UNIVERSITY OF CALIFORNIA

Los Angeles

Discovery and characterization of human exonic transcriptional regulatory elements

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Molecular and Medical Pharmacology

By

Arshad H Khan

2012

ABSTRACT OF THE DISSERTATION

Discovery and characterization of human exonic transcriptional regulatory elements

by

Arshad H Khan

Doctor of Philosophy in Molecular and
Medical Pharmacology
University of California, Los Angeles, 2012
Professor Desmond J. Smith, Chair

We sought regulatory elements by shotgun cloning human exonic DNA fragments into luciferase reporter vectors and assessing transcriptional regulatory activity in liver cells. Seven elements within coding regions and three within 3’ UTRs were discovered. Putative regulatory elements were generally but not consistently evolutionarily conserved, enriched in known transcription factor binding sites (TFBSs) and associated with several histone modifications. Evidence of cis-regulatory potential of an element within a TUBA1B exon was established by correlating expression of TUBA1B with activation of transcription factors predicted to have binding sites within this element. Nevertheless, no clear rules defining coding regulatory elements emerged. We estimate that hundreds of exonic regulatory elements exist, an unexpected finding that highlights a surprising multi-functionality of sequences in the human genome.

The dissertation of Arshad H Khan is approved.

Huiying Li

Thomas Graeber

Aldons J. Lusis

Desmond J. Smith, Committee Chair

University of California, Los Angeles

2012

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

APOE	Apolipoprotein E
ChIP	Chromatin immunoprecipitation
FDR	False discovery rate
GEO	Gene expression omnibus
H3K27ac	Histone 3 lysine 27 acetylation
H3K4me1	Histone 3 lysine 4 mono-methylation
H3K4me2	Histone 3 lysine 4 di-methylation
H3K27me3	Histone 3 lysine 27 mono-methylation
PPARA	Peroxisome proliferator-activator receptor alpha
PPREs	PPAR response elements
RORA	Retinoic acid receptor-related orphan receptor alpha
TFBSs	Transcription factor binding sites
TUBA1B	Tubulin alpha 1B
TUBB4	Tubulin beta 4
UTRs	Untranslated regions

ACKNOWLEDGEMENT

I would first like to thank Dr. Desmond Smith for being a thoughtful and supportive advisor in my research career. Dr. Desmond Smith's consistent support prepared me to become a better scientist. His intelligence and guidance helped me to think independently and solve problems thoughtfully and meticulously. He challenged me to constantly question the results of my work, and his encouragement built-up my confidence.

I also would like to thank Dr. Thomas Graeber, Dr. Aldons J. Lusis and Dr. Huiying Li for agreeing to be committee members for my dissertation project. Your kind support and advice were thoughtful and appreciated. Dr. Huiying Li showed me how to be an expert in my research. Dr. Thomas Graeber helped me to formulate a clear plan for my dissertation. Dr. Jake Lusis's exceptional kindness and support provided me with the confidence to complete my project.

I would have not been able to finish this dissertation without the help of my colleagues in Smith lab. Most importantly, my foremost thanks go to Andy Lin, a post doctoral fellow in our lab. His help in every step preparing for this project, particularly in analyzing data, were critical to complete this dissertation. My current understanding of statistical could not been acquired without Andy's help. I would also like to thank Christopher Park, Albert Chen and Richard Wang for their support and suggestions for this project.

Lastly, my completion of this dissertation would not be possible without the love and support of my family. They have been consistent in their support and love over the course of my studies. They have shown me the value of commitment and belief in my abilities. My humble thanks go to my wife Taneema who has always encouraged me to follow my dreams

and never complained when I was not around. Also, thanks go to my parents, my brothers, my in-laws and my friends who were always supportive in every step of my life.

VITA

1997	B.S., Biochemistry University of California Los Angeles Los Angeles, California
1997-1999	Staff Research Associate I Howard Hughes Medical Institute/UCSD San Diego, California
1999-2000	Staff Research Associate I Department of Molecular and Medical Pharmacology University of California, Los Angeles
2000-2002	Staff Research Associate II Department of Molecular and Medical Pharmacology University of California, Los Angeles
2002-2005	Staff Research Associate III/Lab Manager Department of Molecular and Medical Pharmacology University of California, Los Angeles
2006-2007	Staff Research Associate IV/Lab Manager Department of Molecular and Medical Pharmacology University of California, Los Angeles
2007-2012	Graduate Student Researcher Department of Molecular and Medical Pharmacology University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

A.H. Khan, D.M. Sayah, T.L. Gasperoni, and D.J. Smith, A genetic screen for novel behavioral mutations in mice. Mol Psychiatry 5 (2000) 369-77.

V.M. Brown, A. Ossadtchi, A.H. Khan, S.S. Gambhir, S.R. Cherry, R.M. Leahy, and D.J. Smith, Gene expression tomography. Physiol Genomics 8 (2002) 159-67.

V.M. Brown, A. Ossadtchi, A.H. Khan, S.R. Cherry, R.M. Leahy, and D.J. Smith, High-throughput imaging of brain gene expression. Genome Res 12 (2002) 244-54.

V.M. Brown, A. Ossadtchi, A.H. Khan, S. Yee, G. Lacan, W.P. Melega, S.R. Cherry, R.M. Leahy, and D.J. Smith, Multiplex three-dimensional brain gene expression mapping in a mouse model of Parkinson's disease. Genome Res 12 (2002) 868-84.

A. Ossadtchi, V.M. Brown, A.H. Khan, S.R. Cherry, T.E. Nichols, R.M. Leahy, and D.J. Smith, Statistical analysis of multiplex brain gene expression images. Neurochem Res 27 (2002) 1113-21.

R.P. Singh, V.M. Brown, A. Chaudhari, A.H. Khan, A. Ossadtchi, D.M. Sforza, A.K. Meadors, S.R. Cherry, R.M. Leahy, and D.J. Smith, High-resolution voxelation mapping of human and rodent brain gene expression. J Neurosci Methods 125 (2003) 93-101.

A.H. Khan, A. Ossadtchi, R.M. Leahy, and D.J. Smith, Error-correcting microarray design. Genomics 81 (2003) 157-65.

Liu D., Singh R.P., Khan A.H., Lusis A. J., Davis R.C., Smith D.J. Identifying loci for behavioral traits using genome-tagged mice. *The Journal of Neuroscience Research* 74 (2003) 562-9

Liu D., Singh R.P., Khan A.H., Lusis A. J., Davis R.C., Smith D.J. Mapping behavioral traits using genetically tagged mice. *American Journal of Geriatric Psychiatry* 12 (2004)158-65.

H. Wang, W.J. Qian, H.M. Mottaz, T.R. Clauss, D.J. Anderson, R.J. Moore, D.G. Camp, 2nd, A.H. Khan, D.M. Sforza, M. Pallavicini, D.J. Smith, and R.D. Smith, Development and evaluation of a micro- and nanoscale proteomic sample preparation method. J Proteome Res 4 (2005) 2397-403.

H. Wang, W.J. Qian, H.M. Mottaz, T.R. Clauss, D.J. Anderson, R.J. Moore, D.G. Camp, 2nd, A.H. Khan, D.M. Sforza, M. Pallavicini, D.J. Smith, and R.D. Smith, Characterization of the mouse brain proteome using global proteomic analysis complemented with cysteinyl-peptide enrichment. *Journal of Proteome Research* 5 (2006):361-9.

M.H. Chin, A.B. Geng, A.H. Khan, W.J. Qian, V.A. Petyuk, J. Boline, S. Levy, A.W. Toga, R.D. Smith, R.M. Leahy, D.J. Smith. A genome-scale map of expression for a mouse brain section obtained using voxelation. *Physiol. Genomics* 30 (2007) 313-21

M.H. Chin, W.J. Qian, H. Wang, V.A. Petyuk, J.S. Bloom, D.M. Sforza, G. Lacan, D. Liu, A.H. Khan, R.M. Cantor, D.J. Bigelow, W.P. Melega, D.G. Camp, 2nd, R.D. Smith, and D.J. Smith, Mitochondrial dysfunction, oxidative stress, and apoptosis revealed by proteomic and transcriptomic analyses of the striata in two mouse models of Parkinson's disease. J Proteome Res 7 (2008) 666-77.

C.C. Park, S. Ahn, J.S. Bloom, T. Wu, A. Lin, R.T. Wang, A. Sekar, A.H. Khan, C.J. Farr, A.J. Lusis, R.M. Leahy, K. Lange, D.J. Smith. Fine mapping of regulatory loci for mammalian gene expression using radiation hybrids. *Nat Genet* 40 (2008) 421-429.

H.P. Chen, A. Lin, J.S. Bloom, A.H. Khan, C.C. Park, and D.J. Smith, Screening reveals conserved and nonconserved transcriptional regulatory elements including an E3/E4 allele-dependent APOE coding region enhancer. Genomics 92 (2008) 292-300.

G.D. Gale, R.D. Yazdi, A.H. Khan, A.J. Lusis, R.C. Davis, and D.J. Smith, A genome-wide panel of congenic mice reveals widespread epistasis of behavior quantitative trait loci. Mol Psychiatry 14 (2009) 631-45.

C.C. Park, G.D. Gale, S.D. Jong, A. Ghazalpour, B.J. Bennett, C.R. Farber, P. Langfelder, A. Lin, A.H. Khan, E. Eskin, S. Horvath, A.J. Lusis, R.A. Ophoff, D.J. Smith. Gene networks associated with conditional fear in mice identified using a systems genetics approach. *BMC Systems Biology* 5(2011) 43-

R.T. Wang, S. Ahn, C.C. Park, A.H. Khan, K. Lange, D.J. Smith. Effects of genome-wide copy number variation on expression in mammalian cells. BMC Genomics 12 (2011) 562-

Chapter 1 Introduction

1.1 Background

An important key to deciphering the human genome is to identify the regulatory elements that control gene expression. Indeed, disruption of these elements has been linked to a number of human diseases including cancers [1], preaxial polydactyly [2], Van Buchem disease [3], and facioscapulohumeral muscular dystrophy [4]. Nevertheless, the vast majority of regulatory elements remain unidentified. One major hurdle to annotating transcriptional regulatory elements is that they are ubiquitous and are found in both intergenic [3; 5], and intronic regions [2; 5]. More surprisingly, isolated examples of transcriptional regulatory elements have recently been found in exons, both coding [5; 6; 7] and non-coding [8; 9]. These coding regulatory elements, though critically important given their dual function, are poorly understood and almost completely uncatalogued.

Because regulatory elements can be found anywhere in the genome, large-scale, high-throughput screens are needed to identify them efficiently. Genome-wide searches have met with some success by exploiting several features of regulatory elements, for example their enrichment in transcription factor binding sites [10; 11; 12; 13], and their association with histone modifications [12; 13; 14; 15]. Unfortunately, coding regions have the same properties, complicating the identification of regulatory elements within coding regions.

1.2 Comparative genomics

Comparative genomics were among the first approaches used to search for functional elements by identifying sequences more conserved across species than would be expected by chance [16]. Although successfully used in intergenic regions, this strategy is not viable for finding regulatory elements within coding regions as both types of sequences are expected to be highly conserved and thus indistinguishable. In fact, early attempts to identify regulatory elements genome-wide intentionally masked coding regions [6]. Recently, it has been shown that regulatory elements within coding regions may be even more conserved than flanking coding regions [7] , presumably due to dual selective pressure to retain both regulatory and coding function. Whether or not coding regulatory elements are super conserved as a rule is unknown.

1.3 Transcription factor binding sites

The current view of transcriptional regulatory elements is that they are clusters of transcription factor binding sites (TFBSs), which when bound by complexes of transcription factors (TFs) can recruit or block various critical components of the transcriptional machinery such as RNA polymerase II [17]. By identifying such clusters, genome-wide computational methods have been used to predict the locations of 118,000 regulatory elements [11]. However, TFBS sequences are typically short, 5-15 bp, and degenerate, creating a substantial false positive problem when only computational methods are used. Alternatively, TFBSs can be identified genome-wide experimentally, via ChIP-chip or ChIP-seq [12; 13]. Although less efficient and much more laborious than computational methods, these methods can at least verify TF binding to predicted elements.

1.4 Histone Modification

Histone modification is a means by which gene expression can be controlled independently of the DNA sequence. Regulatory elements are often associated with particular chromatin states marked by a number of histone variants, particularly those that are methylated and/or acetylated [14; 15; 17] . Genome-wide maps of histone modification have been used to predict a set of 55,000 enhancers [14]. As transcribed regions are themselves associated with their own histone modifications, how these modifications might change in regions of overlap with regulatory elements is unclear.

1.5 Aim

Here we report an unbiased search for exonic regulatory elements active in liver. We expand on previous work in which we evaluated genomic DNA from the ApoE gene cluster on chromosome 19 for regulatory elements [5]. In that investigation, we shotgun cloned DNA into luciferase reporter vectors to assay regulatory activity. For the present study, we interrogated putative regulatory sequences only from exonic DNA. We assessed the properties of these coding regulatory elements by characterizing their degree of evolutionary conservation, TFBS enrichment, GC-content, and association with histone modifications.

Chapter 2 Experimental Methods

2.1 Cell culture and cDNA synthesis

To normalize transcript levels used to generate cDNA, RNA was pooled in equal amounts from three human cell lines; HEK-293 (kidney), C3A (liver) and SvGp12 (astrocyte) (all from ATCC). Cell lines were grown in Eagle's Minimum Essential medium (ATCC 30-2003) and 10% fetal bovine serum (Invitrogen) until 75% confluency was reached. For each cell line, the Oligotex mRNA mini kit (Qiagen) was used to isolate and purify mRNA. Extracted mRNA from all cell lines was then pooled together to synthesize cDNA using the Just cDNA Double Stranded cDNA Synthesis Kit (Agilent). Random hexamers were chosen as primers to avoid the 3’ bias of oligo-dT primers. Quality of RNA and cDNA were assessed using spectrophotometry and gel electrophoresis, respectively.

2.2 Library construction

Samples of pooled cDNA were digested by either Sau3AI or AluI (New England Biolabs) and sub-cloned into the pGL3-promoter vector (Promega), digested with SmaI or BglII, respectively. Vectors were then transformed into MAX efficiency DH5-alpha chemically competent bacteria (Invitrogen), clones isolated, and plasmid DNA purified using 96 Plasmid Miniprep Kit (Qiagen).

2.3 Control clones

The pGL3-promoter and pGL3-basic vectors, both from Promega, served as neutral (promoter, but neither enhancer nor silencer) and negative controls (no promoter, enhancer or silencer), respectively. The reporter gene for both vectors was firefly lucifersase. For a positive control, we used the previously identified human APOE liver-specific enhancer HCR1 inserted into the pGL3- promoter vector [18].

2.4 Transfection and reporter gene activity assays

For each clone, 100 ng of firefly experimental luciferase plasmid and 10 ng of control *Renilla* luciferase plasmid (phRL-TK, Promega) were co-transfected into C3A human liver cells (ATCC) using the Effectene transfection reagent (Qiagen) in 96-well plates. The *Renilla* plasmid serves as a control for transfection efficiency. Transfection was performed when cells had reached 80% confluency. Cells were then grown in Eagle's Minimum Essential medium (ATCC 30-2003) and lysed after 24 hours. Luciferase reporter gene activity was assayed using the Dual-Luciferase Assay Kit (Promega).

2.5 Screens and sequencing

Relative luciferase activity, the $\log_{10}$ ratio of firefly to *Renilla* luciferase signal, was used as a measure of expression relative to transfection efficiency. Raw activity ratios were quantile normalized across 96-well plates. Clones were chosen for further screening upon demonstrating activity two standard deviations away from the mean after normalization. Sequencing of putative clones was performed at GenoSeq, the UCLA genotyping and sequencing core.

Chapter 3 Results

3.1 cDNA library creation and luciferase assays

Three human cell lines, C3A (liver), HEK-293(Kidney) and SVGp12 (Astrocytes) were selected for cDNA synthesis. To maximize transcript coverage, mRNA extracted from all three lines was pooled together. Because we were only interested in exonic sequences, we restricted our assays to cDNA rather than whole genomic DNA. Pooled cDNA was digested independently by Sau3AI or AluI and subcloned into the multiple cloning site upstream of the basal SV40 early promoter of the pGL3-promoter vector. A total of 1932 clones were created, 1008 from Sau3AI and 924 from AluI with an average fragment size of ~167 base pairs based on sequencing.

All clone-containing firefly luciferase vectors were co-transfected with *Renilla* luciferase vectors into C3A cells in 96-well plates. Expression of each luciferase was assayed independently, and the regulatory activity of the putative element estimated from the $\log_{10}$ ratio of firefly to *Renilla* reporter gene activity. This measure evaluated expression of the tested element relative to transfection efficiency. Transfection efficiency measured using the CMV-GFP construct (pEGFP-N3, Clontech) was uniform at approximately seven percent [5].

3.2 Screening for regulatory elements

Quantile normalization was used to compare luciferase activities across plates, and the activities of clones produced by Sau3AI and AluI digestion were normalized separately (Fig. 3-1(A) and Fig. 3-1(B)). Controls acted as expected: vectors with neither a promoter nor enhancer had low activity, vectors with a promoter but no enhancer had moderate activity, and vectors with both a promoter and the known liver enhancer element HCR1 [18] had high activity. The distribution of non-normalized relative luciferase activities was nearly normal and negatively skewed, as described in a previous study [5] (Fig. 3-1(C) and Fig. 3-1(D)). The distribution's unimodality reinforces our previous findings that the distinction between regulatory and nonregulatory sequences is not hard and fast, particularly in the case of enhancers, while the extended negative tail suggests that silencers have a wider range of effect sizes than enhancers [5].

The overall screening procedure was designed to identify coding fragments that reliably show strong regulatory signals, with more stringent thresholds for inclusion at each step (Fig. 3-1(E)).

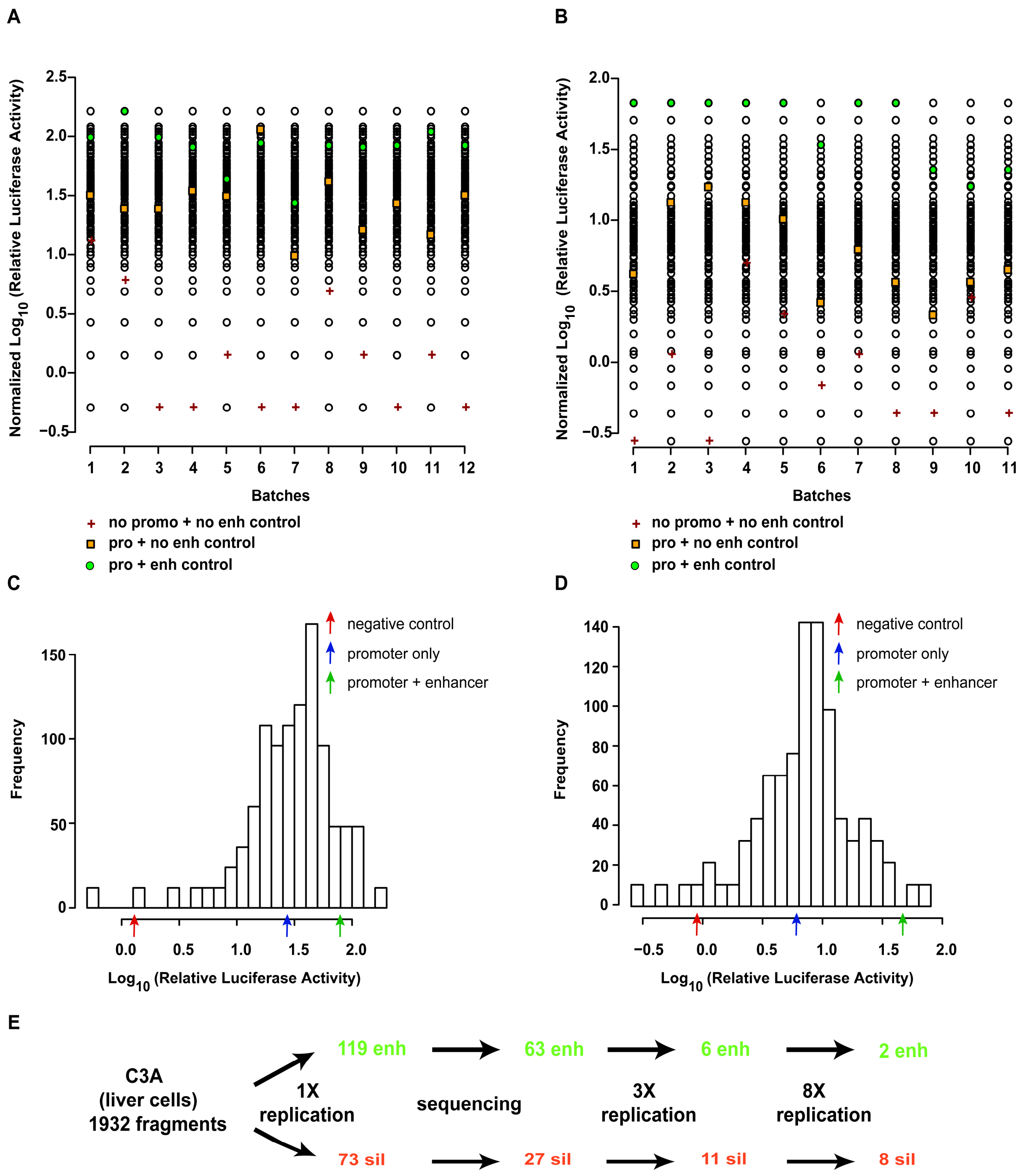

Figure 3-1. Distribution and workflow. (A) Quantile normalized relative luciferase activity for Sau3AI-digested exonic fragments in liver C3A cells compared within and between plates. Relative luciferase activity is the $\log_{10}$ ratio of firefly luciferase to *Renilla* luciferase. Batch number indicates corresponding 96-well plate. (B) Quantile normalized relative luciferase activity for AluI-digested exonic fragments in C3A liver cells. (C) Distribution of relative luciferase activities for Sau3aI-digested fragments in liver C3A cells. (D) Distribution of relative luciferase activities for AluI-digested fragments in liver C3A cells. (E) Workflow for identifying regulatory elements.

From the initial, unbiased screen of all 1,932 fragments, we selected for additional evaluation clones with luciferase activity beyond two standard deviations from the mean. Each of these clones was then sequenced, and its sequence aligned with the human genome (NCBI build 37.2) using BLAT [19]. Non-exonic clones were culled by retaining only those clones whose top BLAT match resided in coding exons, 3' untranslated regions (UTRs) or 5' UTRS (all matches had 100% identity, except one unusually long 305 bp fragment with 98% identity). Exonic clones were then subjected in C3A liver cells to two subsequent rounds of testing for regulatory activity, the first round consisting of three replicate assays and the second round consisting of eight. Clones were removed from consideration if they did not demonstrate luciferase activity significantly different from the pGL3-promoter control in each round of assays as determined by one-sample t-tests (DF = 2, DF = 7, for three and eight replicates, respectively) controlled by false discovery rates (FDR < 5% used as threshold for inclusion). We were confident that eight replicates would provide a robust signal of regulatory acitivity, as luciferase signals across replicates were highly correlated (Pearson correlation of + 0.978, $p < 10^{-300}$).

3.3 Putative regulatory elements

Two clones that showed significantly higher activity than the promoter control across the eight replicates were deemed putative enhancers, while eight clones with lower activity were deemed putative silencers (Figs. 3-2(A) and 3-2(B)).

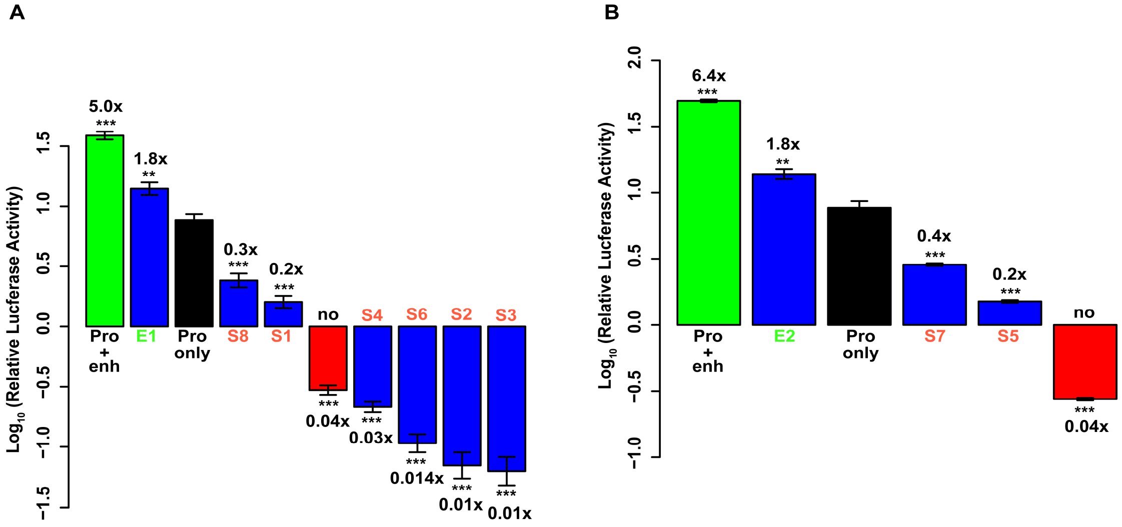

Figure 3-2. Regulatory activity of putative elements. (A) Mean activities of 8 replicates of Sau3AI-digested putative regulatory elements in C3A liver cells. Log_{10} changes relative to promoter-only construct shown. Error bars, standard error of the mean. (**, $p<0.01$ and ***, $p<0.0001$ compared to promoter-only construct, both figures.) (B) Mean activities of 8 replicates of AluI-digested putative regulatory elements in C3A liver cells.

Genomic locations, lengths and host genes of putative elements are provided in table 3-1. Sequences for each element are provided in table 3-2. Of the ten putative elements, six resided in coding regions, three in 3’ UTRs, and one resided in the single non-coding exon of a mitochondrial gene (Fig. 3-3).

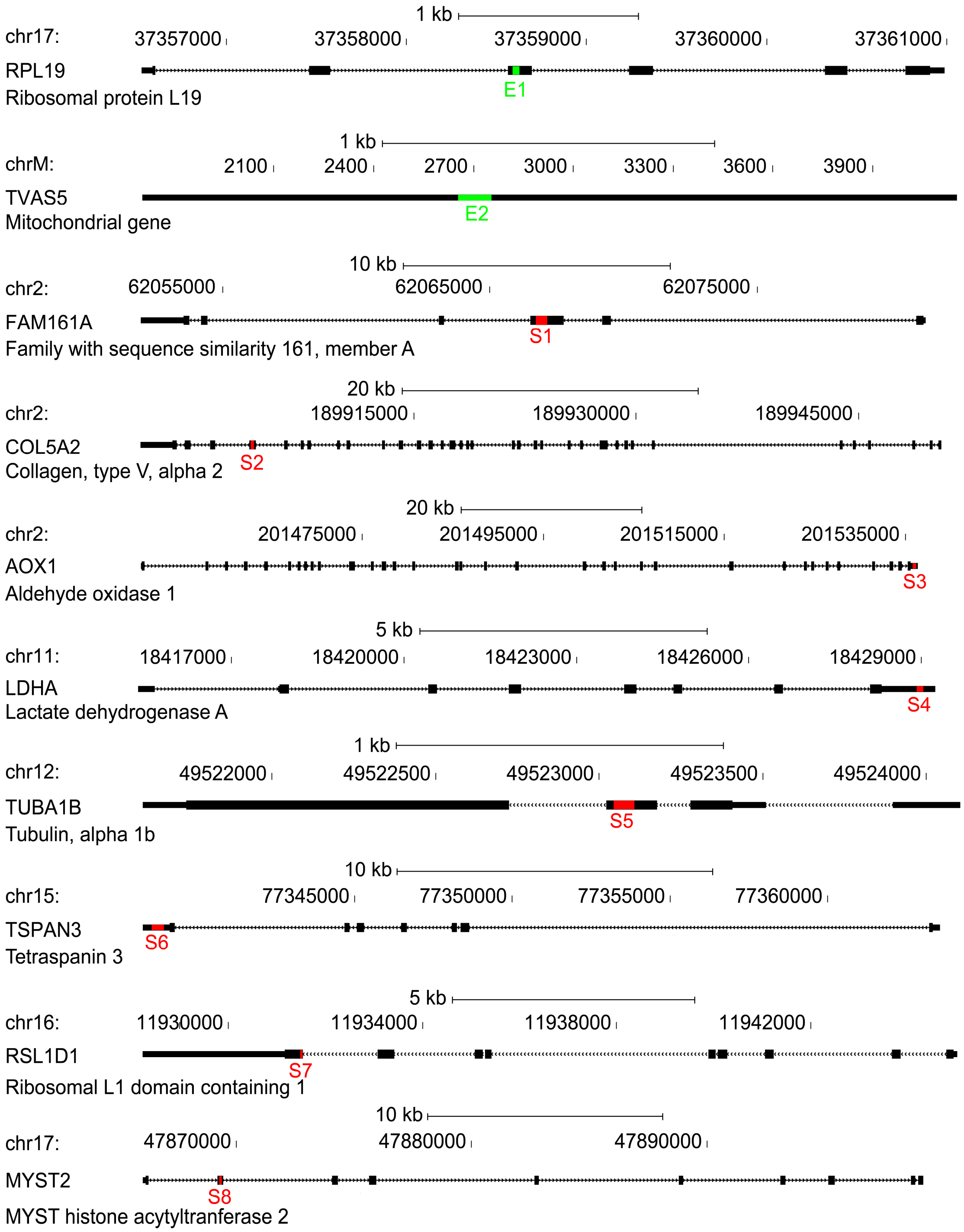

Figure 3-3. Genomic locations of exonic regulatory elements. Positions of fragments within exons, including coding regions (thick boxes) and 3' UTRs (thin boxes).

3.4 Evolutionary conservation of putative regulatory elements

As judged by phastCons (UCSC genome browser, [16]) seven of the ten putative regulatory element sequences were strongly conserved across all vertebrates (mean base-by-base phastCons score for element > 0.5), two were somewhat conserved (score > 0.1), and one was not conserved at all (score < 0.1) (Figure 3-4 and Table 3-3). Generally, regulatory elements found in coding exons were more highly conserved than those elements found in 3' UTRs (mean coding exon score = 0.707 ± 0.143, mean 3'UTR score = 0.392 ± 0.196). As a whole, regulatory element conservation scores preserve both amino acid sequence and regulatory function [7].

We compared the mean phastCons conservation score of each regulatory element to the mean conservation scores of all other exons within the same gene using one-sample t-tests (Table 3-3). Surprisingly, none of the putative elements were significantly more conserved than their neighboring exons, whereas two coding elements, S7 and S8, appeared to be significantly less conserved than their neighboring exons at FDR < 5%. Fragments S3 and S4 were also less conserved, but were in non-coding regions so would not be expected to be superconserved.

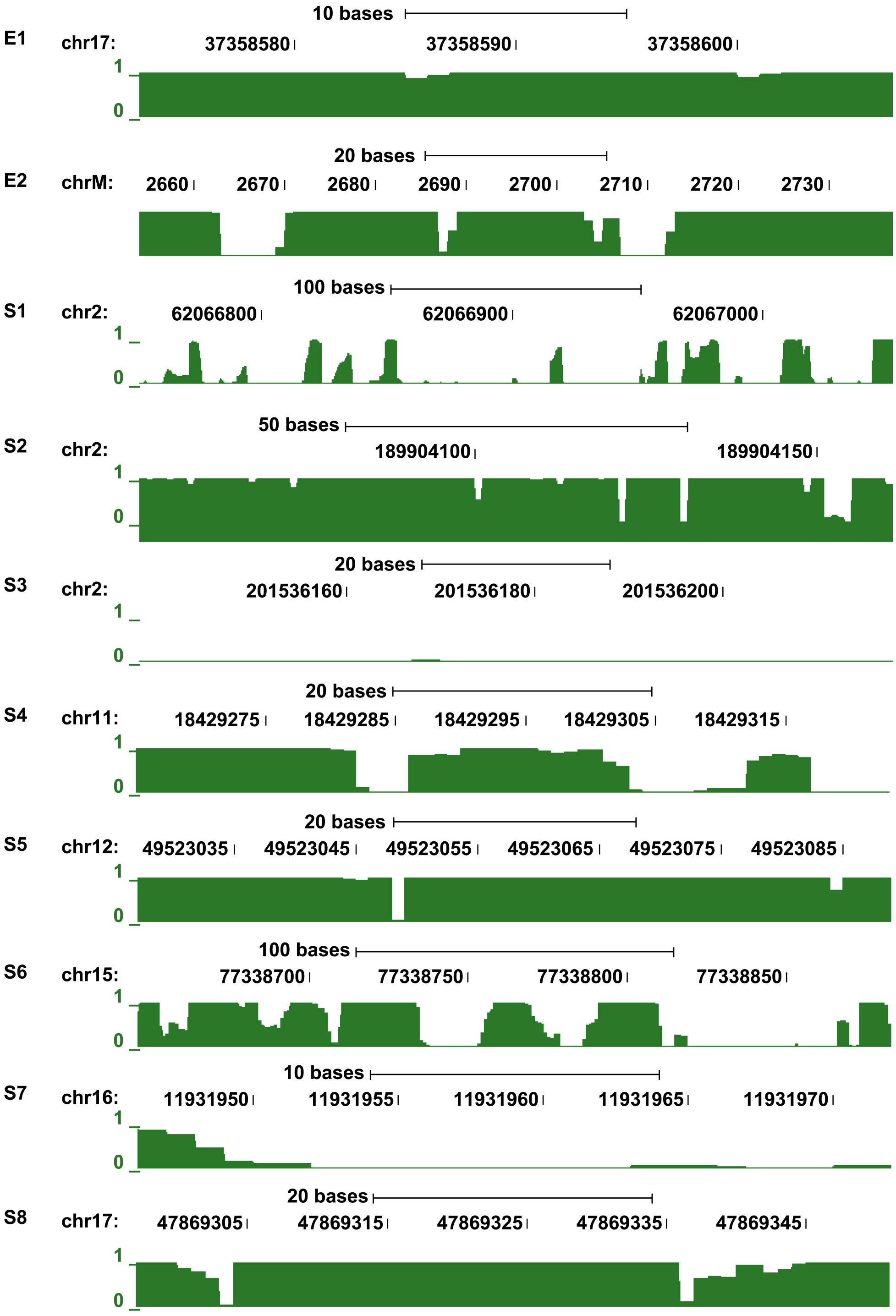

Figure 3-4. Conservation of fragments. PhastCons scores, which represent the probability that a base is conserved across vertebrates, for all bases in each fragment sequence.

3.5 Transcription factor binding sites within putative regulatory sequences

We searched for TFBSs using the UCSC Genome Browser ENCODE/HAIB Transcription Factor Binding Sites "peaks" track, which annotates sites with the best evidence ($p < 10^{-5}$) for TFBS along the entire human genome as determined by ChIP-seq. Because most DNA-protein interactions were tested in HepG2 liver cells, we confined our search to that cell line. In addition, HepG2 cells and the C3A cells used in our study both originate from liver. Transcription factor binding site peaks found within the putative element sequences are listed in table 3-4. Four of our putative elements had known TFBSs as determined by ChiP-seq, two with multiple sites. The most common binding site was for HNF4A, hepatocyte nuclear factor 4α. HNF4A is known to be liver-enriched and to target at least 260 genes, possibly thousands of genes covering a wide array of functions [20].

We also employed the UCSC Genome Browser HMR Conserved Transcription Factor Binding Sites track, which uses comparative genomics to predict TFBS location conserved ($p < 0.01$) across human, mouse, and rat. TFBSs are computationally determined, but not experimentally verified, using TFBS sequence data in the Transfac Matrix Database [10]. Three of our putative silencers contained conserved TFBS (Fig. 3-5). The S5 silencer located within the TUBA1B gene contained two overlapping TFBS for RORA1 ($z = 2.45$, $p = 0.0071$) and PPARA ($z = 2.64$, $p = 0.0041$). We explore this protein-DNA interaction in more detail below.

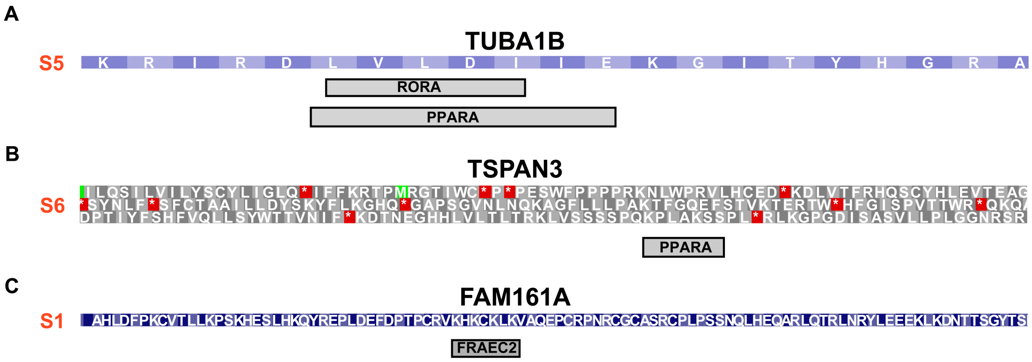
A
TUBA1B
S5
K R I R D L V L D I I E K G I T Y H G R A
RORA
PPARA
B
TSPAN3
S6
IILQSILVILYSCYLIGLQ*IFFKRTPMRGTIWC*P*PESWFPPPPRKNLWPRVLHCED*KDLVTFRHQSCYHLEVTEAG
*SYNLF*SFCTAAILLDYSKYFLKGHQ*GAPSGVNLNQKAGFLLLPAKTFGQEFSTVKTERTW*HFGISPVTTWR*QKQA
DPTIYFSHFVQLLSYWTTVNIF*KDTNEGHHLVLTLTRKLVSSSSPQKPLAKSSPL*RLKGPGDISASVLLPLGGNRSR
PPARA
C
FAM161A
S1
AHLDFPKCVTLLKPSKHESLHKQYREPLDEFDPTPCRVKHKCKLKVAQEPCRPNRCGCASRCPLPSSNQLHEQARLQTRLNRYLEEEKLKDNTTSGYTS
FRAEC2

Figure 3-5. Conserved transcription factor binding sites. Positions of transcription factor binding sites conserved across human, mouse and rat relative to amino acid sequence of fragment.

3.6 GC-content of putative regulatory sequences

GC-rich CpG islands (CGIs) are most often found in the core promoter region immediately upstream of the transcription start site. However, the discovery of a CGI in the intron of the PAX6 gene that may act as an alternative transcription start site has introduced the notion that CGIs not associated with the core promoter may also play a role in transcriptional regulation [21]. To determine whether the putative exonic regulatory elements were found within CGIs, we employed the USCS Genome Browser CpG island prediction track, which identifies sequences at least 200 bp long consisting of > 50% GC-content arranged as CpG dinucleotides at least 60% as frequently as expected from GC-content. None of the putative elements were found in CGIs.

It is possible that regulatory elements may actually be less likely to be in CGIs than coding exons in general. We therefore compared the GC-content of each element to the GC-content of all exons within the same gene using one-sample t-tests (Table 3-5). Three of the ten elements had significantly less GC-content than their neighboring exons. This result is expected for the two elements located within 3’ UTRs, as non-coding exons typically have low GC-content. However, the lower GC-content of coding element S2 is somewhat surprising, as S2 has the highest GC-content of all putative elements, although its host gene also has the highest GC-content of all genes.

3.7 Histone modification signatures of regulatory elements.

Well-studied histone modifications associated with enhancers include H3K4me1 [14; 15], H3K4me2 [15], H3K27ac [14] and H2A.Z [14]. Histone modifications that predict silencers are not as well known, but several combinations of modifications at promoters have been found to be correlated with low expression, most of which contain H3K27me3 [15]. To determine whether the putative exonic regulatory elements were associated with histone modifications, we aligned our regulatory sequences with the ENCODE/Broad Histone Modification track of the UCSC genome browser, which maps histone modifications across the genome as determined by ChIP-seq across several cell lines including liver. Because only a portion of tested histone modifications were mapped in HepG2 liver cells, we used tracks from all cell types.

True to their versatility, fragments varied in the number and types of histone modifications with which they were associated (Table 3-6). The E1 enhancer was associated with all 3 known enhancer modifications, H3K4me1, H3K4me2, and H3K27ac, the latter two in liver. On the other hand, four of eight silencers were also associated with H3K4me1 and H3K27ac, although none of them in liver. Repressive signature H3K27me3 was associated with 7 of 8 silencers as well as enhancer E1. Other modifications associated with a majority of fragments include H3K79me2 and H3K20me1.

3.8 Cis-regulation of host genes

If exonic enhancers and silencers cis-regulate expression of their host genes, then manipulation of transcription factors that bind to the regulatory element should alter the expression of the host gene. To test this hypothesis, we searched the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) for studies in which the relevant transcription factor was perturbed and the target gene's expression was measured. We focused on the gene target tubulin alpha 1b (TUBA1B) and the putative silencer S5, located inside one of the coding exons of the gene. Two transcription factors, peroxisome proliferator-activated receptor α (PPARA) and retinoic acid receptor-related orphan receptor α (RORA) have overlapping binding sites within the boundaries of the S5 silencer. Both PPARA and RORA are involved in lipid metabolism and so are both highly expressed in liver [22; 23].

We first analyzed data from a pair of studies in which global gene expression was measured in wild-type and PPARA-null mice after administration of either the PPARA agonist WY1463 [22] or after fasting [24]. Fasting is known to induce PPARA expression in small intestine [24]. To ensure that expression changes due to PPARA induction were specific, we also tested for association between PPARA activation and expression of tubulin beta-4, TUBB4. Tubulin beta belongs to the same protein family as tubulin alpha, but has no known PPARA binding site.

Figure 3-6(A) shows the effects of PPARA activation on TUBA1B expression in small intestine in wild type and PPARA-null mice treated with or without a PPARA agonist. Although there were significant main effects of PPARA-genotype (F = 12.292, DF = 1, p = 0.008) and agonist (F = 16.548, DF = 1, p = 0.004), the interaction of these two factors was not significant (F = 2.8643, DF = 1, p = 0.129). Both main effects appeared to be driven by the decrease in expression in the wild-type/agonist condition (Fig. 3- 6(A)), suggesting that the PPARA-agonist actually affects TUBA1B expression only in wild-type mice. Indeed, post-hoc t-tests suggest that TUBA1B expression is attenuated in wild type mice treated with the PPARA-agonist compared to wild type alone (t = 4.358, DF = 4, p = 0.012), but is not attenuated in PPARA-null mice treated with agonist compared to PPARA-null mice alone (t = 1.583, DF = 4, p = 0.189). In contrast, for expression of tubulin, beta 4 (TUBB4), there were no significant effects of PPARA-genotype (F = 0.066, DF = 1, p = 0.803), PPARA-agonist (F = 0.095, DF = 1, p = 0.766), or their interaction (F = 0.095, DF = 1, p = 0.7663) (Fig. 3-6(B)).

Results from the fasting study were similar (Fig. 3-6(C) and 3-6(D)). Although the fasting effect on TUBA1B expression was significant (F = 19.392, DF = 1, p = 0.002), the effects of PPARA-genotype (F = 3.256, DF = 1, p = 0.109) and the interaction of fasting and genotype were not (F = 1.228, DF = 1, p = 0.3) (Fig. 3-6(C)). Once again, wild-type mice that fasted had lower expression of TUBA1B than mice who did not fast (t = 4.836, DF = 4, p = 0.008), while PPARA-null mice showed no difference when fasting (t = 2.005, DF = 4, p = 0.119). The expression of TUBB4 (Fig. 3-6(D)) did not depend on PPARA genotype (F = 2.52, DF = 1, p = 0.151), fasting (F = 0.727, DF = 1, p = 0.419), or their interaction (F = 4.674, DF = 1, p = 0.063). Together, the results of this pair of studies suggest that PPARA is a repressor of TUBA1B.

To test for association between RORA activity and TUBA1B expression, we used data from a study in which global gene expression was compared in skeletal muscle taken from wildtype mice and mice with a RORA dominant negative mutation. Mean transcript levels of TUBA1B in wildtype and RORA dominant negative mice are shown in Fig. 3-6(E)). TUBA1B was expressed less in RORA dominant negative mice than in wildtype mice (t = 4.5516, DF = 4, p =0.013), suggesting that RORA is an activator of TUBA1B. No data for TUBB4 expression were available from this study [23].

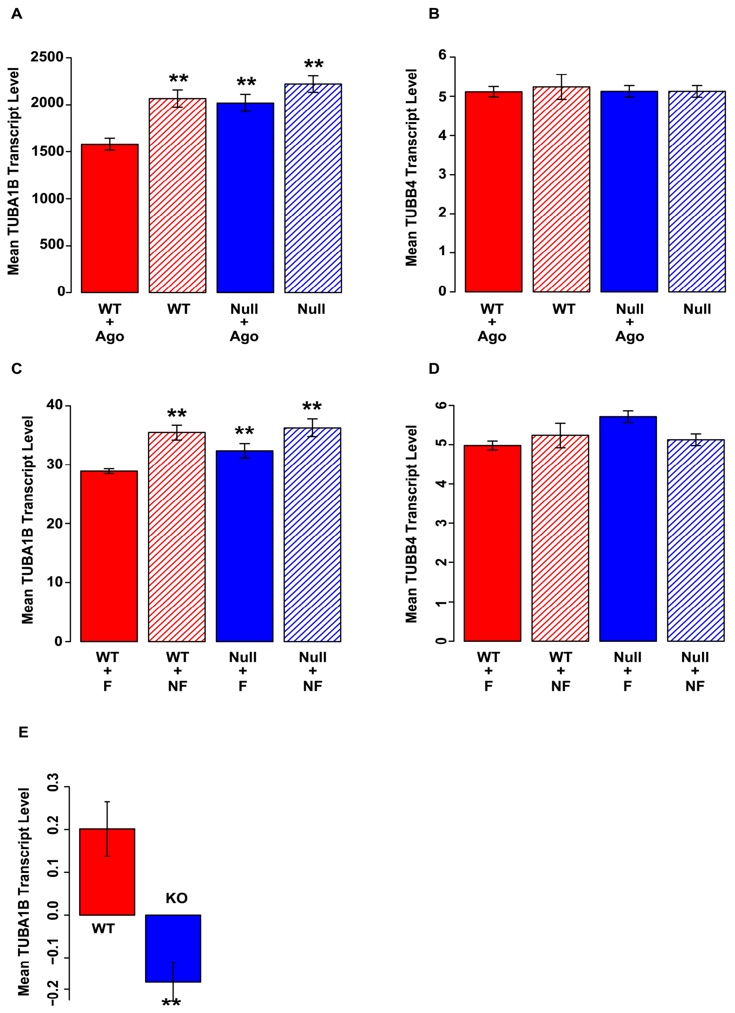

Figure 3-6. PPARA and RORA regulate TUBA1B expression. For all figures, n = 3 per bar. (A) Effects of PPARA genotype (wildtype = red, null = blue) and PPARA agonist WY14643 (present = filled, absent = dashed) on TUBA1B expression in murine small intestine. Error bars, standard error of the mean. (**, p<0.01 compared to wild type + agonist condition). Only PPARA wild-type mice receiving the PPARA agonist show a reduced TUBA1B expression. Data from [22]. (B) PPARA genotype and PPARA agonist do not affect TUBB4 expression. Data from [22]. (C) PPARA genotype and fasting (PPARA activation, fasting = filled, no fasting = dashed) effects on TUBA1B expression in murine small intestine (**, p<0.01 compared to wild type + fasting condition). Only PPARA wildtype mice that fasted show reduced TUBA1B expression. Data from [24]. (D) PPARA genotype and fasting do not affect TUBB4 expression. Data from [24]. (E) RORA activates TUBA1B expression in murine skeletal muscle (**, p<0.01). Data from [23].

We propose that PPARA and RORA compete to bind the TUBA1B regulatory element, wherein PPARA represses TUBA1B when bound, while RORA activates (Fig. 3-7). Consistent with the opposing effects of PPARA and RORA on TUBA1B expression, published reports demonstrate an antagonistic relationship between the two transcription factors. A number of peroxisome proliferated activated receptors, including PPARA and PPARG, as well as orphan nuclear receptors like RORA have highly similar carboxyl terminal extensions in their DNA binding domains that recognize a conserved 5'-extended sequence of some PPAR response elements (PPREs) [25]. PPARs and orphan nuclear receptors compete to bind for overlapping sites, such as those found within the TUBA1B exon. For example, the response element RevDR2 has been shown to mediate repression of its host gene by orphan nuclear receptor Rev-ErbA but activation by PPARA [25]. Coexpression of Rev-ErbA and PPARA inhibits activation by PPARA [25]. Similarly, RORA and PPARG have overlapping binding sites in the PPRE located in the promoter of the perilipin gene. RORA blocks induction of perilipin through PPARG activation [26].

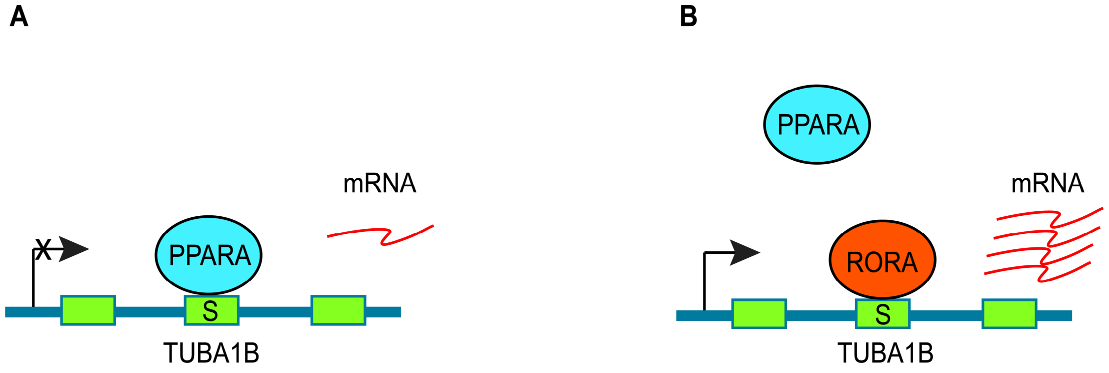

Figure 3-7. Competition model for PPARA and RORA regulation of TUBA1B expression.
(A) TUBA1B expression is repressed by binding of PPARA to putative element S5.
(B) TUBA1B gene expression is activated by displacement of PPARA from S5 by RORA.

Table 3-1. Putative Regulatory Elements

Element[a]	Restriction enzyme	Gene	Region	Start position	Length (bp)
E1	Sau3AI	RPL19	Coding	chr17:37,358,574	34
E2	AluI	TVAS5	Coding	chrM:2,655	83
S1	Sau3AI	FAM161A	Coding	chr2:62,066,752	305
S2	Sau3AI	COL5A2	Coding	chr2:189,904,052	110
S3	Sau3AI	AOX1	3'UTR	chr2:201,536,139	80
S4	Sau3AI	LDHA	3'UTR	chr11:18,429,266	58
S5	AluI	TUBA1B	Coding	chr12:49,523,028	62
S6	Sau3AI	TSPAN3	3'UTR	chr15:77,338,647	237
S7	AluI	RSL1D1	Coding	chr16:11,931,947	26
S8	Sau3AI	MYST2	Coding	chr17:47,869,298	54

[a] Elements labeled “E” are putative enhancers; elements labeled “S” are putative silencers

Table 3-2. Fragment Sequences

Element	Sequence
E1	GATCAGCCCATCTTTGATGAGCTTCCGGATCTGC
E2	CTGTCTCTTACTTTTAACCAGTGAAATTGACCTGCCCGTGAAGAGGCGGGCATGACACAGCAAGACGAGAAGACCCTATGGAG
S1	TCTACTTATGGTTCAACTACCAATGACAAGTTAAAAGAAGAAGAAGCTCTATCGAAACCTTAGGACACAGCTGAGAGCCCAGGAGCATTTACAGAACTCATCTCCTCTGCCTTGTAGGTCAGCTTGCGGATGCAGGAACCCCAGGTGTCCTGAACAGGCTGTAAAGTTGAAGTGTAAACACAAGGTTAGGTGCCCAACTCCTGATTTTGAGGACCTTCCTGAGAGATACCAGAAACACCTCTCAGAACACAAGTCTCCAAAACTCTTAACAGTGTGTAAACCATTTGATCTGCTGATCTGCATCTC
S2	TCAGGCGGCTCCTGATGACAAAAACAAAACGGACCCAGGGGTTCATGCTACCCTGAAGTCACTCAGTAGTCAGATTGAAACCATGCGCAGCCCCGATGGCTCGAAAAAGC
S3	GATCATTTAACATTCTGTGTATGTAACAAAATATCACATGCATAAATATTATGTATCAATAAAATT TTTTAATGGGCAAA
S4	AGATCTTTTTACATTATATGGTAATGTACACTACTGATATAGTTCACAAAATAAGATC
S5	CCCGAGGGCACTACACCATTGGCAAGGAGATCATTGACCTTGTGTTGGACCGAATTCGCAAG
S6	GATCCTACAATCTATTTTAGTCATTTTGTACAGCTGCTATCTTATTGGACTACAGTAAATATTTTTTAAAAGGACACCAATGAGGGGCACCATCTGGTGTTAACCTTAACCAGAAAGCTGGTTTCCTCCTCCTCCCCGCAAAAACCTTTGGCCAAGAGTTCTCCACTGTGAAGACTGAAAGGACCTGGTGACATTTCGGCATCAGTCCTGTTACCACTTGGAGGTAACAGAAGCAGG
S7	AGATTCAAAAACATGCCACAGGAAAG
S8	AGATCTCGAGCACACAGACAGTTCAGAAAGTGATGGCACATCCCGACGATCTGC

Table 3-3. Conservation of Fragments and Neighboring Exons

Element	Mean phastCons score for element	Mean phastCons score for all exons	t	df	p-value	FDR
E1	0.989	0.892	-1.672	5	0.892	0.892
E2	0.820	0.564[a]	N/A	N/A	N/A	N/A
S1	0.211	0.496	2.238	5	0.075	0.134
S2	0.932	0.934	0.155	37	0.878	0.988
S3	0.004	0.705	20.840	34	6.088e-21	5.479e-20
S4	0.631	0.756	1.129	6	0.302	0.453
S5	0.977	0.708	-1.177	2	0.360	0.463
S6	0.540	0.820	3.523	6	0.012	0.028
S7	0.110	0.422	3.362	8	0.009	0.027
S8	0.921	0.979	18.027	9	2.265e-08	1.017e-7

[a] only one exon for this gene

Table 3-4. Transcription Factor Binding Sites Determined By ChIP-seq

E1	HNF4A
E2	CTCF, HNF4A, p300, YY1, ZBTB33
S4	FOSL2, FOXA1, HEY1, HNF4A, JunD, SP2
S5	HEY1

Table 3-5. GC Content of Fragments and Neighboring Exons

Element	GC content of element	GC content of all exons	t	df	p-value	FDR
E1	49.738	54.703	1.816	5	0.129	0.290
E2	49.606	44.158	N/A	N/A	N/A	N/A
S1	44.014	42.991	-0.192	5	0.855	0.855
S2	52.906	55.970	2.609	37	0.0130	0.039
S3	22.362	46.618	21.541	34	2.135e-21	1.921E-20
S4	25.054	47.607	5.283	6	0.001	0.005
S5	50.537	53.546	1.590	2	0.253	0.325
S6	43.102	48.778	1.274	6	0.250	0.375
S7	39.764	41.904	0.764	8	0.467	0.525
S8	51.575	47.625	-1.465	9	0.177	0.319

Table 3-6. Histone Modification Associated with Fragment Sequences[a]

	H2A.Z	H3K4me1	H3K4me2	H3K27ac	H3K27me3	H3K79me2	H3K20me1
E1(RPL19)	other	other	liver	liver	other	liver	liver
E2(TVAS5)[b]	-	-	-	-	-	-	-
S1(FAM161A)	other	-	-	-	other	-	-
S2(COL5A2)	liver	-	-	other	other	other	-
S3(AOX1)	other	-	-	-	other	-	other
S4(LDHA)	-	other	-	other	-	-	liver
S5(TUBA1B)	other	other	liver	other	other	liver	liver
S6(TSPAN3)	-	-	-	-	other	liver	liver
S7(RSL1D1)	-	other	-	other	other	liver	liver
S8(MYST2)	-	other	other	other	liver	liver	liver

[a] "liver" signifies histone modifications associated with fragment in HepG2 cells; "other" signifies histone modifications in cell types other than HepG2

[b] No histone modification data for mitochondrial DNA

Chapter 4 Discussion

4.1 Coverage

From a pool of 1932 random fragments we discovered 10 exonic regulatory elements active in liver within coding regions and 3' UTRs (as well as a non-coding exon of a mitochondrial gene). A previous screen of 1,798 random fragments from a BAC containing both genic and intergenic DNA from the ApoE gene cluster on chromosome 19 also yielded 10 regulatory elements active in liver [5], suggesting that regulatory elements are as common in exons as they are in the genome as a whole. Since we screened a total of ~325 kb of transcribed sequences, and there is a total of ~30 mb of expressed regions, our work suggests there are at least hundreds of exonic regulatory elements for liver cells in the human genome.

In both our present study of exons and our previous study of the chromosome 19 genome region [5], silencers constituted a substantial portion of the uncovered regulatory elements. Since most assays specifically seek enhancers [27; 28], a large number of regulatory elements may well be missed by current approaches.

4.2 Conservation of putative elements

Nine out of ten of the exonic regulatory fragments were conserved across vertebrates, with seven strongly conserved (phastCons score > 0.5). The evolutionary conservation of the regulatory fragment was correlated with conservation of the host gene as a whole, and fragments within coding regions were more conserved than those in 3' UTRs. Others have shown that computationally predicted exonic regulatory elements have lower nucleotide substitution rates than other coding exons within the same host gene, presumably because of dual selective pressure to preserve both protein-coding sequence and TFBSs [7]. However, we found that none of our fragments were more conserved than other protein-coding exons within the host gene, whereas four fragments, two coding and two non-coding, were significantly less conserved. Fragment S6 contains a conserved TFBS so its lower conservation compared to neighboring coding exons is especially unexpected. Perhaps, then, some exonic regulatory elements are in fact released from selective pressure, possibly as a means to allow for transcriptional control while still preserving protein composition.

4.3 CpG islands and GC content

Traditionally, high GC-content has been associated with core promoter sequences, while thus far evidence of association between distal regulatory elements and higher GC-content is scarce [21]. Most exonic regulatory fragments had GC-content higher than the genome as a whole, but much like with conservation, it is difficult to separate whether high GC-content is associated with coding or regulatory function or both. In contrast, we found that no exonic regulatory fragments resided within CpG islands (CGI) and some fragments had lower GC-content than neighboring exons within the same host gene. Currently, it seems GC-content and CGI residence would be best left to predict promoters only.

4.4 Transcription factor binding sites enrichment in putative elements

Six fragments had predicted TFBSs determined by ChiP-seq and comparative genomics. Three fragments had binding sites for HNF4A, hepatocyte nuclear factor 4α. HNF4A is known to be a master regulator of the expression of a wide variety of genes in liver [20], so this finding is unsurprising. Enhancer E2 was found to have five predicted transcription factor binding sites. E2 resides within the single coding exon of mitochondrial gene TVAS5. Unfortunately, no conservation, GC-content, or histone modification data were available for mitochondrial DNA, limiting our ability to both characterize and verify E2. Nevertheless, five predicted TFBSs provide strong evidence that E2 is a true enhancer and may suggest that mitochondrial gene expression is regulated much like nuclear gene expression.

4.5 Regulation of TUBA1B gene

Overlapping conserved TFBSs for PPARA and RORA were predicted within a putative silencer for TUBA1B through comparative genomics. We verified the cis-regulatory potential of this silencer by positively and negatively correlating expression of TUBA1B with RORA and PPARA activation, respectively. Nuclear receptors, like RORA and PPARs, have been shown to have opposing effects on downstream gene expression [25; 26]. One interesting implication of this relationship is that TUBA1B silencer S5 may have been discovered as an enhancer had the complement of TFs in the cell assay been different, for example if RORA were overexpressed relative to PPARA. It is possible many regulatory elements may have bidirectional effects, depending on TF interactions, and what were once known as "enhancers" and "silencers" may be more appropriately called "regulators".

4.6 Histone signatures significance on putative elements

The versatility and difficulty in identifying regulatory elements is reflected in the diversity of histone modifcations associated with them. Most of the signatures previously used to identify enhancers, such as H3K4me1, were also associated with several exonic silencers. Because not all histone modifications were mapped in HepG2 liver cells, we also looked at the chromatin state at the position each of our putative exonic regulatory elements in all other cell types tested for the UCSC Genome Browser ENCODE/Brorad Histone Modification track. Cross-cell type inferences should be made cautiously, as histone modifications at enhancers are known to vary considerably between cell types [14]. Nevertheless, no modification clearly delineated the boundary between regulatory and non-regulatory exon fragments, or between enhancer and silencer. Indeed, histone modifications often correlate with each other, suggesting that rather than individual modifications, modules consisting of many interacting modifications are the true markers of regulatory activity [15].

4.7 Conclusion

In the same way, coding regulatory elements appear to be sequences that are concurrently conserved, enriched in TFBSs and associated with several histone modifications. This likely reflects the biology: a given TFBS motif appears many times in the genome, and only a fraction are likely true binding sites, most likely those which are clustered together, in which access to the TFBSs is permitted by histone modifications and those where the TFBSs are conserved across species. No feature correlates perfectly with regulatory activity, so single-feature based approaches are likely to fail. Integrated approaches have already been successfully used to predict the locations of coding regulatory elements. For example, a search for clusters of TFBS conserved spatially and evolutionary across human, mouse, and rat was used to predict ~700,000 regulatory elements, including an experimentally-verified coding enhancer for the gene ADAM metallopeptidase with thrombospondin type 1 motif, 5 (ADAMTS5) [6].

In sum, the variability of coding regulatory elements requires that genome-wide searches will have to define an appropriately multifaceted signature. The complexity of interactions between DNA, transcription factors and chromatin state should be integrated into tools used to search for the sequences where these interactions occur.

References

[1] G.A. Maston, S.K. Evans, and M.R. Green, Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet 7 (2006) 29-59.

[2] L.A. Lettice, S.J. Heaney, L.A. Purdie, L. Li, P. de Beer, B.A. Oostra, D. Goode, G. Elgar, R.E. Hill, and E. de Graaff, A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12 (2003) 1725-35.

[3] G.G. Loots, M. Kneissel, H. Keller, M. Baptist, J. Chang, N.M. Collette, D. Ovcharenko, I. Plajzer-Frick, and E.M. Rubin, Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease. Genome Res 15 (2005) 928-35.

[4] D. Gabellini, M.R. Green, and R. Tupler, Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell 110 (2002) 339-48.

[5] H.P. Chen, A. Lin, J.S. Bloom, A.H. Khan, C.C. Park, and D.J. Smith, Screening reveals conserved and nonconserved transcriptional regulatory elements including an E3/E4 allele-dependent APOE coding region enhancer. Genomics 92 (2008) 292-300.

[6] K.K. Barthel, and X. Liu, A transcriptional enhancer from the coding region of ADAMTS5. PLoS One 3 (2008) e2184.

[7] X. Dong, P. Navratilova, D. Fredman, O. Drivenes, T.S. Becker, and B. Lenhard, Exonic remnants of whole-genome duplication reveal cis-regulatory function of coding exons. Nucleic Acids Res 38 (2009) 1071-85.

[8] M. Chiquet, U. Mumenthaler, M. Wittwer, W. Jin, and M. Koch, The chick and human collagen alpha1(XII) gene promoter--activity of highly conserved regions around the first exon and in the first intron. Eur J Biochem 257 (1998) 362-71.

[9] A.S. McLellan, T. Kealey, and K. Langlands, An E box in the exon 1 promoter regulates insulin-like growth factor-I expression in differentiating muscle cells. Am J Physiol Cell Physiol 291 (2006) C300-7.

[10] V. Matys, E. Fricke, R. Geffers, E. Gossling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A.E. Kel, O.V. Kel-Margoulis, D.U. Kloos, S. Land, B. Lewicki-Potapov, H. Michael, R. Munch, I. Reuter, S. Rotert, H. Saxel, M. Scheer, S. Thiele, and E. Wingender, TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31 (2003) 374-8.

[11] M. Blanchette, A.R. Bataille, X. Chen, C. Poitras, J. Laganiere, C. Lefebvre, G. Deblois, V. Giguere, V. Ferretti, D. Bergeron, B. Coulombe, and F. Robert, Genome-wide

computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. Genome Res 16 (2006) 656-68.

[12] R.M. Myers, J. Stamatoyannopoulos, M. Snyder, I. Dunham, R.C. Hardison, B.E. Bernstein, T.R. Gingeras, W.J. Kent, E. Birney, B. Wold, and G.E. Crawford, A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol 9 (2011) e1001046.

[13] E. Birney, J.A. Stamatoyannopoulos, A. Dutta, R. Guigo, T.R. Gingeras, E.H. Margulies, Z. Weng, M. Snyder, E.T. Dermitzakis, R.E. Thurman, M.S. Kuehn, C.M. Taylor, S. Neph, C.M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J.A. Greenbaum, R.M. Andrews, P. Flicek, P.J. Boyle, H. Cao, N.P. Carter, G.K. Clelland, S. Davis, N. Day, P. Dhami, S.C. Dillon, M.O. Dorschner, H. Fiegler, P.G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K.D. James, B.E. Johnson, E.M. Johnson, T.T. Frum, E.R. Rosenzweig, N. Karnani, K. Lee, G.C. Lefebvre, P.A. Navas, F. Neri, S.C. Parker, P.J. Sabo, R. Sandstrom, A. Shafer, D. Vetrie, M. Weaver, S. Wilcox, M. Yu, F.S. Collins, J. Dekker, J.D. Lieb, T.D. Tullius, G.E. Crawford, S. Sunyaev, W.S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A. Sandelin, I.L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H.A. Hirsch, E.A. Sekinger, J. Lagarde, J.F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermuller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korbel, O. Emanuelsson, J.S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M.C. Dickson, D.J. Thomas, M.T. Weirauch, J. Gilbert, et al., Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447 (2007) 799-816.

[14] N.D. Heintzman, G.C. Hon, R.D. Hawkins, P. Kheradpour, A. Stark, L.F. Harp, Z. Ye, L.K. Lee, R.K. Stuart, C.W. Ching, K.A. Ching, J.E. Antosiewicz-Bourget, H. Liu, X. Zhang, R.D. Green, V.V. Lobanenkov, R. Stewart, J.A. Thomson, G.E. Crawford, M. Kellis, and B. Ren, Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459 (2009) 108-12.

[15] Z. Wang, C. Zang, J.A. Rosenfeld, D.E. Schones, A. Barski, S. Cuddapah, K. Cui, T.Y. Roh, W. Peng, M.Q. Zhang, and K. Zhao, Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet 40 (2008) 897-903.

[16] A. Siepel, G. Bejerano, J.S. Pedersen, A.S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L.W. Hillier, S. Richards, G.M. Weinstock, R.K. Wilson, R.A. Gibbs, W.J. Kent, W. Miller, and D. Haussler, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15 (2005) 1034-50.

[17] C.T. Ong, and V.G. Corces, Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet 12 (2011) 283-93.

[18] N.S. Shachter, Y. Zhu, A. Walsh, J.L. Breslow, and J.D. Smith, Localization of a liver-specific enhancer in the apolipoprotein E/C-I/C-II gene locus. J Lipid Res 34 (1993) 1699-707.

[19] W.J. Kent, BLAT--the BLAST-like alignment tool. Genome Res 12 (2002) 656-64.

[20] E. Bolotin, H. Liao, T.C. Ta, C. Yang, W. Hwang-Verslues, J.R. Evans, T. Jiang, and F.M. Sladek, Integrated approach for the identification of human hepatocyte nuclear factor 4alpha target genes using protein binding microarrays. Hepatology 51 (2010) 642-53.

[21] D.A. Kleinjan, A. Seawright, A.J. Childs, and V. van Heyningen, Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription. Dev Biol 265 (2004) 462-77.

[22] M. Bunger, H.M. van den Bosch, J. van der Meijde, S. Kersten, G.J. Hooiveld, and M. Muller, Genome-wide analysis of PPARalpha activation in murine small intestine. Physiol Genomics 30 (2007) 192-204.

[23] S. Raichur, R.L. Fitzsimmons, S.A. Myers, M.A. Pearen, P. Lau, N. Eriksson, S.M. Wang, and G.E. Muscat, Identification and validation of the pathways and functions regulated by the orphan nuclear receptor, ROR alpha1, in skeletal muscle. Nucleic Acids Res 38 (2010) 4296-312.

[24] T.C. Leone, C.J. Weinheimer, and D.P. Kelly, A critical role for the peroxisome proliferator-activated receptor alpha (PPARalpha) in the cellular fasting response: the PPARalpha-null mouse as a model of fatty acid oxidation disorders. Proc Natl Acad Sci U S A 96 (1999) 7473-8.

[25] M.H. Hsu, C.N. Palmer, W. Song, K.J. Griffin, and E.F. Johnson, A carboxyl-terminal extension of the zinc finger domain contributes to the specificity and polarity of peroxisome proliferator-activated receptor DNA binding. J Biol Chem 273 (1998) 27988-97.

[26] N. Ohoka, S. Kato, Y. Takahashi, H. Hayashi, and R. Sato, The orphan nuclear receptor RORalpha restrains adipocyte differentiation through a reduction of C/EBPbeta activity and perilipin gene expression. Mol Endocrinol 23 (2009) 759-71.

[27] L.A. Pennacchio, N. Ahituv, A.M. Moses, S. Prabhakar, M.A. Nobrega, M. Shoukry, S. Minovitsky, I. Dubchak, A. Holt, K.D. Lewis, I. Plajzer-Frick, J. Akiyama, S. De Val, V. Afzal, B.L. Black, O. Couronne, M.B. Eisen, A. Visel, and E.M. Rubin, In vivo enhancer analysis of human conserved non-coding sequences. Nature 444 (2006) 499-502.

[28] A. Visel, E.M. Rubin, and L.A. Pennacchio, Genomic views of distant-acting enhancers. Nature 461 (2009) 199-205.

Appendix

Introduction

In addition to my thesis work I participated in other projects not directly related to the thesis but for which my contribution was significant. Involvement in these side projects provided me with knowledge and expertise that were necessary to successfully complete my dissertation. This appendix summarizes the results of some published side projects as well as a project that was not published.

Error-correcting microarray design

Arshad H. Khan, Alex Ossadtchi, Richard M. Leahy, and Desmond J. Smith. Genomics 81 (2003) 157-65.

cDNA microarrays are a powerful technology that allows a researcher to measure thousands of gene expression levels simultaneously and to compare gene expression profiles between different biological samples. Many factors can affect array quality, such as irregularities in DNA spot deposition, efficiency of hybridization and RNA quality. Many of these factors can be resolved by simultaneously hybridizing experimental and control samples labeled with distinct fluorophores to the arrays and taking the ratio of expression intensities. However, missing or damaged spots in the array results in the irretrievable loss of gene expression information. One way to overcome this problem is to print each gene more than once at different positions on the array. However, the number of spots that an array can hold is limited, so printing the entire spectrum of human transcripts in multiple replicates is not feasible. To resolve the issue of missing information from cDNA arrays because of spot drop-out, we used error correcting principles from digital communication to develop a microarray design in which multiplexing of more than one gene onto each spot was employed. Computational decoding of these multiplex spots allowed us to retrieve the expression information for each individual gene even in the presence of corrupted spots.

To evaluate the concept of error correcting codes in microarrays, we first investigated the effect of multiplexing four genes with six spots in a combinatorial fashion and compared the decoded expression intensity to simplex mode (one gene per spot) expression intensity. The genes chosen for this array design were based on already known expression values in kidney relative to brain. To ascertain the sensitivity of the multiplexing scheme and the decoded value of each gene expression, 10 different concentrations (using a serial dilution) of kidney RNA was co-hybridized with a constant concentration of brain RNA. Normalized intensity values of each spot were then decoded to get the intensity for individual genes. Comparison of the intensity of simplex spots with the decoded intensities from multiplexed spots showed a strong correlation. In a subsequent analysis we used the information from four spots out of the six multiplexed spots, dropping information from two of the multiplexed spots on purpose. We were still able to decode the intensity of each individual gene and these intensities were highly similar to the values from simplex spots, indicating the robustness of the error correcting principle for uncovering lost gene expression information.

Further validation of the application of the error correcting principle to microarray design was done using two additional quadruplet gene sets as well as a set of six different genes encoded in multiplex. All three sets showed good agreement with simplex spot intensities. For all three sets, loss of up to one third of the spots still permitted accurate decoding of the intensity of individual genes. For each of the genes tested in the microarray designs, RT-PCR was done to measure the expression levels of each gene. Expression levels from RT-PCR were found to be highly similar to expression values from the decoded intensity level of each gene from the multiplexing schemes. This high replicabilty further indicates the robustness of the error correcting principle applied to microarrays.

In this project, I performed the RNA extraction from brain and kidney, all hybridization experiments, all validation experiments using different sets of genes, as well as the validation using RT-PCR. Analysis of the data was partly performed by me and by Alex Ossadtchi.

A genome-scale map of expression for a mouse brain section obtained using voxelation
Mark H. Chin, Alex B. Geng, **Arshad H. Khan**, Wei-Jun Qian, Vladislav A. Petyuk, Jyl Boline, Shawn Levy, Arthur W. Toga, Richard D. Smith, Richard M. Leahy, and Desmond J. Smith.
Physiol. Genomics 30 (2007) 313-21

Understanding neurological diseases is a daunting task. These disorders are reflected in the corresponding gene expression signatures that exist in the brain. To understand the structure of the brain transcriptome, we investigated the expression of approximately 20000 genes in a coronal slice of the mouse brain using cDNA microarrays. The coronal slice was taken at the level of striatum and 1 mm^3 voxels were generated from the slice. Each voxel was then analyzed for genes expression signatures using the microarrays. Gene expression data from each voxel was employed to reconstruct two dimensional images of gene expression in the brain. Multiple replicates of the gene expression studies from the coronal section were performed and strong replicability was confirmed. Further validation of the gene expression data from the section was confirmed by RT-PCR, mass spectrometry and from publicly available in situ hybridization data.

Using this voxelation approach, we validated known and novel gene expression patterns in the brain. Additionally we identified a set of genes that showed a gradient of dorsal/ventral expression. This study using the voxelation method combined with microarray technology will be a valuable resource to better comprehend neurological disease processes.

My contribution to this project was to dissect and generate voxels from the coronal sections of the mouse brain, extract RNA from each voxel and measure the RNA concentration. Hybridization of RNA samples from each voxel to the cDNA arrays was done using the Vanderbilt core facility. RT-PCR was done partly by me and partly by Mark Chin. Analysis of the data was primarily done by the other authors of this paper.

Fine mapping of regulatory loci for mammalian gene expression using radiation hybrids
Christopher C. Park, Sangtae Ahn, Joshua S. Bloom, Tongtong Wu, Andy Lin, Richard T. Wang, Aswin Sekar, **Arshad H. Khan**, Christine J. Farr, Aldons J. Lusis, Richard M. Leahy, Ken Lange, Desmond J. Smith. Nat Genet 40 (2008) 421-429.

Using expression analysis combined with high resolution genotyping of a large panel of mouse-hamster radiation hybrid cell lines, we mapped regulatory loci for most protein coding genes. The large numbers of breakpoints in the radiation hybrid cell lines and the dense genotyping allowed very sharp mapping (<150 kb) of the regulatory loci. We identified approximately 30,000 trans ceQTLs (copy number expression quantitative trait loci) at a false discovery rate < 0.4. Of the *trans* ceQTLs, 13 of them acted as hotspots, each regulating more than 4,100 genes in *trans*. Additionally, we found that 2,761 *trans* ceQTLs had no known genes associated with them suggesting the importance of gene deserts in regulation. Analysis also revealed that genes on the X chromosome had significantly weaker *cis* ceQTLs than genes on the autosomes, suggesting dosage sensitive autoregulation of X chromosome genes independent of X chromosome inactivation.

My contribution to this project was to grow each hybrid cell line, extract RNA and DNA from each cell line and measure the concentration of RNA and DNA. Also, hybridization of RNA samples from the hybrid cell lines to the microarrays was partly performed by me. Data analysis was primarily done by other authors of this paper.

Screening reveals conserved and nonconserved transcriptional regulatory elements including an E3/E4 allele-dependent APOE coding region enhancer
Hsuan Pu Chen, Andy Lin, Joshua S. Bloom, **Arshad H. Khan**, Christopher C. Park, Desmond J. Smith. Genomics 92 (2008) 292-300.

In this project we screened random DNA fragments from a human BAC (153 kb) containing the APOE gene cluster to search for enhancer and silencer regulatory elements. We identified 14 regulatory elements; 9 enhancers and 5 silencers that were active in liver or astrocyte cells. Two previously known enhancers in the APOE gene cluster regions were also validated. Surprisingly we identified one enhancer element that resided within coding sequence in exon 4 of the APOE gene. This enhancer sequence harbored a single nucleotide polymorphism, the E4 allele is known to be associated with Alzheimer's disease, but not E3. Analysis of the two alleles, showed that the E4 allele had enhancer activity, but E3 did not. This finding may explain the known higher expression level of the APOE E4 allele compared to E3.

Our finding of an enhancer within a coding sequence suggests that there are perhaps several types of transcriptional regulatory elements that share overlapping function with other elements, including coding sequences. This finding, along with other isolated instances of regulatory elements within coding regions described in the literature, prompted me to perform the research described in this dissertation, in which I screened for regulatory elements that are specifically located within coding regions, genome-wide.

My contribution to this paper was to help in the transfection experiments and in performing the assays for reporter gene activity.

Mapping genetic interaction in Drosophila melanogaster using synthetic enhancement genetics

Arshad H. Khan, Andy Lin, John R. Merriam and Desmond J. Smith

Introduction

Genetic interactions underlie the relationship between an organism’s genotype and phenotype (1). However, genetic interaction profiles for various species have been poorly explored to date. In yeast, a systematic deletion approach demonstrated that 80 percent of its genes are not required for viability when tested individually (1). This raises the question of why there is an excess number of genes in an organism above those required for viability. This question can be addressed with the reasoning that genes hardly act alone; rather, their effects depend on their functional relationships with other genes. Determining the genetic interactions between genes in an organism that together influence viability may help answer questions about the nature of this functional co-operation.

Evidence of genetic interaction

Several approaches have been employed to map genetic interaction in eukaryotes. For example, synthetic genetic array (SGA) technology (2) permitted high density arrays of double mutants and enabled researchers to map 30% of all possible synthetic lethal interactions in yeast (1, 3). Similarly, using RNA interference methods to generate mutants and crossing the alleles allowed researchers to investigate the effect of double mutants on survival in C. elegans. A total of 0.03 % of all possible interactions were mapped in C.elegans using this approach (4).

Significance of interactions

Approximately 1000 out of 6000 yeast genes were found to be essential for survival when tested individually suggesting an extensive buffering against genetic perturbation (3). However, recent studies using synthetic enhancement genetics in yeast identified 170,000 synthetic lethal interactions (5). The effects of interactions may explain the apparently limited effects of individual genes on viability. A better understanding of gene-gene interaction networks will thus provide us with deeper insights into normal and abnormal cellular processes.

Aims

Previous studies of genetic interactions in radiation hybrid panels in our laboratory identified over 7 million gene-gene interactions by analyzing the co-retention pattern of two regions of triploid DNA (6). Similar work had also been done in yeast, as described above, allowing researchers to create an interaction map between genes (5). These recent lines of evidence prompted us to evaluate the frequency of interactions between mutations in different genes in Drosophila melanogaster. More specifically, we looked for the effect of gene- gene interactions in the progeny of double mutant crosses. Any deviation in the number of progeny from Mendelian inheritance would suggest the presence of interaction between the two genes. Mutant genes that interact in Drosophila may allow us to infer the results of the corresponding human gene-gene interactions. We hoped to show from this work that our strategy to detect genetic interaction in flies is valid, eventually permitting construction of a gene-gene interaction map for all Drosophila genes.

Experimental design

Two sets of P element mutants on chromosome 2L (16 female) and chromosome 2R (17 male) were crossed (272 crosses) each other (Fig.1) (7-9). Each mutant was homozygous inviable and balanced with the CyO balancer chromosome which contains the Curly wing dominant marker (*Cy*). Approximately 500 progeny from each cross was scored in 4 broods (original, 1st transfer, 2nd transfer and 3rd transfer) and progeny from each brood was collected at four different days (every other day). Scoring consisted of counting the number of wild-type and *Cy* winged flies. The expected genotype and phenotype of the progeny from these crosses are shown in Table 1 and Fig. 1.

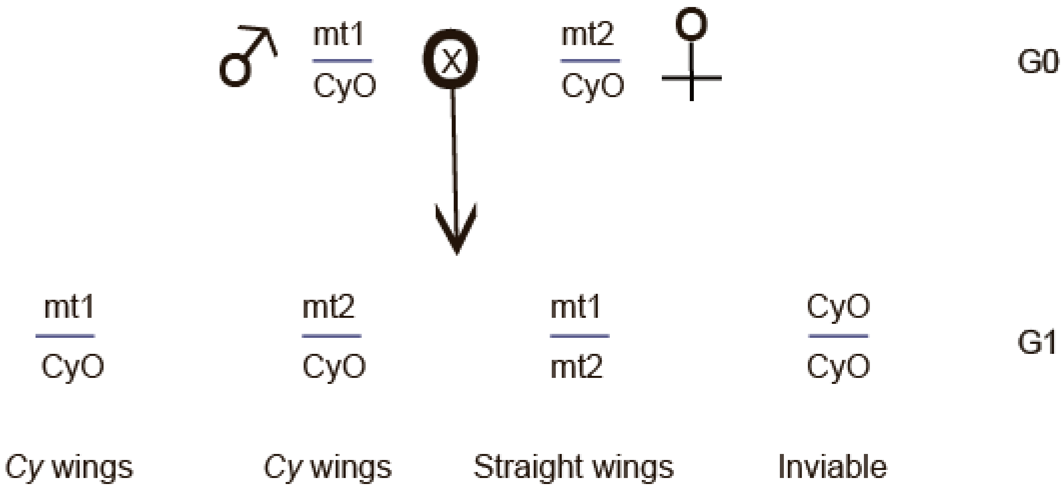

Fig.1. Expected genotypes and phenotypes from double mutant crosses

Table 1. Expected progeny classes

		Gametes	
		mt2	Curly
Gametes	mt1	mt1/mt2 (Straight wing flies)	mt1/Curly (Curly wing flies)
	Curly	mt2/Curly (Curly wing flies)	Curly/Curly (dies)

The curly and straight wing progeny are expected to follow a 2:1 ratio based on Mendelian inheritance. A significant deviation from the 2:1 *Cy*/straight ratio in the progeny from each cross is interpreted as an interaction between the mutants.

Results and discussion

After collecting the data from the initial crosses, the ratio of curly/straight wings flies was calculated for each cross. The Mean ratio of *Cy*:straight progeny from all crosses was 1.98, which followed the expected Mendelian inheritance of 2. The frequency of curly to straight progeny from all crosses showed a normal distribution (Fig. 2), with an extreme outlier at the right likely representing a strong interaction between two mutant genes.

A Chi square test on the data identified twenty two individual crosses (approximately 9 % of the crosses) where the progeny deviated significantly from the expected ratio of 2 (p<0.0035, FDR < 5%) (Fig. 3, Table 2). Of these 22 crosses, four were found to have synergistic effects in which the number of double mutant (straight wing) flies was less than expected. The remaining

18 crosses showed epistatic effects in which the number of double mutant flies was greater than expected.

Table 2. Interaction between mutant genes

Gene name (Female stock)	Gene symbol (Female stock)	Female stock number (Bloomington)	Gene name (Male stock)	Gene symbol (Male stock)	Male stock number (Bloomington)	Interaction type[a]
cropped	crp	10362	*ken and barbie*	ken	10420	Epi
cropped	crp	10362	*walrus*	wal	10447	Epi
cropped	crp	10362	*I(2)06496*	I(2)06496	10450	Epi
cropped	crp	10362	*CG8078*	CG8078	10468	Epi
Cyclin E	CycE	10384	*ken and barbie*	ken	10420	Epi
Cyclin E	CycE	10384	*CG30496*	CG30496	10434	Epi
Star	S	10418	*ken and barbie*	ken	10420	Epi
no mitochondrial derivative	nmd	10435	*overgrown hematopoietic organs 55DE*	oho55DE	10200	Epi
no mitochondrial derivative	nmd	10435	*inscuteable*	insc	10373	Epi
no mitochondrial derivative	nmd	10435	*Sec61β*	Sec61β	10376	Epi
no mitochondrial derivative	nmd	10435	*charlatan*	chn	10380	Epi
no mitochondrial derivative	nmd	10435	*ken and barbie*	ken	10420	Epi
turtle	tutl	10451	*blistered*	bs	10413	Syn
taiman	tai	10453	*charlatan*	chn	10380	Syn
Ribosomal protein S21	RpS21	10457	*overgrown hematopoietic organs 55DE*	oho55DE	10200	Epi
Ribosomal protein S21	RpS21	10457	*Tfb1 i*	Tfb1	10398	Syn
Ribosomal protein S21	RpS21	10457	*ken and barbie*	ken	10420	Epi
Ribosomal protein S21	RpS21	10457		CG8078	10468	Epi
spitz	spi	10462	*ken and barbie*	ken	10420	Epi
V-ATPase 69 kDa subunit 2	Vha68-2	10463	*I(2)06496*	I(2)06496	10450	Epi
CG9302	CG9302	10475	*blistered*	bs	10413	Syn
CG9302	CG9302	10475	*I(2)06496*	I(2)06496	10450	Epi

[a] Synergistic = Syn and Epistatic = Epi

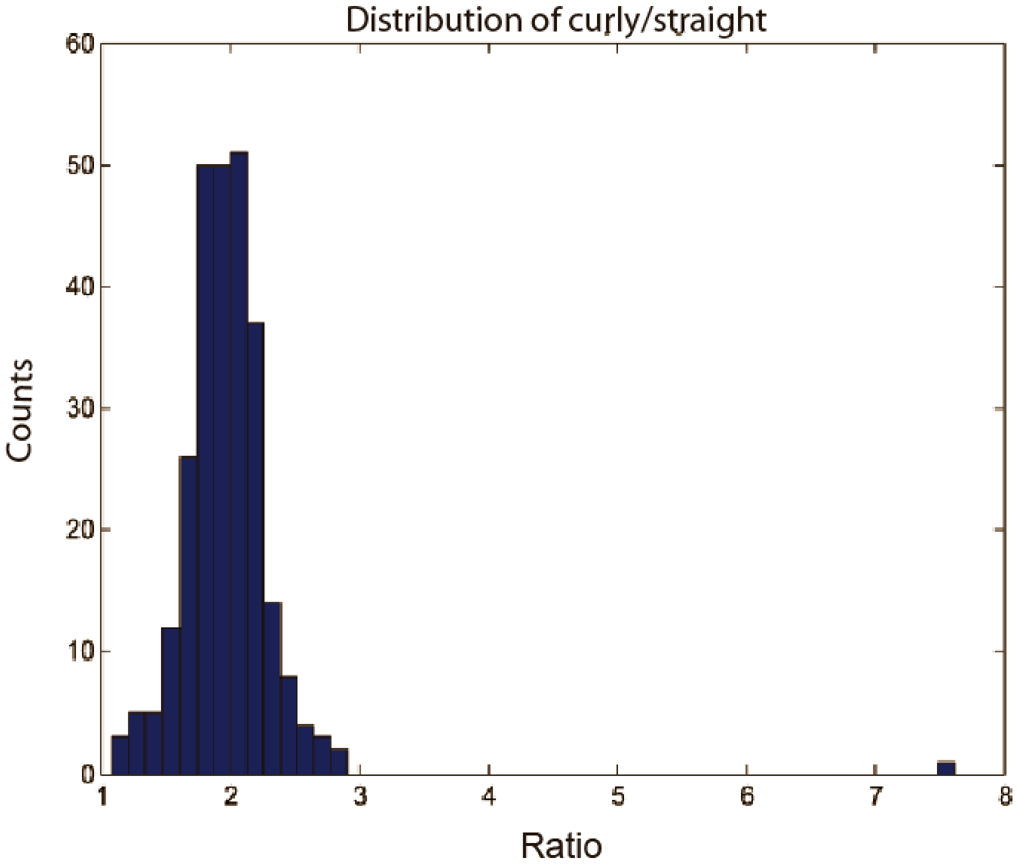

Figure 2. Frequency distribution of curly/straight progeny ratio from 272 crosses

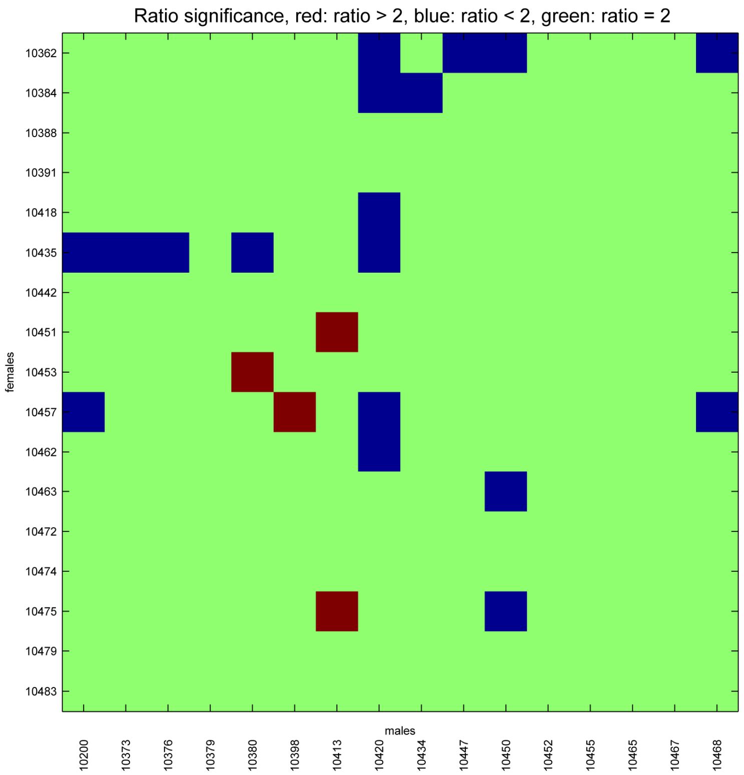

Figure 3. Interaction map of 272 crosses. Blue squares represent significant epistatic interactions. Red square represents synergistic effects. Green square represents no interaction between mutants. Numbers represent Bloomington stock numbers.

To investigate the effects of cross direction and the replicability of the gene-gene interactions, reciprocal (sex reversal) matings were performed for the 22 crosses that showed significant interactions. Very high replicability was obtained ($r = 0.95$ and $p<8.3\times10^{-12}$) (Fig. 4). The direction of the interaction (synergistic or epistatic) was preserved in 20/22 of the reciprocal crosses, suggesting high replicability of the gene-gene interactions and only minor effects of cross direction.

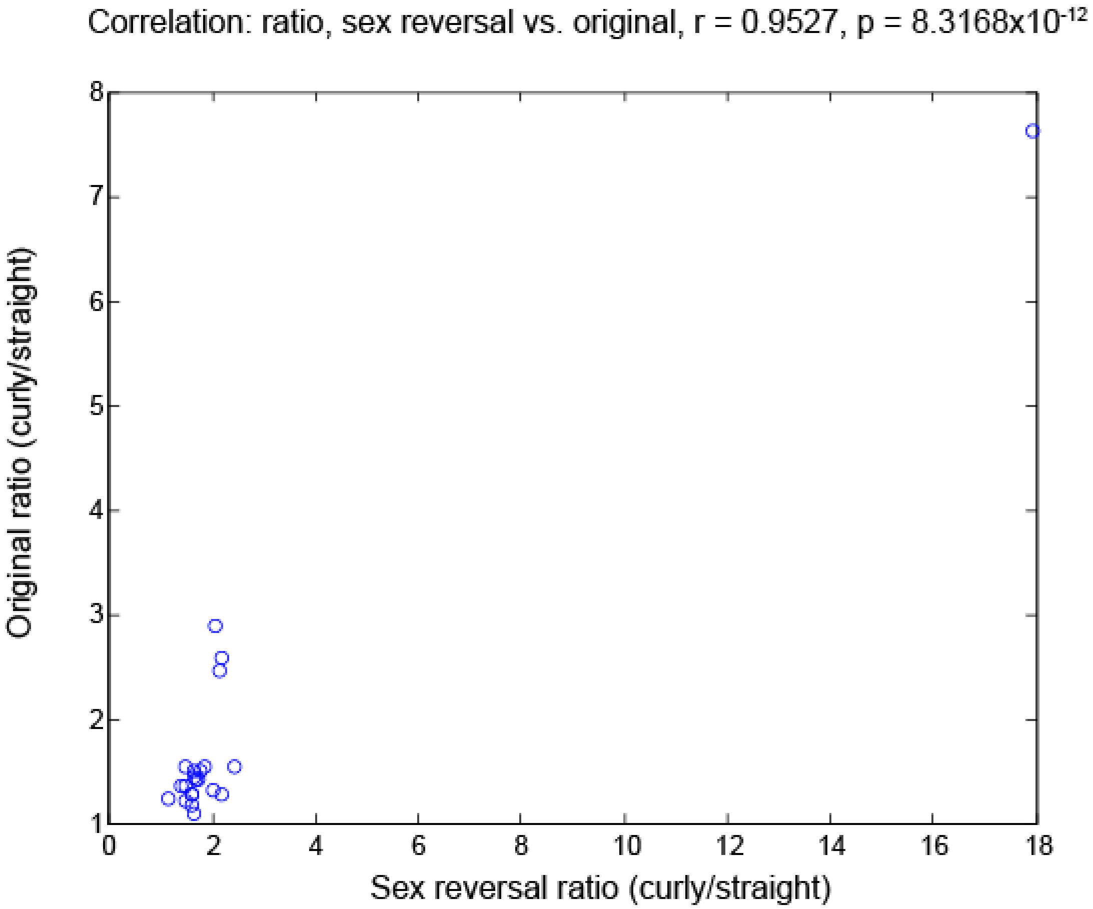

Fig 4. Shows high replicability between replicates (r =0.95 and p<8.3x10^{-12})

We further extended our search to seek interactions in mutant genes that are medically relevant to human disease. To select the fly genes we used homoloGene database (http://www.ncbi.nlm.nih.gov/homologene), GO (Gene Ontology) (http://www.geneontology.org/) and published articles that studied orthologs of genes in flies for diseases such as cancer, Alzheimer's disease, Parkinson's disease, and Huntington's disease (10-13). We obtained each fly stock from the Bloomington Drosophila stock center (http://flystocks.bio.indiana.edu/). We performed similar crosses as described above using 19 medically relevant genes in 171 crosses and sought interactions. From this study we identified 11 statistically significant interactions (P<0.0031, FDR < 5%), 5 synergistic and 6 epistatic (Table 3). The effects of cross direction were tested for these medically relevant mutants as described above. The replicability of the data based on changing the direction of the crosses was preserved (r = 0.93 and $p<3.6x1^{-9}$). The direction of the interaction (synergistic or epistatic) was preserved in 9/11 of the reciprocal crosses.

Table 3. Interaction between medically relevant mutant genes

Gene name (Female stock)	**Gene symbol (Female stock)**	**Female stock number (Bloomington)**	**Gene name (Male stock)**	**Gene symbol (Male stock)**	**Male stock number (Bloomington)**	**Interaction type[a]**
Death caspase-1	Dcp-1	10390	*thickveins*	tkv	11191	Syn
Death caspase-1	Dcp-1	10390	*baboon*	babo	11207	Epi
Suppressor of variegation 2-10	Su(var)2-10	11344	*Cyclin E*	CycE	11396	Epi
numb	numb	11278	*Calcium ATPase at 60A*	Ca-P60A	12389	Syn
Death caspase-1	Dcp-1	10390	*Rho1*	Rho1	12185	Epi
brain tumor	brat	10601	*Posterior sex combs*	Psc	10688	Syn
RNA polymerase II 33kD subunit	RpII33	10575	*thickveins*	tkv	11191	Syn
longitudinals lacking	lola	10946	*thickveins*	tkv	11191	Syn
Src oncogene at 42A	Src42A	10969	*Rho1*	Rho1	12185	Epi
Posterior sex combs	Psc	10688	*dacapo*	dap	11377	Epi
Calmodulin	Cam	10379	*Rho1*	Rho1	12185	Epi

[a] Synergistic = Syn and Epistatic = Epi

If the interaction between two mutant genes is gene specific, then different alleles of the same gene should replicate the interaction. To investigate this issue we crossed the interacting medically relevant genes using independent alleles. Surprisingly, replicability of the interactions with statistical significance was not observed using the independent alleles. This finding suggests that the interactions that we initially observed may be allele specific. Since Drosophila mutants are not, in general, created in isogenic backgrounds, random mutations in other regions of the genome in these strains could also provide an explanation for the change in the interaction behavior. Thus currently available resources may not allow us to map genetic interaction in flies at present. However, similar genetic interaction studies are more feasible in higher organisms such as the mouse, because of the availability of mutant alleles on inbred backgrounds.

In this project, I carried out all the experimental work, including all the double mutant crosses, and examination of progeny fly phenotypes. Andy Lin assisted with statistical analysis of the data. Dr. John Merriam provided advice and guidance on fly genetics. The project idea was conceived by Dr. Desmond Smith.

References:

1 Dixon, S. J. et al. *Annu Rev Genet* 43, 601 (2009).

2 Tong, A. H. et al., *Science* 294 (5550), 2364 (2001).

3 Tong, A. H. et al., *Science* 303 (5659), 808 (2004).

4 Fortunato, A. and Fraser, A. G., *Biosci Rep* 25 (5-6), 299 (2005).

5 Costanzo, M. et al., *Science* 327 (5964), 425.

6 Lin, A. et al., *Genome Res* 20 (8), 1122.

7 Deak, P. et al., *Genetics* 147 (4), 1697 (1997).

8 Newsome, T. P et al., *Development* 127 (4), 851 (2000);

9 Xu, T. and Rubin, G. M., *Development* 117 (4), 1223 (1993).

10 Vidal, M. and Cagan, R. L., *Curr Opin Genet Dev* 16 (1), 10 (2006).

11 Brumby, A.M. and Richardson, H. E., *Nature Reviews Cancer* 5, 626 (2005).

12 Bonner JM. and Boulianne GL., *Exp Gerontol* 2011 46 (5), 335 (2011).

13 Jeibmann A. and Paulus W., *Int J Mol Sci* 2009 February; 10 (2), 407 (2009).