing not the treatment but complete abstinence from addictive substances (37, 61).

A large number of studies including genome wide association investigations has uncovered potentially relevant allelic contributors to the genetic and molecular basis of addiction phenotypes(62-68). Genetic studies such as those conducted on alcohol dehydrogenase variants and the *ADH* family has been important for advancing our understanding of the genetic basis of addiction phenotypes(69, 70). Literature points to two main molecular areas underlying addiction phenotypes. The first is the analysis of disruption of normal ranges of neurological function (71-73). This has been characterized well both in the literature and in the canonical pathways that have been identified as participating in addictions, such as the alcoholism, cocaine, and dopaminergic abuse (74-76). The second is addressed at the level of metabolic function(77) such as the characterization of the ADH gene family and its function in the metabolism of alcohol products (78, 79).

Association studies are yet to provide a complete picture for elucidating the complex pathways and mechanisms underlying addiction phenotypes. The population subtype dependent mechanisms of addiction have not been fully explored (80-82) and so are epigenetic factors involved in addiction (36, 40, 83, 84). A big part of the problem is that tissues involved in addiction, neurons and liver tissue, cannot be experimentally investigated in the human. Hence animal models and in vitro experiments have been used to quantify epigenetic changes along with those related to transcriptome (85-87) , metabolome (88-90), and proteome (91-93) in the neuron synoptic system and in the liver tissue under addiction disorder conditions. These animal studies have shown that there is a robust set of homologous genes and pathways involved in addiction (94).

Studies on the human tended to focus attention on the predominantly European and African American addicted cohorts with few studies examining other populations (95, 96). Familial addiction behavior patterns have long been observed to vary within and between ethnic communities (55, 97) . This has limited our ability to understand how addiction gene variation plays out in the geographically and ethnically diverse human populations that are afflicted by addiction disorders. Using a diverse array of ethnic populations representing major human geographical regions and ancestries would be particularly useful in understanding the role that variation at potential addiction complex sites in the genome plays.

In this study we focus on opiate, dopamine, and GABA addiction disorders. We map literature curated addiction genes onto human chromosomes. Clusters these genes form identify hotspot locations on the genome. Addiction hotspots contain genes previously not linked to addiction as well as regulatory motifs with population subtype dependent polymorphisms. Our bioinformatics based discovery and subsequent investigation of addiction hotspots reveal their roles in addiction as well as the dependence of addiction motifs on population subtypes. Results point out multiple hotspots participating in pathways of addiction. Our findings also point to possible roles of SLC membrane transport proteins on population type dependence of addiction.

Specific Aim: Characterize the role of genome locality on addiction disorders and how that variation is partitioned in human populations

I hypothesize that genes involved in the characterized opiate, dopamine and GABA addiction disorders will form genomic regions of with functional specificity. These regions should exist above and beyond that small subset of genes that are shared between these gene lists. This hypothesis derives from the anecdotal observation that individuals with chronic substance addiction have shared traits that appear to be maintained, despite the divergent mechanisms underlying metabolism of these substances. I further hypothesize that addiction hotspots have ethnic and regional specificity. The rationale behind pursuing this approach is that genes involved in common addiction phenotypes should have common variants underlying them. In order to understand the ways in which genetic variation segregates in human populations and

its potential implications in personalized treatment regimens to addiction, we need to survey how variation at regions of the genome responsible for addiction phenotypes is partitioned in human populations.

METHODS

Addiction linked genes and genome hotspots

A simple search of *NCBI Gene* based on strings of two words shown in Fig. 1 was used in order to generate a list of genes with biological relevance to addiction. We focused on addiction linked to dopamine (cocaine and crystal methamphetamine), opiate (heroine and morphine), and GABA (Alcohol and GHB). Then we mapped the resulting list of genes on human chromosomes and considered the clusters they form (98-100). Then, we used the output BED files to visualize the cluster genes at UCSD's Genome browser as a custom track using the HG-19 build of the genome.

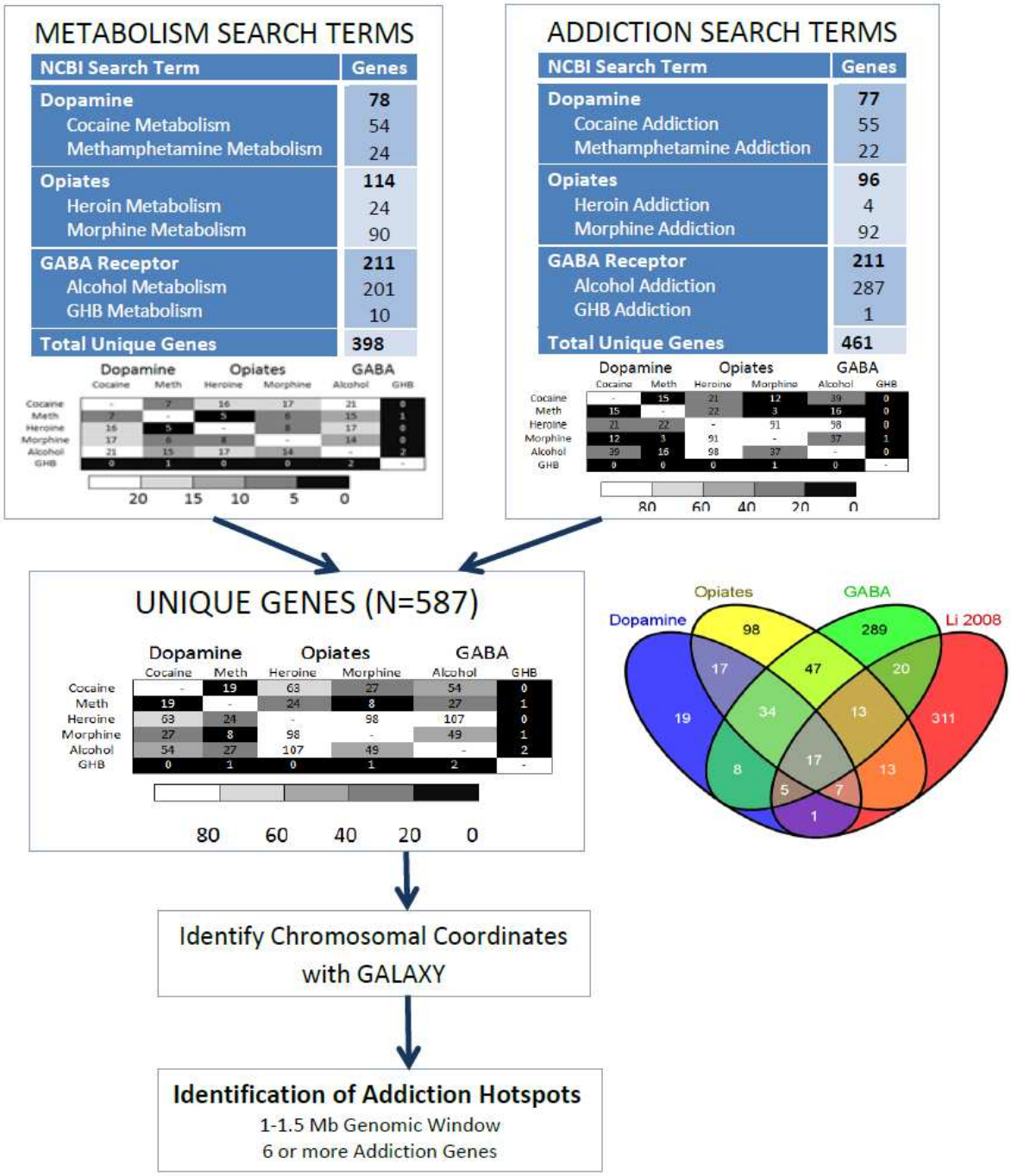


Figure 3: Flowchart for identifying biologically relevant addiction genes. The search terms used at NCBI Gene to populate a list of unique addiction genes are shown. The heat maps indicate intersection of gene sets in three classes of addiction: dopamine, opiate, and GABA. Addiction hotspots were defined at a genomic region with six or more addiction genes within a 1-1.5Mb genomic window. A Venn diagram shows the comparison with the 387 genes identified through an alternate addiction study(1).

Hotspots were defined as genic regions approximately 1- 1.5 Mb in length along the genome, which contained six or more genes identified from our combined addiction gene list. Each hotspot contained genes not currently associated with the addiction phenotypes. We included such genes into our analysis due to high probability of common regulation patterns within a hotspot. Next we investigated the statistical significance of observing six or more genes as hotspots, given a starting number of 587 addiction associated genes. We ran *in silico* computations that choose 587 genes randomly in 10,000 simulations and counted the number of times one saw at least the same number of hotspots we found with our list.

Functional Annotation of Hotspot Genes

Genes located within hotspots were considered in two ways in statistical enrichments: all genes in the hotspot window, and only those previously linked to addiction. All genes in the hotspots were annotated using DAVID's Bioinformatics Resources Tool software (101, 102) for biological process, molecular function, cell compartment and KEGG pathways (46, 49). Functional enrichments were quantified using Benjamini score analysis cutoffs of 0.01 (103). A MATLAB code was written to multi-color the nodes in KEGG pathways to differentiate between genes belonging to different hotspots.

Genetic variation in hotspots in population subtypes

We examined the hotspot associated polymorphisms identified in 11 HAPMAP sample populations with distinct geographical occupation: East Asian ancestry [Japanese-JPT, Chinese (collected in Beijing)-CHB, Chinese (collected in Denver)-CHD], African ancestry populations [Yoruba-YRI, Masaai-MKK, Luhya-LWK, and African Americans-

ASW], European ancestry populations [Europeans of Northern and Western Ancestry-CEU and Toscana- TSI], a South Asian ancestry population [Guajarati in Houston-GIH]; and an admixed American population [Mexicans in Los Angeles-MEX] (104-106) (Table1). To exclude the possibility of confounding effects of population-specific demography and to set up an empirically derived neutral estimate of allelic variation, we analyzed 20 concatenated autosomal loci across the human genome identified as neutrally evolving (107). It was assumed that the polymorphism variation undergoing selection will have non-neutral allele frequency patterns. SNP frequencies were trimmed to exclude SNPs that were almost fixed in populations (> 0.9) or of low frequency (< 0.15). Average SNP frequencies were calculated across the window and compared in all populations.

Geographical Region	Population	Population Code	DNA Samples (2N)
African	Yoruba in Ibadan, Nigeria	YRI	220
	Maasai in Kinyawa, Kenya	MKK	205
	Luhya in Webuye, Kenya	LWK	122
	African Ancestry in SW USA	ASW	98
East Asian	Han Chinese in Beijing, China	CHB	162
	Japanese in Tokyo, Japan	JPT	131
	Chinese in Denver, CO, USA	CHD	129
Western European	Northwestern European Ancestry in UT, USA	CEU	180
	Toscana in Italia	TSI	114
South Asian	Gujarati Indians in Houston, TX, USA	GIH	117
American	Mexican Ancestry in LA, CA, USA	MEX	104

Table 1: Population subtypes considered at Addiction Hotspots. The eleven HapMap populations grouped into geographical regions. Populations are described by ethnic identity and by their sample collection location.

To test for polymorphism deviations from neutral demographic patterns, we used the neutral autosomal regions and then used these combined autosomal region to give us a sense of the average allele frequencies per population in the absence of selective forces. The Kolmogorov-Smirnov test (108) was performed to test for significance between the frequencies of neutral regions and the seven addiction hotspots. Comparisons were also made between sample ethnic populations for each hotspot. Populations were clustered by regional ethnic origin: Africans (Luhya (LWK), Masaai (MKK), Yoruba (YRI), and African Americans (ASW)); Asians (Chinese-Beijing (CHB), Chinese-Denver (CHD), and Japanese (JPT)); and Europeans (Western Europeans (CEU) and the Toscana (TSI)).

Populations were compared African to European, African to Asian and European to Asian in order to identify the pairwise significantly different polymorphisms among all comparisons using a chi squared test which tested the divergence of the frequency ratios from 1. The significant polymorphisms were then sorted by their genomic location, as coding, intronic or intergenic polymorphisms.

Additionally, we examined common polymorphisms with significantly different allele frequencies using GWAS3D, a web-based software that identifies regulatory elements, long range linkage and cross chromosome interactions (109). We also identified 17 significant polymorphisms from six addiction studies curated by GWAS Central(110). These polymorphisms were then identified in the 11 sample populations of the HapMap dataset.

RESULTS

Molecular functions and biological processes of addiction genes

A list of biologically relevant genes for addiction was gathered from literature curated sources. *Figure 1* shows the flowchart employed to identify biologically relevant addiction genes and the resulting number of genes in each addiction gene set. Three classes of addiction genes were obtained using the NCBI gene search: dopamine addiction genes (N=108), opiate addiction, (N=246) and GABA addiction genes (N= 433). The search word chains consisted of 'metabolism' (N=398 genes) or 'addiction' (N=461) followed by the names of three addictive substances: dopamine, opiates and GABA receptor. The respective heat maps in Figure 3 illustrate the intersections within the search terms for both metabolism and addiction gene lists. Additionally, we compared our gene list to the one reported by Li, Mao and Wei (1), containing 387 genes involved in four addiction disorders. This comparison is shown using a Venn diagram with the bulk of genes (N=311) not identified in our analyses belonging to nicotine addiction, an addiction disorder we do not address here. The rest was added to our gene list. A set of addiction genes (N=587), compiled from the union of all search terms, was thus determined.

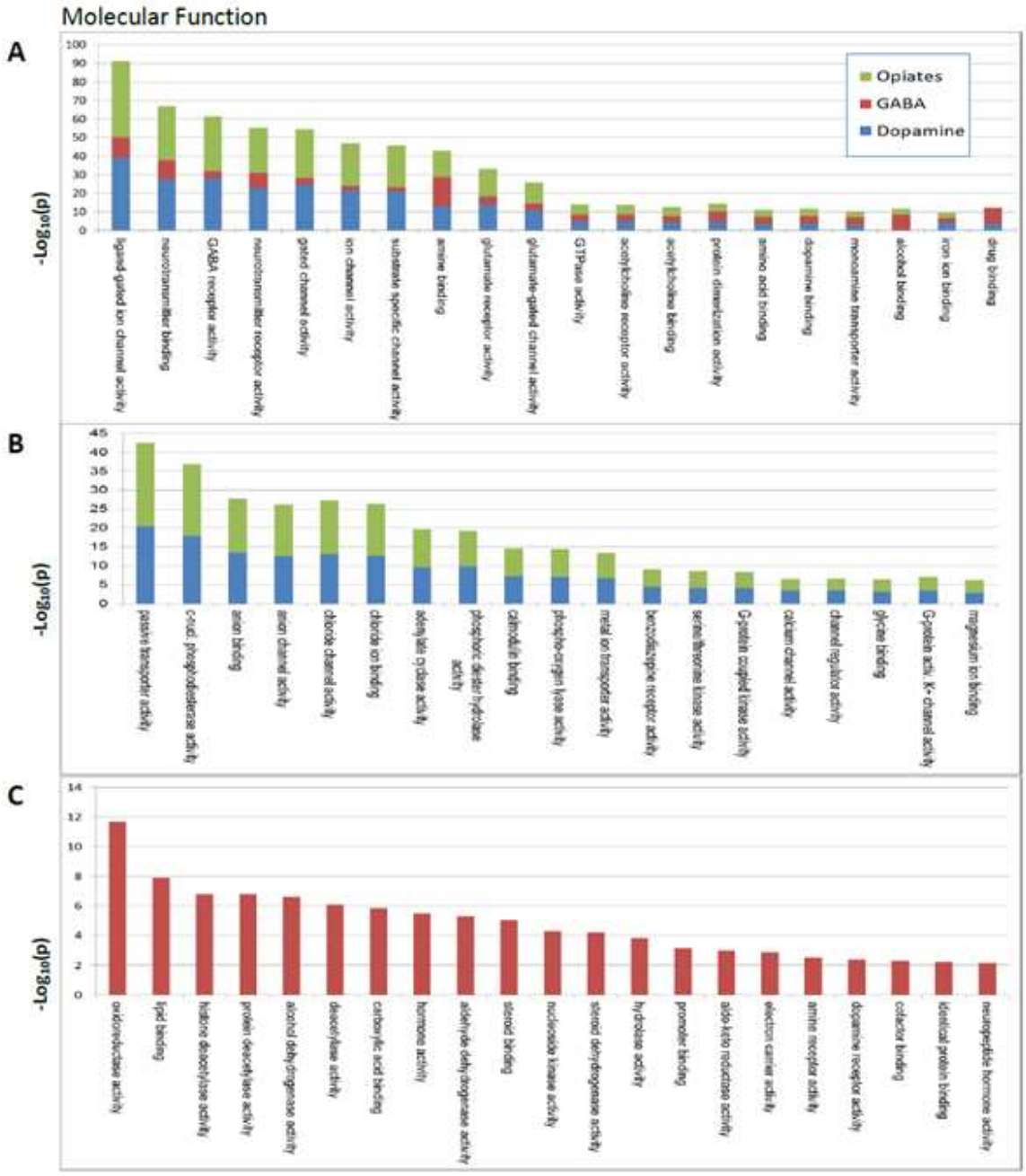


Figure 4: Gene ontology molecular function annotations for dopamine, GABA and opiate addictions. The top 75 significant functional annotations were obtained (Benjamini > 0.01) for gene ontology molecular function for each addiction gene class. The vertical axis of each graph shows the $-\log_{10}$ (p) while the horizontal axis indicates the functional annotations. The gene lists are color coded for opiates (green), GABA (red), and dopamine (blue).

Functional annotation was undertaken for the list of addiction genes, the GABA, opiate, and dopamine addiction classes, and their pairwise intersections. *Figure 4* shows the top 75 statistically enriched molecular functions for the dopamine, opiate, and GABA gene sets. Functional annotation shows a wide range of biological processes dominated by cell-cell signaling. Dopamine and opiate addiction gene lists shared many more molecular functions with each other than they did with the GABA addiction gene set. The genes common to three types of addiction under consideration included those with ligand-gated ion channel activity, neurotransmitter binding, and amine binding activities. The biological processes enriched included response to organic substances, synaptic transmission processes, and response to endogenous stimulus process (not shown).

Figure 5 shows the overlap in the gene sets associated with the opiate, dopamine, and GABA classes of addiction. The figure provides the gene symbols for each intersection subset. Substantial sets of genes were shared amongst all addiction types (N=51 genes). This subset included CREB genes, leucine zipper family of DNA binding proteins; GRIN genes, which code for glutamate-gated ion channels; and SLC genes, which are sodium: neurotransmitter transporters.

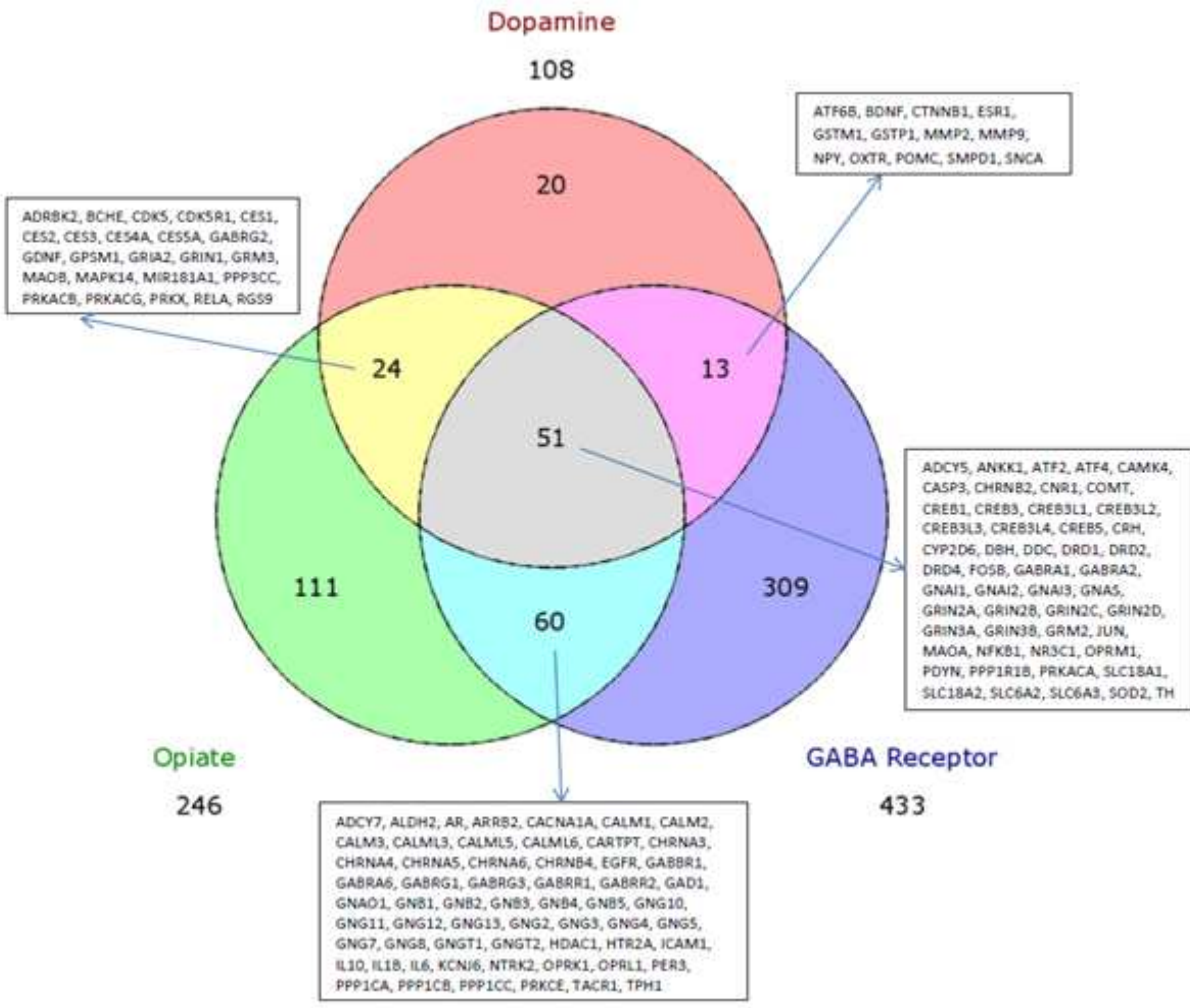


Figure 5: Genes shared between dopamine, opiate and GABA addiction sets. The figure provides the gene symbols for intersections of gene sets corresponding to dopamine, opiate, and GABA addictions.

Genes sitting in the intersection of opiates and GABA addiction included genes of the GABR family, mediating the fastest inhibitory synaptic transmission in the central nervous system. Also present at this intersection subset were genes from the GNAO/B/G family, known to integrate signals between receptors and effector proteins, and the PPP1C genes involved in the regulation of cellular processes. Genes at the intersection of dopamine and opiate addictions included three PRKAC genes coding

cAMP-based signaling molecules; five genes from the CES gene family, responsible for the hydrolysis of xenobiotics; and two genes from the CDK family, cyclin-dependent kinases, implicated in neuronal growth and repair. The overlap of the dopamine and GABA addiction gene sets included two genes from the MMP family, involved in the breakdown of extracellular matrix. These results indicate that literature curated addiction genes under consideration were involved in diverse processes required of addiction

Identification and annotation of addiction hotspots on the human genome

The addiction gene list consisting of 587 genes for the three types of addiction was mapped onto the human genome. Most of the addiction genes were dispersed through the genome but 63 addiction genes mapped onto seven genomic hotspot regions, all less than 1.5 Mb, the typical length for a genetic recombination unit on human chromosomes (Figure 6). All these hotspots had at least six or more addiction genes. The statistical significance of finding the number of genes in each hotspot was calculated using hypergeometric tests for each chromosome.

Three of the seven hotspots contained genes exclusively associated GABA addiction while the remaining four hotspots contained genes involved in GABA, dopamine, and opiate addictions (Figure 6). For each hotspot, we cataloged the co-located genes in the hotspots not yet identified as addiction-related. These additional genes represent candidates for further investigation. Hotspot windows contained 14 to 58 genes, with the number of addiction genes ranging from 6 to 19 curated genes located in a hotspot interval. The lists of genes in each hotspot are shown in Figure 6, with ones already linked to addiction shown in bold.

In addition to examining the distribution of those genes that existed in relatively short linear spaces, denoting the possibility for shared regulation across the genome, we also conducted hypergeometric tests to determine the statistical significance associated with these cluster of genes. Table 2 reports the hypergeometric values per hotspot location.

hotspot	Genes of Chr	Successes in population	Sample size	# of successes	P (X>1)
4 100Mb-100.9Mb	757	19	6	1	0.008438435460426
6 25.7Mb- 26.4Mb	2000-2057	40	19	1	0.174098577977048
6 27.8Mb- 28.9Mb	2000-2057	14	9	1	0.00619955765330504
10 5Mb- 6Mb	800-1200	19	6	1	0.00757992388448803
11 67Mb- 68.5Mb	1,524	43	6	1	0.0242776787539259
19 48.8Mb-49.9Mb	2,072	61	8	1	0.080814625284644
16 55.5Mb-57Mb	850	39	8	1	0.048231263268633

Table 2: Hypergeometric Analysis of the Addiction Hotspots identified along the Genome. The seven hypergeometric values were calculated as a measure of genomic hotspots significance.

Figure 6: Hotspots participating in acute and chronic alcoholism pathways. The alcohol addiction pathway contains genes located in all seven of the addiction hotspots, each are colored according to the legend. They participate in post-synaptic processes in both acute and chronic alcohol signaling. Additionally each component in the alcoholism pathway has a bar above the gene/component indicating their involvement in dopamine,

The hotspot gene lists were used to identify gene ontology biological processes and molecular functions enriched for each hotspot. Table 3 summarizes the results. The hotspot on chromosome 4 for GABA addiction is dominated by metabolic processes and enzymes. The GABA hotspot at chromosome 6 is crowded by genes involved in nucleosome assembly and DNA packaging. The GABA hotspot adjacent is also dominated by DNA binding proteins. The genes crowding the mixed hotspot on Chromosome 10 are involved in oxidation reduction and steroid metabolic processes. This hotspot has a series of neurological function genes that regulate appetite during stress in the brain and neuro-epithelial remodelers.

Region		Biological Processes		Molecular Function	
	Chr Location	Functional Annotation	Benjimani	Functional Annotation	Benjimani
4q23	4:100Mb-100.9Mb	Ethanol metabolic process	5.0E-9	Alcohol dehydrogenase activity	2.9E-7
GABA Only		Ethanol oxidation	5.0E-9		
		Alcohol metabolic process	5.0E-9		
6p22.2	6:25.7Mb-26.4Mb	Nucleosome assembly	3.2E-9	DNA binding	5.6E-3
GABA Only		DNA packaging	3.4E-9	Ion membrane transporter activity	7.5E-5
		Chromosome organization	1.8E-6	Alkali metal ion binding	7.5E-3
6p22.1	6:27.8Mb-28.9Mb	Nucleosome assembly	2.9E-8	DNA binding	1.0E-3
GABA Only		DNA packaging	2.2E-8		
		Chromosome organization	1.3E-5		
10p15.1	10:5Mb-6Mb	**Oxidation reduction**	**1.8E-2**	Steroid dehydrogenase activity	1.5E-6
Mixed		**Steroid metabolic process**	**2.8E-2**	Trans- 1,2-dihydrobenzene -1,2-diol dehydrogenase activity	5.7E-6
				Aldo-keto reductase activity	7.5E-5
11q13.2-3	11:67Mb-68.5Mb	None	NA	Aldehyde dehydrogenase activity	3.6E-2
Mixed					
16q22.1	16:55.5Mb-57Mb	None	NA		
Mixed				Cadmium ion binding	1.8E-16
				Cation binding	2.2E-5
				Transition metal ion binding	2.7E-3
19q13.33 Mixed	19:48.8Mb-49.9Mb	**Cell-cell signaling**	**0.02**	Hormone activity	0.03

Table 3: Addiction genes within hotspots show distinct functional classification. The table shows the gene ontology molecular functions and biological processes statistically enriched in addiction hotspots (Benjamini coefficient < 0.03). Annotations in bold correspond to the addiction genes identified within a cluster and not the entire genome region.

The hotspot in Chromosome 11 has multiple genes involved in actin creation/dynamics, cell development/differentiation, and lipid metabolism. The 16q22.1 locus has metallothionein genes involved in metabolism of xenobiotics. The 19q33.33 locus contains genes mediating spermatogenesis, hormone activity and signaling. It is clear from these results that addiction hotspots contain genes with complementary functions.

The genes at the seven addiction hotspots were mapped onto KEGG cellular pathways. We found consistent participation of hotspot genes in pathways involved in neurological signal transmission. Figure 7 illustrates the roles that hotspot addiction genes play in canonical addiction pathways. The acute and chronic alcoholism pathways are heavily influenced at the post-synaptic neuronal cells by genes contained in all seven of the hotspot regions. The pathway contains genes that carry out the two major mechanisms of addiction regulation: neurotransmission remodeling and epigenetic modifications. The major pathways involved in dopaminergic and morphine addiction also show participation of genes with neurotransmitter molecular functions and synaptic transmission biological processes. It is clear from Figure 7 that genes belonging to different hotspots coordinate to function in cellular pathways involved in addiction.

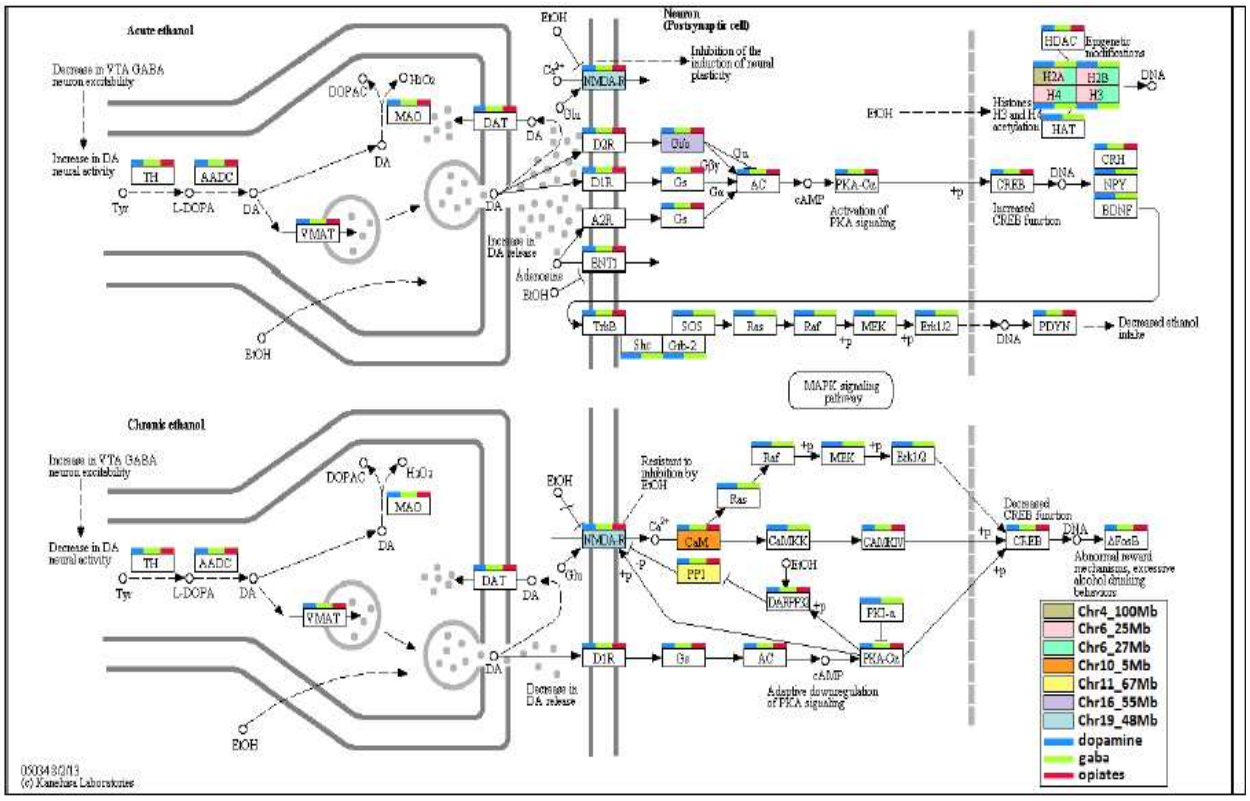


Figure 7: Hotspots participating in acute and chronic alcoholism pathways. The alcohol addiction pathway contains genes located in all seven of the addiction hotspots, each are colored according to the legend. They participate in post-synaptic processes in both acute and chronic alcohol signaling. Additionally each component in the alcoholism pathway has a bar above the gene/component indicating their involvement in dopamine, GABA or opiate addiction.

Addiction hotspots exhibit population subtype dependent polymorphisms

The presence of addiction hotspots on human chromosomes allowed us to identify genomic variation at shared polymorphism within a hotspot for the HapMap populations. We found significant differences in allelic distributions both among populations and from the population specific versus neutral comparisons. When population specific allele frequencies were sorted into frequency bins, we were able to see whether the shape of their distributions were uniform, had an excess of high or low frequency alleles as compared to the neutral distribution (Figure 8). As explained in the methods section,

the hotspot polymorphism distributions were then tested against the natural loci using the nonparametric Kolmogorov-Smirnov Test. Only the distributions with gray-scaled lines were not significantly different than the polymorphisms found in the neutral loci ($p < 0.01$). To better visualize these patterns, we generated the heat maps shown in Figure 9 based on the p values obtained in population comparisons for each hotspot. The hotspot identified at 6p21.2 (25.7 – 26.4 Mb) in this figure shows considerable similarity on polymorphisms for populations from the same origin, Africa, Asia, Europe. This appears to be particularly true for the Chinese of Beijing (CHB), Chinese of Denver (CHD) and Japanese of Tokyo (JPT) populations who are significantly different in polymorphism expression from all non-Asian populations. These results indicate addiction hotspots at chromosome 6 as loci with polymorphisms highly dependent on population subtypes.

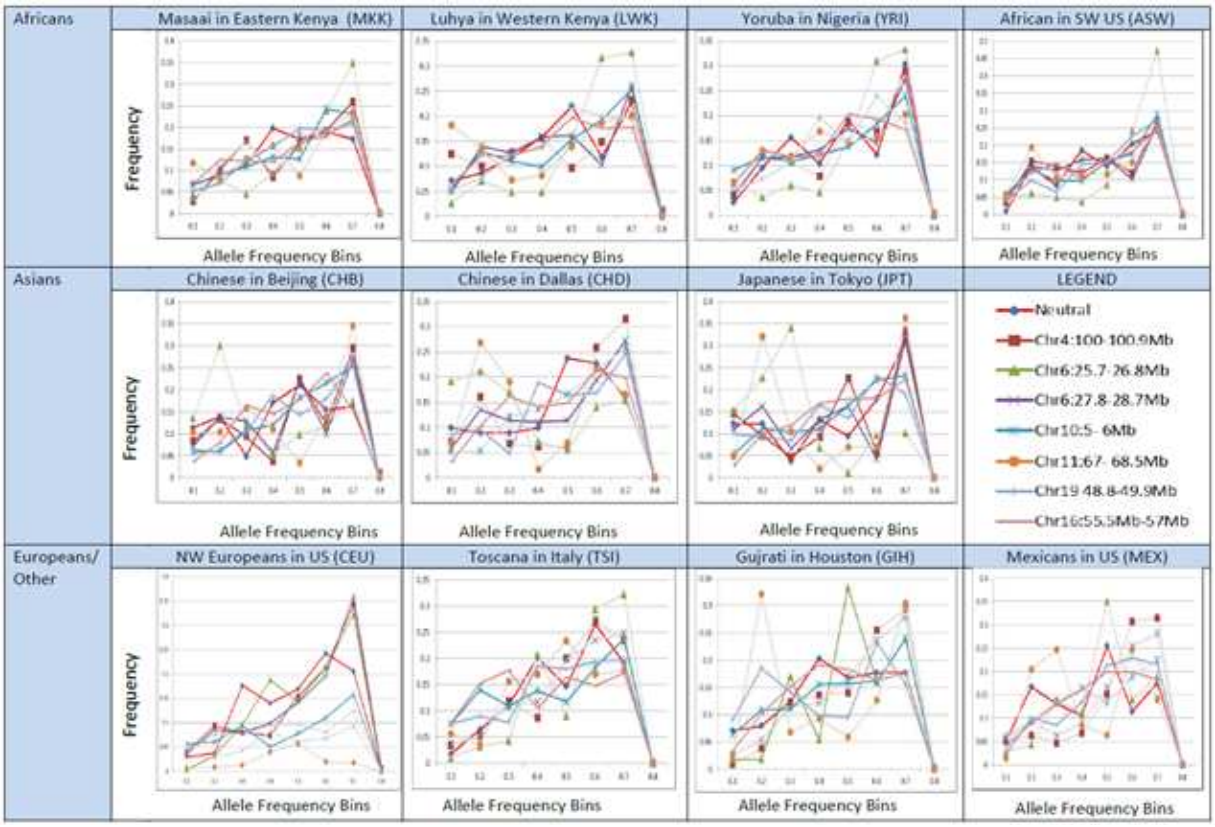


Figure 8: Allelic distributions at addiction hotspots show population and regional variation. Each graphic considers polymorphism distributions of addiction hotspots against the neutral distribution (in red). Hotspot distributions were tested against the neutral loci using the non-parametric Kolmogorov-Smirnov test. Distribution graphs with grey-scaled lines are those not significantly different from the neutral distribution.

Responding to results observed in Figure 7, shared polymorphisms at each addiction hotspot region was grouped by population regional ancestry – Asian, African, and European so that we could make comparisons between these ethnic types. The population comparison is illustrated in Figure 8 for SNPs along the hotspots. Figure 10 shows quite a few SNPs in all addiction hotspots deciphered in this study have altered allele frequencies between Asian, African, and European populations, with the chi-squared test ($p < 0.01$) in region level comparisons. The accompanying Table 4 identifies significant SNPs along the hotspot 6p21.2 (25.7 – 26.4 Mb) for the three

population comparisons. A large portion of these SNPs falls onto the intron regions of histone genes known to have roles in addiction. But three others fall onto the gene SLC17A4, which codes a sodium/phosphate co-transporter in the intestinal mucosa. The protein plays an important role in the absorption of phosphate from the intestine and its possible role in addiction is yet to be determined. Table 4 also shows five intergenic variants identified as significantly varying between these populations (rs6906576, rs6924948, rs7740793, rs9348699, rs933199). When the frequencies of significant polymorphisms identified through GWAS studies (Table 5) were examined, we found that there are strong allele frequency differences when populations were considered by region of origin. In particular we noted that rs8040009, rs2154294, rs2827312, rs6701037, and rs1109501 showed allele frequencies that differentiated East Asians from non-East Asian populations. Two alleles differentiate Europeans from non-European populations: rs2140418 (alcoholism- alcohol use disorder) and rs10908907 (alcoholism- heaviness of drinking).

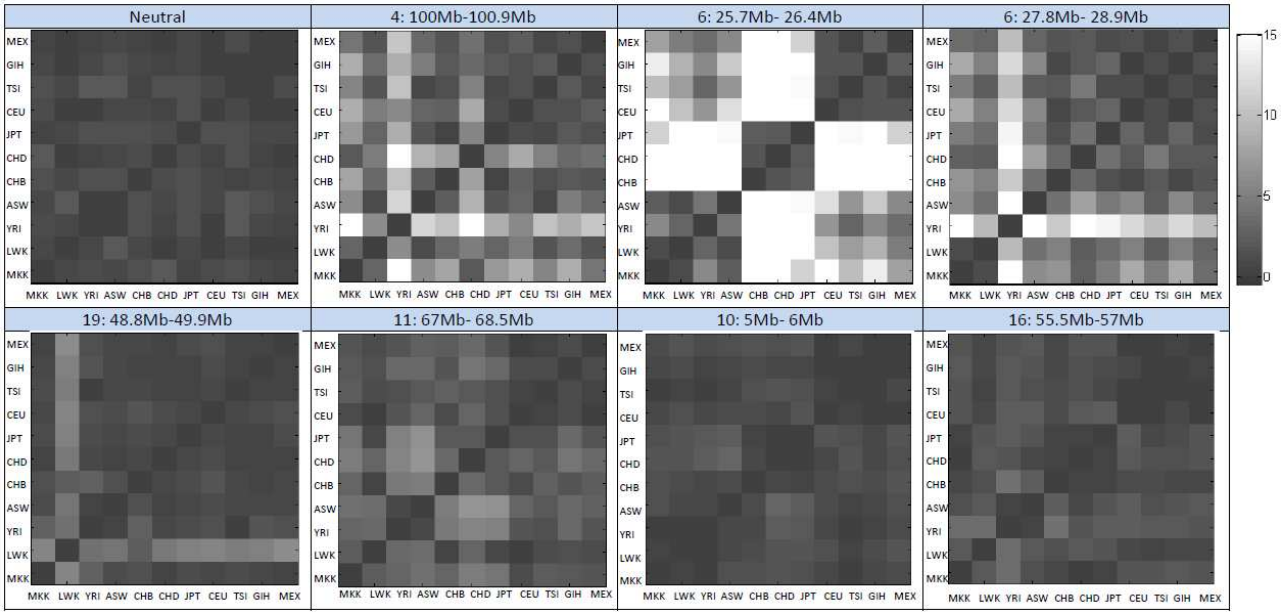


Figure 9: Heat maps of SNP distribution on hotspots show population and geographical Patterning. Comparisons of the average reference allele frequencies are shown for each addiction hotspot across populations using -log10 of the p values obtained from the Kolmogorov-Smirnov test. The variation in neutral regions shows very little demographic patterning while the hotspot located at Chr6: 25.7-26.4 shows considerably more regional blocks.

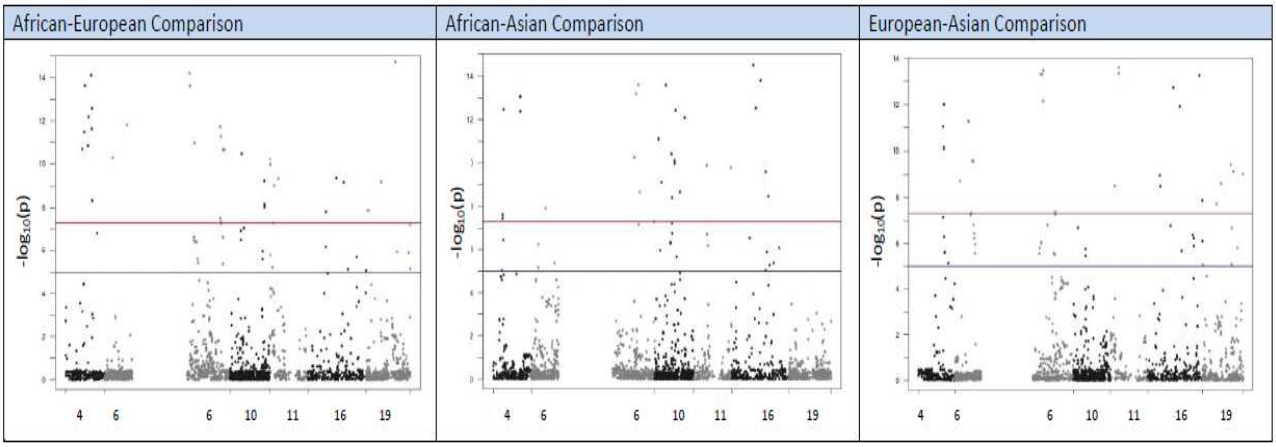


Figure 10: Hotspot SNPs with significantly altered expression in regional population subtypes. Significant polymorphisms were found for regional population comparisons between African-Asian (N=112), African-European (N=126), and European-Asian (N=122) using chi squared test with a p value cutoff of 1e-5 as shown by the blue line in the figure. The value for the red line cut off is 5e-8.

SNP			Africans				Asians			European		Other		P values		
	GENE	TYPE	MKK	LWK	YRI	ASW	CHB	CHD	JPT	CEU	TSI	GIH	MEX	African-Asian	African-European	Asian-European
rs1800708	HFE	intron	0.029	0.083	0.088	0.07	0.456	0.445	0.319	0.08	0.05	0.2	0.149	4.04E-01	9.69E-01	1.47E-07
rs9366637	HFE	Intron	0.045	0.077	0.075	0.079	0.493	0.495	0.385	0.066	0.059	0.173	0.216	3.96E-01	9.17E-01	2.57E-10
rs2237231	HIST1H1A	downstr. 500B	0.054	0.077	0.075	0.088	0.489	0.5	0.389	0.084	0.059	0.193	0.216	4.01E-01	9.78E-01	5.82E-08
rs9393682	HIST1H1C	upstr. 2KB	0.042	0.077	0.075	0.079	0.485	0.5	0.389	0.066	0.059	0.203	0.216	3.95E-01	9.27E-01	2.48E-10
rs2051542	HIST1H1T	missense	0.029	0.082	0.088	0.07	0.46	0.445	0.319	0.08	0.054	0.198	0.164	4.04E-01	9.97E-01	3.59E-07
rs3830054	HIST1H2AB	upstr. 2KB	0.042	0.077	0.075	0.088	0.496	0.5	0.389	0.084	0.059	0.203	0.216	3.97E-01	9.89E-01	4.85E-08
rs6908263	HIST1H2AC	intron	0.141	0.114	0.136	0.105	0.46	0.45	0.319	0.08	0.064	0.198	0.172	4.86E-01	4.70E-01	2.73E-06
rs7760713	HIST1H2AC	intron	0.138	0.114	0.139	0.105	0.463	0.449	0.321	0.076	0.064	0.193	0.216	4.85E-01	4.40E-01	1.11E-06
rs9467684	HIST1H2BD	intron	0.179	0.132	0.153	0.149	0.011	0.005	0.08	0.013	0.025	0.035	0.017	1.51E-04	1.60E-12	4.94E-01
rs4145878	HIST1H2BF	upstr.2KB intron	0.391	0.405	0.412	0.404	0.106	0.092	0.018	0.5	0.495	0.381	0.414	4.28E-06	8.49E-01	3.92E-01
rs1892252	SLC17A4	intron	0.362	0.232	0.248	0.316	0.051	0.032	0.075	0.133	0.098	0.03	0.172	6.90E-06	1.32E-01	5.86E-01
rs3734525	SLC17A4	utr 3'	0.106	0.186	0.211	0.193	0.007	0.005	0.075	0.004	0.01	0.005	0.017	5.73E-07	0.00E+00	1.67E-03
rs3823151	SLC17A4	intron	0.388	0.336	0.35	0.272	0.099	0.087	0.111	0.04	0.049	0.03	0.078	1.64E-02	5.32E-11	2.21E-01
rs199738	TRIM38	utr 5'	0.337	0.403	0.445	0.36	0.074	0.046	0.053	0.19	0.157	0.183	0.147	1.21E-08	2.20E-01	5.04E-01
rs6906576	intergenic	genomic	0.074	0.114	0.175	0.14	0.391	0.408	0.35	0.058	0.039	0.183	0.241	5.02E-01	1.11E-01	5.31E-12
rs6924948	intergenic	genomic	0.023	0.028	0.028	0.035	0.485	0.495	0.386	0.076	0.054	0.215	0.207	3.49E-01	5.74E-01	1.91E-09
rs7740793	intergenic	genomic	0.122	0.114	0.19	0.149	0.391	0.408	0.35	0.058	0.039	0.175	0.241	5.32E-01	4.95E-02	5.31E-12
rs9348699	Intergenic	genomic	0.122	0.114	0.19	0.149	0.391	0.408	0.35	0.058	0.039	0.183	0.241	5.32E-01	4.95E-02	5.31E-12
rs933199	intergenic	genomic	0.03	0.083	0.091	0.07	0.459	0.445	0.319	0.082	0.054	0.203	0.17	4.05E-01	9.94E-01	5.88E-07

Table 4: SNPs identified as significant in comparisons between Africans, Asians and Europeans for the 6p21.2 addiction hotspot. All SNPs were tested for significant differences in allele frequencies using a chi square test (cutoff < 1e-5) in pairwise regional comparisons: Africans (ASW, YRI, LWK, and MKK), Europeans (CEU and TSI), and Asians (CHB, CHD, JPT). The Allele frequencies of variants in populations are given with the p values of the grouped regional ethnic populations.

							Africans				Asians			Europeans		S Asian	American
	GWAS ID	SNP ID	CHR	Location	pValue	Allele	MKK	LUH	YRI	ASW	CHB	CHD	JPT	CEU	TSI	GIH	MEX
Alcohol Dependence	HGVRS1526	rs6701037	1	1.75E+08	2E-07	C	0.369	0.372	0.3	0.316	0.098	0.065	0.1	0.518	0.369	0.318	0.27
	HGVRS1526	rs750338	11	1.25E+08	0.000001		0.332	0.417	0.52	0.378	0.598	0.559	0.54	0.177	0.244	0.273	0.18
Alcoholism, alcohol use disorder factor score	HGVRS1536	rs6716455	2	1.51E+08	7E-07	G	0.895	0.978	1	0.969	0.817	0.869	0.76	0.845	0.787	0.898	0.83
	HGVRS1536	rs9556711	13	98016416	0.000002	G	0.5	0.356	0.34	0.449	0.78	0.735	0.87	0.929	0.943	0.909	0.91
	HGVRS1536	rs2140418	6	34975415	0.000004	C	0.44	0.283	0.39	0.541	0.439	0.359	0.48	0.774	0.784	0.568	0.76
	HGVRS1536	rs768048	18	50285398	0.000008	C	0.476	0.376	0.53	0.592	0.988	0.97	0.97	0.838	0.947	0.892	0.91
Alcoholism, heaviness of drinking	HGVRS1537	rs9512637	13	27920611	1E-07	C	0.591	0.411	0.39	0.347	0.232	0.282	0.16	0.673	0.591	0.443	0.42
	HGVRS1537	rs8040009	15	93044339	3E-07	T	0.461	0.443	0.31	0.385	0.812	0.859	0.84	0	0.807	0.949	0.76
	HGVRS1537	rs1109501	4	71329490	0.000005	G	0.88	0.906	0.97	0.929	0.573	0.559	0.55	0.741	0.636	0.756	0.68
	HGVRS1537	rs10908907	9	92249584	0.000006	G	0.413	0.356	0.26	0.357	0.527	0.421	0.5	0.739	0.619	0.339	0.44
	HGVRS1537	rs2827312	21	23631676	0.000008	G	0.542	0.689	0.56	0.664	0.915	0.946	0.91	0.628	0.551	0.648	0.63
	HGVRS1537	rs195204	1	1.16E+08	0.000009	T	0.86	0.9	0.91	0.857	0.537	0.624	0.5	0.765	0.773	0.562	0.65
Alcoholism, 12-mth wkly alc. consum.	HGVRS1538	rs2154294	14	42655275	0.000003	G	0.462	0.456	0.56	0.561	0.744	0.835	0.84	0.549	0.545	0.591	0.57
Alcohol dependence	HGVRS1539	rs9556711	13	98016416	8E-07	G	0.5	0.356	0.34	0.449	0.78	0.735	0.87	0.929	0.943	0.909	0.91
	HGVRS1539	rs2140418	6	34975415	0.000004	C	0.44	0.283	0.39	0.541	0.439	0.359	0.48	0.774	0.784	0.568	0.76
	HGVRS1539	rs10253361	7	1.21E+08	0.000006	T	0.727	0.917	0.83	0.776	0.293	0.376	0.47	0.562	0.523	0.477	0.51
	HGVRS1539	rs933769	15	96052742	0.000007	T	0.839	0.872	0.96	0.898	0.634	0.72	0.61	0.867	0.807	0.642	0.77
Alcohol dependence	HGVRS1742	rs8062326	16	19156016	0.000004	T	0.115	0.078	0.08	0.082	0	no data	0	0.031	0.034	0.028	0.06

Table 5: GWAS derived Addiction Alleles showed significant differences in allele frequencies between Asians and Non-Asian populations. Alleles identified as significant in GWAS studies were typed in the 11 HapMap populations.

Distant connections of hotspot polymorphisms

We followed these frequency-based analyses of hotspots with analysis uncovering linkage to distant sites. For this purpose we used the web platform GWAS3D developed recently by Jun Li M et al. (109). The platform identifies genetic variants affecting regulatory pathways and underlying disease/trait associations by integrating chromatin state, functional genomics, sequence motif, and conservation information given a variant list. In the addiction case under study, we examined the distant regulatory landscape through linkage of the significant SNPs identified at hotspots in cross population comparisons.

Figure 11 shows the significant common variants between Africans and Europeans projected onto the Yoruba population. In the outer ring, polymorphisms or genomic regions are identified. The second ring identifies the genes or chromosomal regions these polymorphisms sit in and finally the red lines indicate the strength of local or long range interactions. Thus, with the use of emerging bioinformatics web platforms, deciphering addiction hotspots on the human genome show potential for further discovery of DNA motifs distant to the hotspots.

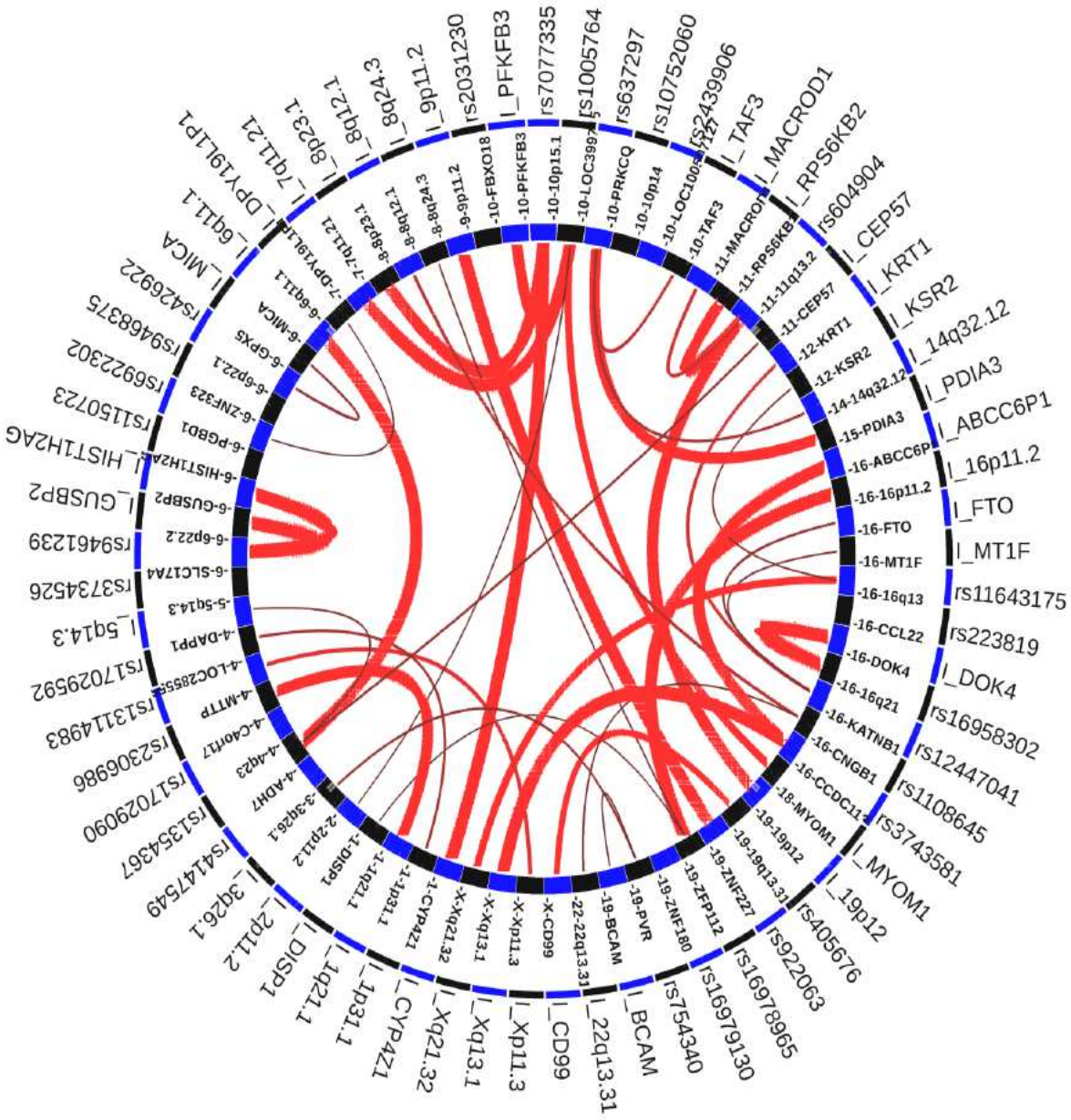


Figure 11: Significant SNPs characterized in the Yoruba population show local and long range interactions. Polymorphisms identified as significant in African to European comparison were projected onto the GWAS3D platform. At least six long-range trans-chromosomal interactions were identified and numerous local interactions were observed.

DISCUSSION

The NCBI Gene web platform recently began providing gene lists associated with disease or disorders, ordered according to prevalence in literature searches. Some of the genes in the list were inferred from genome wide association studies. The others were derived from research focusing specific targets. Using NCBI curated gene lists, we investigated the biological processes and cellular pathways involved in three types of addiction: addiction linked to dopamine (cocaine and crystal methamphetamine), opiate (heroine and morphine), and GABA (Alcohol and GHB). Biological process annotations could be sorted into three categories. Those involving all addiction classes included neurological, behavioral and cell signaling function. Processes common to dopamine and opiate addiction included signal transduction, neurological system process, and transport. GABA specific processes, on the other hand, consisted of chromatin assembly, organization and homeostatic processes.

The gene molecular functions involving all addiction classes included ion channel activity, receptor function and binding. Those involved in dopamine and opiate addiction largely consisted of anion binding, kinase activity and channel regulation. Molecular functions specific to GABA addiction involved DNA binding, hormone activity, and deacetylase activity. These findings indicate that the differentiating mechanisms of GABA addiction from opiate and dopamine addictions all involved chromatin remodeling. Some of the addiction genes specific to GABA addiction, histone genes, are known to mediate methylation events, suggesting epigenetic modifications be an important component of alcohol addiction.

Recent studies point to hotspots on human chromosomes, where disease related genes clustered (111). Similarly, a meta-analysis of GWAS studies on aging led to discovery of a number of disease hotspots on the human genome. Other studies indicate the presence of foci for susceptibility to autism (112). Taking inspiration from these studies, we mapped the addiction gene lists onto human chromosomes and demonstrated the existence of addiction hotspots on the human genome. About 11 percent of literature-curated addiction genes formed seven clusters of 6 or more genes within a span of 1.5 Mb, a typical length for genetic variation, along the chromosomes. The DNA segments containing the clusters were deemed as hotspots.

The addiction hotspots fell onto chromosomes 4, 6, 10, 11, 16, and 19. Chromosome 6 contained two hotspots in close vicinity at 6p21.2. All hotspots also contained genes not previously linked to adhesion, gene regulatory motifs and polymorphisms. Three of the seven addiction hotspots (ones on chromosomes 4 and 6) were related exclusively to GABA addiction, one on chromosome 4 linked to alcohol metabolic processes, and the two on chromosome 6 related to chromatin packaging and ion membrane transport. The rest were mixed. The hotspot on chromosome 10 was related to steroid metabolism whereas the hotspot on 19 exhibited cell-cell signaling through hormone activity. Overall, the genes in addiction hotspots crowded the canonical addiction pathways. The addiction hotspot gene list was also enriched in pathways for systemic lupus erythematous, viral

carcinogenesis, and steroid hormone biosynthesis, suggesting susceptibility of the addicted to other disease and disorders. The functional pleiotropic nature of these hotspots helps to explain how addiction creates a spectrum of disorders that are not limited to particular addictive substance (113, 114).

The presence of hotspots on the human genome provides an opportunity for discovery of addiction related genes. Addiction genes accounted for a significant percentage of total genes (> 36 %) in the three hotspots exclusively related to GABA addiction with nearly all genomic regions containing near statistically significant or statistically significant numbers of NCBI identified genes within the hotspot region. The other four mixed hotspots contained a vast majority of genes not currently linked to addiction. Consider for example, the gene RG9MTD2 in the GAMA only hotspot on Chromosome 4. This gene codes a RNA transmethylase, expressing an acetaminophen-binding site. Acetaminophen has been shown to increase feelings of intoxication in combination with ethanol in a human cohort and does not mitigate subjective feelings of alcohol intoxication (115).

Additionally, our mixed addiction hotspot located at Chr19 sits adjacent to the Killer cell Immunoglobulin-like Receptor (KIR) genomic region. This foci displays extensive diversity through allelic polymorphism within individual KIR genes(116). It is possible to speculate that the regulatory changes in our identified hotspot could change the regulatory landscape for the KIR region. Another family of genes with potentially important roles (yet to be defined) is the SLC17A1-4 genes coding organic ion

transporters in a GABA only hotspot on Chr6. These genes were identified as having storage activity (117) for neurotransmitters playing crucial roles in addiction (32).

Addiction hotspots showed ethnic population dependent polymorphisms. We have identified polymorphism in hotspots, which significantly varied among the 11 ethnic HapMap populations. Pooling these populations by their geography of origin (East Asian, African, and European) allowed us to achieve a regional perspective of the importance of ethnic origin in future analyses at the genetic and epigenetic interactions. Our population-based variation analyses found significant population variation at the 6p21.2 region, which differentiated HapMap Chinese and Japanese populations from all other non-Asian populations. Analysis brought to light SNPs that fall onto histone and SLC genes, as the major differentiating factors. It is clear that the 6p21.2 locus is a potentially potent source of genetic variation that could help explain the phenotypes variation between East Asian and non-Asian populations. Additionally the Yoruba (Nigerian population) have unique signals of population variation at 4q23 and 6p21.1. Additional analyses of addiction GWAS polymorphisms in HapMap populations showed similar patterns to those we saw in our dataset. This pattern showed strong East Asian to non-East Asian population differences in commonly shared allele frequencies. We believe that this observation supports our finding that the polymorphism analyses we perform here are representing some genomic phenomenon in Asian populations. The list of candidate

polymorphisms presented in Table 3 may become substrates future association analyses.

The gene contents of GABA only addiction hotspots suggest an epigenetic role in alcohol addiction (84, 118-120). The two hotspots on Chr6 are highly enriched with histone genes. It is well established that methylation state of histone proteins are directly related to genes in DNA being "off" or "on". Functional annotation shows histone involvement in non-addiction related KEGG pathways such as systemic lupus, viral carcinogenesis and transcriptional mis-regulation of cancer. In fact, histone protein modification through methylation has been identified as one of the possible reasons for the diverse phenotypes seen in addicted individuals (114, 121, 122). The SNPs in histone genes on Chr6 may lead to alternate epigenetic modifications among ethnic population, a result previously seen in other datasets(123). Moreover, our GWAS3D analyses found that 6p21.2 SNPs identified as significant in our cross population assessments were also involved in local and trans-chromosomal interactions. Potential links between hotspots and addiction genes dispersed on human chromosomes merit further investigation.

Specific Aim II: Addiction and Mental Health

Introduction

Addiction disorders represent a major economic and health cost to human populations (51). Addiction is loosely defined as a chronic relapsing spectrum disorder characterized by loss of control over substance taking (52-54). Similarly, neurological disorders such as depression, bipolar disorder, and schizophrenia represent a significant strain on health and judicial entities. Both disorder classes have long been identified as co-morbidities (124). Both Substance addiction and mental health are behavior-based phenomenon representing a diverse array of psychological, biological, and genetic attributes and environmental and cultural factors (32, 33, 55-57).

A number of studies have identified the clinical intersection of mental health and substance addiction traits (125, 126). This has been shown in a diverse variety of mental health conditions such as depression (127, 128), bipolar disorder (129-131), and schizophrenia (132-137). To date, the crosstalk of addiction and mental health genetic contributors has not been fully understood based on association studies (68, 138) and functional genomics (41, 42). Studies show that drugs targeting specific mental health conditions such as schizophrenia have

In this study we focus on the addiction disorders for opiate, dopamine, and GABA addiction as well as the mental health disorder identified as depression, bipolar disorder and schizophrenia. We map literature curated addiction and mental health genes onto human chromosomes. Clusters of these genes identify hotspot locations

within the genome. These hotspots contain genes previously not linked to addiction and mental health as well as regulatory motifs with drug binding sites. Our bioinformatics based discovery and subsequent investigation of combined addiction and mental health hotspots reveal their functional roles. Results point out multiple hotspots participating in dual addiction and mental health phenotypes.

Specific Aim II: Assessing the genomic regions sitting at the intersection of Addiction and Schizophrenia, Bipolar Disorder and Depression disorders and their interplay with drug binding sites.

I hypothesize that when Addiction and mental health loci are jointly considered, we can identify hotspots for their interactions. A number of studies have identified the clinical intersection of mental health and substance addiction traits (125, 126). This has been shown in a diverse variety of mental health conditions such as depression (127, 128), bipolar disorder (129-131), and schizophrenia (132-137). To date, the crosstalk of addiction and mental health genetic contributors has not been fully understood based on association studies (68, 138) and functional genomics (41, 42).

Methods

Addiction linked genes and genome hotspots

A search of NCBI Gene based on strings of disease words was used in order to generate a list of genes with biological relevance to addiction. We focused on dopamine, opiate and GABA addiction; bipolar disorder; depression; and schizophrenia. We mapped the resulting list of genes onto human chromosomes

and considered the clusters they form. Then, we wrote a Matlab based program to cluster genes in the genome and report those hotspot regions back as an output file.

Hotspots were defined as genic regions approximately 2 Mb in length along the genome, which contained 13 or more genes identified from our combined addiction and mental health gene list. These hotspot regions were elongated from those identified in specific aim I to accommodate the differences in recombination rate found in genomic regions and the starting number of genes in the initial gene list. Additionally we collapsed multiple small genomic hotpot windows into larger ones that spanned the 2Mb or smaller region. Each hotspot contained genes not currently associated with these joint addiction and mental health phenotypes. We included such genes into our analysis due to high probability of common regulation patterns within a hotspot. Next we investigated the statistical significance of observing 18 or more genes as hotspots, given a starting number of addiction and mental health associated genes.

Functional Annotation of Addiction and Mental Health Hotspot Genes

Genes located within hotspots were considered in two ways in statistical enrichments: all genes in the hotspot window, and only those previously linked to addiction. All genes in the hotspots were annotated using DAVIDs Bioinformatics Resources Tool software (101, 102) for biological process, molecular function, cell compartment and KEGG pathways (46, 49). Functional enrichments were quantified using Benjamini score analysis cutoffs of 0.01 (103). Finally, all genes in hotspots were annotated for drug interactions using Drug Bank (139-142).

Results

Overlap in Gene Lists for Addiction, Schizophrenia, Bipolar Disorder and Depression

We conducted searches to identify genes belonging to NCBI curated mental health and addiction genes lists. The bipolar disorder gene list comprised 626 genes, depression comprised 357 genes, schizophrenia comprised 1121 genes and Addiction comprised 587 genes garnered from opiate, GABA and dopamine addictions. When considered together this genes list represented 1968 genes encoding putative addiction and mental health

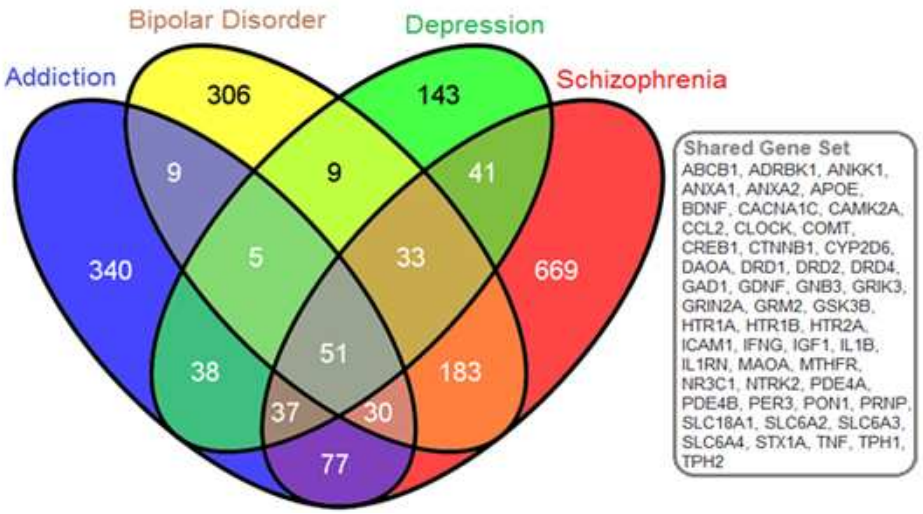


Figure12: Genes shared between dopamine, opiate and GABA addiction sets with common mental health conditions. The Venn diagram provides the gene symbols (N=1968 genes) for intersections of gene sets corresponding to dopamine, opiate, and GABA addictions with bipolar disorder, depression, and schizophrenia. The gray boxed genes are those genes common to all disorders.

targets. Figure 12 is a Venn diagram that shows the overlap in genes from these four gene lists. Interestingly, there were 51 genes that were shared between all gene lists but only four of these shared genes were represented in the eight hotspots. These are indicated in red in Table 6. This overlap gene set contained the DRD,

HTR and SLC6A gene families as well as a host of immune function genes including: ICAM1, IFNG, IGF1, IL1B, IL1RN, and TNF. Of this combined set of addiction and mental health genes, 192 genes fell into cluster regions within the genome. Our analyses identified these eight genomic regions with significant numbers of genes involved in the dual disorders.

Chr	Location	N	NCBI Identified Genes	Candidate Interspersed Genes
2	98444858- 99400475	13	INPP4A, MGAT4A, KIAA1211L, TSGA10, C2orf15, LIPT1, MITD1, MRPL30, LYG2, LYG1, TXNDC9, EIF5B, REV1	C2orf64, UNC50
3	51,741,081-53,752,625	33	**GRM2**, ALAS1, TLR9, TWF2, PPM1M, WDR82, GLYCTK, DNAH1, BAP1, PHF7, SEMA3G, TNNC1, NISCH, STAB1, NT5DC2, PBRM1, GNL3, SNORD19, SNORD19B, SNORD69, GLT8D1, SPCS1, NEK4, ITIH1, ITIH3, ITIH4, MUSTN1, TMEM110, SFMBT1, RFT1, PRKCD, TKT, CACNA2D3	IQCF6, IQCF1, IQFC4, IQFC3, IQFC2, IQFC5, RRP9, PARP3, PCBP4, ACY1, ABHD14B, LINC00696, DUSP7, POC1A
6	25782897-27450742	28	SLC17A3, HIST1H3A, HIST1H4B, HIST1H3B, HIST1H3C , HIST1H2AE, HIST1H3E,HFE, HIST1H3E, , HIST1H2BE, HIST1H4D, HIST1H4C, HIST1H3D, HIST1H3F, HIST1H4E, HIST1H4F HIST1H3G, BC079832, HIST1H4H BTN3A2, BTN2A2, BTN3A1, HIST1H2BJ, HIST1H2AG, PRSS16, POM121L2, ZNF184	SLC17A2, TRIM38, HIST1H2AB, HIST1H1C, HIST1H1T, HIST1H2BC, HIST1H1E, HIST1H2AC, HIST1H2BDHIST1H2AD, HIST1H2BF, , HIST1H2BG, HIST1H1D, , HIST1H4G, HIST1H2BH, HIST1H2BI,BTN1A1, ABT1, ZNF322A, HIST1HBK, HIST1H41
6	30,184,455-33621379	31	HLA-E, TUBB, LINC00243, DDR1, HLA-C, HLA-B, MICB, LTA, **TNF**, PRRC2A, BAG6, HSPA1L, HSPA1A , HSPA1B, TNXB, ATF6B, RNF5, AGER, NOTCH4, HCG23, HLA-DRA, HLA-DRB1, HLA-DQA1, HLA-DQB1, RXRB, KIFC1, PHF1, SYNGAP1, ITPR3	ABCF1, PPPI110, MRPS18B, C6ORF134, DHX16, KIAA1949, MDC1,FLOT1, IER3, GTF2H4, VARS2, SFTA2, DPCR1, MUC21, CCHCR1, PSORS1C1, CDSN, TCF19, POU5F1, HCG27,
11	64,803514- 66,033,706	22	MEN1, CAPN1, CDC42EP2, RELA, CFL1, GAL3ST3, SF3B2, PACS1, KLC2, RAB1B, CNIH2, YIF1A, TMEM151A, CD248, RIN1, BRMS1, B3GNT1, SLC29A2, NPAS4, MRPL11, PELI3, DPP3	EHD1, ATG2A, PPP2R5B, GPHA2, BATF2, ARL2, ARL2-SNX15, SAC3D1, NAALADL1, ZFPL1, TM7SF2, ZNHIT2, MRPL49, SYVN1, CDCA5, FAU, SPDYC, SLC22A20, POLA2, DPF2, TIGD3, FRMD8, SLC25A45, MALAT1,SCYL1, KCNK7, LTBP3,SSSCA1, FAM89B, EHB1L1, MAP3K11, PCNXL3, SIPA1, KAT5, CATSPER1, EIF1AD, TSGA10IP, C11orf68, FOSL1, EFEMP2, MUS81, SART1, FIBP, SNX32, AP5B1, BANF1, CST6
11	66,034,695- 68008578	22	BBS1, ZDHHC24, ACTN3, CTSF, CCDC87, CCS, RBM14, RBM4, RBM4B, SPTBN2, C11orf80, RCE1, PC, LRFN4, C11orf86, SYT12, **ADRBK1**, CABP4, GSTP1, NDUFV1, ALDH3B2, ALDH3B1	TBX10, NUDT8, DOC2GP, X15673, AK129926, ACY3, C11orf72, AIP,CDK2AP2, PITPNM1, TMEM134, CORO1B, GPR152, TBC1D10C, PPP1CA, RAD9A, POLD4, CLCF1, 7SK, CARNS1, PTPRCAP, RPS6KB2, SSH3, ANKRD13D, KDM2A, RHOD
19	48047843- 49429398	27	SULT2A1, PLA2G4C, GRIN2D, SULT2B1, FAM83E, RPL18, SPHK2, DBP, CA11, NTN5, FUT2, MAMSTR, RASIP1, IZUMO1, FUT1, FGF21, BAX, FTL, LHB, CGB, CGB2, CGB1, CGB5, CGB8, CGB7, NTF4, SLC17A7	GRWD1, KCNJ14, CYTH2, LMTK3, SPACA4, BCAT2, HSD17B14, PPP1R15A, TULP2, PLEKHA4, NUCB1, GYS1, RUVBL2, KCNA7, LIN7B, PPFIA3, HRC, TRPM4, SLC16A16,CD37, TEAD2, DKKL1, PTH2
22	18906223-21207972	18	PRODH, DGCR2, DGCR14, UFD1L, CLDN5, TBX1, GNB1L, **COMT**, ARVCF, DGCR8, TRMT2A, RANBP1, ZDHHC8, RTN4R, MED15, PI4KA, GGT2	TSSK2, SLC25A1, CLTCL1, HIRA, MRPL40, C22ORF39, CDC45L, SEPT5, GP1BB, C22ORF29

Table 6: Addiction, bipolar disorder, depression, and schizophrenia form eight genomic hotspots. When genes were mapped to the genome, we found eight regions smaller than 1.5 Mb each with 11- 38 genes identified from NCBI Gene curated gene lists. Additionally we consider the genes interspersed with our set of addiction and mental health genes. Red gene names are those that were among the 51 genes shared by all addiction and mental health disorders.

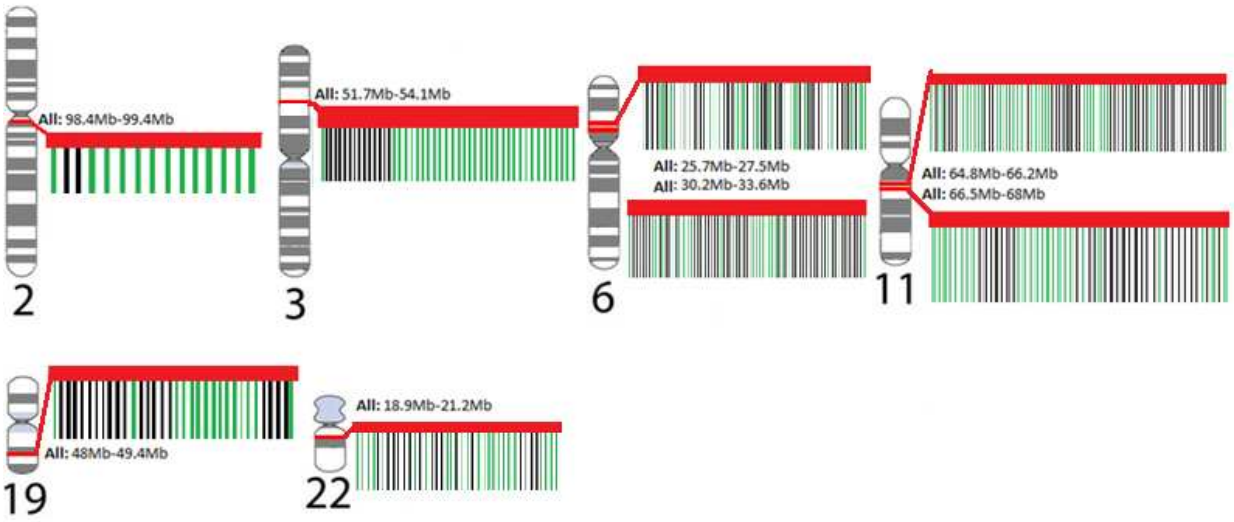


Figure 13: Eight bipolar, depression, schizophrenia and addiction hotspots identified on the human genome. Each karyotype of a chromosome shows the location of the genes that are represents as hotspots. Genes that lie within each hotspot were identified through USCS Genome Browser. The insets for each hotspot show genes identifies curated genes (green) and interspersed genes (black) with a red band showing the entire hotspot region.

These eight regions are shown in Figure 13 with their associated NCBI-identified genes (green) and the candidate genes (black) that are interspersed. Hotspots ranged from having 13- 33 literature curated genes in a genomic window approximately 2Mb or smaller.

Functional annotation of Schizophrenia, bipolar, depression and addiction genes
We performed a functional annotation on the complete list of NCBI identified genes in order to determine the major functional roles for addiction and mental health genes. Functional annotation of the combined set of depression, schizophrenia, bipolar and addiction genes found commonalities as identified in Figure 14, with shared function revolving around neurological function, response organic

substances and cell-cell signaling. Addiction and depression overlapped in three functional regulatory roles: cellular localization, cyclic nucleotide biosynthesis and metabolism. Additionally when pairwise comparisons were made (not shown), depression and schizophrenia gene lists overlapped in homeostatic processes, while bipolar and schizophrenia genes jointly participated in neuronal development and cellular differentiation. Hotspot gene lists were used to identify gene ontology biological processes and molecular functions enriched for each hotspot. When we considered the identified literature curated genes that identified each hotspot and then those genes that were not previously identified, but could be addiction and mental health candidates, we found no real differences in the functional annotation of these sets.

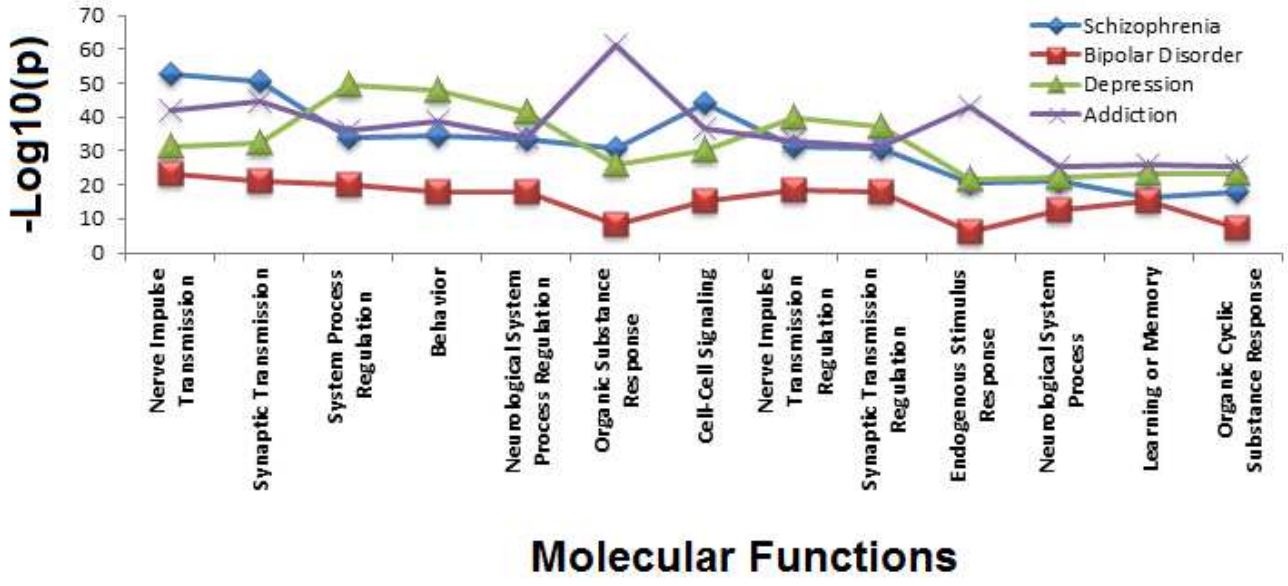


Figure 14: Gene ontology molecular function annotations for addiction, bipolar disorder, depression and schizophrenia. The top 25 significant functional annotations were obtained (Benjamini > 0.01) for gene ontology molecular function for each addiction, bipolar, depression, schizophrenia gene class. The vertical axis of each graph shows the $-\log_{10}$ (p) while the horizontal axis indicates the functional annotations. The gene lists are color coded for depression (green), bipolar disorder (red), schizophrenia (blue). Addiction genes are a composite of opiate, GABA, and dopamine addiction genes (purple). Thirteen functions are shared between the gene sets.

Chr	Gene Target	Drug	Drug Name	Mechanism of Action	Treatment Targets
3:51	TLR9	DB05463	ISS-1018	immunostimulatory activity.	*Hepatitis B, b-cell or non-hodgkin's lymphoma.
	TLR9	DB04900	Thymalfasin	unspecified	Influenza Hepatitis B
11:64	GSTP1	DB04339	Carboxymethylene-cysteine	unspecified	unspecified
	MTR, MTRR, **MMACHC,** MTHFR	DB00115	Cyanocobalamin	unspecified	pernicious anemia vitamin B 12 deficiency
11:66	TOP2A, TOP2B	DB00773	Etoposide	unspecified	refractory testicular tumors, small cell lung cancer, Lymphoma non-lymphocytic leukemia glioblastoma multiform
	DNA	DB01008	Busulfan	selective immunosuppressive effect on bone marrow	conditioning regimen prior to allogeneic hematopoietic, progenitor cell transplantation for chronic myelogenous
19:48	bglA	DB04658	(1S,2R,3S,4R,5S)-8-AZABICYCLO [3.2.1]OCTANE-1,2,3,4-TETROL	unspecified	unspecified
	BCL2, BAD BBC6, BCL2L8	DB05764	**ABT-263**	It blocks some of the enzymes that keep cancer cells from dying.	lymphomas and other types of cancer
	CYP17A1	DB05812	**Abiraterone**	derivative of steroidal progesterone	hormone refractory prostate cancer.
	KCNJ PTGS1 COX11,	DB00350	**Minoxidil**	direct-acting peripheral vasodilator	reduces peripheral resistance produces a fall in blood pressure
	BRAF1, RAFB1 FLT4, FLT3, VEGFR3, VEGFR2,	DB00398	**Sorafenib**	small molecular inhibitor of Raf kinase, PDGF (platelet-derived growth factor), VEGF receptor 2 & 3 kinases and c Kit the receptor for Stem cell factor	advanced renal cell carcinoma advanced hepatocellular carcinoma
	PTGS2, COX2, PTGS1, COX1,	DB00316	**Acetaminophen**	analgesic and antipyretic effects	Therapeutic effects are similar to salicylates, but it lacks anti-inflammatory, antiplatelet, and gastric ulcerative effects.
22:18	COMT	DB00323	Tolcapone	inhibits the enzyme catechol-O-methyl transferase (COMT)	Parkinson's disease
	COMT	DB00494	Encapone	reversible inhibitor catechol-O-methyl transferase (COMT)	Parkinson's disease
	COMT	DB01454	3,4-Methylenedi-oxymethamphetamine	classified as a hallucinogen and causes marked, long-lasting changes in brain serotonergic systems It is commonly referred to as MDMA or ecstasy	post-traumatic stress disorder (PTSD) and anxiety associated with terminal cancer
	COMT	DB01235	L-DOPA	naturally occurring form of dihydroxyphenylalanine and the immediate precursor of dopamine. Unlike dopamine itself, it can be taken orally and crosses the blood-brain barrier. It is rapidly taken up by dopaminergic neurons and converted to dopamine	idiopathic Parkinson's disease (Paralysis Agitans), postencephalitic parkinsonism, symptomatic parkinsonism which may follow injury to the nervous system by carbon monoxide intoxication, and manganese intoxication

	COMT	DB00668	Epinephrine	active sympathomimetic hormone from the adrenal medulla	anaphylaxis and sepsis

Table 7: Drug binding sites were annotated for the genes in the hotspots. When genes in the hotspots were annotated for drug interactions, 16 drug binding sites were found for both commercially developed and illicit substances.

While not all hotspots had drug binding sites located within their domains, five of the hotspots did. When we identified the types of drugs found at these five hotspots we found some thematic divisions in drug targets. Five drug binding sites are associated with cancer therapies: Sorafenib**,** Abiraterone, ABT-263, Etoposide, and Thymalfasin. Four drugs were associated with the mental health condition Parkinson's disease, post-traumatic stress disorder or anxiety: L-DOPA, 3,4-Methylenedi-oxymethamphetamine, Encapone, and Tolcapone. One drug binding sites was specifically targeted ecstasy.

Discussion

We have observed that addiction and mental health genes create eight hotspots in the genome. Their shared functional annotation and drug binding site annotation support the idea that this clustering is meaningful. This approach helps to unify the disparate clinical, genetic, and functional observations about addiction and mental health co-morbidity. Our analyses demonstrate that genes curated for their involvement in opiate, GABA and dopamine addictions share significant genomic position overlap with genes involved in the bipolar disorder, depression and schizophrenia. Additionally when we perform functional annotation on these gene sets, we find that they share core molecular processes such as cell-cell signaling, synaptic transmission and responses to organic substances. These genes also share more amorphous processes such as learning, memory and behavior. Finally we found that when these addiction and mental health genetic hotspots were annotated for their drug interactions, they identified binding sites for illicit drugs, cancer drugs, and neurological degenerative disorders such as Parkinson's disease.

While there has been significant identification of the clinical synergies between addiction and some mental health phenotypes (32, 33, 82), we show here that a subset of these addiction and mental health genes actually sit in genomic windows. These genomic windows contain four of the 51 genes that are common to all gene lists (Table 6); instead we found that genes sitting in genomic hotspots were often shared by only two of the comparisons. When mapped to the genome, these regions were not random with respect to genome locality. These eight hotspot regions contain 192 of the 1968 unique genes identified as participating in addiction and

mental health genetics and represent nearly 10% of genes associated with addiction and mental health.

Functional annotation of genes residing in the windows provides a verification of the role that these genes play within the wider context of the genomic region they inhabit. It is reassuring that our analyses return functional categories emphasizing the shared role that neurological system regulation, cell-cell signaling, stimulus response and organic substance response play in the development of mental health and addiction disorders. These findings support the complex functional interactions between opiate, dopamine and GABA addictions with schizophrenic, depressive and bipolar mental health disorders.

Previous studies have identified that drugs involved with addiction and mental health may have pleiotropic effects (82). It is therefore of significant interest to characterize the drug targets in these hotspot regions. We found that drug binding sites fall in to three major therapeutic categories, those related to cancers, those related to up regulation of immune responses and those involved in neurodegenerative disorders. Of particular interest was the hotspots located on chromosome 19 (ABT-263, Abiraterone, Minoxidil, and Sorafenib) and chromosome 22 (Acetaminophen, Tolcapone, Encapone, 3,4-Methylenedi-oxymethamphetamine, L-DOPA, and Epinephrine).

The observation of increased anxiety behaviors and morphine consumption in a Sprague-Dawley rat model suggests that genetic variation in the epinephrine-mediated norepinephrine signaling pathway may be a novel mechanism for affective behavior such as anxiety and addiction(143). In a human study of African American young adults, levels of urinary epinephrine were predictive of drug use a year later(144). The drug binding site for epinephrine is located in the chr22 hotspot gene COMT. Sorafenib, a hepatocellular carcinoma treatment, is mediated by cellular signaling mechanisms. This drug target annotation is consistent with the genomic functional annotation of the genes underlying addiction and mental health disorders. There is some evidence of genetic variant mediated drug resistance. Sorafenib can activate addiction switches leading to reduced drug efficacy (145). While the mechanisms of this are not fully elucidated, the localization of Sorafenib among known addiction genes could be a reason for this trigger and bears further investigation.

We hypothesized that we could identify genomic sites that sat at the intersection of addiction to opiate, GABA and dopamine with mental health disorders identified as depression, schizophrenia, and bipolar disorder. We have shown that genes involved separately in these two disorders are co-located at eight genomic regions. Additionally we hypothesized that each of these regions might have drug binding sites that share functional annotations with the genes identified in the region. Our analyses have identified 16 drug binding target regions located in five of the eight hotspot regions which share functional or therapeutic activity with addiction and mental health disorder phenotypes. This finding compels us to speculate on the role

that functional ontology plays in the primary of counter-indicative phenotypes that these drugs present.

Specific Aim III: Addiction and Immunity

Introduction

Human immunological interactions with their environment have long been considered the substrate for natural selection(6). While this has been the acknowledged paradigm in evolutionary medicine, few connections have been made between concerning the evolutionary relationships between the complex chronic diseases underlying substance addiction and selection for immunity (39, 83). For example, addiction phenotypes in chronically substance abusing individuals bear striking similarities to immunological compromised individuals. This is thought to be because of an as yet under-characterized interplay between addiction phenotypes and immunity phenotypes. While the relationship has been documented between chronic substance addiction and degenerative immunity, the converse reaction is less well characterized. For example, opioids are known to effect host defenses (38) with heroin addicts presenting higher prevalence of infectious disease than those non-heroin abusers.

Our previous observations of immune genes sharing genome locality with addiction genes at addiction hotspots motivated our interest in further understanding whether addiction alleles might arise in ethnic populations as a result of natural selection against infectious disease at immune loci. Specifically, we hypothesize that alleles associated with addiction that lie within hotspots adjacent to immune functioning genes will be correlated to infectious disease response and by extension will have allele frequencies that are tied to global distributions of infectious disease prevalence, where the associations are of a sufficient duration to be the substrate for

natural selection. To date this has been a challenging endeavor due to the lack of global population sets within which to test these hypotheses.

One approach to studying natural selection in humans has been to examine single genes in a population to directly assess selection caused by some environmental effect (i.e. *HBB* and malaria, *SLC24A5* and UV exposure) (146, 147). An analysis conducted on a global population set has demonstrated success in identifying variation in allele frequencies between populations taking into account diet, subsistence strategy and ecoregions (148). Indeed this approach can be seen as a major innovation in multivariate analysis for factors affecting gene frequencies in human populations. No portrait of environmental selection pressures would be complete without the inclusion of disease status and pathogen density. The interplay of nutrition, location, and exposure to pathogens are compelling external forces that impact individual survival.

By examining correlations between environmental and disease conditions with allele frequencies, we will be able to search for allele frequencies consistent with signatures of selection on multi-locus traits(148). This method is useful in identifying adaptations (whether tolerance or resistance focused) for complex infectious diseases. Asian populations provide an opportunity to examine how environmental effects affect complex traits because they exhibit a variety of subsistence strategies, population histories and exposures to infectious disease. They also live in a variety

of ecological regions, nutritional contexts and latitudes. These various conditions make them a living laboratory for studies in natural selection.

Specific Aim III: Determine whether of infectious disease burden plays a role as an evolutionary driver of addiction genetics.

I hypothesize that environmental factors such as geographical location, relative pathogen burden and infection rates will identify allele frequency differences in ethnic population for immune and addiction gene hotspots. We further hypothesize that these hotspots are consistent with natural selection within human populations living predominantly in tropical environments. This finding would establish a critical link between the agent of natural selection, the genetic process and a complex disease that lies within the close proximity to an allele under selection.

Methods

Identification of Immune and addiction gene clusters

To test whether there are correlative relationships between diet, ecoregions, disease and population allele frequencies, we use genes contained in addiction and immune gene step curated by NCBI. Immune associated genes were identified from NCBI gene lists using search terms, ‘adaptive immunity,’ ‘innate immunity,’ ‘autoimmunity,’ and ‘Th1/Th2.’ These terms were added to 587 genes previously identified as being involved in opiate, dopamine, and GABA reception addiction. These genes were then projected onto the genome to identify cluster regions of genetic importance for

immunity and addiction. Clusters were defined as regions of the genome with more than 25 genes within a 2.5Mb linear genomic window.

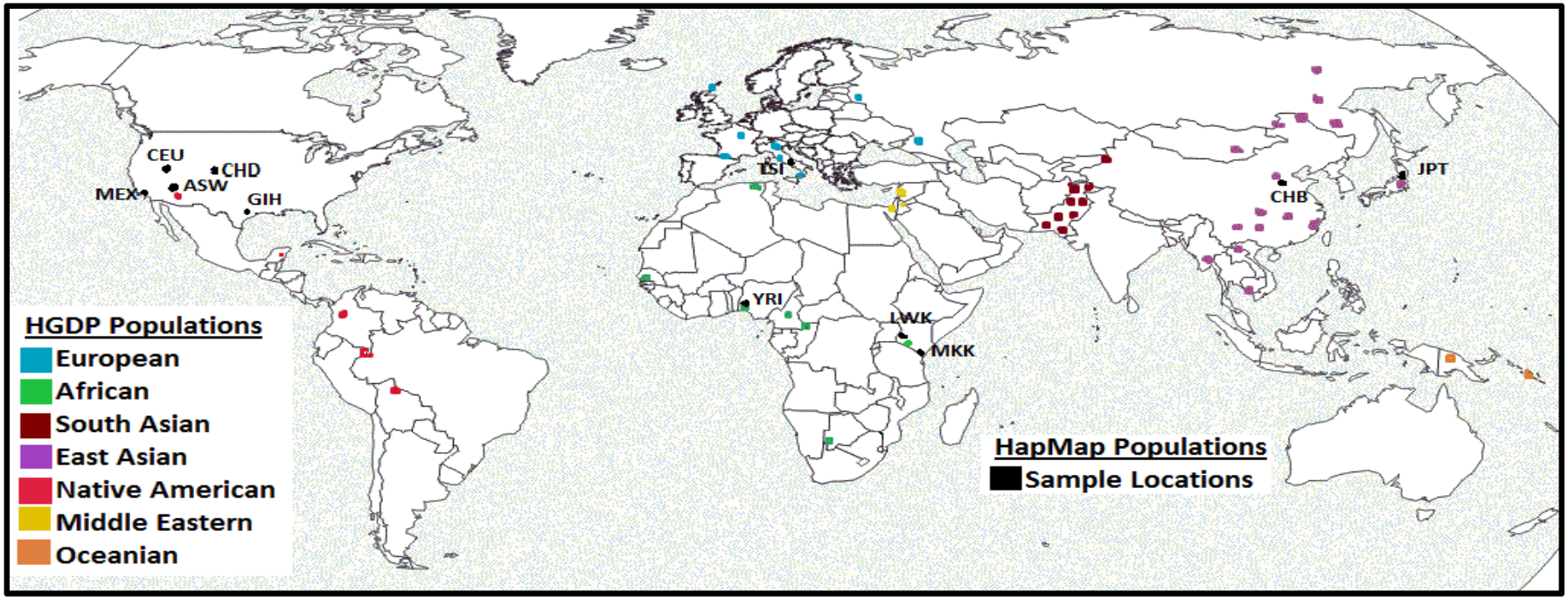
Annotation of Genes in the Addiction and Immunity Windows

To determine the functional role that these NCBI Genes for addiction and immunity as well as candidate genes hotspots play, we performed functional annotation for these three hotspot regions. Genes located within hotspots were considered in two ways in statistical enrichments: all genes in the hotspot window, and only those previously linked to addiction. All genes in the hotspots were annotated using DAVID's Bioinformatics Resources Tool software (101, 102) for biological process, molecular function, and KEGG pathways (46, 49). Functional enrichments were quantified using $pV < 0.05$ and Benjamini score analysis cutoffs of 0.01 (103). KEGG pathways were colored using a MATLAB code to differentiate between genes belonging to different hotspots.

Populations Studied and Determination of Environmental Context

In order to understand how populations vary at these key addiction and immunity clusters, we surveyed human polymorphism data from the 1054 individuals comprising 51 populations of the Human Genome Diversity Panel (HGDP), and the 1078 individuals comprising the 11 sample populations of the HapMap project dataset. The set of polymorphism shared amongst these populations was identified and cross population comparisons were made on allele frequencies. Figure 16 shows the sample populations.

To determine how populations fit into ecological regions, a map with population sampling locations was cross-referenced to global Bailey's ecoregions maps (149). This map characterizes the environmental conditions under which study populations live. Figure 16 also makes a comparison between those populations living in tropical environments as identified by the Bailey's ecoregion map for a global population set. We used sampling locations as the determinant for the locality of the population. For this analysis, the Human Genome Diversity Panel populations were more appropriate to differentiate between those populations Population sample locations were cross-referenced with the World Health Organizations' public health indicators (available online at: http://apps.who.int/gho/data/?theme=main). Disease prevalence and pathogen load data for various infectious diseases including: malaria (*Plasmodium vivax* and *P. falciparum*), Cholera (*Vibrio cholera*), Polio (*Poliovirus spp.*), Schistosomiasis (*Schistosoma japonicum* and *S. mansoni*) and Yellow Fever (*Yellow fever virus*) were obtained from World Health Organization datasets to characterize the epidemiological environments within which population samples live along with additional health indicators which might determine the role that these pathogens might play in determining the in allele frequency variation.



	Populations	SNPs typed	N. of Individuals	# of Tropical populations
HapMap Populations	11	~10 million through Phase III	1078	1
Human Genome Diversity Panel	51	650,000	1054	15

Figure 15: Ethnic populations surveyed for immunity and addiction crosstalk from the HapMap and Human Genome Diversity Project. The HapMap populations (populations described in Table 1) and the HGDP populations: HGDP Africans (green)- Moazibite, Mandenka, Yoruba, Biaka, Mbuti, San, NE Bantu, and SAf Bantu. HGDP Asians (purple); N. Asia: Oroquen, Daur, N. Han, Hezhen, Japanese, Uygur, Xibo); C. China: Han, Yi, She, Tu; S.E. Asia: Naxi, Lahu, Dai, Miao, Cambodian), S. Asian (burgundy): HGDP Europeans (blue): Adygei, Basque, French, North Italian, Orcadian, Russian, Sardinian, and Tuscan ; and HGDP Oceanians (orange): Melanesia, Papua

Geographical mapping of correlated variants

To assess the effect of disease on the Acute Inflammatory Response Pathway, we surveyed 10 genes associated with this process (*AHSG, APCS, C2, CEBPB, CFHR1, KRT1, LBP, MBL2, ORM1, and SIGIRR*) were all part of our immune NCBI Gene lists. This subset of genes was identified using the Molecular Signatures database housed in the Gene Sets Enrichment Analysis database (available online at: http://www.broadinstitute.org/gsea/index.jsp). Acute Inflammatory Response functions in short lived antigenic challenge as is demonstrated from infectious diseases such as Hepatitis B and Hepatitis C. All single nucleotide polymorphisms (SNPs) typed in the 10 genes were assessed for their allele frequencies following the Hancock methodology (148). This refers specifically to SNPs typed in the HGDP populations using the Illumina 650Y platform. SNPs were filtered to exclude those that had minor allele frequencies that fell above 0.90 and below 0.10. These sites were then sorted by geographical region of sample origin into major sub continental regions (Sub-Saharan Africa, Central Asia, Northern Asia, Central China, Southern Asia, and Oceania).

To test whether there is a correlation between Hepatitis B, acute inflammatory response and geography, a Pearson correlation analysis was performed between the variables of mean acute inflammatory response frequency and disease (150). Candidate and neutral SNP sets were combined and a Spearman Rank Correlation was performed to determine whether candidate SNPs were found to be statistical outliers.

Results

Addiction and Immunity Hotspots

Our analyses identified three hotspots containing genes from both addiction and immunity NCBI Gene lists. We note that these three genomic hotspots were identified despite the lack of intersection of gene sets between the addiction and immunity. Addiction and immunity hotspots were located on chromosomes 11, 17and 19. Table 8 identifies the genes in this study with NCBI identified genes bolded and those gene names were color coded to represent the category of gene list with which these genes were initially identified. Hotspots all shared genes from immunity and addiction gene sets. The chromosome 11 hotspot contained 10 genes from autoimmune, cocaine, alcohol and innate immunity lists along with seven genes previously unidentified as participating in addiction and immunity phenotypes. The chromosome 17 hotspot contains 11 genes from innate immunity, alcohol, morphine, and Th1 along with 44 previously unassociated genes. And finally the chromosome 19 hotspot locus contains 18 genes from alcohol, GHB, autoimmune, innate immunity and Th2.

Addiction and Immunity Hotspot Annotation

We used David's Functional Annotation web tool to annotate the genes that were identified in each hotspot and to determine the role that each hotspot played in addiction and immunity disorders. The chromosome 19 hotspot is located between 47.8 and 49 Mb of the Hg18 build of the human genome.

Chr	Location	Genes in Window
11	46,406,640-47,606,115	**CREB3L1, DGKZ, MDK**, CHRM4, **AMBRA1, HARBI1, ATG13**, ARHGAP1, ZNF408, SNORD67, **F2** (autoimmune and innate**)**, LRP4, C11orf49, **CKAPS**, ARFGAP2, **ACP2**(Autoimmune and alcohol), **NR1H3**
17	39,393,369-41,277,468	**KRT16,** KRT42P, EIF1, GAST, HAP1, JUP, LEPREL4, KLHL10, FKBP10, ACLY, TTC25, CNP, DNAJC7, NKIRAS2, ZNF385C, **DHX58,** KAT2A, HSPB9, RAB5C, **HCRT, GHDC,** STAT5B, **STAT5A**(TH2 and Autoimmune**), STAT3,** PTRF, ATP6V0A1, NAGLU, **HSD17B1,** MLX, COASY, PSMC3IP, FAM134C, TUBG1, TUBG2, CCR10, PLEKHH3, CNTNAP1, EZH1, RAMP2, VPS25, WNK4, CNTDN1, **BECN1,** PSME3, AOC3, **AOC2, G6PC,** AARSD1, PTGES3L, RPL27, IF135, RUNDC1, **VAT1,** RND2, **BRCA1**
19	47,870,466-49,085,208	**SULT2A1,** BSPH1, ELSPBP1, CABP5, PLA2G4C, LIG1, **CARD8,** ZNF114, CCDC114, TMEM143, EMP3, SYGR4, KDELR1, **GRIN2D**(alcohol and cocaine), GRWD1, KCNJ14, CYTH2, **SULT2B,** FAM83E, SPACA4, RPL18, SPHK2, DBP, CA11, NTN5, **FUT2,** MAMSTR, RASIP1, IZUMO1, **FGF21** PLEKHA4, PPP1R15A, TULP2, BCAT2, HSD17B14, DHDH, **BAX, FTL,** GYS1, RUVBL2, **LHB, CGB, CGB2, CGB1, CGB5, CGB8, CGB7,** NTF4, KCNA7, **SNRNP70,** LIN7B , PPFIA3, HRC, TRPM4, SLC6A16, CD37, TEAD2, DKKL1, CCDC155, ALD16A1, SLC17A7, PIH1D1, **FLT3LG** (Th1 and Adaptive), RPL13A, SNORD32A, RPS11, FCGRT, RCN3, NOSIP, PRRG2, RRAS, PRR12, SCAF1**, IRF3**

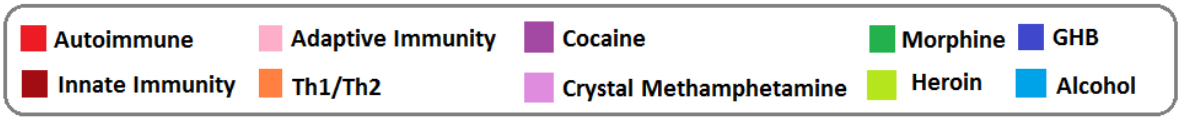


Table 8: Genes identified at the intersection of Addiction and Immunity. Three hotspot regions were identified located on chromosome 17, 11, and 19. Each hotspot had genes identified through multiple addiction and immunity gene lists.

.

	Biological Processes		Molecular function		KEGG Pathways
CHR	Process	P	Function	P	Pathways
17	JAK-STAT Cascade	5.4^{-5}	Phosphate binding	0.01	-Acute myeloid leukemia -Adipocytokine signaling pathway
	Growth hormone receptor signaling pathway	2.5^{-4}	GTPase activity	3.6^{-3}	
	Response to growth hormone	6.6^{-3}	Steroid hormone receptor binding	5.8^{-3}	
	Homeostatic process	0.01	Ion binding	0.04	
	eating behavior	0.02			
11	catalytic activity	0.01	nucleotide regulator activity	0.01	-Regulation of actin cytoskeleton -Neuroactive ligand-receptor interaction -Cholinergic synapse -Hepatitis B -Complement and coagulation cascades -Peroxisome -Hypertrophic cardiomyopathy -Nucleotide excision repair -Dilated cardiomyopathy -Alcoholism
	protein import	0.02	enzyme activator activity	3.0^{-3}	
	regulation of cellular protein metabolism	0.04			
	Negative regulation of endocytosis	0.02			
	Regulation of phosphorylation	0.04			
19	fertilization	3.4^{-3}	Hormone activity	6.6^{-4}	-Ribosome
	Neurotransmitter transport	4.0^{-3}			
	Cell-cell signaling	7.7^{-3}			
	Single fertilization	0.02			

Table 9: Functional Annotation of Addiction and Immunity Hotspot Crosstalk Regions. Genes identified in each hotspot were annotated from biological process, molecular function and pathway participation. Annotation was undertaken using DAVID functional annotation with a P value cutoff of 0.05.

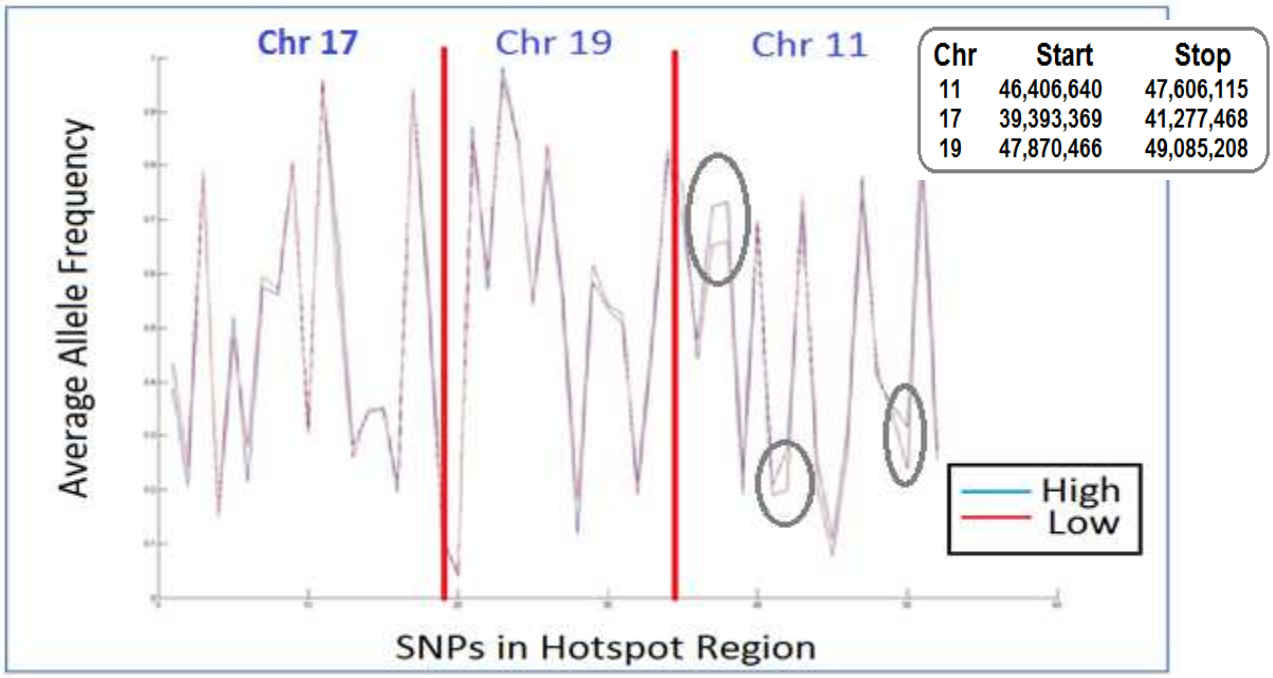


Figure 16: Five SNPs underlying the Chr 11 hotspot vary between high tropical versus low tropical living HGDP populations. Allele frequencies were plotted in the tropical living climates (blue) versus those living in temperate climates (red). The inset shows the hotspot locations.

Annotation of Tropical identified SNPs

We further annotated the top five polymorphisms that showed allele frequency differences between populations that lived in temperate environments and those living in tropical environments as illustrated in Figure 2. These five SNPs were identified in the HGDP populations: rs11818969, rs17790342, rs12417519, rs752849, and rs901746. Annotation of these five SNPS varying showed that they exhibit addiction, mental health and immune function. These SNPs are further characterized in Table 7. Additionally we typed these five polymorphisms in the 11 HapMap populations to understand whether these populations showed similar patterns to the HGDP population datasets. Our analyses confirm that they HapMap

and HGDP populations show allelic congruence at these five sights, suggesting this pattern is a true representation of what is happening in human populations.

					Africans				Europeans		East Asian		
SNPs	Chr	Location	Gene	Type	LWK	MKK	YRI	ASW	CEU	TSI	CHB	CHD	JPT
rs11819869	11	46500680	AMBRA1	Intron	0.644	0.451	0.566	0.592	0.159	0.153	0.049	0.041	0.047
rs17790342	11	47067706	C11orf49	intron	0.089	0.091	0.062	0.173	0.088	0.091	0.037	0.047	0.041
rs12417519	11	47069397	C11orf49	intron	0.828	0.818	0.845	0.837	0.781	0.812	0.354	0.333	0.314
rs752849	11	47115327	C11orf49	intron	0.533	0.563	0.442	0.490	0.775	0.727	0.927	0.935	0.930
rs901746	11	47200319	DDB2	intron	0.483	0.486	0.606	0.571	0.301	0.222	0.744	0.759	0.738

Table 10: Top five SNPs have shared patterns of allele frequency differences in HapMap populations as in HGDP populations. SNPs are identified with their location type and gene. We then typed these genes in the HapMap populations and found congruence in the tropical populations of the HapMap and those of the HGDP.

The rs11818969 polymorphism found in the *autophagy/beclin-1 regulator 1* gene (*AMBRA1*) is intimately involved in the development of the nervous system (151). It has been shown to be a component of a complex network between autophagy, cell growth and cell death of crucial importance to neural development (152). In particular, the rs11818969 polymorphisms have been identified as Schizophrenia associated polymorphism (153). A follow up study of this polymorphism has extended this work to find that it specifically alters impulsivity-related behavioral and neural traits (154). We note that the AMBRA2 gene is more than four times more prevalent in African populations than in Europeans and more than 15 times more prevalent than in East Asian populations of the HapMap. In an African American HapMap sample collected in the southwestern United States (ASW) the frequency was found to be 0.592

There were three polymorphisms, rs17790342, rs12417519, and rs752849, which are located in the Chromosome 11 open reading frame 49 (*C11orf49*). While open reading frame regions in the genome are not well characterized, these particular SNPs have been identified as participating in the liver interactome (155). The liver is the site of metabolism of xenobiotics so these SNPs may have metabolic function.

Finally the rs901746 polymorphism is located in the Damage-Specific DNA Binding Protein 2 (*DDB2). DDB2* is involved in the repair of UV damage to DNA. The *DDB2* gene participates in a complex that mediates the ubiquitylation of histones H3 and H4, which facilitates the cellular response to DNA damage. Additionally, the DDB2 genes has been implicated in lung cancer susceptibility (156), and most importantly in the destabilized the Hepatitis B viral protein X (157). This destabilization of the viral protein X is thought to be a key component in the prevention of viral particle proliferation.

We used GWAS3D software to look for local and long range interactions identified from genome wide association studies, regulatory variation. In particular it is good for characterizing noncoding phenotypically associated variants that underlie the molecular mechanisms of complex traits. Our analysis identified three long range interactions with AMBRA1 (see Figure 17). The first occurs between the AMBRA1 gene, and 2 genes located in a chromosome 8 interaction window: *ZNF705D* is a zinc finger protein that is thought to be involved in transcriptional regulation, while *FAM66D*, is a human specific gene that is known to interact with

Tetrachlorodibenzodioxin. According to the Comparative Toxicogenomics Database, Tetrachlorodibenzodioxin has more than 1000 documented interactions with *AHR* (N=3255 Interactions) and *CYP1A19* (N=1795 Interactions) and is also used as a treatment for drug-induced liver damage.

dbSNP ID	GENE	SNP Functional Annotation	**Interacting dbSNP ID**	CHR:location	GENE	Location	Transcription Factor**s**?
rs11819869	AMBRA1	Schizophrenia associated	rs7130141	11:46499874	AMBRA1	Intronic	Yes- EBF1
rs17790342	C11orf49	Liver interactome	rs12576831	11:47082255	C11orf49	Intronic	No
rs12417519	C11orf49	Liver interactome	rs11601798	11:47158392	C11orf49	Intronic	No
rs752849	C11orf49	Liver Interactiome	rs7940473	11:47182353	C11orf49	Intronic	Yes- CTCF
rs901746	DDB2	Inhibition of Hepatitis B Protein X	rs2167079	11:47270255	ACP2	**Coding**	Yes- 62 TFs

Table 11: GWAS3D analysis of the tropical segregating polymorphisms showed interacting SNPs, transcriptions factors and ACP2, a gene involved in de-phosphorylation. We characterized the set of SNPs that were frequency divergent in tropical and non-tropical living populations. The set had interacting SNPs that were locally located, often in the same gene with the exception of rs901746 which interacted with rs2167079 at ACP2, an adjacent gene. Three of the interacting SNPs sat within transcription factors, with the ACP2 gene SNP (rs2167079) sat at the intersection of 62 different transcription factors.

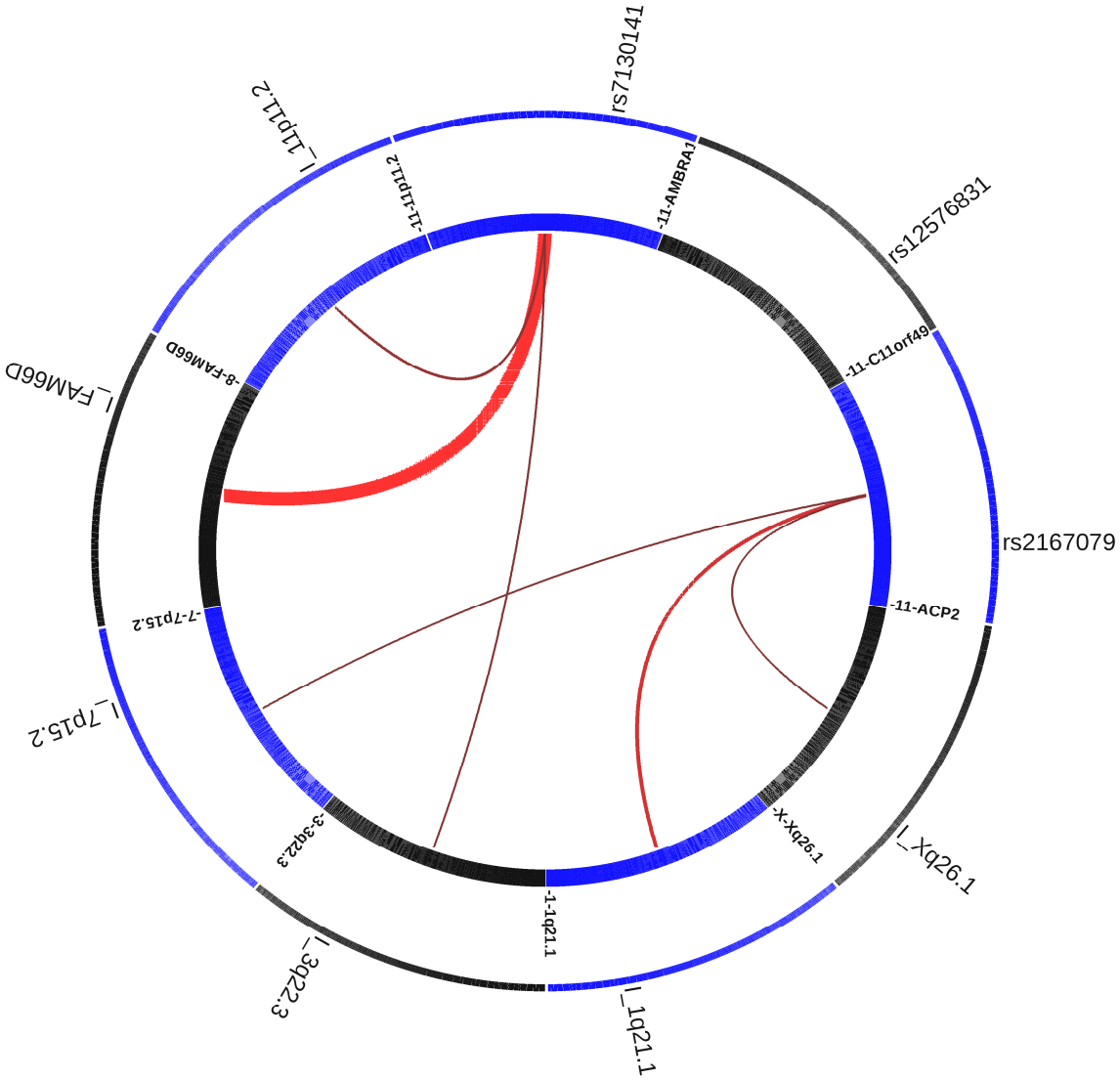


Figure 17: GWAS3D analysis of the five single nucleotide polymorphisms found to differ in tropical versus non-tropical living populations show that SNPs have local and long range interactions. AMBRA1 showed 2 long range interactions on chromosomes 3 and 8 along with a local interaction on chromosome 11. The ACP2 gene (rs2167079), an interacting partner of rs901746, had three interactions: at Xq26.1, 1q21.1, and 7p15.2.

The second region with *AMBRA1* interactions is a 10Kb chr3 interaction region (chr3:136520001-136530000) that lies just upstream of the *SLC35G2* gene. While this gene is not well annotated, there are 13 ENCODE-identified different transcription factors found bundled together (FOS, KAP1, JUN, MEF2, NFIC, BATF, ATF2, USF1, USF2, CTCF, GATA3, and RUNX3). The third regional interaction is a local one located between the *AMBRA1* gene and a local chr11 region (chr11:46180001-46190000). There is one polymorphism identified as rs7128538 which has been associated with Systemic Sclerosis (158).

Two regions were shown to have strong enhancer signatures: chr11:46046791-46539727 and chr11:46446962-46516078. These signatures are generated from the ENCODE analysis of H3K4me1, H3K27ac, P300, and DHS. These enhancers represent epigenetic enhancers of on genomic sequences. Finally there was one identified Conservation Region of GERP++ Elements located at chr11:46499803-46500030. GERP++ Elements are constrained elements in multiple alignments and are estimates potential functional constraint. These are summarized in Table 11.

Acute Inflammatory Response and Hepatitis B

Following up on this observation that rs901746 is a polymorphism involved in Hepatitis B prevention, and our hypothesis that infectious disease might be the driving force shaping allele frequencies at adjacent addiction and mental health sites, we wanted to identify whether we could identify if there was a correlation between Hepatitis B infection prevalence in a global set of human populations and these populations. When we studied 10 genes associated with acute inflammation in

HGDP Africans (*Africa*- Mandenka, Yoruba, Biaka, Mbuti, San, NE Bantu, SAf Bantu), Asians (*N. Asia*: Oroquen, Daur, N Han, Hezhen, Japanese, Uygur, Xibo); *C. China:* Han, Yi, She, Tu; *S. Asia:* Naxi, Lahu, Dai, Miao, Cambodian) and Oceanians (*Oceania:* Melanesia, Papua), we found that there was a trend towards significance (pV= 0.08). Alleles were then correlated to a map of hepatitis B prevalence described from sentinel surveillance conducted in 2004 (159).

Mean allele frequencies were calculated within an ethnic group sample and then across a geographical region (Figure 18) . These candidate regions were compared to neutral SNPs identified using HOMINID coordinates representing 71 regions of the human genome that are far from genes/motifs and are thought to be consistent with neutral evolutionary processes. On the x axis – the mean minor allele frequency (MAF) for African populations was 0.3418, while the mean pan Asian MAF mean was 0.3781. This was significantly different (p = 0.00436). When Asian populations were grouped regionally, Central Chinese sample locations (Han, Yi, She, and Tu) had a MAF mean almost identical to African populations (0.3428), and the central Chinese population differed significantly from N. Asia and S. Asia regional means (p=0.0045 and p=0.0022, respectively) using the Student's T-test statistic to determine significance.

To test whether there is a correlation between Hepatitis B, acute inflammatory response and geography, we performed a Pearson correlation analysis between the variables of mean acute inflammatory response frequency and disease (150). This result is consistent with the central Chinese region being a nexus of Hepatitis B transmission. The pattern observed in genes of the acute inflammatory response

pathway is consistent with the observation that acute inflammation is an integral part of the Hepatitis B infection process (155). When HGDP sampling locations are overlaid with Hepatitis B prevalence maps, there is trend towards congruence between high Hepatitis B prevalence and low MAF frequencies (159).

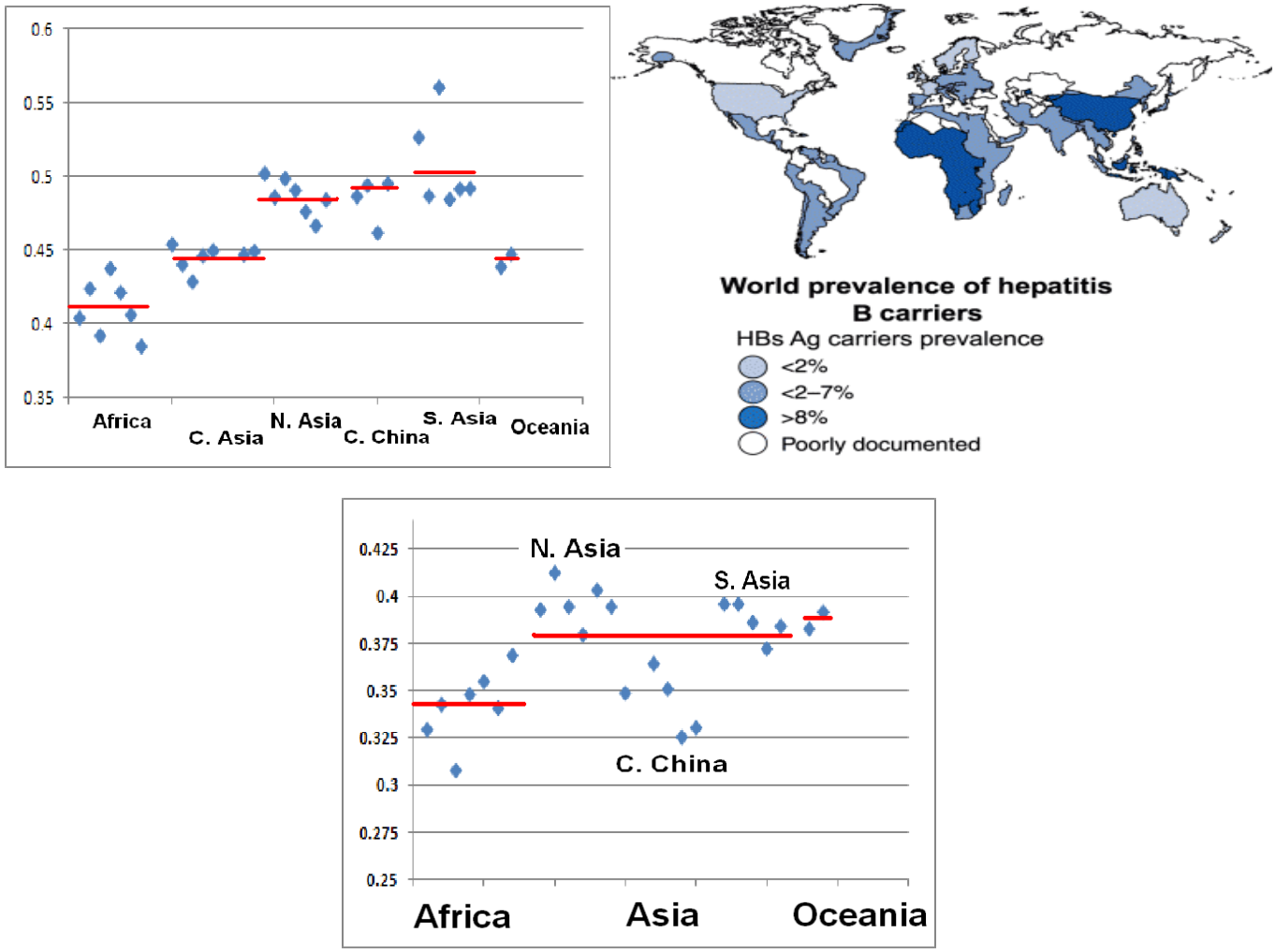


Fig 18: Neutral SNPs in global human populations and the Acute Inflammatory response gene set surveyed in a diverse set of African and East Asian populations. Neutral SNPs only show significant differences between Africans and non-African populations. A Pearson correlation of Hepatitis B prevalence and mean acute inflammatory response allele frequency shows a trend toward correlation (p=0.08).

Discussion

Our findings show that genes involved in opiate, dopamine and GABA addiction do form three genomic hotspot regions in conjunction with immunity regulating genes within the human genome. These positions are located on three separate chromosomes: chr 9, chr 11, and chr19. Functional annotation of these joint addiction and immunity hotspot regions confirmed that the genes previously identified in either addiction or immune surveys share broad functional classifications with those candidate genes that are their hotspot neighbors. Finally when we survey genetic polymorphisms that underlie these three genomic regions in a global distribution of human populations we find that when populations are grouped by their locality in tropical ecological zones, we identified polymorphisms that significantly differ at the chromosome 11 hotspot region. Further annotation of these climate and frequency divergent polymorphisms showed that they have been identified as playing key roles in liver interactome function, schizophrenia and Hepatitis B infection. A follow up analysis of the relationship between Hepatitis B and the acute inflammation process points to a correlation between Hepatitis B and acute inflammation pathways in African and Asian populations living in those areas of the continent that have high hepatitis burden. Taken together, our results point to a functional trifecta existing between immunity, addiction and mental health related SNPs at the chromosome 11 locus. We can infer that hepatitis B is at least a contributing factor in the addiction and mental health allele frequencies that are seen in tropical living populations.

When these addiction and immunity genomic hotspots were functionally annotated using the David's functional annotation web tool (101, 102, 160, 161), we found that while hotspots were not well annotated with common computational tools, the genes underlying these regions represent candidates for both mental health (and infectious disease genes. Interestingly, alcoholism was one of the KEGG pathways identified as participating in the metabolic dynamics of these genomic regions. This gives additional support to our conjecture that the SNPs for C11orf49 may indeed play a role in the liver interactome in a manner that potentially creates susceptibility to alcoholic substances.

We considered the relationship between tropical-living populations, as a climatological surrogate for multiple pathogen loads. These analyses determined that those populations living in highly tropical environments showed the highest polymorphism frequency differences to their temperate climate-living population comparisons. This was true when populations were considered without respect to ethnic origin, confirming that human population locality and more specifically proximity to pathogens was sufficient to explain the observed allele frequency differences. This finding further supports the idea that local climate and specifically potential disease burden appear to play a significant role in shaping both immunological but also co-located non-immunological phenotypes.

Previous studies examining the role of climate – either as a surrogate for pathogen load (162) or as a function of climate change (163) have shown that local climate affects the availability of the pathogen substrates that are thought to drive natural selection in human populations. It is therefore not surprising that selection for

immune fitness can lead to hitchhiking of addiction associated SNPs. The proximity of these polymorphisms indicates that their evolutionary histories are intertwined. This concept now gives us a fertile ground on which to build further analyses between the addiction alleles and immune-regulatory elements under infectious disease based selection. Furthermore, there is evidence that opioid peptides may have first arisen as modulators of cellular immune function- where morphine down regulates immune processes in addiction, an action/function that it appears to normally perform (39). This strengthens our argument that these putatively adaptive features that increased human fitness in disease rich environments may now be causing secondary effects when the immune pressure is removed, or when xenobiotics such as the addictive substances we study here are given in non-homeostatic doses.

Our assessments of GWAS3D interacting polymorphisms showed that the five polymorphisms identified have both local and long distance interactions. These interactions include rs2167079, a coding SNP that sits in at least 62 transcription factors. These analyses demonstrate that addiction genes do indeed form genomic hotspots with immune genes, that these immune hotspots share functional cohesiveness, and that when SNPs at these addiction and immunity hotspots were compared between HGDP populations living in tropical versus temperate climates, they identified SNPs associated with hepatitis B, the liver interactome, and schizophrenia. A follow on analysis of a subset of immune genes involved in the acute inflammatory response show congruence with reported hepatitis B endemicity geographic distributions in central Asia and sub Saharan Africa. This analysis begs

follow up in understanding how pervasive this phenomenon is in complex disorders and their associated immune responses.

Conclusions

Summary of Findings

Addiction disorders represent a major economic and health cost to human populations (51). Addiction is loosely defined as a chronic relapsing spectrum disorder characterized by loss of control over substance taking (52-54). It is a behavior-based phenomenon representing a diverse array of psychological, biological, and genetic attributes and environmental and cultural factors (55-57). As with addiction, neurological disorders such as depression, bipolar disorder, and schizophrenia represent a significant strain on health and judicial entities. Both addiction and mental health disorder classes have long been identified as co-morbid conditions (124). While we can make clear inferences about the role that genes play in Mendelian genetic disorders, characterizing the genetic underpinnings of complex disease has been significantly more challenging. The intersection of two complex diseases such as addiction and mental provides an additional level of genetic complexity. Both Substance addiction and mental health are behavior-based phenomenon representing a diverse array of psychological, biological, and genetic attributes and environmental and cultural factors (32, 33, 55-57). Schizophrenia is a mental illness that affects 1% of the global human population(132). It is identified as a disorder that disrupts brain neural networks and is characterized by hallucinations, delusions, lack of willpower, and cognitive deficits (164). Genealogy studies have shown that there is a strong genetic component to schizophrenia with genetic components accounting for as much as 80% of the risk variance (165). Linkage studies and candidate gene approaches have identified over 1000 candidate genes associated with schizophrenia. By examining correlations between environmental,

addiction, and disease conditions with allele frequencies, I searched for allele frequencies consistent with signatures of selection on multi-locus traits. This method is useful in identifying adaptations (whether tolerance or resistance focused) for complex infectious diseases. Tropical populations provide an opportunity to examine how environmental effects affect complex traits because they exhibit a variety of subsistence strategies, population histories and exposures to infectious disease.

A list of biologically relevant genes for addiction was gathered from literature curated sources. *Figure 3* shows the flowchart employed to identify biologically relevant addiction genes and the resulting number of genes in each addiction gene set. Three classes of addiction genes were obtained using the NCBI gene search: dopamine addiction genes (N=108), opiate addiction, (N=246) and GABA addiction genes (N= 433). The search word chains consisted of 'metabolism' (N=398 genes) or 'addiction' (N=461) followed by the names of three addictive substances: dopamine, opiates and GABA receptor. The respective heat maps in Figure 3 illustrate the intersections within the search terms for both metabolism and addiction gene lists. Additionally, we compared our gene list to the one reported by Li, Mao and Wei (1), containing 387 genes involved in four addiction disorders. This comparison is shown using a Venn diagram with the bulk of genes (N=311) not identified in our analyses belonging to nicotine addiction, an addiction disorder we do not address here. The rest was added to our gene list. A set of addiction genes (N=587), compiled from the union of all search terms, was thus determined.

We hypothesized that genes involved in the characterized opiate, dopamine and GABA addiction disorders will form genomic regions of with functional specificity. These regions should exist above and beyond that small subset of genes that are shared between these gene lists. Our analyses demonstrate that genes associated with dopamine, opiate and GABA addictions cluster into seven regions of the genome that have a significant or near significant overabundance of addiction genes based on a hypergeometric test (Table 2). These seven hotspots were split between GABA specific (4q23, 6p22.2, 6p22.1) and mixed addiction hotspots, containing all genes, (10p15.1, 11q13.2-3, 16q22.1, and 19q13.33).

Functional annotation was undertaken for the list of addiction genes, the GABA, opiate, and dopamine addiction classes, and their pairwise intersections. These genes when functionally annotated both with and independent of their interspersed neighbor genes shared functions. This finding supports the idea that these NCBI-curated hotspot genes are truly identifying regions with functional genomic signatures for addiction. Functional annotation shows a wide range of biological processes dominated by cell-cell signaling. Dopamine and opiate addiction gene lists shared many more molecular functions with each other than they did with the GABA addiction gene set. The genes common to the three types of addiction under consideration included those with ligand-gated ion channel activity, neurotransmitter binding, and amine binding activities. The biological processes enriched included response to organic substances, synaptic transmission processes, and response to endogenous stimulus process. This approach has identified additional candidate

genes such as those in the SLC17A family, which code a sodium/phosphate co-transporter in the intestinal mucosa. The protein plays an important role in the absorption of phosphate from the intestine and its possible role in addiction is yet to be determined.

We examined the hotspot associated polymorphisms identified in 11 HAPMAP sample populations with distinct geographical occupation: East Asian ancestry [Japanese-JPT, Chinese (collected in Beijing)-CHB, Chinese (collected in Denver)-CHD], African ancestry populations [Yoruba-YRI, Masaai-MKK, Luhya-LWK, and African Americans-ASW], European ancestry populations [Europeans of Northern and Western Ancestry-CEU and Toscana- TSI], a South Asian ancestry population [Guajarati in Houston-GIH]; and an admixed American population [Mexicans in Los Angeles-MEX] (104-106). To exclude the possibility of confounding effects of population-specific demography and to set up an empirically derived neutral estimate of allelic variation, we analyzed 20 concatenated autosomal loci across the human genome identified as neutrally evolving (107). It was assumed that the polymorphism variation undergoing selection will have non-neutral allele frequency patterns. When we considered genetic polymorphisms in 11 HapMap populations that span three major ethnic regional populations (East Asian, European, and African), we identified polymorphisms (as shown in Figure 9 that varied between East Asians and Africans/Europeans populations (rs6906576, rs6924948, rs7740793, rs9348699, rs933199).

We followed these frequency-based analyses of hotspots with analysis uncovering linkage to distant sites. For this purpose we used the web platform GWAS3D developed recently by Jun Li M et al. (109). The platform identifies genetic variants affecting regulatory pathways and underlying disease/trait associations by integrating chromatin state, functional genomics, sequence motif, and conservation information given a variant list. In the addiction case under study, we examined the distant regulatory landscape through linkage of the significant SNPs identified at hotspots in cross population comparisons. Figure 11 shows the significant common variants between Africans and Europeans projected onto the Yoruba population. In the outer ring, polymorphisms or genomic regions are identified. The second ring identifies the genes or chromosomal regions these polymorphisms sit in and finally the red lines indicate the strength of local or long range interactions. Thus, with the use of emerging bioinformatics web platforms, deciphering addiction hotspots on the human genome show potential for further discovery of DNA motifs distant to the hotspots.

These addiction findings made us question whether our computational approach could be used to identify those genes that are sitting at the intersection of multiple complex disorders. A number of studies have identified the clinical intersection of mental health and substance addiction traits (125, 126). This has been shown in a diverse variety of mental health conditions such as depression (127, 128), bipolar disorder (129-131), and schizophrenia (132-137). To date, the crosstalk of addiction

and mental health genetic contributors has not been fully understood based on association studies (68, 138) and functional genomics (41, 42).

The bipolar disorder gene list comprised 626 genes, depression comprised 357 genes, schizophrenia comprised 1121 genes and Addiction comprised 587 genes garnered from opiate, GABA and dopamine addictions. When considered together this genes list represented 1968 genes encoding putative addiction and mental health targets. Figure 12 is a Venn diagram that shows the overlap in genes from these four lists. Interestingly, there were 51 genes that were shared between all gene lists but only four of these shared genes were represented in the eight hotspots. These are indicated in red in Table 6. This overlap gene set contained the DRD, HTR and SLC6A gene families as well as a host of immune function genes including: ICAM1, IFNG, IGF1, IL1B, IL1RN, and TNF. Of this combined set of addiction and mental health genes, 192 genes fell into cluster regions within the genome. Our analyses identified these eight genomic regions (with significant numbers of genes involved in the dual disorders of addiction and mental health.

We performed a functional annotation on the complete list of NCBI identified genes in order to determine the major functional roles for addiction and mental health genes. Functional annotation of the combined set of depression, schizophrenia, bipolar and addiction genes found functional commonalities. These functions are characterized in Figure 14.

Drug target annotation was performed to determine the role that pharmacological products interacted to binding sites within these hotspots. When the hotspot regions were annotated for drug interactions 16 drugs were found to have binding sites with shared function. This function revolves around neurological function, response organic substances and cell-cell signaling. Addiction and depression overlapped in three functional regulatory roles: cellular localization, cyclic nucleotide biosynthesis and metabolism. While not all hotspots had drug binding sites located within their domains, five of the hotspots did. When we identified drugs that identified these five hotspots, we found some thematic division in the types of drug binding site targets. Five drug binding sites are associated with cancer therapies: Sorafenib, Abiraterone, ABT-263, Etoposide, and Thymalfasin. Interestingly, a survey of the literature regarding these drug's effects showed that they were involved in secondary addiction and mood disorder phenotypes. Four drugs were associated with the mental health condition Parkinson's disease, post-traumatic stress disorder or anxiety: L-DOPA, Encapone, Tolcapone, 3,4-Methylenedi-oxymethamphetamine. One drug binding sites were specifically directed at an illicit drug, ecstasy. These findings are summarized in Table 7.

We wanted to follow up our early observations that addiction hotspots genes at two hotspots the 6p21.1 and 4q23 were interspersed with immune function genes, and that polymorphisms at these hotspots showed allele frequency differences in the Yoruba population, a sub-Saharan tropical living African population. We hypothesized that environmental factors such as geographical location, relative

pathogen burden and infection rates will identify allele frequency differences in for immune and addiction gene hotspots between populations that are consistent with natural selection within human populations living predominantly in tropical environments.

To test whether there are correlative relationships between ecoregions, disease and population allele frequencies, we use genes contained in addiction and immunity curated by NCBI. Immune associated genes were identified from NCBI gene lists using search terms, ‘adaptive immunity,’ ‘innate immunity,’ ‘autoimmunity,’ and ‘Th1/Th2.’ These terms were added to 587 genes previously identified as being involved in opiate, dopamine, and GABA reception addiction (submitted in Jackson et al). These genes were then projected onto the genome to identify cluster regions of genetic importance for immunity and addiction. Clusters were defined as regions of the genome with more than 15 genes within a 1.5Mb linear genomic window.

When addiction and immunity gene lists were combined, we found that they created three hotspots located on chromosomes 11,17, and 19. In order to understand how populations vary at these key addiction and immunity clusters, we surveyed human polymorphism data from the 1054 individuals comprising 51 populations of the Human Genome Diversity Panel (HGDP), 1148 individuals comprising the 11 sample populations of the HapMap Project and the 1092 individuals representing the 1000 Genomes dataset. The set of polymorphism shared amongst these datasets was identified and cross population comparisons were made on allele frequencies.

To determine how populations fit into ecological regions, a map with population sampling locations was cross-referenced to global Bailey's ecoregions maps(149). This map characterizes the environmental conditions under which study populations live. We grouped populations into groups of tropical environment, moderate environment and non-tropical environment. This sorting process split African, Asian and American populations, thereby ensuring that the effect that we saw was not the result of ancestry but instead was eco-climate related. Our analyses demonstrate that when we grouped a global set of 51 human populations into tropical versus non-tropical living groups (the distal ends of our sorting spectrum), we found significant differences in allele frequencies at the hotspot located on chromosome 11 for 5 polymorphisms (Table 10). Finally we used GWAS3D to identify local and long range interactions between our 5 significant SNPs and their cis and trans-chromosomal partners.

These analyses have demonstrated that NCBI identified addiction genes form hotspots in the genome. These genomic hotspots for opiate, dopamine and GABA addiction share functional cohesiveness. Analyses of the SNPs typed in HapMap populations demonstrated that this approach can identify genomic variants of interest within these hotspot regions. We then extended our analyses to the intersection of addiction with three mental health conditions: schizophrenia, depression and bipolar disorder. Each of these mental health conditions has a strong genetic basis and we identified 7 hotspot regions sitting in the intersection of

addiction and mental health. We characterized the 16 drug binding sites in five of the seven hotspots which recapitulated the functional intersections of addiction and mental health, while also indicating that again, infectious disease therapies had binding sites within our regions. Finally we directly studied addiction and immunity genes sets to identify three genomic hotspots. These genomic hotspot again shared functional cohesiveness and when typed in the HGDP populations, we discovered that five polymorphisms had statistically different allele frequencies in tropical living versus non tropical living populations. These variants point to hepatitis B as a potential selective agent. Following this line of inquiry, we studied genes involved in acute inflammatory response, a subset of our immune gene set. We saw convergence between our geographic regions and those regions of the world with endemic hepatitis B infection. This result points to a strong link between infectious disease and addiction. We therefore propose that hepatitis B driven infection may be a major factor in the changes in allele frequencies in seen in five alleles on chromosome 11, of which three are involved with the liver interactome and one is involved with schizophrenia phenotypes. These alleles present compelling evidence for further selection analyses to identify whether hitchhiking, the rise of alleles in proximity to a selected allele, is the mechanism for these variant differences.

Bibliography

1. Li CY, Mao X, Wei L. Genes and (common) pathways underlying drug addiction. PLoS computational biology. 2008;4(1):e2.
2. Darwin C. On the origin of species by means of natural selection, or the preservation of favored races in the struggle for life. Champaign, Ill.: Project Gutenberg,. Available from: http://libproxy.rpi.edu/login?url=http://www.netlibrary.com/urlapi.asp?action=summary&v=1&bookid=1057000.
3. Darwin C. The descent of man, and selection in relation to sex. New York: D. Appleton and company; 1873.
4. Lederberg J. J. B. S. Haldane (1949) on infectious disease and evolution. Genetics. 1999;153(1):1-3.
5. Haldane JB. Mathematical Darwinism: A discussion of the genetical theory of natural selection. The Eugenics review. 1931;23(2):115-7.
6. Haldane JB. Natural selection in man. Acta genetica et statistica medica. 1956;6(3):321-32.
7. Wright S. Evolution in Mendelian Populations. Genetics. 1931;16(2):97-159.
8. Wright S. The Distribution of Gene Frequencies in Populations. Proceedings of the National Academy of Sciences of the United States of America. 1937;23(6):307-20.
9. Kemper JT. The evolutionary effect of endemic infectious disease: continuous models for an invariant pathogen. Journal of mathematical biology. 1982;15(1):65-77.
10. Ikwueke K. The changing pattern of infectious disease. Br Med J (Clin Res Ed). 1984;289(6455):1355-8.
11. Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, et al. Common west African HLA antigens are associated with protection from severe malaria. Nature. 1991;352(6336):595-600.
12. Andreasen V, Christiansen FB. Disease-induced natural selection in a diploid host. Theoretical population biology. 1993;44(3):261-98.
13. Levin BR, Lipsitch M, Bonhoeffer S. Population biology, evolution, and infectious disease: convergence and synthesis. Science. 1999;283(5403):806-9.
14. Karlsson EK, Harris JB, Tabrizi S, Rahman A, Shlyakhter I, Patterson N, et al. Natural selection in a bangladeshi population from the cholera-endemic ganges river delta. Science translational medicine. 2013;5(192):192ra86.
15. Elhassan AA, Hussein AA, Mohamed HS, Rockett K, Kwiatkowski D, Elhassan AM, et al. The 5q31 region in two African populations as a facet of natural selection by infectious diseases. Genetika. 2013;49(2):279-88.
16. Antonovics J, Boots M, Ebert D, Koskella B, Poss M, Sadd BM. The origin of specificity by means of natural selection: evolved and nonhost resistance in host-pathogen interactions. Evolution; international journal of organic evolution. 2013;67(1):1-9.
17. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. Nature reviews Genetics. 2014;15(6):379-93.
18. Allison AC. The distribution of the sickle-cell trait in East Africa and elsewhere, and its apparent relationship to the incidence of subtertian malaria. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1954;48(4):312-8.
19. Kwiatkowski DP, Luoni G. Host genetic factors in resistance and susceptibility to malaria. Parassitologia. 2006;48(4):450-67.
20. Kwiatkowski DP. How malaria has affected the human genome and what human genetics can teach us about malaria. American journal of human genetics. 2005;77(2):171-92.

21. Duggan CW. The PARASITE of MALARIA in the FEVERS of WEST AFRICA. British medical journal. 1898;1(1933):139-40.
22. Edington GM, Lehmann H. Sickle-cell trait and malaria in Africa. Bulletin of the World Health Organization. 1956;15(3-5):837-42.
23. Joy DA, Feng X, Mu J, Furuya T, Chotivanich K, Krettli AU, et al. Early origin and recent expansion of Plasmodium falciparum. Science. 2003;300(5617):318-21.
24. Tishkoff SA, Verrelli BC. Patterns of human genetic diversity: implications for human evolutionary history and disease. Annual review of genomics and human genetics. 2003;4:293-340.
25. Tishkoff SA, Verrelli BC. Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping. Current opinion in genetics & development. 2003;13(6):569-75.
26. Kwiatkowski D. Genetic susceptibility to malaria getting complex. Current opinion in genetics & development. 2000;10(3):320-4.
27. Kwiatkowski DP. The complexity of genetic variation in a simple immune system. Trends in genetics : TIG. 2005;21(4):197-9.
28. Costabile M, Quach A, Ferrante A. Molecular approaches in the diagnosis of primary immunodeficiency diseases. Human mutation. 2006;27(12):1163-73.
29. Hudson ME. Sequencing breakthroughs for genomic ecology and evolutionary biology. Molecular ecology resources. 2008;8(1):3-17.
30. Lee JH, Jeon JT. Methods to detect and analyze copy number variations at the genome-wide and locus-specific levels. Cytogenetic and genome research. 2008;123(1-4):333-42.
31. Muldrew KL. Molecular diagnostics of infectious diseases. Current opinion in pediatrics. 2009;21(1):102-11.
32. Tomkins DM, Sellers EM. Addiction and the brain: the role of neurotransmitters in the cause and treatment of drug dependence. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2001;164(6):817-21.
33. Kamerow DB, Pincus HA, Macdonald DI. Alcohol abuse, other drug abuse, and mental disorders in medical practice. Prevalence, costs, recognition, and treatment. JAMA : the journal of the American Medical Association. 1986;255(15):2054-7.
34. Crews FT. Immune function genes, genetics, and the neurobiology of addiction. Alcohol research : current reviews. 2012;34(3):355-61.
35. Mayfield J, Ferguson L, Harris RA. Neuroimmune signaling: a key component of alcohol abuse. Current opinion in neurobiology. 2013;23(4):513-20.
36. Schumacher A, Petronis A. Epigenetics of complex diseases: from general theory to laboratory experiments. Current topics in microbiology and immunology. 2006;310:81-115.
37. Hoffman RS, Goldfrank LR. The impact of drug abuse and addiction on society. Emergency medicine clinics of North America. 1990;8(3):467-80.
38. Alonzo NC, Bayer BM. Opioids, immunology, and host defenses of intravenous drug abusers. Infectious disease clinics of North America. 2002;16(3):553-69.
39. Stefano GB, Fricchione G, Goumon Y, Esch T. Pain, immunity, opiate and opioid compounds and health. Medical science monitor : international medical journal of experimental and clinical research. 2005;11(5):MS47-53.
40. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature. 2010;465(7299):721-7.
41. Vercauteren FG, Bergeron JJ, Vandesande F, Arckens L, Quirion R. Proteomic approaches in brain research and neuropharmacology. European journal of pharmacology. 2004;500(1-3):385-98.
42. Ho MK, Goldman D, Heinz A, Kaprio J, Kreek MJ, Li MD, et al. Breaking barriers in the genomics and pharmacogenetics of drug addiction. Clin Pharmacol Ther. 2010;88(6):779-91.

43. Hamosh A, Scott AF, Amberger J, Bocchini C, Valle D, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic acids research. 2002;30(1):52-5.
44. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic acids research. 2005;33(Database issue):D514-7.
45. McKusick VA. Mendelian Inheritance in Man and its online version, OMIM. American journal of human genetics. 2007;80(4):588-604.
46. Kanehisa M. The KEGG database. Novartis Foundation symposium. 2002;247:91-101; discussion -3, 19-28, 244-52.
47. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. Nucleic acids research. 2008;36(Database issue):D480-4.
48. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic acids research. 2010;38(Database issue):D355-60.
49. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. Nucleic acids research. 2002;30(1):42-6.
50. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic acids research. 2004;32(Database issue):D277-80.
51. Lynskey MT, Strang J. The global burden of drug use and mental disorders. Lancet. 2013;382(9904):1540-2.
52. Peele S. Redefining addiction. I. Making addiction a scientifically and socially useful concept. International journal of health services : planning, administration, evaluation. 1977;7(1):103-24.
53. Girard DE, Carlton BE. Alcoholism. Earlier diagnosis and definition of the problem. The Western journal of medicine. 1978;129(1):1-7.
54. Goodman A. Addiction: definition and implications. British journal of addiction. 1990;85(11):1403-8.
55. Goodwin DW. Genetic determinants of alcohol addiction. Advances in experimental medicine and biology. 1975;56:339-55.
56. Truan F. Addiction as a social construction: a postempirical view. The Journal of psychology. 1993;127(5):489-99.
57. Naranjo CA, Bremner KE. Behavioural correlates of alcohol intoxication. Addiction. 1993;88(1):25-35.
58. Dean JC, Rud F. The drug addict and the stigma of addiction. The International journal of the addictions. 1984;19(8):859-69.
59. Cunningham JA, Sobell LC, Chow VM. What's in a label? The effects of substance types and labels on treatment considerations and stigma. Journal of studies on alcohol. 1993;54(6):693-9.
60. Room R. Stigma, social inequality and alcohol and drug use. Drug and alcohol review. 2005;24(2):143-55.
61. Lorman WJ. Maintaining sobriety and recovery. The Nursing clinics of North America. 2013;48(3):437-44, vi.
62. Popendikyte V, Laurinavicius A, Paterson AD, Macciardi F, Kennedy JL, Petronis A. DNA methylation at the putative promoter region of the human dopamine D2 receptor gene. Neuroreport. 1999;10(6):1249-55.
63. Messer CJ, Eisch AJ, Carlezon WA, Jr., Whisler K, Shen L, Wolf DH, et al. Role for GDNF in biochemical and behavioral adaptations to drugs of abuse. Neuron. 2000;26(1):247-57.
64. Shi J, Hui L, Xu Y, Wang F, Huang W, Hu G. Sequence variations in the mu-opioid receptor gene (OPRM1) associated with human addiction to heroin. Human mutation. 2002;19(4):459-60.

65. Kreek MJ, Nielsen DA, LaForge KS. Genes associated with addiction: alcoholism, opiate, and cocaine addiction. Neuromolecular medicine. 2004;5(1):85-108.
66. Le Foll B, Gallo A, Le Strat Y, Lu L, Gorwood P. Genetics of dopamine receptors and drug addiction: a comprehensive review. Behavioural pharmacology. 2009;20(1):1-17.
67. Olfson E, Bierut LJ. Convergence of genome-wide association and candidate gene studies for alcoholism. Alcoholism, clinical and experimental research. 2012;36(12):2086-94.
68. Treutlein J, Rietschel M. Genome-wide association studies of alcohol dependence and substance use disorders. Current psychiatry reports. 2011;13(2):147-55.
69. Pietruszko R. Mammalian liver alcohol dehydrogenases. Advances in experimental medicine and biology. 1975;56:1-31.
70. Crabb DW, Bosron WF, Li TK. Ethanol metabolism. Pharmacology & therapeutics. 1987;34(1):59-73.
71. Gardner EL. Addiction and brain reward and antireward pathways. Advances in psychosomatic medicine. 2011;30:22-60.
72. Maldonado R. The neurobiology of addiction. Journal of neural transmission Supplementum. 2003(66):1-14.
73. Chao J, Nestler EJ. Molecular neurobiology of drug addiction. Annual review of medicine. 2004;55:113-32.
74. Commons KG. Neuronal pathways linking substance P to drug addiction and stress. Brain research. 2010;1314:175-82.
75. Sun J, Zhao Z. Functional features, biological pathways, and protein interaction networks of addiction-related genes. Chemistry & biodiversity. 2010;7(5):1153-62.
76. Hillemacher T. Biological mechanisms in alcohol dependence--new perspectives. Alcohol and alcoholism. 2011;46(3):224-30.
77. Chiang YC, Lo YN, Chen JC. Crosstalk between dopamine D(2) receptors and cannabinoid CB(1) receptors regulates CNR1 promoter activity via ERK1/2 signaling. Journal of neurochemistry. 2013;127(2):163-76.
78. Ehlers CL, Liang T, Gizer IR. ADH and ALDH polymorphisms and alcohol dependence in Mexican and Native Americans. The American journal of drug and alcohol abuse. 2012;38(5):389-94.
79. Cotton RW, Goldman D. Review of the molecular biology of the human alcohol dehydrogenase genes and gene products. Advances in alcohol & substance abuse. 1988;7(3-4):171-82.
80. Pasche B, Yi N. Candidate gene association studies: successes and failures. Current opinion in genetics & development. 2010;20(3):257-61.
81. Yuferov V, Levran O, Proudnikov D, Nielsen DA, Kreek MJ. Search for genetic markers and functional variants involved in the development of opiate and cocaine addiction and treatment. Annals of the New York Academy of Sciences. 2010;1187:184-207.
82. Uhl GR, Drgon T, Johnson C, Li CY, Contoreggi C, Hess J, et al. Molecular genetics of addiction and related heritable phenotypes: genome-wide association approaches identify "connectivity constellation" and drug target genes with pleiotropic effects. Annals of the New York Academy of Sciences. 2008;1141:318-81.
83. Zhang R, Miao Q, Wang C, Zhao R, Li W, Haile CN, et al. Genome-wide DNA methylation analysis in alcohol dependence. Addict Biol. 2013;18(2):392-403.
84. Ponomarev I. Epigenetic control of gene expression in the alcoholic brain. Alcohol research : current reviews. 2013;35(1):69-76.
85. Saito M, Szakall I, Toth R, Kovacs KM, Oros M, Prasad VV, et al. Mouse striatal transcriptome analysis: effects of oral self-administration of alcohol. Alcohol. 2004;32(3):223-41.

86. Contet C. Gene Expression Under the Influence: Transcriptional Profiling of Ethanol in the Brain. Current psychopharmacology. 2012;1(4):301-14.
87. Mulligan MK, Ponomarev I, Hitzemann RJ, Belknap JK, Tabakoff B, Harris RA, et al. Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(16):6368-73.
88. Zhang GC, Vu K, Parelkar NK, Mao LM, Stanford IM, Fibuch EE, et al. Acute administration of cocaine reduces metabotropic glutamate receptor 8 protein expression in the rat striatum in vivo. Neuroscience letters. 2009;449(3):224-7.
89. Adkins DE, McClay JL, Vunck SA, Batman AM, Vann RE, Clark SL, et al. Behavioral metabolomics analysis identifies novel neurochemical signatures in methamphetamine sensitization. Genes, brain, and behavior. 2013;12(8):780-91.
90. Lee AM, Messing RO. Protein kinases and addiction. Annals of the New York Academy of Sciences. 2008;1141:22-57.
91. Freeman WM, Brebner K, Amara SG, Reed MS, Pohl J, Phillips AG. Distinct proteomic profiles of amphetamine self-administration transitional states. The pharmacogenomics journal. 2005;5(3):203-14.
92. Li KW, Jimenez CR, van der Schors RC, Hornshaw MP, Schoffelmeer AN, Smit AB. Intermittent administration of morphine alters protein expression in rat nucleus accumbens. Proteomics. 2006;6(6):2003-8.
93. Brown JN, Ortiz GM, Angel TE, Jacobs JM, Gritsenko M, Chan EY, et al. Morphine produces immunosuppressive effects in nonhuman primates at the proteomic and cellular levels. Molecular & cellular proteomics : MCP. 2012;11(9):605-18.
94. Klee EW, Schneider H, Clark KJ, Cousin MA, Ebbert JO, Hooten WM, et al. Zebrafish: a model for the study of addiction genetics. Human genetics. 2012;131(6):977-1008.
95. Roman S, Zepeda-Carrillo EA, Moreno-Luna LE, Panduro A. Alcoholism and liver disease in Mexico: genetic and environmental factors. World journal of gastroenterology : WJG. 2013;19(44):7972-82.
96. Levran O, Londono D, O'Hara K, Randesi M, Rotrosen J, Casadonte P, et al. Heroin addiction in African Americans: a hypothesis-driven association study. Genes, brain, and behavior. 2009;8(5):531-40.
97. Cheung YW. Ethnicity and alcohol/drug use revisited: a framework for future research. The International journal of the addictions. 1990;25(5A-6A):581-605.
98. Goecks J, Nekrutenko A, Taylor J, Galaxy T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome biology. 2010;11(8):R86.
99. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, et al. Galaxy: a web-based genome analysis tool for experimentalists. Current protocols in molecular biology / edited by Frederick M Ausubel [et al]. 2010;Chapter 19:Unit 19 0 1-21.
100. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy: a platform for interactive large-scale genome analysis. Genome research. 2005;15(10):1451-5.
101. Huang da W, Sherman BT, Zheng X, Yang J, Imamichi T, Stephens R, et al. Extracting biological meaning from large gene lists with DAVID. Curr Protoc Bioinformatics. 2009;Chapter 13:Unit 13 1.
102. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4(1):44-57.
103. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behavioural brain research. 2001;125(1-2):279-84.

104. International HapMap C. A haplotype map of the human genome. Nature. 2005;437(7063):1299-320.
105. International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449(7164):851-61.
106. International HapMap C, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467(7311):52-8.
107. Wall JD, Cox MP, Mendez FL, Woerner A, Severson T, Hammer MF. A novel DNA sequence database for analyzing human demographic history. Genome research. 2008;18(8):1354-61.
108. Fuerst PA, Chakraborty R, Nei M. Statistical studies on protein polymorphism in natural populations. I. Distribution of single locus heterozygosity. Genetics. 1977;86(2 Pt. 1):455-83.
109. Li MJ, Wang LY, Xia Z, Sham PC, Wang J. GWAS3D: Detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. Nucleic acids research. 2013;41(Web Server issue):W150-8.
110. Beck T, Hastings RK, Gollapudi S, Free RC, Brookes AJ. GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. European journal of human genetics : EJHG. 2013.
111. Bradley WE, Raelson JV, Dubois DY, Godin E, Fournier H, Prive C, et al. Hotspots of large rare deletions in the human genome. PloS one. 2010;5(2):e9401.
112. Vincent JB, Choufani S, Horike S, Stachowiak B, Li M, Dill FJ, et al. A translocation t(6;7)(p11-p12;q22) associated with autism and mental retardation: localization and identification of candidate genes at the breakpoints. Psychiatric genetics. 2008;18(3):101-9.
113. Beitner-Johnson D, Nestler EJ. Morphine and cocaine exert common chronic actions on tyrosine hydroxylase in dopaminergic brain reward regions. Journal of neurochemistry. 1991;57(1):344-7.
114. Schmidt HD, McGinty JF, West AE, Sadri-Vakili G. Epigenetics and psychostimulant addiction. Cold Spring Harbor perspectives in medicine. 2013;3(3):a012047.
115. Pickworth WB, Klein SA, George FR, Henningfield JE. Acetaminophen fails to inhibit ethanol-induced subjective effects in human volunteers. Pharmacology, biochemistry, and behavior. 1992;41(1):189-94.
116. Hsu KC, Chida S, Geraghty DE, Dupont B. The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. Immunological reviews. 2002;190:40-52.
117. Reimer RJ, Edwards RH. Organic anion transport is the primary function of the SLC17/type I phosphate transporter family. Pflugers Archiv : European journal of physiology. 2004;447(5):629-35.
118. Miller CA, Campbell SL, Sweatt JD. DNA methylation and histone acetylation work in concert to regulate memory formation and synaptic plasticity. Neurobiology of learning and memory. 2008;89(4):599-603.
119. Glahn A, Riera Knorrenschild R, Rhein M, Haschemi Nassab M, Groschl M, Heberlein A, et al. Alcohol-Induced Changes in Methylation Status of Individual CpG Sites, and Serum Levels of Vasopressin and Atrial Natriuretic Peptide in Alcohol-Dependent Patients during Detoxification Treatment. European addiction research. 2013;20(3):143-50.
120. Bali P, Im HI, Kenny PJ. Methylation, memory and addiction. Epigenetics : official journal of the DNA Methylation Society. 2011;6(6):671-4.
121. Bilinski P, Wojtyla A, Kapka-Skrzypczak L, Chwedorowicz R, Cyranka M, Studzinski T. Epigenetic regulation in drug addiction. Annals of agricultural and environmental medicine : AAEM. 2012;19(3):491-6.

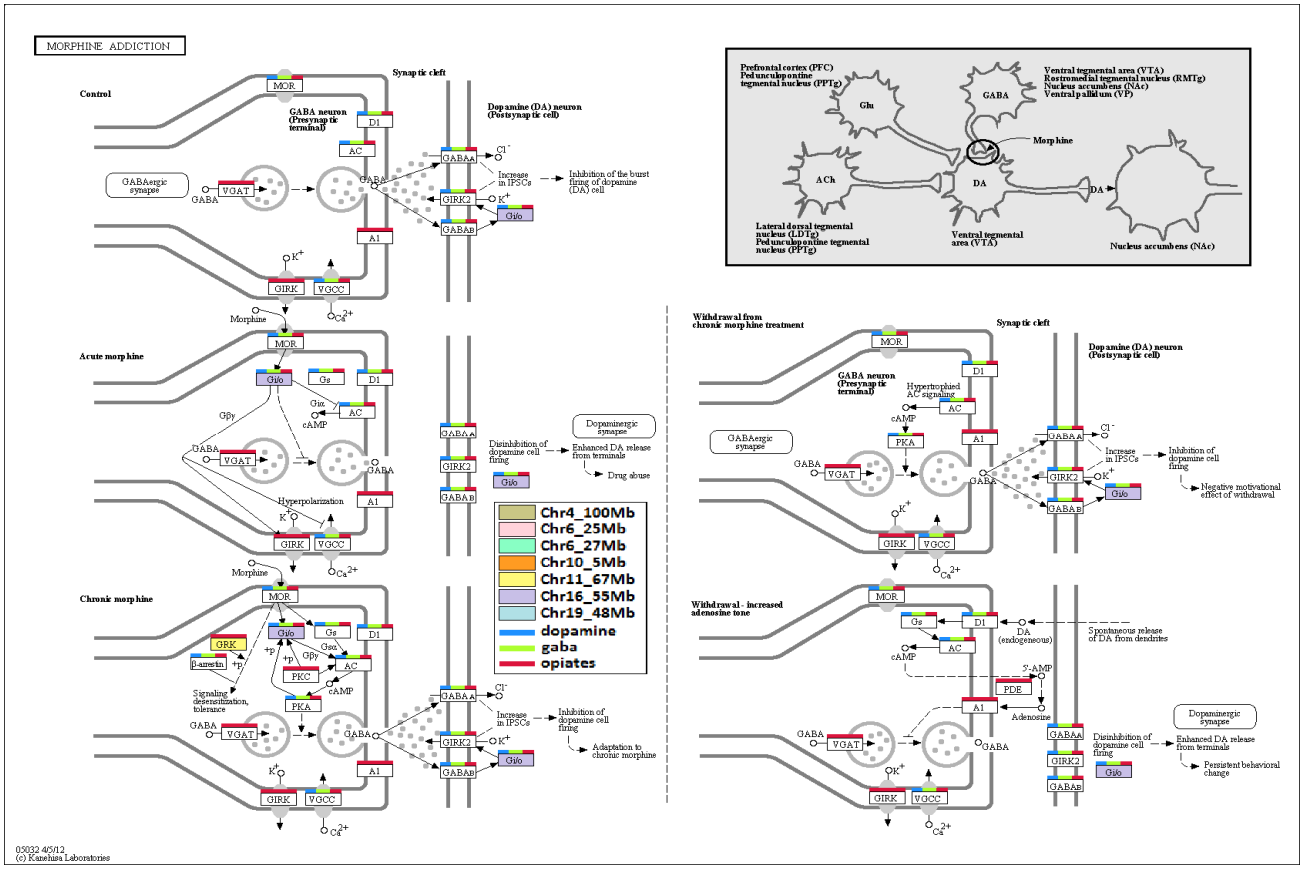
122. Schifano ED, Li L, Christiani DC, Lin X. Genome-wide association analysis for multiple continuous secondary phenotypes. American journal of human genetics. 2013;92(5):744-59.
123. Fraser HB, Lam LL, Neumann SM, Kobor MS. Population-specificity of human DNA methylation. Genome biology. 2012;13(2):R8.
124. Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LL, et al. Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. JAMA : the journal of the American Medical Association. 1990;264(19):2511-8.
125. Blum K, Oscar-Berman M, Badgaiyan RD, Palomo T, Gold MS. Hypothesizing dopaminergic genetic antecedents in schizophrenia and substance seeking behavior. Medical hypotheses. 2014;82(5):606-14.
126. McKetin R, Lubman DI, Najman JM, Dawe S, Butterworth P, Baker AL. Does methamphetamine use increase violent behaviour? Evidence from a prospective longitudinal study. Addiction. 2014;109(5):798-806.
127. Jegede RO. Depressive symptomatology in drug addicts: therapeutic and aetiological implications. African journal of medicine and medical sciences. 1978;7(3):171-4.
128. Woody GE, O'Brien CP, Rickels K. Depression and anxiety in heroin addicts: a placebo-controlled study of doxepin in combination with methadone. The American journal of psychiatry. 1975;132(4):447-50.
129. Altamura AC. Bipolar spectrum and drug addiction. Journal of affective disorders. 2007;99(1-3):285.
130. Geoffroy PA, Goddefroy G, Rolland B, Cottencin O. Efficacy of aripiprazole in comorbid addiction in bipolar disorder. CNS neuroscience & therapeutics. 2012;18(4):359-60.
131. Maremmani I, Perugi G, Pacini M, Akiskal HS. Toward a unitary perspective on the bipolar spectrum and substance abuse: opiate addiction as a paradigm. Journal of affective disorders. 2006;93(1-3):1-12.
132. Batel P. Addiction and schizophrenia. European psychiatry : the journal of the Association of European Psychiatrists. 2000;15(2):115-22.
133. Dubertret C, Bidard I, Ades J, Gorwood P. Lifetime positive symptoms in patients with schizophrenia and cannabis abuse are partially explained by co-morbid addiction. Schizophrenia research. 2006;86(1-3):284-90.
134. Falkai P, Moller HJ. The functional sequelae of schizophrenia: consequences of long-term pharmacotherapy and the neurobiology of addiction. European archives of psychiatry and clinical neuroscience. 2011;261(2):83-4.
135. Laviolette SR, Grace AA. The roles of cannabinoid and dopamine receptor systems in neural emotional learning circuits: implications for schizophrenia and addiction. Cellular and molecular life sciences : CMLS. 2006;63(14):1597-613.
136. Schmidt WJ, Beninger RJ. Behavioural sensitization in addiction, schizophrenia, Parkinson's disease and dyskinesia. Neurotoxicity research. 2006;10(2):161-6.
137. Seibyl JP, Brenner L, Krystal JH, Johnson R, Charney DS. Mazindol and cocaine addiction in schizophrenia. Biological psychiatry. 1992;31(11):1179-81.
138. Spanagel R, Bartsch D, Brors B, Dahmen N, Deussing J, Eils R, et al. An integrated genome research network for studying the genetics of alcohol addiction. Addiction biology. 2010;15(4):369-79.
139. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic acids research. 2011;39(Database issue):D1035-41.
140. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic acids research. 2014;42(Database issue):D1091-7.

141. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic acids research. 2008;36(Database issue):D901-6.
142. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic acids research. 2006;34(Database issue):D668-72.
143. Brody GH, Yu T, Mackillop J, Miller GE, Chen E, Obasi EM, et al. Catecholamine levels and delay discounting forecast drug use among African American youths. Addiction. 2014.
144. Zhai B, Sun XY. Mechanisms of resistance to sorafenib and the corresponding strategies in hepatocellular carcinoma. World journal of hepatology. 2013;5(7):345-52.
145. McElligott ZA, Fox ME, Walsh PL, Urban DJ, Ferrel MS, Roth BL, et al. Noradrenergic synaptic function in the bed nucleus of the stria terminalis varies in animal models of anxiety and addiction. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology. 2013;38(9):1665-73.
146. Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, Aros MC, et al. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science. 2005;310(5755):1782-6.
147. Wood ET, Stover DA, Slatkin M, Nachman MW, Hammer MF. The beta -globin recombinational hotspot reduces the effects of strong selection around HbC, a recently arisen mutation providing resistance to malaria. American journal of human genetics. 2005;77(4):637-42.
148. Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, et al. Adaptations to climate in candidate genes for common metabolic disorders. PLoS genetics. 2008;4(2):e32.
149. Bailey RG. Identifying ecoregion boundaries. Environmental management. 2004;34 Suppl 1:S14-26.
150. Babor F, Grund S, Siepermann M, Oommen PT, Kuhlen M, Schuster FR, et al. Epidemiology and clinical characteristics of pandemic (H1N1) 2009 influenza infection in pediatric hemato-oncology and hematopoietic stem cell transplantation patients. Transplant infectious disease : an official journal of the Transplantation Society. 2012;14(6):589-94.
151. Fimia GM, Stoykova A, Romagnoli A, Giunta L, Di Bartolomeo S, Nardacci R, et al. Ambra1 regulates autophagy and development of the nervous system. Nature. 2007;447(7148):1121-5.
152. Behrends C, Sowa ME, Gygi SP, Harper JW. Network organization of the human autophagy system. Nature. 2010;466(7302):68-76.
153. Rietschel M, Treutlein J. The genetics of alcohol dependence. Annals of the New York Academy of Sciences. 2013;1282:39-70.
154. Heinrich A, Nees F, Lourdusamy A, Tzschoppe J, Meier S, Vollstadt-Klein S, et al. From gene to brain to behavior: schizophrenia-associated variation in AMBRA1 alters impulsivity-related traits. The European journal of neuroscience. 2013;38(6):2941-5.
155. Wang J, Huo K, Ma L, Tang L, Li D, Huang X, et al. Toward an understanding of the protein interaction network of the human liver. Molecular systems biology. 2011;7:536.
156. Hu Z, Zhang Z, Kim JW, Huang Y, Liang TJ. Altered proteolysis and global gene expression in hepatitis B virus X transgenic mouse liver. Journal of virology. 2006;80(3):1405-13.
157. Bergametti F, Sitterlin D, Transy C. Turnover of hepatitis B virus X protein is regulated by damaged DNA-binding complex. Journal of virology. 2002;76(13):6495-501.
158. Allanore Y, Saad M, Dieude P, Avouac J, Distler JH, Amouyel P, et al. Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. PLoS genetics. 2011;7(7):e1002091.
159. Lavanchy D. Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. Journal of viral hepatitis. 2004;11(2):97-107.

160. Huang da W, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome biology. 2007;8(9):R183.
161. Huang da W, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic acids research. 2007;35(Web Server issue):W169-75.
162. Graczyk TK, Lucy FE, Tamang L, Mashinski Y, Broaders MA, Connolly M, et al. Propagation of human enteropathogens in constructed horizontal wetlands used for tertiary wastewater treatment. Applied and environmental microbiology. 2009;75(13):4531-8.
163. Mills JN, Gage KL, Khan AS. Potential influence of climate change on vector-borne and zoonotic diseases: a review and proposed research plan. Environmental health perspectives. 2010;118(11):1507-14.
164. Winchester CL, Pratt JA, Morris BJ. Risk genes for schizophrenia: Translational opportunities for drug discovery. Pharmacology & therapeutics. 2014;143(1):34-50.
165. Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, et al. Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. Archives of general psychiatry. 1999;56(2):162-8.

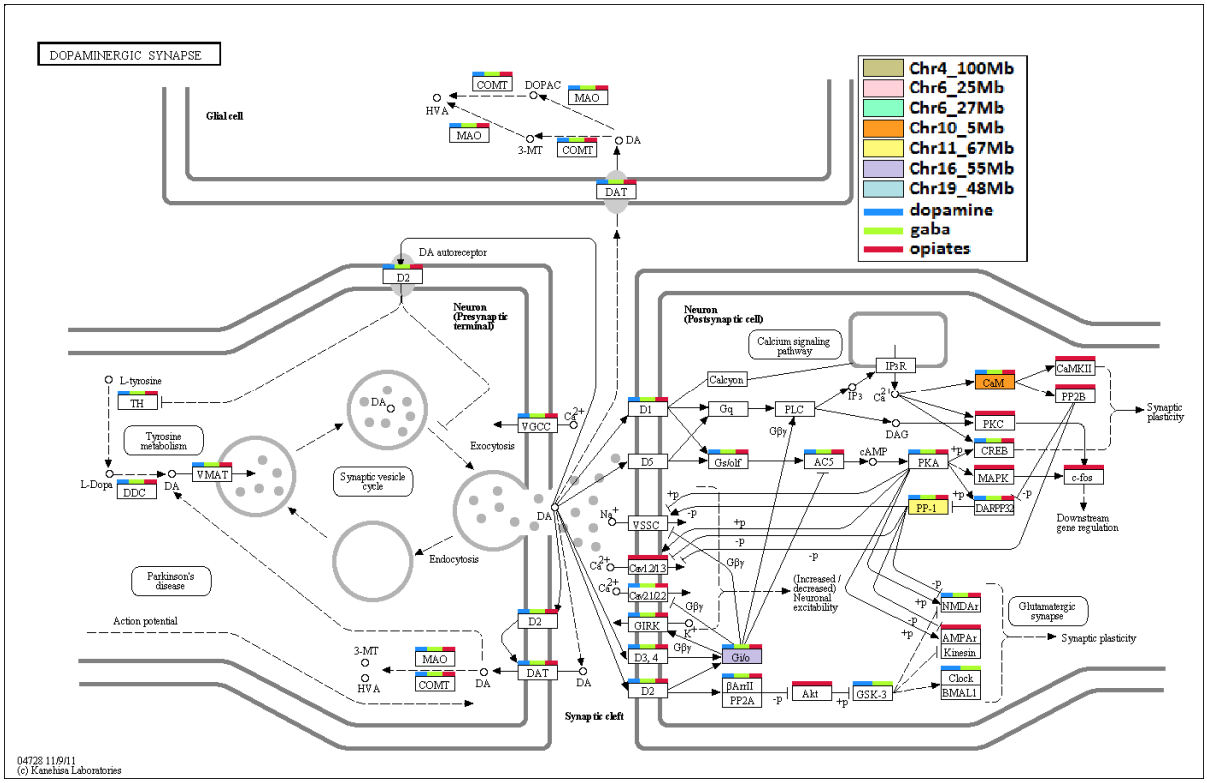
Appendices

1. KEGG pathways

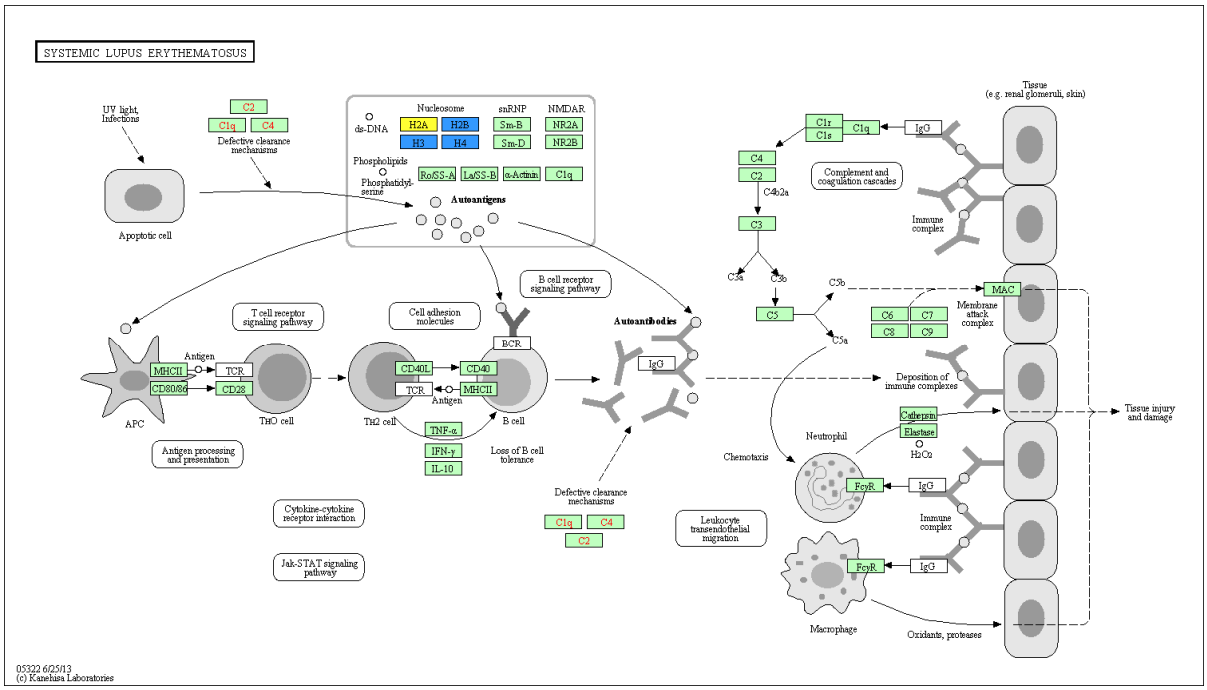


Appendix A: KEGG Pathways of Morphine Addiction in humans.

Appendix B: KEGG Pathway of Dopaminergic Synapse in Humans.



Appendix C: KEGG Pathway of Systematic Lupus Erythematosus in humans.



Appendix D: Final SNP Table- All SNPs generated for the all allelic polymorphisms found in the genome.

dbSNP ID	Gene	Position
rs1800708	HFE	intron
rs9366637	HFE	Intron
rs2237231	HIST1H1A	downstr. 500B
rs9393682	HIST1H1C	upstr. 2KB
rs2051542	HIST1H1T	missense
rs3830054	HIST1H2AB	upstr. 2KB
rs6908263	HIST1H2AC	intron
rs7760713	HIST1H2AC	intron
rs9467684	HIST1H2BD	intron
rs4145878	HIST1H2BF	upstr.2KB intron
rs1892252	SLC17A4	intron
rs3734525	SLC17A4	utr 3’
rs3823151	SLC17A4	intron
rs199738	TRIM38	utr 5’
rs6906576	Intergenic	genomic
rs6924948	Intergenic	genomic
rs7740793	Intergenic	genomic
rs9348699	Intergenic	genomic
rs933199	intergenic	Genomic
rs11819869	AMBRA1	intron
rs17790342	C11orf49	intron
rs12417519	C11orf49	intron
rs752849	C11orf49	intron
rs901746	DDB2	intron

Appendix E: GWAS3D- Identification of the European long distance interactions

Curriculum Vitae

LATIFA F. JACKSON
3514 Spring Garden Street, A34
Philadelphia, PA 19104
United States of America
Cell: 520.820.7405
E-mail: lfj27@drexel.edu

EDUCATION

Drexel University *Philadelphia, PA*
June 2014 (expected)
Ph.D. in Biomedical Science
Addiction, Mental Health and Infectious Disease: A Complex Web of Genetic Interactions
Advisor: Dr. Aydin Tozeren
KEYWORDS: Bioinformatics, Evolutionary Population Genetics, Systems Biology, Evolutionary Biology, Functional Genomics, Biomedical Science, Human Genetics, Complex Disease

University of Arizona *Tucson AZ*
Dec 2011
M.S. in Ecology and Evolutionary Biology

Høje Taastrup Language School *Høje Taastrup, Denmark*
June 2004
Advanced Danish Language Certification

University of Maryland *College Park, MD*
May 1998
B.S. degree Cell/Molecular Biology and Genetics
B.A. degree French Language and Literature

Middle East Institute *Washington, DC*
July 2002
Intensive Arabic I Course

EMPLOYMENT HISTORY

Graduate Research Associate 9/12-present
Drexel University, Philadelphia PA

- Conducted independent bioinformatics research on addiction, mental health and immunity genetics in ethnic human populations.
- Presented research findings at international, national and university wide research conferences.
- Prepared, lectured and assisted instructor for discussion sections of ~10 students in Genome Information Engineering and Systems Biology.

- Wrote presentations, manuscripts to share research findings to science and non-science audiences.

Online Science Content Editor 7/10-8/11
Elsevier Press, Cambridge MA
- Managed and *edited* web-based translational research articles Cell's online Daily News Aggregator.
- Collaborated with Elsevier Staff to manage scientific conference announcements content.

Graduate Research Associate
8/08-5/10
University of Arizona, Department of Ecology and Evolutionary Biology, Tucson AZ
- Conducted research human evolution using computational, functional and genomics based approaches.
- Prepared, taught, and led multiple course discussion sections of ~30 students in Evolution, Genetics, and Darwinian Medicine.
- Presented independent research at professional conferences and university research forums.
- Recruited and mentored undergraduate students.

International HIV/AIDS Technical Consultant 1/06-9/06
National Alliance of State and Territorial AIDS Directors, Washington DC
- Developed HIV/AIDS Prevention training materials for US-CDC/UNAIDS collaboration with the Cambodian government.
- Collaborated with NASTAD-Cambodia to develop training-of-trainer materials.

Global AIDS International Technical Assistance Coordinator 12/98- 11/02
National Alliance of State and Territorial AIDS Directors, Washington DC
- Worked with international donor agencies and foreign governments to build the technical capacity of foreign governments for HIV prevention and care program management.
- Developed and recruited cohort of HIV prevention and care technical experts to provide peer-based technical assistance to state AIDS programs and foreign governments.
- Managed all aspects of technical assistance logistics including the negotiation of scopes of work.
- Coordinated NASTAD's technical assistance activities with other federally funded technical assistance providers, CDC and USAID.
- Wrote policy issue briefs and white papers to educate congressional delegations about state AIDS policy issues.
- Planned National AIDS Conference and the HIV/AIDS Community Planning Conference as a steering committee member.
- Developed and led workshops on community planning, grant writing, project management and HIV prevention to community groups, state health departments, NGOs and foreign governments.

Molecular Genetics Research Assistant 9/98-3/99
Howard University Cancer Research Center, Washington DC

- Sequenced human genomes of Liberian origin via advanced molecular biology techniques to determine variants of their mitochondrial DNA variable D-loop region.
- Participated in research scholars' journal group.

Faculty Biology Research Assistant 8/97-8/98
University of Maryland Department of Biology, College Park MD

- Managed research environment in an academic research setting.
- Conducted independent research sub-project to further elucidate the role of *semushi* gene in embryonic development via molecular biology techniques.
- Performed literature based research to formulate overall research methods strategy.
- Supervised undergraduate honors program students.
- Presented research findings with academic research settings.
- Managed laboratory supplies and equipment, ensuring laboratory safety and radioactive chemical certification.

Undergraduate Teaching Assistant 1/97-5/97
University of Maryland Departments of Biology and Anthropology, College Park MD

- Taught and evaluated upper level Human Evolution laboratory section to approximately 25 students.
- Collaboratively developed a laboratory manual comprising new experimental procedures for testing human evolutionary theory.

Conference Planner 1/97-7/97
University of Maryland Center for the Study of Troubling Behavior College Park MD

- Responsible for all administrative activities for a statewide conference on special education needs for children with emotional disabilities and adjudicated youth.
- Created logo and publicity materials, organized mailings and created a database of presenters and attendees.

Campus Events Planner 9/93-5/97
University of Maryland Student Entertainment Enterprises, College Park MD

- Organized major concerts, lectures, and cultural events for the University of Maryland Student community. Managed a $250,000 + budget annually.
- Oversaw the grant making for external cultural events.
- Directly supervised twelve program and logistical support directors.

Resident Assistant 8/93-5/96
University of Maryland Department of Resident Life, College Park MD

- Mentored, counseled, administered, and supervised approximately 70 students per term.

- Created, planned and promoted programs involving faculty, health educators and university administrators.

Laboratory Assistant 10/92-6/93
University of Maryland College of Nutrition and Food Science, College Park MD
- Worked with Faculty researchers to maintain the laboratory infrastructure.
- Responsible maintenance of rat colony and primary data collector for research study on the role of corticosteroids on hunger management.

PUBLICATIONS (*Borgelin was my married name)
1. **L. Jackson,** C. Tanes, and A. Tozeren. Whole Genome Analysis of Addiction Pathways and their Population Subtypes (submitted to Addiction, ***Impact Factor***: 4.746)
2. **L. Jackson**, A. Tozeren. Addiction and Mental Health Genes form Genomic Hotspots with Strong Drugable targets. (in prep)
3. **L. Jackson**, M. Shestov, and A. Tozeren. Genes and Geography: Addiction and Immune Pathway Analysis in a Global Human Sample (in prep)
4. **L. Jackson,** J. Saini, C. Tanes, M. Shestov and U. Hershberg. Multilevel Selection Reveals Coding Bias in B cell populations at evolutionary timescales. (in prep)
5. F.L.C. Jackson, and ***L.F. J. Borgelin****. Chapter: How Genetics Can Provide Detail to the Transatlantic African Diaspora in The African Diaspora and the Discipline. Indiana University Press, Bloomington, IN. 2010
6. S.O.Y. Keita, F.L.C. Jackson, **L. Borgelin***, and K.N. Maglo. Letter to the Editor: Commentary on the Fulani—History, Genetics, and Linguistics, an Adjunct to Hassan et al., 2008 *American Journal of Physical Anthropology* Published Online: 20 Jan 2010
7. F.L.C. Jackson, K.M. Jackson, **L. Jackson**, S. Khan, L. Heywood, M. Raslan, X. Johnson, and R. A. Kittles. Strategies for overcoming current limitations on comparative genetic studies of African Atlantic Diaspora. *American Journal of Physical Anthropology* 590): 187-188 2000
8. **L. Jackson** and J.G. Rendon. Bright Ideas II: Innovative or Promising Practices in HIV Prevention and HIV Prevention Community Planning. CDC, AED, and NASTAD Publication. March 2001
9. **L. Jackson** and J.G. Rendon (15%). Bright Ideas: Innovative or Promising Practices in HIV Prevention and HIV Prevention Community Planning. CDC, AED, and NASTAD Publication. March 2000
10. **L. Jackson**. Technical Assistance and Capacity Building Provided by Health Departments to Community Based Organizations. NASTAD Issue Brief. March 2000
11. **L. Jackson**, L. Greabell, and J. Marin. HIV Prevention Community Planning- Co-chair's Perspectives. NASTAD Issue Brief. May 1999

PRESENTATIONS

1. "Addiction, Mental Health and Infectious Disease: A Complex Web of Genetic Interactions" Drexel University Office of Graduate Studies Books and Bagels Lecture Series May 14, 2014
2. "Addiction and Immunity Intersections-a genomic approach" Drexel University 7th Annual Student Conference on Global Challenges -Gender 2014
3. "Genetic Implication of Infectious Diseases in African Populations" ASWAD Barbados 2007
4. "Lessons Learned in International Technical Assistance" National HIV Prevention Conference 2001
5. "Technical Assistance and Capacity Building between US Health Departments and Community Based Organizations" HIV/AIDS Prevention Community Planning Leadership Conference 2000
6. "Youth Participation in HIV/AIDS Prevention and Care" United States Conference on AIDS 1999

Posters

1. **Jackson, L.**, Liu, Y., and Tozeren, A. "Whole Genome Analysis of Addiction and Their Population Subtypes" Am Soc. Human Genetics International Meetings 2013
2. **Borgelin, L.*** and Hammer, M.F. "What is structuring STRUCTURE? Understanding the underlying implications to a population structure algorithm" Society for Molecular Biology and Evolution International Conference 2010
3. **Borgelin, L.*****, Keck, M, and Morales, D.A.M. "Using Entropy to Identify Functional Motifs on the Y-Chromosome in Humans" National Science Foundation IGERT Program Grantees Meeting 2006
4. **Jackson, L.,** Rastogi, R. and Sakolsky, N. "Technical Assistance and Capacity Building between US Health Departments and Community Based Organizations" HIV/AIDS Prevention Community Planning Leadership Conference 2001

Academic Service

Bioinformatics Mentor, Master Charter High School Internship Program Spring2014

Mentor, Sunnyside High School Students College Application Program 2007-2008

Executive Secretary, University of Arizona Black Graduate Student Association 2007-2008

Panel Member, University of Arizona African American Recruitment Weekend 2007

Contributor, Book: Min Første Dansker Fall 2004

Steering Committee Member, United States Conference on AIDS 1999-2001

Professional Associations

- Member, American Society for Human Genetics

- Member, Society for Molecular Biology and Evolution
- Member, Society for Mathematical Biology
- Past Member, Association for the Study of the Worldwide African Diaspora
- Past Member, Genetics Society of America
- Past Member, National Science Teachers Association

Awards, Fellowships and Grants

- Rice University XSEDE High-throughput Computer Science Fellow
  2012-14
- FASEB MARC Travel Award to ASHG 2013 2013
- Drexel University Dept. of Education GAANN Fellow
  2012-14
- University of Arizona Outstanding Student Service Nominee
  2007-8
- NSF G-K12 Biology from Molecules to Ecology Fellow
  2007-8
- NSF Interdisciplinary Graduate Education and Research Traineeship Grant
  2005-7
- University of Copenhagen Scholarship Award
  2004-5
- Office of Multicultural Student Education Academic Excellence Award
  1993
- University of Maryland-Africa in the Americas Seminar Grant Recipient 1994

Laboratory Competencies

Molecular Biology:

DNA/RNA: PCR, DNA/RNA purification, cloning, Southern/Western/Northern blotting, genomic/cDNA library construction and screening, mutagenesis, DNA sequencing, gene and promoter analysis and many other common approaches for detection and manipulation of nucleic acids.

Bioinformatics:

- Gene identification/ annotation tools, gene prediction, motif prediction and analysis
- Competancies in software packages including: Strider, Oligo, Primer3, Galaxy, STRUCTURE, ADMIXture, DNAsp, Arlequin, Netlogo, MrBayes.

Computer Skills

- Microsoft Office suite (such as Microsoft Word, Access, Excel, and Outlook)
- Macintosh Apple word processing and graphics suite
- Presentation Software: PowerPoint and Claris Impact
- Desktop Publishing Software: Adobe PageMaker, Claris Works, Ofoto, Adobe Acrobat Writer, and Adobe Photoshop
- Internet Software: Netscape and Internet Explorer

Extended International Travel

Tanzania, Liberia, Lithuania, Poland, Hungary, Spain, Norway, Sweden, Denmark, Guyana, Bahamas, Dominican Republic, Egypt, England, Nigeria, Burkina Faso, Ivory Coast, Italy, Czech Republic, Turkey, Barbados, Mexico and France.

Languages

Professionally Fluent in Danish Professionally Proficient in French Professionally Fluent in American English Conversational Rudimentary Arabic

Programming language competancies: Matlab, PERL, and some familiarity with Python/R