MQP-BC-DSA-7218

THE ETHANOL PREFERENCE PHENOTYPE

A Major Qualifying Project Report

Submitted to the Faculty of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

in

Biochemistry

by

Caitlin Lally

April 24, 2008

APPROVED:

George Acquaah-Mensah, Ph.D.
Dept of Pharmaceutical Sciences
Massachusetts College of Pharmacy
Major Advisor

David S. Adams, Ph.D.
Biology and Biotechnology
WPI Project Advisor

ABSTRACT

Alcoholism is a complex disease with a number of components. Understanding ethanol preference, the tendency to choose ethanol over another drink, is the focus of this report. Genes whose activities constitute a basis for the ethanol preference phenotype have been previously reported. With the use of computational tools, high throughput microarray data was analyzed to identify transcription regulating relationships between subsets of specific genes of interest, including Carm1, Ube2m, Creb1, Crebbp, Stat3, Nfkbib and Atf2, among others. These findings confirm the importance of previously identified genes, and identify complex inter-connections between the regulations of many different pathways in the expression of the ethanol preference phenotype.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First, I would like to thank George Acquaah-Mensah and the rest of the Massachusetts College of Pharmacy and Health Sciences for welcoming me into their school. Without the continued guidance of Dr. Acquaah-Mensah, this project would not be possible. I would also like to thank Carolyn Friel of MCPHS for allowing me to use the computer in her lab throughout the entire year. And finally, I would like to thank Dave Adams, my WPI advisor, for helping initiate the project, keeping me on track, and assisting me throughout the lengthy editing process.

BACKGROUND

Alcoholism is a complex disease with multiple components. Among current adult drinkers in the United States, more than one-half say they have a blood relative who is or was an alcoholic or problem drinker (GDCADA, 2006). The desire to drink excessive amounts of alcohol may not only be due to lack of self-control, because alcoholism has been linked to both hereditary and environmental influences, and it is estimated to have a high heritability rate of 50-60% (Enoch, 2003).

Ethanol Preference

Ethanol preference is a difficult trait to understand due to the genetic complexity of alcoholism and the environmental variability of the subjects involved. For these reasons, it is extremely difficult to study alcoholism in humans. Some rodents, such as certain strains of rats and mice, show a tendency to prefer alcohol over water when given the choice of either. Among the many different types of ethanol preference testing procedures, one common type is a two-bottle choice paradigm where the rodent is given the choice between two solutions (Figure-1). The mouse is given two sources of liquid, one of which contains an unsweetened concentration of ethanol, and one of which contains another drink, usually water or a sucrose solution. Mice bred with the ethanol preference phenotype significantly will choose the ethanol over the water (Green and Grahame, 2008).

Figure 1: Test for Ethanol Preference with Mouse (Singh, 2006)

A group of rodent genes has been identified as contributing to the ethanol preference phenotype, although there has been little discovered concerning how these genes affect the phenotype (Mulligan et al., 2006). A Quantitative Trait Locus, QTL, is a region of DNA that has been associated with a particular phenotype (Williams, 2004). A QTL has been identified on chromosome 9 of mice that is associated with the ethanol preference phenotype. This QTL contains a number of candidate genes for the study of the phenotype. These genes have been previously associated with the regulation of alcohol consumption and served as a starting point for the study of the ethanol preference phenotype (Mulligan et al., 2006).

Inbred Mouse Strains

Inbred mouse strains facilitate a better control of experiments because they provide an opportunity to study the effects of the environment on a specific genotype by repeatedly accessing a genetically fixed population (Bogue, 2004). Such animals play an

important role in the study of the ethanol preference phenotype due the disease's ability to be influenced by both the environment and genetics.

Gene Regulatory Networks

Gene regulatory networks are a collection of DNA segments which interact, not necessarily physically, with one another in the cell. These interactions control the rate at which genes are transcribed into RNA. The RNA then proceeds to making a specific protein through translation. As part of this process, a signal molecule contacts a receptor protein, which then triggers a cascade of events throughout the cell. An inactive transcription factor will then become active, and bind to a segment of DNA usually upstream from the gene controlled by it, which then results in the transcription of RNA. Once the RNA is translated into a protein, the process is complete (U.S. Department of Energy Office of Science, 2006). Figure 2 shows the simplest type of network in which DNA segments encoding a transcription factor and RNA polymerase affect the transcription of a target gene.

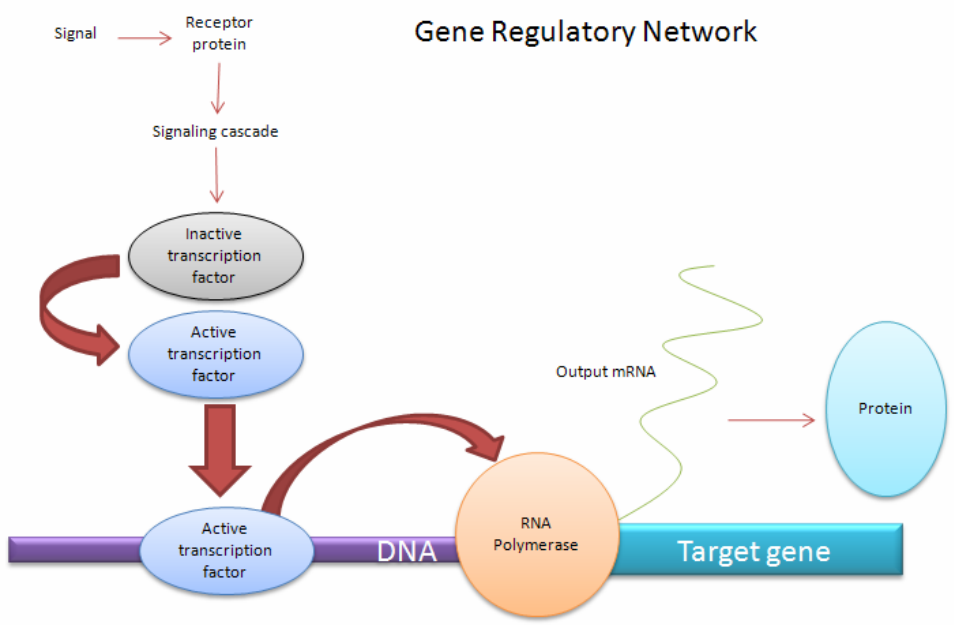

Figure 2: Diagram of a Simple Gene Regulatory Network. Based on one created by the U.S. Department of Energy Office of Science, 2006.

Regulatory networks can include three-gene relationships referred to as feed-forward loops (Figure-3). To summarize this process, consider three genes: X, Y, and Z. Gene X is a transcription factor for gene Y. Once gene Y has been activated, both genes X and Y together activate gene Z. This is commonly the case when gene Z requires both X and Y to activate it (Kalir et al., 2005).

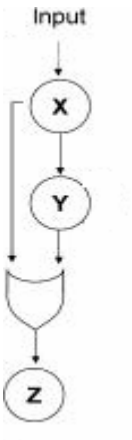

Figure 3: Diagram of a Gene Network Feed-Forward Loop (Kalir, 2005)

Obtaining an understanding of the regulatory networks involved in alcohol preference would greatly increase the possibility of new treatments for alcoholism, as the identified networks could potentially be disrupted. Although few networks have been discovered that affect alcoholism, many individual genes have been identified to be possibly related (Mulligan et al., 2006).

Microarray Analysis

Microarray analysis is a technique that has become common among researchers in the past 5-10 years. This technique allows for the study of the expression of thousands of genes simultaneously. The process involves using gene-specific probes that represent thousands of individual genes within the genome being studied. These synthetic DNA oligos, representing specific genes, are spotted by the commercial vendor onto membranes. Two membranes are then hybridized to fluorescently labeled cDNAs synthesized from two different RNA populations. For example, one experiment might involve comparing the RNAs expressed in control mouse brain versus RNAs expressed in the brains of mice that prefer alcohol. The levels of fluorescence at each spot location are assayed, a higher fluorescence (hybridization) implies a higher relative degree of expression of that specific gene represented by the spot. The information obtained from the microarray analyses differs depending on the statistical analysis applied to the data (Quackenbush, 2006).

There are two common types of microarray data, two-color arrays and single color arrays. In the two-color array, the test sample is labeled in one color and the

control sample is labeled in another, thus the two RNA (cDNA) samples can be compared on one membrane while measuring fluorescence at two different wavelengths. The expression of the test sample is then compared to that of the control sample. In the single color array, the expression is measured by fluorescence at one wavelength but on two different membranes, each hybridized to a different RNA (cDNA) population (Quackenbush, 2006). The arrays are commonly analyzed using either Euclidean distance or Pearson's correlation-coefficient distances. Euclidean distance is used when the magnitude of gene expression is important, and Pearson's distance is used when the pattern of expression is important (Quackenbush, 2006).

Computational Pharmacology

Among other things, computational pharmacology utilizes machine-based learning to analyze high throughput microarray data. Both Biochemical and Computer Science concepts have been combined to create algorithms that analyze data through a process known as reverse engineering, the process of taking a massive set of gene expression data and inferring the corresponding protein network. Two reverse engineering algorithms were used in this MQP report. Computer programs play an essential role in reverse engineering, they analyze and inform for efficient knowledge discovery.

R (www.r-project.org) is a computational analysis tool which utilizes a unique computer language. For this experiment, R was run using Linux, a UNIX-like platform. Within R, specific functions are manually defined to allow the user to have control over their data and algorithmic choices. Bioconductor (www.bioconductor.org) provides a number of different packages to extend the statistical techniques available within R,

including Affy, which analyzes information provided by Affymetrix (www.affymetrix.org).

The Algorithm for the Reconstruction of Gene Regulatory Networks (ARACNE) (Margolin et al., 2006) is an algorithm which groups together genes with similar transcriptional responses indicating their regulation of one another. This algorithm is able to determine if the expression of any set of two genes are statistically dependent upon one another. It can also distinguish between direct and indirect regulatory relationships, using Data Processing Inequality (DPI). Two genes with a dependence upon each other have mutual information that is statistically significantly greater than zero.

Mutual Information

Mutual Information (MI) is a measure of the statistical dependency of two genes that are co-regulated, directly bound to one another, that participate in the same pathway, or are co-expressed. Thus MI indicates a biological relationship between the genes. Similar to a correlation coefficient, MI ranges between zero and one, with zero indicating no relationship and one signifying the strongest relationship. In this report, the p-value at which an MI value was deemed statistically different from zero was set at 1e10-7 for the algorithm. The formula for MI between two random variables is defined as:

$$I(X;Y) = \sum_{i,j} P(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

where $P(x_i)$ is the probability that X = x_i. For genes, X and Y could represent a transcription factor and its potential target gene, and x_i and y_j represent particular expression levels (Faith et al., 2007).

Bayesian Networks

A Bayesian network is a probabilistic graph model that describes the probability distribution for a set of variables (Pearl, 1988). Bayesian Network Inference with Java Objects, Banjo, is a software application that aids in learning the structural function of both static and dynamic Bayesian networks. A Bayesian network is a probabilistic graphical model that represents a set of variables and their dependency upon one another. Banjo loops through four main steps looking for a network with the highest score (Figure-4). Once the initial data has been put into Banjo, the proposer suggests a new network to be considered, the cycle checker looks for cycles within the proposed network, the evaluator computers a network score for the proposed network based on a previously determined metric, and the decider determines whether or not to accept the proposed network (Sladeczek et al., 2007). This process continues for the amount of time chosen by the user, giving reports of the best network every ten minutes. Once the time limit has terminated, the software reports the final determined network with the direction of interaction.

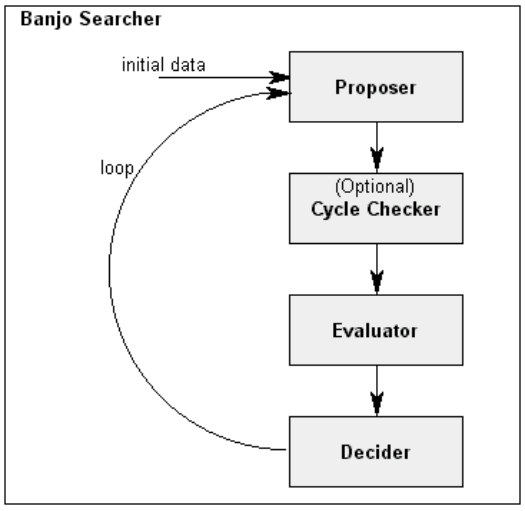

Figure 4: The Four Steps of Banjo (Sladeczek et al., 2007)

All of these computational tools work together to help gain a greater understanding of gene regulatory networks. By using these tools, one is able to study and analyze previously composed microarray data to look for an underlying gene regulatory network. The analysis tools expand the ability to learn about the ethanol preference phenotype without having to be in wet lab settings. This provides an opportunity to study the phenotype without involving more test subjects. One can continue to learn about the genes involved through mice that have already been sacrificed.

PROJECT PURPOSE

It has already been discovered that specific genes including Carm1, Ube2m, Creb1, Crebbp, Stat3, Nfkbib, and Atf2 contribute to the ethanol preference phenotype in mice (Mulligan et al., 2006). This MQP project aims to shed light on potential regulatory relationships among these genes using computational algorithms. By using different computational tools, this project expands the general knowledge about the ethanol preference phenotype, including increasing our knowledge of some of the genes involved, and the transcription factors that regulate their expression. This type of information can potentially be used in the future to disrupt the networks to help treat alcohol dependency.

METHODS

The process of inferring gene regulatory networks requires a long series of steps, utilizing many different tools. Computational pharmacology uses statistics, biology, chemistry, and computer science to solve biological problems at the molecular level. The process applied in this project to learn more about the regulatory networks involved in the ethanol preference phenotype is outlined below.

Obtaining the Data

Phenogen Informatics is a database that serves as a comprehensive toolbox for scientists to store their microarray data for public use. Each hybridized array is the result of a wet lab procedure that has been previously performed in a scientist's laboratory and is then placed online for others to analyze. A total of 569 ".CEL" files were obtained from 705 mouse whole brain microarrays on seven Affymetrix (oligonucleotide) platforms from the Phenogen Informatics database (http://phenogen.uchsc.edu). The 569 samples were sorted according to their platforms. The seven resulting platforms were: Murine Genome U74A, Mouse Expression Set 430A, Mouse Expression Set 430B, Mouse Genome 430 2.0 Array, Murine Genome U74Av2, Murine Genome U74Bv2, and Murine Genome U74Cv2.

RMA Analysis

After obtaining the samples, the Robust Microarray Average (RMA) analysis was performed on each of the seven platforms (Irizarry et al., 2003). With the use of program R and the Bioconductor package "Affy", background correction, normalization, and log-

transformation was performed on the data. Background correction eliminates noise that may have resulted from the scanning laser reflecting off the surface of the array and the normalization step aims to correct any differences between the genes or arrays within each platform (Tarca et al., 2006). The RMA analysis allowed for the compilation of all ".CEL" file information to one file per platform.

Selecting Genes of Interest

Each file contained expression information for over 22,000 genes, only some of which were important in the study of ethanol preference. Based on information obtained in an article written by Mulligan et al. (2006), 75 candidate genes were isolated from each platform using a parser-selector program written by Dr. Acquaah-Mensah. A number of stress sensitive transcription factors were appended to the list of candidate genes.

The parser-selector program searched through the 22,000+ genes located on each of the platforms and secluded the information that corresponded to the genes of interest and the transcription factors. A new file was created from the resulting information for further analysis, and the remaining gene information was discarded.

Algorithm for the Reconstruction of Accurate Cellular Networks

The Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE) is an "algorithm for the reverse-engineering of transcriptional networks from microarray data" that is able to separate direct regulatory interactions from those that are indirect (Margolin et al., 2006). This algorithm iterates through the gene list and, for each possible pair, determines the amount of MI between the two genes.

Once the Mutual Information has been determined, the algorithm evaluates the Data Processing Inequality (DPI). "The DPI states that if genes g1 and g3 interact only through a third gene, g2, then:

$$I(g1,g3) \leq \min [I(g1,g2);I(g2,g3)].\text{"}$$ (Margolin et al., 2006)

Simply stated, in each iteration, the algorithm evaluates three sets of MI values and removes all MI values that indicate an indirect relationship. Discarding indirect relationships (i.e. relationships once removed) eliminates false MI values, such that if gene X and gene Y have MI, gene Y and gene Z have MI, and gene X and gene Z have MI but only because of the relationship between gene Y and gene Z, the MI value for gene X and gene Z will be eliminated. The parameters were set to be conservative to ensure the result to be statistically significant. The p-value for the algorithm was set to 1e-7 (highly significant) and the DPI tolerance was set to 0.15. A DPI tolerance of 0.15 indicates that 15% of the MI estimations would be considered sampling errors.

The resulting information was put into an output file containing the pairs of genes and their Mutual Information. These files were converted into a Simple Interaction File (.sif) for use in Cytoscape. For each gene pair, this file type contains the probeset ID of gene A on one side in column 1, the probeset ID of gene B on the other side in column 3, and their calculated MI in column 2. A list of probeset to genecode conversions were formatted into a file. A parser selector was used to convert the .sif probeset files to .sif genecode files.

Some of the platforms used proved to have no gene pairs with MI statistically significantly different from zero. These platforms were Mouse Expression Set 430B, Murine Genome U74Bv2, and Murine Genome U74Cv2. The three platforms all

originally contained 6 ".CEL" files. Because they did not have genes with mutual information, the three platforms were eliminated from the remainder of the project.

Cytoscape

Cytoscape (www.cytoscape.org) is a bioinformatics platform that allows a user to create visual networks, and integrate these networks with gene expression. These networks can either consist of protein-protein interactions or protein-DNA interactions. This tool visually represents gene regulatory networks in the form of nodes and edges. There are a number of plugins that can be used in Cytoscape to allow the user to obtain different types of networks.

The ARACNE output files were imported into this software, resulting in four different visual networks. Of the four networks, Murine Genome U74Av2 had the smallest network with only nine nodes. The Mouse Expression Set 2.0 platform data yielded the largest inferred network. Due to the differences among the networks, they required different plugins for facilitated analysis. The networks were visually represented in a number of ways and each way was analyzed to search for new regulatory networks involved in the ethanol preference phenotype.

One tool that was used to represent the four different visual networks was the Collective Analysis Between Interactive Networks, CABIN (Singhal and Domico, 2007). CABIN allowed for all four networks to be analyzed together. Subset regulatory networks were then isolated using the genes of interest and their immediate neighbors. Based on these networks, relationships were determined between genes. Once networks had been created with the genes of interest, and the genes with which they were directly related, another algorithm was used to determine the direction of the regulation.

Bayesian Network Inference with Java Objects

A tool called the Bayesian Network Inference with Java Objects, Banjo, was used to determine the direction of regulation between pairs of genes. This was only performed with one gene pair at a time to build on the ARACNE findings while avoiding indirect regulatory relationships. The program was used to determine the direction of interactions between transcription factors and non-transcription factors. The program was run several times, with an average length of approximately one hour for each run. The resulting formation was incorporated into Cytoscape, replacing undirected edges with arrowed edges to indicate direction of regulation.

RESULTS

This project used a combination of different software applications to expand knowledge about ethanol preference phenotype. Once the files underwent the preparatory analysis, they were able to be imported into Cytoscape. This allowed for visual representation of the gene regulatory networks.

One gene that was a focus of the project was Coactivator-Associated Arginine Methyltransferase 1, Carm1. This gene is up-regulated in mice bred for the ethanol preference phenotype, and is located on the quantitative trait locus for ethanol preference on chromosome 9 (Mulligan et al., 2006). Carm1 did not have any direct statistical dependencies on any of the other transcription factors considered to be associated with the ethanol preference phenotype. Carm1 is up-regulated in ethanol-preferring mice. It has direct statistical dependencies with several key ethanol preference genes involved in signal transduction (Figure-5). Figure 5 depicts the resulting Cytoscape network for Carm1 with its regulators and targets. The genes in blue are regulators of Carm1, and include Ubiquitin-conjugating enzyme E2M (Ube2m) and Cathepsin B (Ctsb), among others. The genes in pink are regulated by Carm1, and include Matrix metallopeptidase 17 (Mmp17), Defender against cell death 1 (Dad1), among others. Edge directionality for the genes in grey could not be definitively determined using the Bayesian approach.

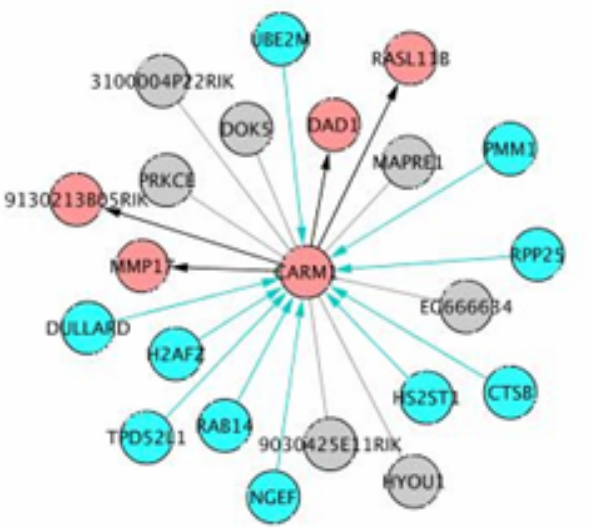

Figure 5: Cytoscape View of Gene Carm1 With its Regulators and Targets. Carm1 is shown in the center of the diagram. Genes shown in blue denote Carm1 regulators, while those shown in pink are regulated by Carm1. Genes shown in grey have an unknown direction of regulation of Carm1.

Carm1 and one of its related genes, Protein Kinase C-epsilon (Prkce) (shown as gray in Figure-5, upper left), have statistical dependencies with a number of the same genes. The relationship between Carm1 and Prkce was predicted to be representative of a feed-forward loop. Prkce is expected to be important to the ethanol preference phenotype because mice in which the Prkce gene was knocked out show a drop in consumption of ethanol drinking compared to mice with Prkce expression (Mulligan et al., 2006).

Figure 6, shows both Carm1 and Prkce with their targets. The horizontally lined group in the middle represents the common targets for both genes, and the groups on the sides are the ones they share statistical dependency with, independent of each other.

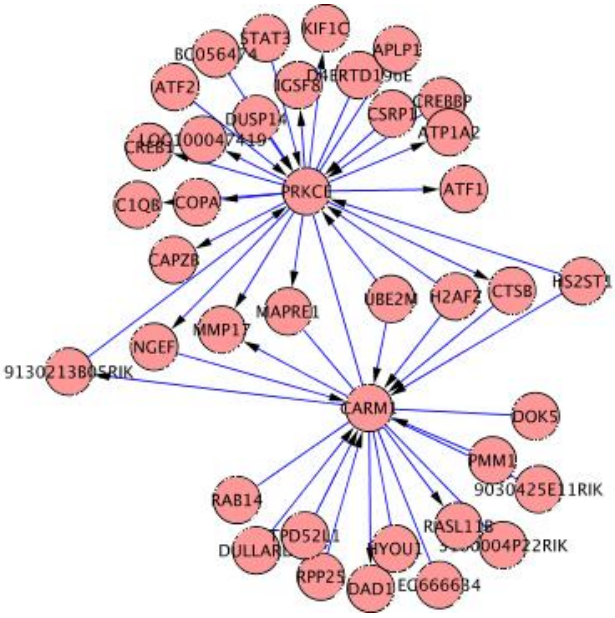

Figure 6: Cytoscape Representation of Genes Carm1 and Prkce With Their Regulators and Targets. Prkce is the center gene for the top cluster, while Carm1 is the center gene for the bottom cluster. The horizontally lined cluster in the middle represents the nodes that interact with both Carm1 and Prkce.

The nodes that interact with both Carm1 and Prkce appear to have a number of different roles. In some instances, Carm1 regulates a gene which then in turn regulates Prkce. One example is Carm1 → 9130213B05RIK → Prkce. The opposite is also true with Prkce → Ngef → Carm1. Three genes regulate both Carm1 and Prkce, H2afz, Ube2m, and Hs2st1. The original hypothesis of a feed-forward loop could be a possible explanation for the relationship between Carm1, Prkce, and Matrix metallopeptidase 17 (Mmp17). The relationships between Carm1, Prkce, and their common regulators and targets are shown in Table 1.

Table 1: Regulation by Carm1 and Prkce of Their Common Targets.

	Mmp17	Mapre1	H2afz	Ube2m	Ngef	Ctsb	Hs2st1	9130213B05RIK
Regulated by Prkce	X	X			X	X		
Regulated by Carm1	X							X

Another gene hypothesized to influence the ethanol preference phenotype is the cAMP responsive element binding protein 1, Creb1. Creb1 is studied for its effect on the brain in processes involving addiction, depression, and memory (Carlezon, et al., 2005). Creb1 is a transcription factor that has regulatory relationships with many of the alcohol preference candidate genes (Figure-7). One gene that has statistically significant Mutual Information with Creb1 is Zinc Finger Protein 143 (Zfp143). This is a transcription factor that has been identified with the ethanol preference phenotype because it is significantly up-regulated in models with high ethanol preference (Mulligan et al., 2006).

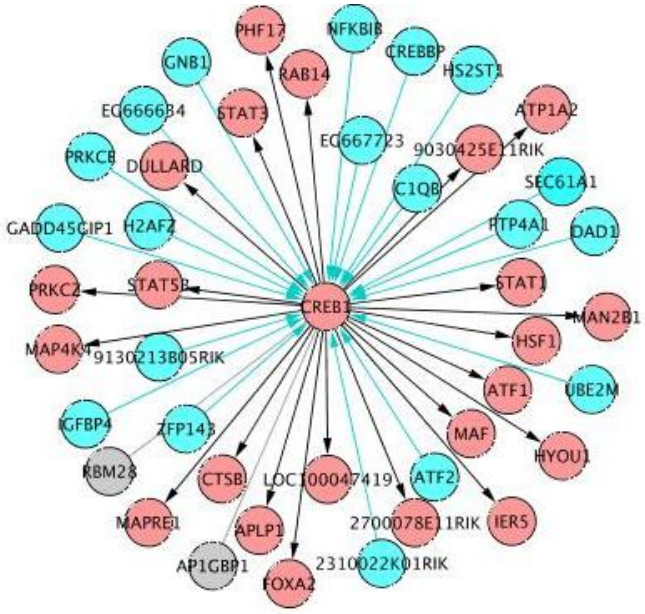

Figure 7: Cytoscape View of Creb1, Crebbp, and the Genes With Which They Share Mutual Information. Creb1 is shown in the center of the diagram. Genes shown in blue denote Creb1 regulators, while those shown in pink are regulated by Creb1. Genes shown in grey have an unknown direction of regulation of Creb1.

Each gene has a number of probeset identifiers, all of which were used in the Banjo run. In some cases, different probeset IDs of the same gene indicated conflicting directions of regulatory control. Table 2 shows the Banjo results for the comparison of

Creb1 and Zfp143. Because the algorithm bases its results on probabilities, the "majority" outcome was used for determining which gene is the regulator and which is being regulated. As an example, for these two genes in Table 2 Zfp143 is a regulating factor for Creb1.

Table 2: Conflicting Directions of Regulatory Control Obtained with Different Probe Sets of the Same Gene Pair

Gene Regulating	Zfp143	Creb1	Zfp143
Gene Being Regulated	Creb1	Zfp143	Creb1

Creb1 and Activating transcription factor 2 (Atf2) have already been shown to have a relationship on the mediating effects of AMP-activated protein kinase for the expression of genes that have a cAMP-response element in their promoters (Thomson et al., 2008). Creb1 and Atf2 had statistically significant Mutual Information and, upon the analysis of the Banjo results, it was determined that Atf2 regulates Creb1 in mouse brain (Figure-8). Atf2 also shared a statistical dependency with adipocyte specific protein 5 (9030425E11RIK) which has been previously identified as having a correlation between gene expression and the ethanol preference phenotype based on recombinant inbred strains (Mulligan et al., 2006). The algorithm results showed that Atf2 is a regulator of 9030425E11RIK.

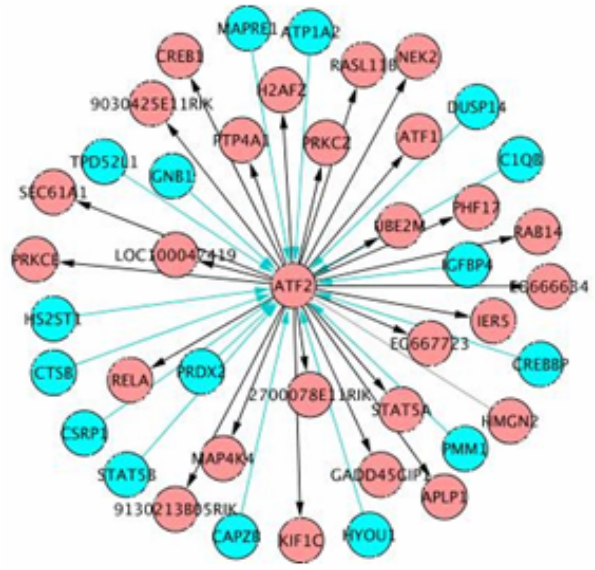

Figure 8: Cytoscape Representation of Atf2 and its Targets. Atf2 is shown in the center of the diagram. Genes shown in blue denote Atf2 regulators, while those shown in pink are regulated by Atf2.

Atf2 is regulated by protein Hypoxia up-regulated 1 (Hyou1), a protein associated with a response to stress. Atf2 has regulatory relationships with a number of the genes targeted by both Carm1 and Prkce. The direction of the regulation between the genes varies. Atf2 regulates 9130213B05RIK, Ube2m, and H2afz, while it is regulated by Mapre1, Ctsb, and Hs2st1 (Table-3). Table 3 shows a comparison of the Prkce, Carm1, and Atf2 with their regulation patterns. Aft2 regulates Ube2m and H2afz, two genes that regulate both Carm1 and Prkce.

Table 3: Regulatory Relationships Between Prkce, Carm1, Atf2, and Common Targets.

Prkce	Carm1	Atf2
9130213B05RIK → Prkce	Carm1 → 9130213B05RIK	Atf2 → 9130213B05RIK
Prkce → Ngef	Ngef → Carm1	---
Prkce → Mmp17	Carm1 → Mmp17	---
Prkce → Mapre1	inconclusive	Mapre1 → Atf2
Ube2m → Prkce	Ube2m → Carm1	Atf2 → Ube2m
H2afz → Prkce	H2afz → Carm1	Atf2 → H2afz
Prkce → Ctsb	Ctsb → Carm1	Ctsb → Atf2
Hs2st1 → Prkce	Hs2st1 → Carm1	Hs2st1 → Atf2

Other genes that have previously been identified as associated with the ethanol preference phenotype also showed dependencies on Creb1. This includes forkhead box A2, Foxa2, a transcription factor that is down-regulated in ethanol-preferring mice (Mulligan et al., 2006). This gene has dependencies on several of the key ethanol preference genes involved in signal transduction. Because it is down-regulated, its lack of expression could allow for the expression of the key ethanol preference genes. This gene also shows statistical dependencies on other transcription factors considered in this project, including signal transducer and activator of transcription 3, Stat3, and nuclear factor of kappa light chain gene enhancer in B-cells inhibitor-beta, Nfkbib. The Banjo results concluded that both Stat3 and Nfkbib are regulated by the Foxa2 gene (Figure-9).

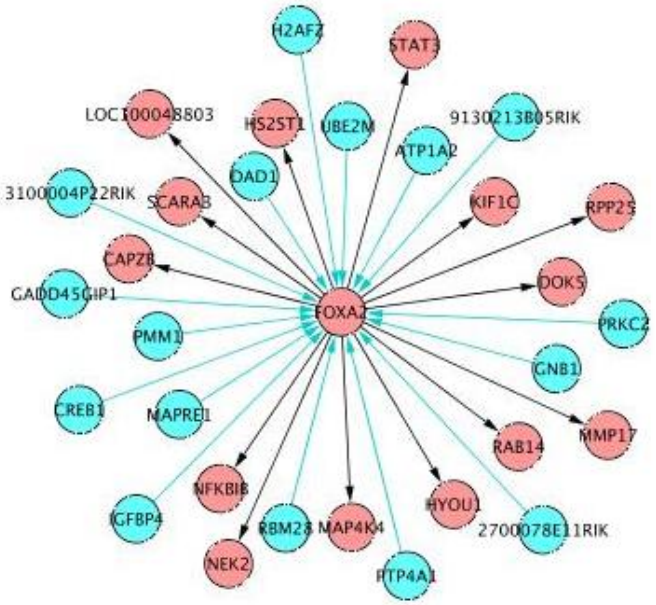

Figure 9: Cytoscape Representation of the Foxa2 Gene and Its Targets. Foxa2 is shown in the center of the diagram. Genes shown in blue denote Foxa2 regulators, while those shown in pink are regulated by Foxa2.

Foxa2 regulates both Hyou1, a gene previously determined to have a regulatory role over Atf2, and Heparan sulfate 2-O-sulfotransferase 1 (Hs2st1), a gene that regulates Carm1, Prkce, and Atf2 (Table-4). There are three instances where a gene regulated by Atf2 in turn regulates Foxa2: 9130213B05RIK, Ube2m, and H2afz. Both Atf2 and Foxa2 are regulated by Mapre1, which is regulated by Prkce.

Table 4: Regulatory Relationship Comparison Between Prkce, Carm1, Atf2, and Foxa2.

Prkce	Carm1	Atf2	Foxa2
9130213B05RIK → Prkce	Carm1 → 9130213B05RIK	Atf2 → 9130213B05RIK	9130213B05RIK → Foxa2
Prkce → Ngef	Ngef → Carm1	---	---
Prkce → Mmp17	Carm1 → Mmp17	---	Foxa2 → Mmp17
Prkce → Mapre1	inconclusive	Mapre1 → Atf2	Mapre1 → Foxa2
Ube2m → Prkce	Ube2m → Carm1	Atf2 → Ube2m	Ube2m → Foxa2
H2afz → Prkce	H2afz → Carm1	Atf2 → H2afz	H2afz → Foxa2
Prkce → Ctsb	Ctsb → Carm1	Ctsb → Atf2	---
Hs2st1 → Prkce	Hs2st1 → Carm1	Hs2st1 → Atf2	Foxa2 → H2st1

A visual representation was created to show the relationships between Carm1, Prkce, Atf2, and Foxa2, along with their shared nodes in Table 4. Figure 10 shows the complex interconnections between these four genes and only some of their common targets. Although this is not a complete gene regulatory network, it is a good starting point for further research. As can be seen in this figure, there is evidence of co-regulation among this set of genes. Prkce has five genes that influence its regulation, and then in turn regulates four genes. Ube2m is only regulated by one gene in this network, Atf2, and it regulates three genes, Carm1, Prkce, and Foxa2.

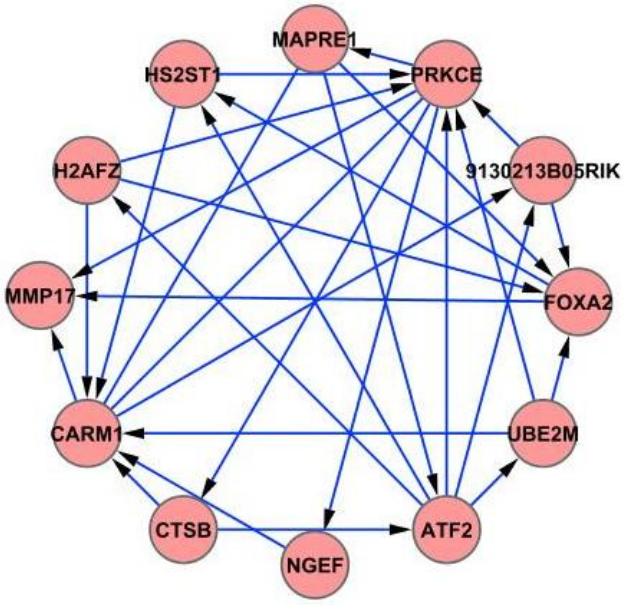

Figure 10: A Combination Diagram of Carm1, Prkce, Atf2, and Foxa2 With Some of Their Common Targets. The arrows indicate direction of regulation.

One gene that shares a number of statistical dependencies with various transcription factors is the ubiquitin-conjugating enzyme E2M, Ube2m. Ube2m is involved in post translational protein modification in the ubiquitin cycle. This gene shares statistical dependencies with most of the transcription factors considered for this project, and a number of the key ethanol preference genes as well. Figure 11 shows Ube2m with its target genes. The blue genes are the ones which have a regulatory role over Ube2m, which includes Atf2, Mapre1, and Stat3. The genes in pink, however, are the genes regulated or influenced by Ube2m, including Creb1, Carm1, Foxa2, Prkce, Nfkbib, and many others. This relationship was interesting in that a non-transcription factor was influencing a number of the transcription factors expected to be associated with the ethanol preference phenotype.

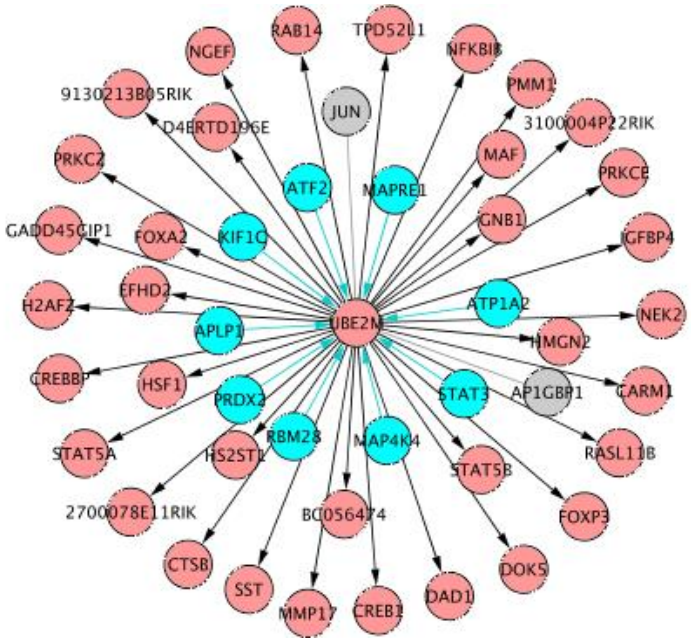

Figure 11: Cytoscape View of the Ube2m Gene and Its Targets. Ube2m is shown in the center of the diagram. Genes shown in blue denote Ube2m regulators, while those shown in pink are regulated by Ube2m. Genes shown in grey have an unknown direction of regulation of Ube2m.

The genes of focus for this MQP project have provided insight to the gene regulatory networks involved in the ethanol preference phenotype. Many of the centers of the gene regulatory networks for the ethanol preference phenotype have been determined. For example, Prkce has been proven to play a role in drinking behaviors of mice, and the results here show its relationship with many of the previously identified ethanol preference genes. These findings, among others, indicate complex inter-connections between the regulations of protein metabolism, stress-responsive, and signal transduction pathways in the expression of the ethanol preference phenotype.

DISCUSSION

Coactivator-Associated Arginine Methyltransferase 1 (Carm1) has been shown to enhance transcriptional activation by nuclear receptors through interactions with a number of coactivators (Miao et al., 2005). This transcriptional activation may play a role in the ethanol preference phenotype. Carm1 is up-regulated in ethanol-preferring mice (Mulligan et al., 2006), is located on a quantitative trait locus on chromosome 9 associated with ethanol preference (Ibid.), and was shown here (Figures 5 and 6) to have Mutual Information with a number of previously identified key ethanol preference genes. This project has shown there is a statistical dependency between Carm1 and Prkce, another gene for which there is evidence of influencing alcohol drinking levels.

Protein Kinase C-epsilon (Prkce) catalyzes the phosphorylation of specific proteins (Pan, et al., 2006). The phosphorylation of proteins affects the shape of the protein, thus either activating or deactivating the newly phosphorylated protein. Prkce may be influencing ethanol consumption through post translational modification. Similarly, Ubiquitin-conjugating enzyme E2M (Ube2m), shown in this project to regulate a number of the ethanol preference genes, has been proven to be involved in post translational modification within the ubiquitin cycle (GO:0006512, 2008).

cAMP responsive element binding protein 1 (Creb1) has been previously studied for its role in brain function including addiction, depression, and memory (Carlezon et al., 2005). It has been determined in this project (Figure-7) that this gene is regulated by Zinc Finger Protein 143 (Zfp143), which is up-regulated in ethanol-preferring mice. This suggests the activation of Zfp143 in ethanol-preferring mice could then in turn regulate

Creb1. The discovery of this relationship links the level of ethanol consumption to brain functions such as addiction and anxiety, both of which Creb1 is known to play a role in.

Activating transcription factor 2 (Atf2) is a member of the MAPKinase signaling pathway in mice (Morton et al., 2004). Many of the Mapk genes are up-regulated in ethanol-preferring mice. This project has determined that Atf2 (Figure-8) is regulated by Hypoxia up-regulated 1 (Hyou1) a protein associated with a response to stress. This gene also has a regulatory relationship with Creb1, Prkce, and Ube2m, among other ethanol preference genes. It regulates Stat3, which is up-regulated in ethanol-preferring mice. The regulatory relationship between Atf2 and a number of the previously identified ethanol preference genes strongly suggests that it has an influence on the ethanol preference phenotype.

Forkhead box 2 (Foxa2) is a down-regulated gene in ethanol-preferring mice (Figure-9). This gene is regulated by Creb1, thus in ethanol-preferring mice the up-regulated Zfp143 regulates Creb1 which in turn regulates the down-regulated Foxa2. Foxa2 regulates the expression of another transcription factor, Signal transducer and activator of transcription 3 (Stat3). The suppression of Foxa2 may play an important role in ethanol consumption and the ethanol preference phenotype.

This project extends the work of Mulligan et al. (2006) in the search for a better understanding of the ethanol preference phenotype. The candidate genes were further analyzed to begin to demonstrate relationships among them, developing a preliminary network for the ethanol preference phenotype. Other than the identification of the candidate genes, little was known about the regulatory networks involved in the ethanol preference phenotype prior to the start of this project. This project uses computational

tools to narrow the search for a regulatory network and determined centers of regulatory networks which should be the focus of any further research.

Because of the previous lack of evidence for the regulatory networks involved in the ethanol preference phenotype, many things can be done to continue with this project. More algorithms can be applied to this data to confirm the networks identified here. Also, wet lab verification can be performed to determine the validity of the algorithmic predictions seen in this project. Ultimately an understanding of the regulatory networks for alcohol preference could lead to methods for disrupting the networks as treatments for alcoholism.

BIBLIOGRAPHY

Bogue, Molly (2004, October) Retrieved March 2008, from Complex Trait Consortium: http://www.complextrait.org/

Carlezon Jr, WA, Duman, RS, Nestler, EJ (2005) The many faces of CREB. *TRENDS in Neurosciences*. **28**: 436-445.

Enoch, MA (2003) Pharmacogenomics of alcohol response and addiction. *American Journal of Pharmacogenomics*. **3**: 217-232.

Faith, JJ, Hayete, B, Thaden, JT, Mogno I, Wierzbowski, J, Cottarel, G, Kasif, S, Collins, JJ, Gardner, TS (2007) Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *Public Library of Science Biology*. **5**: 54-66.

GDCADA. *Alcohol – Facts, Statistics, Resources, and Impairment Charts.* (2006, March). Retrieved February 2008, from Greater Dallas Council on Alcohol & Drug Abuse: http://www.gdcada.org/statistics/alcohol.htm

GO:0006512 [released 15 April 2008] The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics*. **25**: 25-29.

Green, AS, Grahame, NJ (2008) Ethanol drinking in rodents: is free-choice drinking related to the reinforcing effects of ethanol? *Alcohol*. **42**: 1-11.

Irizarry RA, Hobbs, B, Collin, F, Beazer-Barclay, YD, Antonellis, KJ, Speed, T (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. **4**: 249-264.

Kalir, S, Mangan, S, Alon, U (2005) A coherent feed-forward loop with a SUM input function prolongs flagella expression in Escherichia coli. *Molecular Systems Biology*. **1**: 1-6.

Margolin, AA, Nemenman, I, Basso, K, Wiggins, C, Stolovitzky, G, Dalla Favera, R, Califano, A (2006) An Algorithm for the Reconstruction of Gene Regulatory Networks in Mammalian Celluar Context. *Department of Bioinformatics.* **7**: 1-S7.

Miao, F, Li, S, Chavez, V, Lanting, L, Natarajan, R (2005) Coactivator-Associated Arginine Methyltransferase-1 Enhances Nuclear Factor-B-Mediated Gene Transcription through Methylation of Histone H3 at Arginine 17. *Molecular Endocrinology*. **20:** 1562-1573.

Morton S, Davis RJ, Cohen P (2004) Signalling pathways involved in multisite phosphorylation of the transcription factor ATF-2. *FEBS Letters*. **572**: 177-183.

Mulligan, MK, Ponomarev, I, Hitzemann, RJ, Belknap, JK, Tabakoff, B, Harris, RA, Crabbe, JC, Blednov, YA, Grahame, NJ, Phillips, TJ, Finn, DA, Hofman, PL, Iyer, VR, Koob, GF, Bergeson, SE (2006) Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. *Proceedings of the National Academy of Sciences*. **103**: 6368-6373.

Pan Q, Bao LW, Teknos TN, Merajver SD (2006) Targeted disruption of protein kinase C epsilon reduces cell invasion and motility through inactivation of RhoA and RhoC GTPases in head and neck squamous cell carcinoma. *Cancer Research* **66**: 9379-9384.

Pearl, J. "Probabilistic Reasoning in Intelligent Systems: network of plausible inference". (Morgan Kaufmann, San Francisco, 1988).

Quackenbush, J (2006) Microarray Analysis and Tumor Classification. *New England Journal of Medicine.* **354**: 2463-2472.

Singh, S. (2006, June). *Alcoholism Projects*. Retrieved February 2008, from The University of Western Ontario: http://www.uwo.ca/biology/Faculty/singh/alcoholism.htm

Singhal, M, Domico, K (2007) CABIN: Collective Analysis of Biological Interaction Neworks. *Computational Biology and Chemistry*. **31**: 222-225.

Sladeczek, J, Hartemink, AJ, Robinson, J (2007) Bayesian Network Inference with Java Objects Version 2 User Guide. http://www.cs.duke.edu/~amink/software/banjo/documentation/banjo.user.pdf

Tarca, AL, Romero, R, Draghici, S (2006) Analysis of Microarray Experiements of Gene Expression Profiling. *American Journal of Obstetrics and Gynecology.* **195**: 373-388.

Thomson DM, Herway ST, Fillmore N, Kim H, Brown JD, Barrow JR, Winder WW (2008) AMP-activated protein kinase phosphorylates transcription factors of the CREB family. *Journal of Applied Physiology*. **104**: 429-438.

U.S. Department of Energy Office of Science (2006, August). Retrieved March 2008, from Genome Programs of the U.S. Department of Energy Office of Science: http://genomics.energy.gov/

Williams, RW (2004) "QTL." From The WebQTL Glossary--A GeneNetwork Resource. http://www.genenetwork.org/glossary.html