AN ABSTRACT OF THE THESIS OF

Abdulkareem A. AlShaheeb for the Master of Science
(name of student) (degree)

In Biology presented on March 26, 2018

Title: Identification of Unique Regulatory and Coding Regions in Mouse Myosin V Introns: A Comparison of Proteomics- and Genomics-Based Methods

Thesis Chair: Stephen D. Fields

Abstract approved: _____
(Thesis Advisor Signature)

Abstract

Myosins are motor proteins that contribute to many intracellular biological processes in eukaryotic cells, including a major role in transporting various molecules in nerve cells. Cargos delivered by myosin motors contribute to synapse development, a process important in human learning and memory. *Caenorhabditis elegans* uses alternative splicing to code for a motor-less version of myosin V (HUM-2C) that interacts with an RNA binding protein called VIG-1. The objective of this research was to find a homologous version of HUM-2C in mammals. Plasminogen Activator Inhibitor RNA-Binding Protein (PAI-RBP1), a mammalian homologue to VIG-1, was used as a yeast two-hybrid bait to identify an interacting mouse myosin V. A yeast strain carrying PAI-RBP1 was mated to a yeast strain carrying the mouse brain transcriptome library. The mated cells were screened for activation of reporter genes and a protein interaction with

PAI-RBP1. I sequenced 65 randomly selected colonies showing positive interaction. Although no prey gene corresponded to a myosin, I identified some common binding sequences using this method. Multiple sequence alignments were utilized to identify conserved regulatory and coding sequences located in myosin V genes of five mammals, including the common house mouse (*Mus musculus*). Conserved sequence and transcription factor binding sites were identified within one intron of MYO5A. The intronic region may act as a 5' UTR in the expression of a truncated Myo5a protein in mammals. Identifying this version of myosin V in mammalian brain tissue may provide a new target in studying processes that affect neuronal plasticity and learning in humans.

Keywords: PAI-RBP1, Myo5, *C. elegans*, Yeast two-hybrid system, *hum-2C*, MSA, Vig-1, Comparative genomics.

Identification of Unique Regulatory and Coding Regions in Mouse Myosin V Introns: A Comparison of Proteomics- and Genomics-Based Methods

A Thesis Presented to

The Department of Biological Sciences

EMPORIA STATE UNIVERSITY

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Abdulkareem A. AlShaheeb

April 2018

_____

Approved by the Department Chair

_____

Committee Member

_____

Committee Member

_____

Committee Member

_____

Dean of the Graduate School and Distance Education

ACKNOWLEDGEMENTS

This thesis is dedicated to the memory of my beloved grandfather, Abdulkareem, my cousin, Mohammed, and all the Iraqi martyrs.

Foremost, I express my sincere gratitude to my thesis advisor, Dr. Stephen Fields, for his excellent advice, tutelage, and support during every single step of my graduate education. His immense knowledge and enthusiasm encouraged me to work harder towards a strong result. His continuous assistance and editorial comments throughout the writing made my thesis a reality.

Besides my advisor, I thank the rest of my thesis committee members, Dr. David Edds and Dr. Scott S. Crupper, for their patience, insightful comments, and straight-to-the-point questions. Also, thanks to Dr. Tim G. Burnett for providing me the with mice and his guidance thorough my master's program.

Thanks beyond measure to my family as well: to my mother, Bushra, my unlimited inspiration, my father, Ali, my hero and idol, and my siblings, Bassam, Zahraa, and Zainab. They have always motivated me to do whatever it takes in order to achieve my goal. They made me think wisely and positively and helped me venture upon new things and experiences.

Moreover, I thank my Iraqi and international friends for their unconditional support, care, calls, messages, and surprises during my journey towards my master's degree.

As a proud Mesopotamian, I express my profound gratitude to my country for everything, especially the Al-Shuhada'a Establishment for its support with a scholarship that helped me to get this degree.

Unlimited thanks to all the Iraqi heroes who have been protecting my country and its people from ISIS and other terrorist groups until they ultimately defeated them.

PREFACE

The organization of this thesis follows the publication style of Scientific Reports.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF APPENDIX TABLES

Page

LIST OF FIGURES

Page

1. Introduction

The molecular characterization of cytoskeletal elements and motor proteins mediating the rapid and precise positioning of macromolecules and organelles began over 50 years ago with the characterization of the dynein motor in eukaryotic flagella[1]. Genetic and biochemical analyses of cytoskeletal elements and their associated motors continues to this day, with emphases on regulation of gene expression, alternative splicing, and motor-specific cargo transport. In many cases, there is close coordination between various motors as cargos are transported to their destination. Therefore, a brief overview of cytoskeletal-based transport is in order.

Eukaryotic microtubules are composed of α- and β-tubulin subunits polymerized into protofilaments, 13 of which are further assembled into hollow microtubules. Microtubules play important roles in intracellular transport, structural integrity, chromosome segregation during mitosis and meiosis, and cell motility as the major component of flagella and cilia[2]. The kinesin and dynein families of motor proteins use microtubules to transport various cargoes over relatively long distances to specific sites in eukaryotic cells. Kinesins move cargos in an anterograde direction (toward the plus, or growing end, of microtubules) while dyneins move cargoes in the opposite, retrograde direction. In addition, axonemal dynein motor proteins generate the movement of cilia and flagella[1].

Actin microfilaments are composed of globular actin subunits polymerized into 9-nm-thick filaments. These filaments play important roles in muscle contraction, cell shape, general cellular organization, and cell motility through amoeboid movement[3].

Actin filaments also act as tracks for myosin-directed, intracellular transport and positioning of various cellular elements over relatively short distances. The polarity of actin filaments determines the direction of movement of myosin motors as they transport various cargos, with most myosin classes moving towards the plus end of microfilaments[4]. Myosin motors produce movement using energy from adenosine triphosphate (ATP) hydrolysis[4,5].

Phylogenetic alignments of protein sequences of the conserved myosin motor domain from across all eukaryotic taxa have identified 20 myosin classes within this large protein family[6]. Humans have approximately 40 myosin genes, representing 12 of the 20 classes[7, 8]. Myosin classes IV, XI, XII, XIII, XIV, and XVII are found only in lower eukaryotes and fungi, whereas myosin VIII is a plant-specific myosin[7, 9–12]. Myosin II is the muscle myosin that comprises the "thick" filaments that interact with actin and are responsible for muscle contraction. Myosin II is often referred to as "conventional" or "muscle" myosin. The other classes of myosins are categorized as "unconventional" myosins and are not associated with muscle contraction.

There are general structural similarities shared by most unconventional myosin classes, which generally consist of three domains. The head domain serves as the motor, binding to the actin filament and producing force through the hydrolysis of ATP. The neck domain plays a regulatory role, acting as a lever arm for the action of the motor region through amplifying and reversing its activities. The tail domain interacts with various macromolecules, often termed cargos, and is the most variable domain when compared across the myosin family[13, 14].

Myosin classes function in living cells in diverse ways; they serve the purpose of intracellular transport events and cellular activities like cell adhesion, cytokinesis, endocytosis, exocytosis, pigment granule movement, and cell motility and contraction. However, A few other myosins are located within the nucleus. Classes I, II, VI, and Vb contribute to transcription initiation[15–18]. Myosins exist mostly in cytoplasmic regions, are more abundant around actin rich areas, and often are co-localized with various organelles. Some myosins transport specific cargoes such as proteins, mRNAs, vesicles, or organelles, whereas others are rather non-specific and carry a variety of cargoes[19–23]. In general, myosin motor activity is inhibited when the level of calcium is high and enhanced when a cargo binds to the tail or the level of Ca^{2+} is low[24–26].

Myosin V is one of the oldest and most widespread classes of the myosin family[6]. As with most unconventional myosins, myosin V functions as a homodimer through an interaction between the proximal tail domains of two monomers. Because of this dimerization, each functional myosin V protein has two motor domains that work together to "walk" along an actin filament with a step size of 36 nm[23, 27, 28]. The unique C-terminal globular tail domain (GTD) of the myosin V dimer can vary, even within an organism, through tissue-specific alternative splicing[20, 29, 30]. These splice variants allow the protein to bind a diversity of cargos, thereby leading to myosin V's participation in numerous cellular processes, particularly in neurons[7, 31].

Virtually all mammals have three myosin V genes (MYO5A, MYO5B and MYO5C), two of which code for motors that are involved in localizing and recycling subcellular components in neurons. Synaptic vesicles[32], chromaffin granules[33], smooth

ER[34, 35], and endosomal compartments[36, 37] are among the organelles transported by neuronal myosin V. Vertebrate myosin VA also appears to modulate the organization of neurofilaments during neuronal development[38], and myosin V associates with, and may localize, mRNA-binding proteins, including Purα, mStaufen, and FMRP in neuronal processes[39].

The above evidence indicates the significant contribution of myosin V in neuronal development and function. Model organisms ranging from *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Mus musculus* each contain myosin V homologues. The latter three model organisms can be employed in biological laboratories to study the neuronal functions of class V myosins.

Myosin V in C. elegans

Caenorhabditis elegans, the model nematode, has six unconventional myosin classes encoded by the *hum* genes (homologue to unconventional myosin)[7, 40]. The *C. elegans* unconventional myosin genes include *hum-1, hum-2, hum-3, hum-4, hum-5* and *hum-6*, corresponding to myosin classes I, V, VI, VII, IX, and XII, respectively[7].

The Fields Lab has identified three alternatively spliced RNA transcripts derived from the *C. elegans hum-2* (myosin V) locus (Fig. 1). The *hum-2A* and *hum-2B* transcripts code for two slightly different full-length myosin V proteins, called HUM-2A and HUM-2B respectively. The full-length transcripts are expressed in several glandular tissues and in several head neurons and adult tail cells[41]. In addition to these full-length motor variants, a short motor-less myosin V protein, HUM-2C, is also produced by the *hum-2C* transcript[41] (Fig. 1). This alternatively spliced version of myosin V lacks the

typical motor domain and begins with a unique peptide of 51 amino acid residues[41, 42]. The truncated *hum-2C* transcript is controlled by an internal promoter within one of the *hum-2* introns, and its expression is more widespread than that of the full-length transcripts. In fact, *hum-2C* is expressed in most cells of the *C. elegans* nervous system throughout its life cycle[41].

This differential expression provides evidence that the HUM-2A/B motor proteins act independently of the HUM-2C motor-less protein. In addition, disruption of the *hum-2C* gene gives different phenotypic effects than the disruption of the *hum-2A* or *hum-2B* transcripts. RNA interference (RNAi) of *hum-2C* results in locomotion and developmental defects in the worm, whereas a deletion in the *hum-2A/B* gene sequences has no significant effect on locomotion phenotypes[41].

The unique lead peptide of HUM-2C is encoded by an exon that is skipped in the full-length transcripts. This exon codes for a 51-amino acid peptide that does not have significant homology to any non-worm protein. Yeast two-hybrid screens show that this unique motif interacts with a *C. elegans* RNA binding protein called VIG-1[41]. VIG-1 homologues have been shown to play at least a transient role in the RNA-Induced Silencing Complex[43], an important structure in regulating mRNA expression. RNAi of *vig-1* results in a developmental defect similar to that of *hum-2C*-RNAi[41]. The interaction with VIG-1 and the similarities of mutant phenotypes indicate that the short myosin V may play a role in either targeting RNA to specific sites in nerve cells or regulating its expression.

Expressed Sequence Tag (EST) data indicate that a homologous truncated transcript is a conserved feature of the myosin V locus of other rhabditid nematodes, including *Caenorhabditis briggsae* and *C. remanei*[42]. However, there is no empirical evidence for similar, truncated myosin V proteins from higher organisms, including the mammalian proteomes. Given the importance of the truncated myosin V (HUM-2C) in *C. elegans*, it was hypothesized that motorless myosin V proteins play a significant role in mammalian neurons, with potential medical implications.

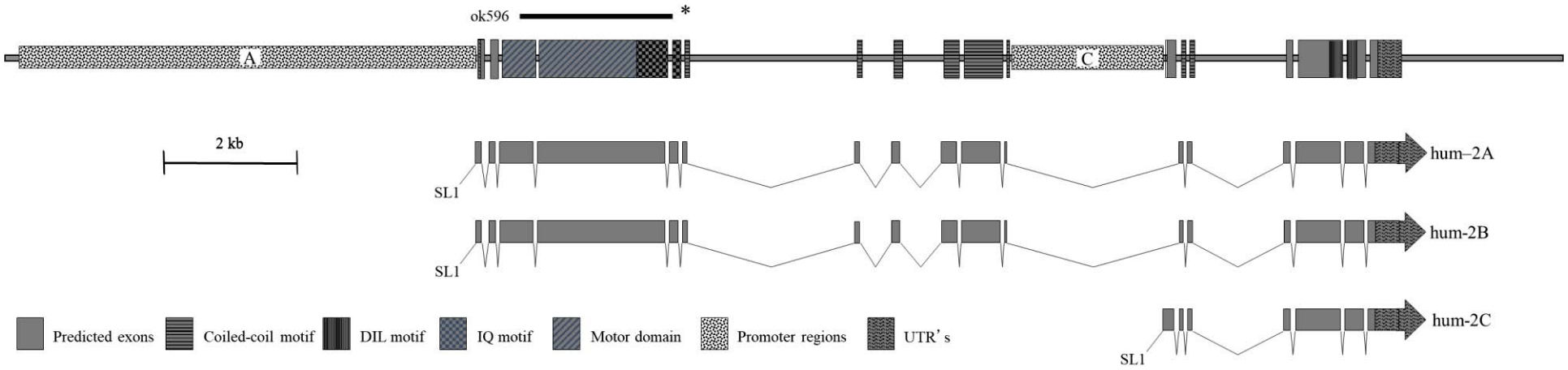

Figure 1. Model of the *C. elegans* myosin V gene, *hum-2*. Based on experimental evidence and predicted structure from the WormBase online biological database[42]. The full-length transcripts *hum-2A* and *hum-2B* encode all of the domains common to other class V myosins, while the truncated *hum-2C* transcript codes for only a portion of the tail. The lead exon of the *hum-2C* transcript codes for a novel peptide. * Region of deletion (ok596 allele).

Experimental Purpose and Design

The central goal of this study was to identify a region within a mouse myosin V gene (either MYO5A, MYO5B or MYO5C) that produces a truncated transcript encoding an orthologue of the *C. elegans* HUM-2C protein. Two different approaches were used to accomplish this objective: a proteomics approach using yeast two-hybrid screens, and a comparative genomics approach using multiple sequence alignments of non-coding regions of mammalian myosin V genes. The effectiveness of each approach was confirmed with RT-PCR to identify the corresponding transcripts in mouse brain tissue.

The yeast two-hybrid (Y2H) screen is a powerful and widely used experimental technique for identifying protein-protein interactions in the context of a cytosolic environment[44–47]. The method involves inserting the gene sequence for a protein of interest into a yeast plasmid that contains an in-frame sequence for the DNA-binding domain of a yeast transcription factor called Gal-4. The resulting fusion protein is referred to as the bait, which is used to search for interacting proteins from a library (pool) of proteins contained on a second yeast plasmid. The library consists of random gene sequences from the organism of interest fused to the gene sequence from the yeast Gal-4 activating domain. These fusion proteins are referred to as prey. In the case of this study, the prey library consisted of all mouse embryo peptides and was purchased from Clontech Laboratories, Inc., whereas the bait was constructed during this study and consisted of a portion of the mouse protein known as plasminogen activator inhibitor 1 RNA binding protein a mammalian homologue to *C. elegans* VIG-1 (see below).

The Y2H screen is designed to detect even weak interactions occurring between the bait and prey fusion proteins when they are co-expressed in a yeast cell[48]. An

interaction between bait and prey brings the DNA-binding and activating domains of the Gal-4 transcription factor in close enough proximity to allow expression of downstream reporter genes contained in the genome of the appropriate yeast strain. Some of the reporter genes are nutritional markers that allow the yeast strain to grow on minimal dropout media. Prey plasmids isolated from these growing yeast colonies should contain nucleotide sequences from the library that code for proteins that interact with the bait protein[49].

In this study, the bait portion of the fusion protein consisted of the carboxy-terminal end of a mouse protein called plasminogen activator inhibitor 1 RNA binding protein (PAI-RBP1). This portion of PAI-RBP1 has homology to the *C. elegans* VIG-1 protein Indicated by high identity (54%) and low E value (8e-16) using NCBI's Basic Local Alignment Search Tools (BLAST). The regressive strategy of these experiments used the *C. elegans* VIG-1 homologue (mouse PAI-RBP1) as a bait to search for interacting mouse proteins, with the underlying hypothesis that a mouse Myo5 peptide would be identified as an interacting partner to PAI-RBP1.

The current study also employed a separate comparative genomics approach to identify important regulatory and coding regions of the mouse myosin V genes. Multiple Sequence Alignment (MSA) of homologous regions from closely to distantly related organisms can identify nucleotide sequences that are under selective pressure to remain conserved. These conserved regions are hypothesized to have either a regulatory function for nearby sequences, or perhaps even a protein-coding function[50]. MSA is a powerful method to identify conserved regions within sequences that may have been mistakenly

determined to be non-coding[51] and is used in parallel with the Y2H system in this study to screen large regions of intronic DNA in search of unique transcripts.

As the results will indicate, there are potential new transcripts that can be identified in the mouse myosin V genes. Whether these transcripts encode proteins that perform functions similar to those of the *C. elegans* HUM-2C protein will take additional empirical evidence.

2. Materials and Methods

Yeast Two-Hybrid

2.1 Brain Extraction

An inbred male mouse (strain C57BL/6J) was obtained from the animal room of the Biological Sciences Department at Emporia State University under The Institutional Animal Care and Use Committee (IACUC) number 15-006 (see Appendix A). The mouse was euthanized using a CO_2 chamber and weighed (Precision™ Balance, Fisher Science Education, United Kingdom, M# AMF 1202).

The mouse was fixed from the limbs with the abdomen down on a foam dissecting board and washed with 95% EtOH (Sigma, USA, Lot#SHBB8633V). The skull was cut with a scalpel along the midsagittal and opened toward the sides by Kelly forceps, and the blood was rinsed away by phosphate-buffered saline (PBS; pH 7.4). The entire brain was extracted as one piece and transferred directly into a 95-mm petri dish containing PBS. The brain was weighed (Denver Instrument, USA, M# P-114) then chopped into four quadrants of less than 0.5 cm thickness, each weighing approximately 130 mg. Each quadrant was stored in a separate 1.5 ml microcentrifuge tube (Eppendorf) containing RNA*later* RNA Stabilization Reagent (QIAGEN, Germany, Lot# 151052596) and stored at -20°C.

2.2 RNA Isolation

Total brain RNA was isolated from 25 mg of tissue (from section 2.1), using the RNeasy Plus Mini Kit (QIAGEN, Germany, Lot# 154016964) and following the manufacturer's protocol.

2.3 Complementary DNA (cDNA) Generation

Reverse transcription polymerase chain reaction (RT-PCR) was used to synthesize the PAI-RBP1-cDNA first strand using 163 ng of isolated brain RNA as template, the PAI-RBP1-cDNA primer, and the M-MLV Reverse Transcriptase kit (Sigma, USA, Lot# 039K0791), according to the manufacturer's instructions (see Appendix C). The first strand synthesis was carried out on the Applied Biosystems™ 2720 Thermal Cycler (Thermo Fisher, Singapore). An RNA*later* RNA Stabilization Reagent was used as RNase inhibitor for the M-MLV Reverse Transcriptase kit.

2.4 PAI-RBP1 Amplification (Bait)

The first strand PAI-RBP1 cDNA (from section 2.3) was amplified with PCR using the PAI-RBP1-OR and PAI-RBP1-OF primers and the PAI-RBP1 PCR cycle (see Appendix C). Amplimers were visualized by gel electrophoresis on the ENDURO™ Gel XL Electrophoresis System (Labnet, USA, M# E0160) using a 1% agarose gel with 0.5 µg/ml ethidium bromide (EtBr) (Sigma, USA, Lot# SLBF7130V)

The product from the PAI-RBP1 amplification process (section 2.4) was purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Germany, Lot# 1608/004) following the manufacturer's protocol.

2.5 pGBKT7 Vector Isolation

The following plasmid miniprep protocol was used to isolate plasmids from *E. coli*. A single colony of *E. coli* carrying plasmid was used to inoculate 5 ml LB_{Kan} or LB_{Amp} medium and incubated overnight at 37°C at 250 RPM on an incubating rotary shaker (Benchmark Scientific, Inc., USA, INCU-SHAKER™ 10L). One and a half ml of

the overnight culture was pipetted into a 1.5-ml microfuge tube and the tube was microcentrifuged at 14,000 RPM for 1 minute. After discarding the supernatant, the pellet was resuspended in 100 µl of Resuspension Buffer [50 mM Tris-Cl at pH 8.0, 10 mM EDTA, 100 mg/ml RNase A. One µl of 7 u/µl RNase A (Sigma, USA, Lot# SLBL6452V)] and 200 µl of lysis buffer (200 mM NaOH, 1% SDS) were added to the tube. The mixture was inverted five times and placed on ice for 5 minutes. The lysate was neutralized with 150 µl of Neutralization Buffer (3.0 M potassium acetate, pH 5.5), inverted five times, and placed on ice for five minutes. The tube was microcentrifuged at 14,000 RPM at 4°C for 5 minutes and 400 µl of the supernatant pipetted into a clean microfuge tube. DNA precipitation was facilitated with 120 µl of 5 M NaCl and 800 µl of cold 100% EtOH. The tube was inverted 5 times, microcentrifuged at 14,000 RPM at 4°C for 5 minutes, and the supernatant decanted. The DNA pellet was washed with 750 µl of room temp 70% EtOH, and the tube was centrifuged at 14,000 RPM at 4°C for 5 minutes. The supernatant was aspirated and the tube was kept open for 5 minutes to dry the pellet at room temperature. The pellet was dissolved with 50 µl of dH_2O. The concentration of the plasmid was measured with a NanoDrop spectrophotometer (Thermo Scientific).

2.6 PAI-RBP1 Cloning

The In-Fusion® HD Cloning Kit (Clontech Laboratories, Inc., USA, Lot# 1603367A) was used to insert the PAI-RBP1 cDNA fragment into the vector by following the manufacture's protocol. Briefly, the isolated pGBKT7 vector (from section 2.5) at 275.7 ng/µl was cut with BamHI-HF (New England BioLabs Inc., USA, Lot# 0101608) and EcoRI-HF (New England BioLabs Inc., USA, Lot# 0131610) restriction enzymes to cut and prepare it to be recombined with the PAI-RBP1 insert (from section

2.4) by following the kit manufactures' protocols. The purified bait was mixed with the linearized plasmid at 275.7 µg/µl and ligated for 2 hours to generate the recombinant bait which was named BAITX.

E. coli was transformed with the recombinant plasmid by mixing 2 µl of the BAITX in-fusion mix with 50 µl of competent *E. coli* (DH5α; New England BioLabs Inc., USA, Lot# 3151703). After incubating on ice for 30 minutes, the cells were heat shocked for 30 secs at 42°C and incubated on ice for 5 minutes. SOC outgrowth medium (950 µl) at 37°C (New England BioLabs Inc., USA, Lot# 3151703) was added to the cells and then incubated at 37°C on a rotatory shaker incubator (Benchmark Scientific, Inc., USA, INCU-SHAKER™ 10L) for 1 hour at 250 RPM. Cells were spread on LB_{kan} agar plates and incubated overnight at 37°C.

2.7 *E. coli* Colony PCR

Transformants were screened for BAITX using colony PCR. Briefly, individual colonies were picked into the *E. coli* colony PCR mixture #1 containing T7 and pGBK-R primers (see Appendix C). After amplifying with the *E. coli* colony PCR cycle #1 (see Appendix C), the PCR products were visualized on a 1% agarose gel with 0.5 µg/ml EtBr. Plasmids with the predicted insert size were isolated as in section 2.5, and Sanger Sequencing of the bait vector was performed by Eurofins Genomics to verify the plasmid sequence using T7 and pGBK-R as forward and reverse primers, respectively (see Appendix C). A BAITX isolate containing the correct sequence was selected to be transformed into the Y2HGold *Saccharomyces cerevisiae* strain with the *MATa* mating type (Clontech Laboratories, see Appendix A for genotype).

2.8 Yeast Transformation with BAITX (Bait + pGBKT7)

A single colony of Y2HGold was used to inoculate 100 ml of YPDA medium and incubated overnight at 30°C at 180 RPM. The overnight culture was divided into two 50-ml tubes, centrifuged at 700 g at room temperature for 5 minutes using a benchtop centrifuge (Hettich, Germany, M# rotofix 32a). After decanting the supernatant, the pellet was resuspended in 30 ml of distilled H_2O, centrifuged at 700 g at room temperature for 5 minutes, and the supernatant was decanted. The pelleted *S. cerevisiae* cells were made competent by resuspension in 1.5 ml of 1.1xTE/LiAc (Clontech Laboratories, Inc., USA, Lot# 1603053A); 800 µl of the suspension was transferred into a new microfuge tube. The mixture was microfuged at 17,000 RPM for 15 seconds, the supernatant was decanted, and the pellet resuspended in 600 µl of 1.1xTE/LiAc.

The transformation mix consisted of 5 µl of 10 µg/µl salmon sperm DNA (Clontech Laboratories, Inc., USA, Lot# 1603053A), 1 µl of BAITX (2217.3 ng/µl), 50 µl of the competent yeast cells (Clontech Laboratories, Inc., USA, Lot# 1603053A), and 500 µl of PEG/LiAc (Clontech Laboratories, Inc., USA, Lot# 1603053A). The transformation mix was incubated at 30°C for 30 minutes with tube inversion each 10 minutes. A 20-µl aliquot of dimethyl sulfoxide (DMSO; Sigma, Lot# SHBF7359V) was added to the tube, gently mixed, and incubated in a water bath (Thermo Fisher, Germany, M# Isotemp 210) at 42°C for 15 minutes with mixing every 5 minutes. The tube was centrifuged at 17,000 RPM for 15 secs and the supernatant was decanted. The pellet was resuspended in 1 ml of YPD-plus medium (Clontech Laboratories, Inc., USA, Lot# 1603053A), mixed, and incubated at 30°C for 25 minutes. The tube was centrifuged at 17,000 RPM for 15 secs, the supernatant was discarded, and the pellet was resuspended

in 1 ml of 0.9% (w/v) NaCl solution. To determine transformation efficiency, a 1/10 and 1/100 dilution was made from the resuspended cultures and 100 µl of each dilution was spread onto SD/-Trp and incubated at 30°C for 3 days. The resulting transformation strain was named Y2HGold(BAITX).

2.9 Autoactivation

An autoactivation test was performed as below to ensure that BAITX did not activate reporter gene expression on its own.

2.9.1 pGADT7 Vector Isolation and Yeast Transformation

The empty pGADT7 library plasmid was isolated from *E. coli* (Clontech Laboratories, Inc., USA, Lot# 1603053A) according to section 2.5. Transformation of the *MATα* mating type Y187 yeast strain (Clontech Laboratories, see Appendix A for genotype) with the pGADT7 vector was carried out according to section 2.8, but SD/-Leu plates were used to select transformants. The resulting transformation strain was named Y187 (pGADT7).

2.9.2 Yeast Mating

A single colony of Y2HGold (BAITX) (section 2.8) was used to inoculate 50 ml of SD/-Trp liquid medium and a single colony of Y187 (pGADT7) (section 2.9.1) was used to inoculate 50 ml of SD/-Leu liquid medium. The flasks were incubated at 250 RPM at 30°C until the optical density (OD_{600}) reached 0.81 and 0.64, respectively.

The cultures were centrifuged at 1000 RPM for 5 minutes. The Y2HGold (BAITX) pellet was resuspended in 5 ml of SD/-Trp while the Y187 (pGADT7) pellet was resuspended in 5 ml of SD/-Leu. A mixture of 2 ml of Y2HGold (BAITX), 2 ml of

Y187 (pGADT7), and 10 ml of 2xYPDA was assembled in two 125-ml flasks. The flask was incubated at 30°C on a rotatory shaker at 45 RPM for 20 hours.

2.9.3 Plate Seeding

After 20 hours, yeast mating was confirmed by microscopic examination. The cell culture was transferred to a 50-ml tube, centrifuged at 1000 RPM for 12 minutes, and the supernatant discarded. The pelleted cells were resuspended in 10 ml of 0.5xYPDA and centrifuged at 1000 RPM for 12 minutes, and the supernatant was discarded. The pellet was resuspended in 5 ml of 0.5xYPDA. Serial dilutions of 10^{-1} to 10^{-4} were made from the resuspended cultures and 100 µl of each dilution was plated on SD/-Trp-Leu plates and incubated at 30°C for 3 days to calculate mating efficiency. Single colonies from each plate were also streaked on SD/-Ade/-His/-Leu/-Trp agar plates and incubated at 30°C to test for autoactivation.

2.10 Toxicity test

Y2HGold competent cells were transformed with BAITX at 2217.3 ng/µl (from section 2.5) following the same method as section 2.9, and yeast cells (from section 2.9) were used in this test. A 1/10 and 1/100 serial dilution was made from the transformed yeast cells and 100 µl of each dilution was spread onto SD/-Trp and incubated at 30°C for 3 days.

2.11 Yeast Strain Mating: pGBKT7(BAITX) with pGADT7(Library)

A mouse cDNA library prepared from 17-day old mouse brains and cloned into the pGADT7 yeast two-hybrid vector, was purchased from Clontech. The frozen library

was contained in the Y187 yeast strain [Y187(LIBRARY)] and stored at -80°C until use. The yeast two-hybrid BAITX screen against the library was performed as follows:

2.11.1 Culture Preparation

Fifty ml of SD/-Trp broth medium in a 250 ml Erlenmeyer flask was inoculated with a fresh colony of pGBKT7(BAITX) (section 2.8) and incubated overnight at 30°C at 250 RPM. The resulting OD_{600} of 0.881 was determined by Nanodrop spectrophotometer (Thermo Scientific, Germany, M# NanoDrop 2000C, SW# NANODROP 2000/2000c V# 1.5) using a 1-cm cuvette (Thermo Fisher, Germany, Lot# 14385942). The cell culture was transferred to 50 ml tubes, centrifuged for 5 minutes at 700g using the Hettich benchtop centrifuge. The supernatant was discarded, and the pellet was resuspended in 5 ml of SD/-Trp medium. Serial dilutions of 1/10, 1/100, 1/1000, and 1/10000 were performed from the resuspended culture and the cell density was measured using a hemocytometer (Hausser Scientific, USA, M# Bright-Line 3110) and compound microscope (Olympus, Philippines, M# CX31RBSFA).

2.11.2 Library Titering

Dilution A was assembled by transferring 10 µl Mate & Plate Library (Clontech Laboratories, Inc., USA, Lot# 1603053A) to 1 ml of YPDA broth in a 1.5-ml microfuge tube and mixed by gentle vortexing. Dilution B was prepared by transferring 10 µl library aliquot from Dilution A to 1 ml of YPDA broth in a 1.5-ml microfuge tube and mixed by gentle vortexing. A mixture of 10 µl of dilution A and 50 µl YPDA broth in a 1.5-ml microcentrifuge was assembled, mixed by gentle vortexing, and spread onto an SD/-Leu

plate. Fifty µl of dilution B was spread onto an SD/-Leu plate and the plates were incubated at 30°C for 3 days.

2.11.3 BAITX and Library Mating

A mating mixture was assembled in a sterile 2-l flask by adding the following: 1 ml of the Mate & Plate Library cells (Clontech Laboratories, Inc., USA, Lot# 1603053A), 4 ml of Y2HGold (BAITX) (from section 2.12.1), and 45 ml of 2xYPDA with 50 µg/ml kanamycin (Sigma, China, Lot# SLBH9905V). The library vial was rinsed twice with 1 ml of 2xYPDA and added to the 2-l flask, and the mating mixture was incubated in a rotatory shaker (Benchmark Scientific, Inc., USA, INCU-SHAKER™ 10L) at 30°C at 45 RPM for 23 hours.

2.11.4 Plating Mated Yeast

A drop from the mating flask (section 2.12.3) was observed with a compound microscope (Olympus, Philippines, M# CX31RBSFA) to confirm that mating had occurred. The cell culture was transferred to a 50-ml tube, centrifuged at 1000 RPM for 12 minutes, and the supernatant discarded. The 2-l flask was rinsed twice with 50 ml of 0.5x YPDA containing 50 µg/ml kanamycin (Sigma, China, Lot# SLBH9905V). The pelleted cells were rinsed twice in dH_2O by resuspending and centrifuging at 1000 RPM for 12 minutes. The final pellet was resuspended in 10 ml of 0.5x YPDA with 50 µg/ml kanamycin, and the total volume was measured. Serial dilutions of 10^{-1} to 10^{-4} were made from the resuspended cultures and 100 µl of each dilution plated onto SD/-Trp and SD/-Leu plates. All of the remaining cell suspension was spread onto 100 mm SD/-Ade/-His/-Leu/-Trp plates (125 µl per plate). The plates were incubated at 30°C for 3 days.

2.12 Yeast Colony PCR

Random individual colonies growing on SD/-Ade/-His/-Leu/-Trp were selected and picked for colony PCR. The reaction mixture was prepared as a yeast colony PCR mixture using T7 and pGAD-R primers as forward and reverse primers, respectively, and run with the yeast colony PCR cycle (see Appendix C). Amplimers were visualized on a 1% agarose gel with 0.5 µg/ml EtBr.

2.13 Sequencing Prey Plasmids

Sanger sequencing of prey inserts was performed by Eurofins Genomics using T7 and pGAD-R as forward and reverse primers, respectively (see Appendix C). The SimpleSeq^{TM} Kit (Eurofins Genomics, USA) was used to document and send samples to the sequencing facility (Eurofins Genomics LLC, USA). All sequencing primers were diluted to 10 µM and mixed with the template according to manufacturer's guidelines.

2.14 Data Analysis

The sequence data obtained from the Y2H screen (section 2.13) were analyzed using the Translate tool by SIB ExPASy Bioformatics Resources Portal to translate the nucleotide data into protein in-frame from the start codon on the pGADT7 vector. BLAST by NCBI was used to compare both the nucleotide and translated data against the online database. The web-based Clustal Omega Multiple Sequence Alignment tool by European Bioinformatics Institute (EMBL-EBI) was used for alignments of nucleotide and amino acid sequences to document any consensus sequences that might be common.

Comparative Genomics with Multiple Sequence Alignment and Analysis

MYO5A, B, and C gene sequences for five mammalian species (*Bos taurus, Canis lupus, Homo sapiens, Macaca mulatta,* and *Mus musculus*) were obtained as FASTA sequences from NCBI (see Appendix A, Table A4). The 10^{th}-30^{th} exon sequences of each gene were identified and highlighted to identify and align the homologous introns of the five species. Alignments of the introns were carried out with the Clustal Omega Multiple Sequence Alignment tool by EMBL-EBI. The alignments with highest similarity within the introns between the homologous exons were identified by counting the semi-continuous identical bases using features of Microsoft Excel, Microsoft Word, and Notepad. Briefly, each aligned series of sequences with their asterisks (*) indicating the same nucleotide for all species (hits) were exported to MS Word and given line numbers. Everything was then deleted except the line numbers and the row of asterisks. This file was exported to Notepad where the all asterisks were replaced by the number 1 and saved as a flatfile. MS Excel was used to import the data from the Notepad file as a table, the number of asterisks was summed for each line; any alignment without a score was eliminated so the alignments could be sorted in descending order to count the highest scores and document them.

Primers were designed for the mouse sequence in the area with the highest identity score, which was located between exon 18 and 19 (95,568 - 95,802 bases),: intronic region 18 of the myosin Va gene. RT-PCR was used to synthesize the first strand using 163 ng of isolated brain RNA as template, the MMyo5A-R1 primer, and the M-MLV Reverse Transcriptase kit (Sigma, USA, Lot# 039K0791), according to the manufacturer's instructions (see Appendix C). Nested-PCR reactions were performed to

amplify the targeted fragments using PCR mixtures #1 and #2 for sample A, #3 and #4 for sample B, and #5 and #6 for the control and PCR cycles #1 and #2 for sample A, #3 and #4 for sample B, and #5 and #6 for the control, respectively (see Appendix C).

The products were purified using the ChargeSwitch® PCR clean-up kit (Invitrogen, USA, Lot# 1878522) and 6-Tube Magnetic Separation Rack (New England BioLabs Inc., USA, Lot# S1506S) then cloned into the pMiniT™ 2.0 vector included with the NEB® PCR Cloning Kit (New England BioLabs Inc., USA, Lot# 12025) following the manufacturers' protocols. The resulting *E. coli* transformants were screened using colony PCR. Briefly, individual colonies were picked into *E. coli* colony PCR mixture #2 containing the pMiniT forward and reverse primers and amplified with the *E. coli* colony PCR cycle #2 (see Appendix C). Amplimers were visualized on a 2% agarose gel with 10 µg/ml EtBr using gel electrophoresis. Isolated purified plasmids were sequenced by Eurofins Genomics and the sequence aligned against the mouse genome using BLASTn (NCBI).

3. Results

Yeast Two-Hybrid System

The first-strand DNA synthesis and amplification of PAI-RBP1 from brain tissue RNA yielded a single band of the expected 500 bp size (Fig. 2), which was ligated into the pGBKT7 plasmid and transformed into *E. coli*. Colony PCR products of 50 transformed *E. coli* cells with the bait showed that 20% of them were carrying the recombinant plasmid (Fig. 3). Sequence of the purified PCR products showed two of the colonies contained a full-length PAI-RBP1 amplicon inserted in frame with the plasmid start codon without any errors (Fig. 4). One recombinant vector was chosen and named BAITX.

Transformation of the yeast strain Y2HGold with BAITX and the empty pGADT7 vector resulted in colonies on SD/-Trp/-Leu medium, but not on SD/-Trp/-Leu/-His/-Ade medium, indicating that BAITX does not activate reporter expression on its own. A comparison of colonies carrying BAITX with colonies carrying the empty pGBKT7 vector showed similar-sized colonies for both strains, indicating the PAI-RBP1 insert is not cytotoxic and could be used for library screening.

After mating and plating the bait strain with the library strain, approximately 10^4 colonies grew on the SD -Trp/-Leu/-His/-Ade media. A total of 210 yeast colonies were randomly selected from these plates and the library plasmid insert amplified by colony PCR using primers specific for the library plasmid. Approximately 30% of the colony PCR reactions yielded bright, single bands (Fig. 5) that could be sequenced using the forward primer for pGADT7. A total of 65 inserts were sequenced and then translated in-frame from the vector start codon (Table 1). Nucleotide sequences were identified with

the NCBI Blastn program, and the protein products were identified with the NCBI Blastp program. As expected, all of the sequenced nucleotide sequences corresponded to mouse DNA sequences. Translation of the nucleotide sequences showed that 57% of the total inserts were cloned in-frame, yielding the mouse protein that corresponded to the cloned nucleotide sequence. The rest of the inserts were out-of-frame with the start codon, resulting in anomalous protein products. These either had more similarity to other organisms (10%) or had no database hits/hypothetical proteins (33%). Twenty three percent of the sequences were identical or partially overlapped within the pool of the prey hits (Table 1).

None of the sequenced prey plasmids contained a nucleotide sequence with any homology to a myosin sequence (either coding or non-coding). Multiple sequence alignments of the translated proteins and peptides, whether anomalous or actual mouse proteins, did not identify any consensus sequences that unified the rather diverse assemblage of prey proteins. Furthermore, the sequences that were represented multiple times in the pool of sequenced prey did not seem to have significant sequence or structural similarity to any motif in a myosin sequence.

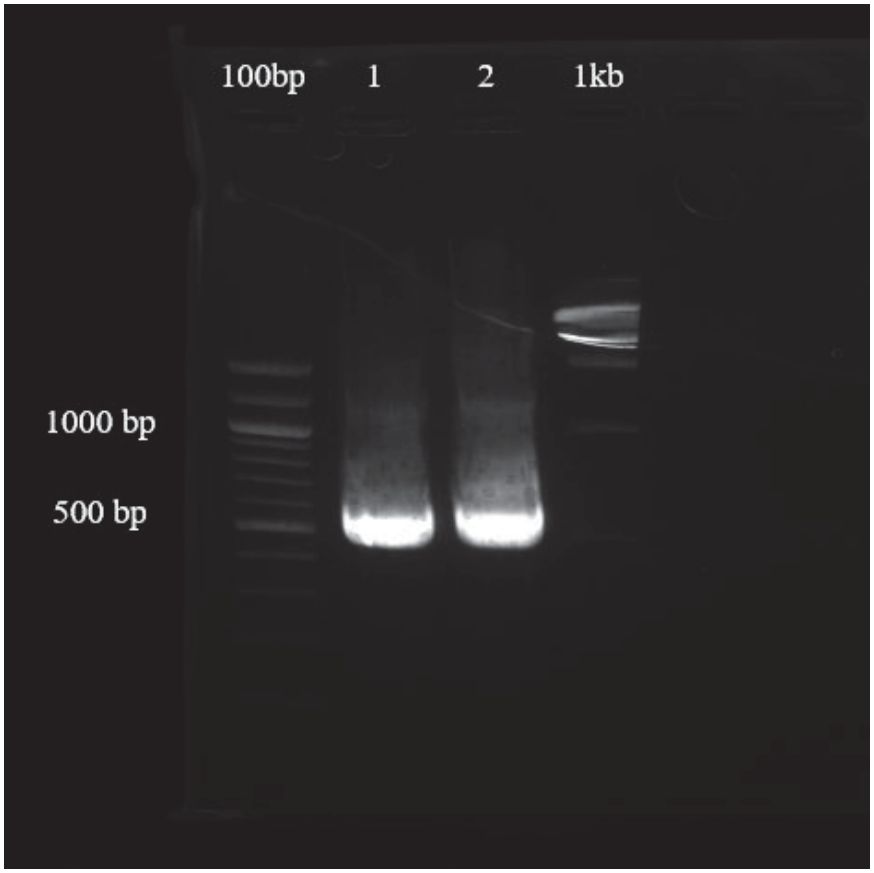

Figure 2. Gel image of PCR product from mouse brain RNA synthesized by using PAI-RBP1-OF and PAI-RBP1-OR primers and PAI-RBP1-cDNA as the first strand cDNA.

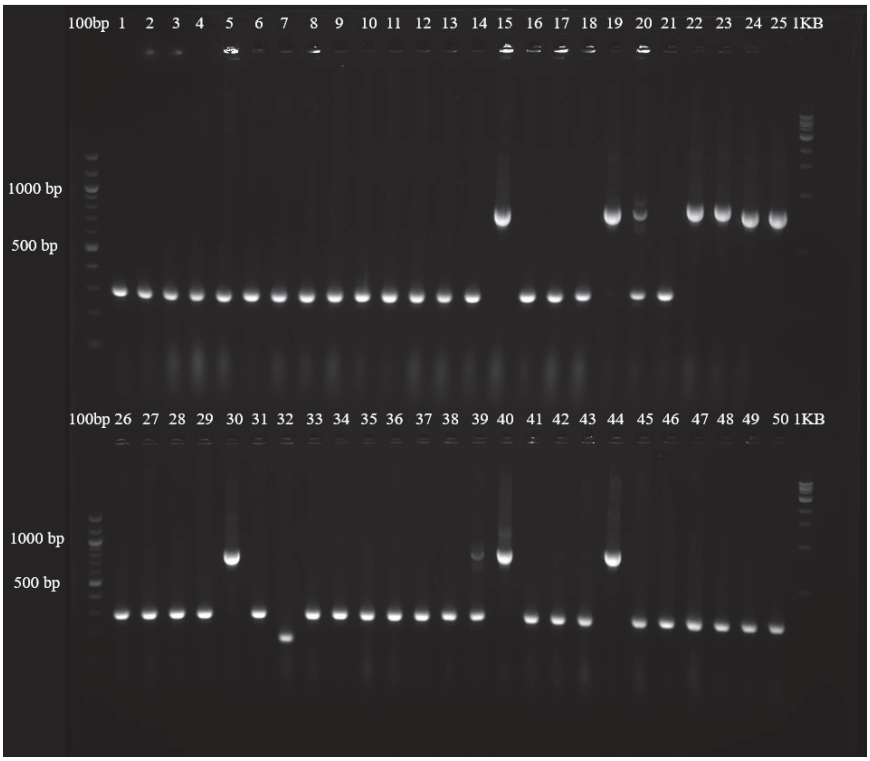

Figure 3. Agarose gel electrophoresis of products from *E. coli* transformed colonies. Colony PCR using T7 and pGBK-R primers an the pGBKT7 vector, wells #15, 19, 22, 23, 24, 25, 30, 40, and 44 are the recombinant plasmid with PAI-RBP1 inserts with the approximate predicted size. Plasmid from #15 was used for subsequent yeast two-hybrid screen.

>pGBKT7_PAI-RBP1
NNNNNNNNNANNNNNNNNTTNNTTNCCCGGNNANGNTNCGGTCCCNTTTTCCNNGTTTNNTTTN
NNNACNNTNNNTCAAGTATACCAAGCATNNATCAATTCCNGGCTGAAGCAAGCCTCCTGAAAGA
TGAAGCTACNNNNTCTNNNGAACAAGCATGCGATANTTGCCGACTTAAAAGCTCNANTGCTCCA
AAGAAAACCGAAGTGCGCCAAGTGTCTGAAGAACAACTGGGAGTGTCGCTACTCTCCCAAAACC
AAAAGGTCTCCGCTGACTAGGGCACATCTGACAGAAGTGGAATCAAGGCTAGAAAGACTGGAAC
AGCTATTTCTACTGATTTTTCCTCGAGAAGACCTTGACATGATTTTGAAAATGGATTCTTTACA
GGATATAAAAGCATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAAAGATGCCGTCACA
GATAGATTGGCTTCAGTGGAGACTGATATGCCTCTAACATTGAGACAGCATAGAATAAGTGCGA
CATCATCATCGGAAGAGAGTAGTAACAAAGGTCAAAGACAGTTGACTGTATCGCCGGAATTTGT
AATACGACTCACTATAGGGCGAGCCGCCATCATGGAGGAGCAGAAGCTGATCTCAGAGGAGGAC
CTGCATATGGCCATGGAGGCCGAATTCAGCGGCTCTCACAACTGGGGAACTGTCAAAGATGAAT
TAACTGATTTGGATCAATCAAATGTGACTGAGGAAACACCTGAAGGTGAAGAGCACCCTGTGGC
AGATACTGAAAATAAGGAGAACGAAGTTGAAGAGGTTAAGGAAGAGGGTCCAAAAGAGATGACT
CTGGATGAGTGGAAAGCTATTCAAAATAAAGACCGAGCAAAAGTAGAATTTAATATCCGAAAAC
CAAATGAAGGCGCCGATGGACAATGGAAAAAGGGATTTGTTCTGCATAAATCAAAAAGTGAAGA
GGCTCATGCTGAAGATTCAGTTATGGACCATCATTTCCGGAAGCCAGCAAATGATATAACATCT
CAACTGGAGATCAATTTTGGAGACTTAGGCCGCCCAGGACGTGGTGGCAGAGGAGGACGTGGTG
GGCGTGGGCGTGGTGGACGTCCAAACCGTGGCAGCAGGACTGATAAGTCAAGNGCNTNNGCTNN

Figure 4. Nucleotide sequence of the cloned PAI-RBP1 (lane #15, Fig. 3) bait in pGBKT7 plasmid. Highlighted sequence is the PAI-RBP1 fragment, unhighlighted represents the vector sequence. This sequence codes for a protein product that is homologous to the VIG-1 region that interacts with HUM-2C.

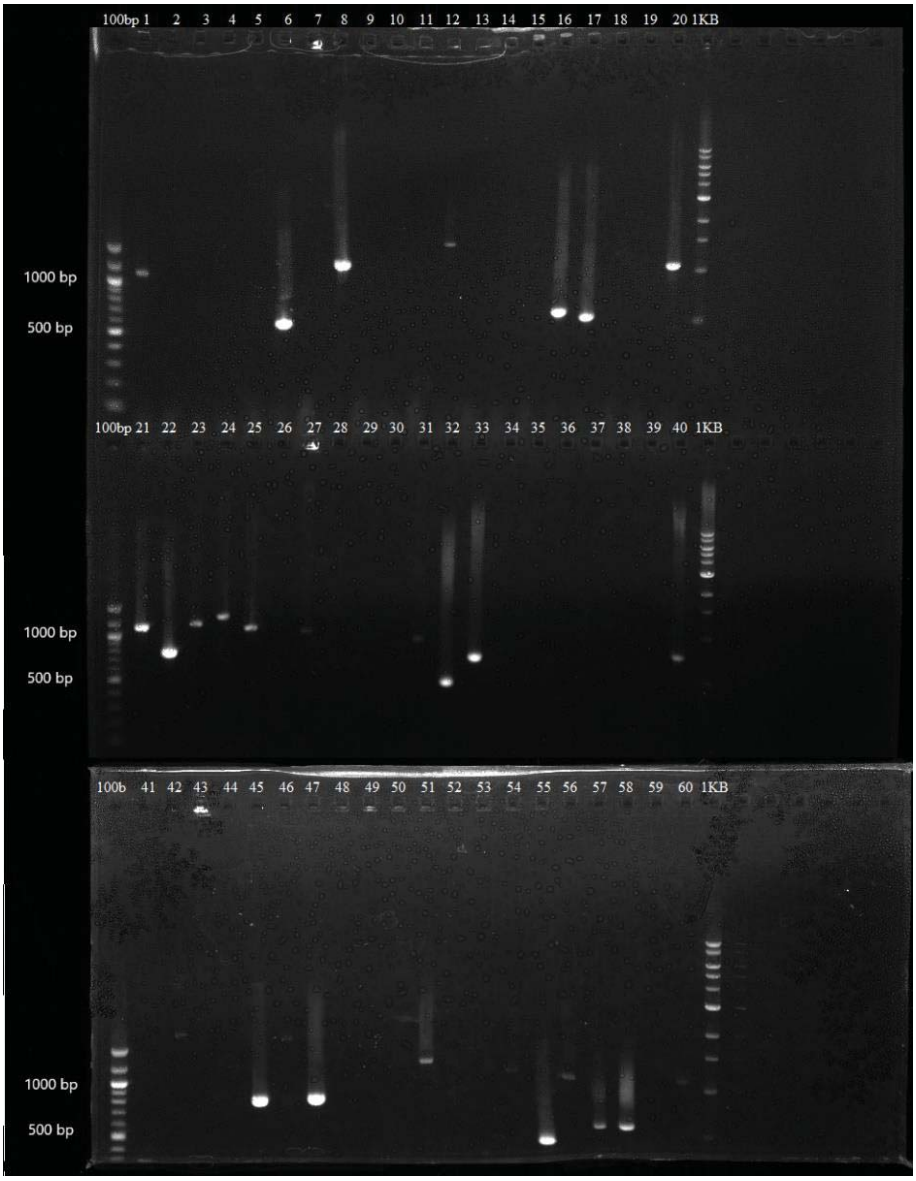

Figure 5. Three of nine agarose gels showing colony PCR products from yeast colonies growing on SD -Ade/-His/-Leu/-Trp from the bait and the prey library mating, using the T7 and pGAD-R as primers. The well numbers reflect the prey plasmid numbers in column 1 of Table 1 and Appendix E. Differences in sizes of the amplicons indicate size of the library insert.

Table 1. Yeast two-hybrid library hits using PAI-RBP1 as bait. Expect values (E value) were derived from Blastn and Blastp searches against the GenBank non-redundant nucleotide database and its derived protein database (NCBI). Nucleotide sequences were translated into their corresponding amino acid sequence in-frame from the start codon of the plasmid, indicated in each sequence shown in Appendix E. Bolded letters and Italicized letters represent identical and overlapped protein sequences, respectively. Shaded rows correspond to in-frame mouse sequences, resulting in a mouse prey protein.

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein
Y2H-006 *Y2H-058* *Y2H-076* *Y2H-086* *Y2H-199*	DNA sequence from clone RP23-99G21 on chromosome 4	*Mus musculus*	7e-125	No significant similarity found	NA	NA	***KSRRTDRTLLHKATLS*** ***RKTKQNKTKNPGSRIQ*** ***IQILMLEGKHLAAPLS*** ***CVLSLPPSPAGSPYSP***
Y2H-008	ATPase Na+/K+ transporting beta 3 polypeptide (Atp1b3) transcript variant 2 mRNA	*Mus musculus*	0	Atp1b3 protein	*Mus musculus*	2e-55	KEENATIATYPEFGVL DLKYFPYYGKKRHVGY RQPLVAVQVKFDSGLN KKEVTVECHIAGTRNL KNKNERDKFLGRVSFK VTARA
Y2H-016	Succinate-Coenzyme A ligase GDP-forming beta subunit mRNA	*Mus musculus*	0	No significant similarity found	NA	NA	IYIHCTHFTLCLNVES TVCLTLSYYQKP
Y2H-017 *Y2H-055*	Ribosomal protein S20 mRNA	*Mus musculus*	1e-170	40S ribosomal protein S20	*Mus musculus*	8e-65 2e-23	TSRNVKSLEKVCADLI RGAKEKNLKVKGPVRM PTKTLRITTRKTPCGE GSKTWDRFQMRIHKRL

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
							IDLHSPSEIVKQITSI SIEPGVEVEVTIADA
Y2H-020	Arrestin domain containing 5 (Arrdc5) mRNA	*Mus musculus*	0	arrestin domain-containing protein 5	*Mus musculus*	4e-109	ESPMSVVKSIEVVLPQ DAVYLAGSIIDGQVVL TLNSTLVDPVVKVELV GRGYVEWNEEIGETRD YSRDVICNNKADYVHK TKAFPIKELRVCRS
Y2H-021	Histone acetyltransferase querkopf mRNA	*Mus musculus*	0	No significant similarity found	NA	NA	MEMPLAFQLLYILFKV TLQHAANVFVSDSAFC STVQVLC
Y2H-022	3-oxoacid CoA transferase 2B (Oxct2b), mRNA	*Mus musculus*	0	Succinyl-CoA:3-ketoacid coenzyme A transferase 2B, mitochondrial precursor	*Mus musculus*	2e-59	DGDLANWMVPGKKVKG MGGAMDLVSSKKTRVV VTMEHCTKTKQPKILK KCTMPLTGKRCVDLII TEKAVFEVNHSKGLTL VELWEGSSVDDIKATT ACSFAVSPNLEPMQQI KLDA
Y2H-032	PREDICTED: CD2-associated protein (Cd2ap) variant X4 mRNA	*Mus musculus*	2e-98	Hypothetical protein	*Mus musculus*	7e-12	TDKKYLPKCSVDFFDE CFMLAFKLEMYEQNET IFTIRIEKCYN
Y2H-033 *Y2H-107*	PREDICTED: Protein arginine	*Mus musculus*	0	PREDICTED: Protein arginine N-	*Mus musculus*	5e-76	***IAPFINQFQVPIRVCL DLSSLPCVPLSQPVEL***

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
	methyltransferase 9 (Prmt9) transcript variant X2 mRNA			methyltransferase 9 isoform X1			***LRLDLMTPYLNTSNKE VKVRVCRSGRVTAVPF WFHLCLDDEVRLDISG EASHWKQAAVVLDNPI QVQAGEELVLSVEHHK SNVSIAVKP***
Y2H-045	FK506 binding protein 3 mRNA	*Mus musculus*	0	FK506 binding protein 3, isoform CRA_b	*Mus musculus*	5e-16	VSEQVKNVKLSDDKPK DSKSEETLDEGPPKYT KSILKKARLEIEPEWA YGKKGQPDAKIPPNTK LIFEVELVDID
Y2H-047	Family with sequence similarity 96 member B mRNA	*Mus musculus*	0	Mitotic spindle-associated MMXD complex subunit MIP18	*Mus musculus*	2e-79	IRSINDPEHPLTLEEL NVVEQVRIQVSDPEST VAVAFTPTIPHCSMAT LIGLSIKVKLLRSLPQ RFKMDVHITPGTHASE HAVNKQLADKERVAAA LENTHLLEVVNQCLSA RS
Y2H-063	Phosphoglucomutase 2 mRNA (cDNA clone	*Mus musculus*	0	Pgm2 protein	*Mus musculus*	8e-39	RLIFADGSRIIFRLSG TGSAGATIRLYIDSYE KDVAKINQDPQVMLAP LISIALKVSQLQERTG RTAPTVIT

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
Y2H-066	Asparagine-linked glycosylation 11 homolog (yeast alpha-1 2-mannosyltransferase)	*Mus musculus*	2e-170	No significant similarity found	NA	NA	SFESTTAEVQIPVYIDNTTVKYPLPSLCIIYLYLIVM
Y2H-075	N-glycanase 1 (Ngly1) mRNA	*Mus musculus*	0	Chain A, Crystal Structure Of Mouse Peptide N-glycanase C-terminal Domain In Complex With Mannopentaose	*Mus musculus*	1e-97	VSDNNINISGWENGVWKMESIFRKVEKDWNMVYLARKEGSSFAYISWKFECGSAGLKVDTVSIRTSSQSFESGSVRWKLRSETAQVNLLGDKNLRSYNDFSGATEVTLEAELSRGDGDVAWQHTQLFRQSLNDSGENGLEIIITFNDL
Y2H-077	Alpha-2-macroglobulin-like (A2ml1) mRNA	*Mus musculus*	0	Ovostatin homolog precursor	*Mus musculus*	1e-101	TVDVEGDGCSFIQATLRYNVPLPKEASGFSLSVKTGKSNSSDEFQTKFELTVTLTYTGARESSVTVLVDVKMLSGFTPVVSSTEELKFNSQVTKTDIKNGHVLFYLENVPKEATSLTFSIEQTNHVANIQPAPVTVYSYEKGEYAFDSYNINSISDSQ

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
Y2H-080	EF-hand calcium binding domain 7 mRNA	*Mus musculus*	0	EF-hand calcium-binding domain-containing protein 7	*Mus musculus*	4e-110	MEANDREGDPLDLWVT LHSMGYNKALELTEAC PFVINIYAERCKPRIK VVHMEACSGQLEKAIC KSVLDRSDAKVMDGYE NIIVHTCNYDTWITSI IENKSDNKVIIHINNE LSKNCVNNRGLNIFAV EVAPRSTMVCQHVMPL NEQEEWIYCCVYSLVA
Y2H-088	PREDICTED: ankyrin repeat domain 17 (Ankrd17) transcript variant X5 mRNA	*Mus musculus*	0	No significant similarity found	NA	NA	SQFSIIIRFYLS
Y2H-097	Mus musculus BAC clone RP24-158K7 from chromosome 2	*Mus musculus*	0	No significant similarity found	NA	NA	QNKTKNLNQRQKSNVV IKIFLILDI
Y2H-098	FK506 binding protein 3 mRNA	*Mus musculus*	0	FK506 binding protein 3, isoform CRA_b	*Mus musculus*	7e-30	IRGWDEALLTMSKGEK ARLEIEPEWAYGKKGQ PDAKIPPNTKLIFEVE LVDID

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
Y2H-099	PREDICTED: Protein phosphatase 1, regulatory subunit 42 (Ppp1r42), transcript variant X3	*Mus musculus*	0	No significant similarity found	NA	NA	DKDTDQTSYRPDWTER FLKVEVISKPASGHAV KVEFSLRV
Y2H-113	Cystatin-related epididymal spermatogenic protein (Cres) mRNA alternatively spliced partial cds	*Mus musculus*	0	cystatin-related epididymal spermatogenic protein	*Mus musculus*	2e-67	ESVGNPEEPQAGVARV RKEPSRTSRWGESQKV ALAVGVDQSKNEVKAQ NYFGSINISNANVKQC VWFAMKEYNKESEDKY VLLVDKILHAKLQITD RMEYQIDVQISRSNCK KPLNNTENCIPQKNPN WKKK
Y2H-114	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 mRNA	*Mus musculus*	0	No significant similarity found	NA	NA	GAGIIRLVLQIQFID
Y2H-121	PREDICTED: fibulin 5 (Fbln5) transcript variant X2 mRNA	*Mus musculus*	0	fibulin-5 isoform X3	*Mus musculus*	3e-63	MDVVSGRSVPADIFQM QATTRYPGAYYIFQIK SGNEGREFYMRQTGPI SATLVMTRPIKGPRDI

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
							QLDLEMITVNTVINFR GSSVIRLRIYVSQYPF
Y2H-124	Ankyrin repeat and sterile alpha motif domain containing 1B (Anks1b) transcript variant 6 mRNA	*Mus musculus*	0	No significant similarity found	NA	NA	EGGSPVLPESLNSLGL SFHGQVSRFEGKVRPV LPHPQSTCKVIVSLST IPCKKTKQPSPGSSVP PHTLDFYICGFYSNQL CQIKETFIE
Y2H-126	PREDICTED: Protein phosphatase 1, regulatory subunit 42 (Ppp1r42), transcript variant X3, mRNA	*Mus musculus*	0	No significant similarity found	NA	NA	VATAAGFGGAWHKDTD QTSYRPDWTERFLKVE VISKPASGHAVKVEFS LRV
Y2H-127 *Y2H-094*	ATPase Na+/K+ transporting beta 1 polypeptide (Atp1b1) mRNA	*Mus musculus*	0	ATPase, Na+/K+ transporting, beta 1 polypeptide, isoform CRA_c	*Mus musculus*	2e-69	DPLMMKYNPNVLPVQC TGKRDEDKDKVGNIEY FGMGG*YYGFPLQYYPY YGKLLQPKYLQPLLAV QFTNLTVDTEIRVECK AYGENIGYSEKDRFQG RFDVKIEIKS*
Y2H-128	Mitochondrial DNA complete	*Mus musculus*	2e-71	NADH dehydrogenase subunit 1	*Mus musculus*	2e-14	HQHSYGSEHLIHASVT INLYIFYEKNFLPLTL ALCM

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
	sequence clone: B82-2748-95						
Y2H-129	PREDICTED: lysine (K)-specific demethylase 2B (Kdm2b) transcript variant X11 mRNA	*Mus musculus*	0	Isoform 4 of Lysine-specific demethylase 2B	*Mus musculus*	6e-54	INDQSINLLTAVGTTT RDSLTEVNLSDCNKVT DLCLSFFKRCGNICHI DLRYCKQATKEGCEQF IAEMSVSVQFGQVEEK LLQKLS
Y2H-130	Aspartoacylase (aminoacylase) 3 (Acy3) transcript variant 3 mRNA	*Mus musculus*	0	No significant similarity found	NA		HLPWFHRYP
Y2H-131	PREDICTED: calpain 7 (Capn7) transcript variant X4 misc_RNA	*Mus musculus*	0	calpain-7	*Mus musculus*	1e-140	VYYPADPPPYIDGIRI NSPHYLTKIKLTTPGT HTFTLVVSQYEKQNTI HYTVRVYSACSFTFSK IPSPYTLSKRINGKWS GQSAGGCGNFQETHKN NPIYQFHIDKTGPLLI ELRGPRQYSVGFEVVA VSIMGDPGPHGFQRKS SGDYRCGFCYLELENI PAGIFNIIPSTFLPKQ EGPFFLDFNSTVPIKI TQLQ

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
Y2H-132	WD repeat domain 61 (Wdr61) transcript variant 2 mRNA	*Mus musculus*	0	WD repeat domain 61, isoform CRA_a	*Mus musculus*	1e-107	EACCAAGCTVNCLAFS ASSLVKEMTNQYSILF KQEQAHDDAIWSVAWE TNKKENIETVVTGSLD DLVKVWKWRDERLELQ WSLEGHQLGVVSVDIS HTLPIAASSSLDAHIR LWDLENGKQMKSIDAG PVDAWTLAFSPDSQYL ATGTHMGKVNIFGVES GKKNILWTLEENSSLV LHIVLMGNTWPAEP
Y2H-133	Mus musculus Ampd3 gene	*Mus musculus*	0	No significant similarity found	NA	NA	LVSGLIPSTPVECSGS SYICSLCMISPFPSLF PETQSIVIKIRLTKPG VVAMP
Y2H-137	RIKEN cDNA 4930423M02 gene long non-coding RNA	*Mus musculus*	0	No significant similarity found	NA	NA	EHTLVQAPLFNTSQGV VIYLRLRLSFSLQASG DTVCSRVK
Y2H-138	NudC domain containing transcript (Nudcd2) variant 1	*Mus musculus*	0	No significant similarity found	NA	NA	GGWLEQPVSTVKDCQV YFNMVVKITFRYGLHR RQSSSNVDR

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
Y2H-142 *Y2H-185*	PREDICTED: teneurin transmembrane protein 3 (Tenm3) transcript variant X18	*Mus musculus*	0	No significant similarity found	NA	NA	*SSLSYSTHCFPLCIMV TNIASEHVVNIKVQVM Y*
Y2H-144	BAC clone RP24-323I5 from chromosome 6	*Mus musculus*	0	No significant similarity found	NA	NA	HQGNLTSHCMSSAPLQ FSHRFCTVTQMHFIFL SLPSTNILISLKSDYS VLWNQKKLTLMMSSDT DTIIKMQE
Y2H-156	PREDICTED: activating signal cointegrator 1 complex subunit 3 (Ascc3) transcript variant X1 mRNA	*Mus musculus*	0	Ascc3 protein	*Mus musculus*	8e-89	ADKRDENKWIKLHADQ EYVLQVSLQRVHFGLP KGKHGNHAVTPRFPKL KDEGWFLILGEVDKRE LMAVKRVGFVRTHHDA SISFFTPETPGRYIFT LYLMSDCYLGLDQRYD IYLNVIKANISTKDSD VFTDLSV
Y2H-158	Subunit of RNA polymerase III transcription initiation factor, BRF1-like	*Mus musculus*	0	No significant similarity found	NA	NA	ETIKVSFVLSDLVSVE GVPVNYLKITTRTIIS

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
Y2H-160	Secreted frizzled-related protein 1 (Sfrp1), transcript variant X1	*Mus musculus*	0	No significant similarity found	NA	NA	NSY
Y2H-161	Chromosome 18 clone RP24-76F12 complete sequence	*Mus musculus*	0	Coiled-coil domain-containing protein 192 isoform 2	*Mus musculus*	1e-73	MGGCHSKKVVTPDIET SARCRSMLGSYQSYIQ SQNRTSGNSPDSGQTV FSLEHLEICLKEAEEK ARALLEQLTASEATKS QLLEKVSMLEGRLEDV NRKNVGGELYENMVLE KDKCIEKLQAEVKASQ EKLNIHLSLSLLKWRR AAEPRL
Y2H-163	Brain cDNA clone MNCb-4285	*Mus musculus*	4e-27	amino acid-binding protein	*Aeromonas veronii*	5.8	ILYVSITLFIT
Y2H-178	PREDICTED: Importin 7 (Ipo7), transcript variant X1	*Mus musculus*	0	hypothetical protein ABT21_06255	*Thiobacillus* sp. *SCN 65-179*	2.9	MATLSVSFFVPKALSV QVDIWIQDLM
Y2H-181 *Y2H-068*	NPC intracellular cholesterol transporter 2 (Npc2)	*Mus musculus*	0	epididymal secretory protein E1 precursor	*Mus musculus*	4e-76	*VIKEVNVSPCPTDPCQ LHKGQSYSVNITFTSG TQSQNSTALVHGILEG IRVPFPIPEPDGCKSG*

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
							INCPIQKDKVYSYLNK LPVKNEYPSIKLVVEW KLEDGKKNNLFCWEIP VQITS
Y2H-186	DEP domain containing MTOR-interacting protein (Deptor) transcript variant 2	*Mus musculus*	0	Depdc6-003	*Mus musculus*	3e-08	GPAAAAGMKVCQFVVS VNGLNVLNVDYRTVSN LILTGPRTIVMEVMEE LDC
Y2H-187	PREDICTED: histone deacetylase 2 (Hdac2) transcript variant X2	*Mus musculus*	0	No significant similarity found	NA	3e-87	VEEATQSGMLPDVGHT RLQLPLIVKFPMSCHI MITLSILDQTSNCILV LQT
Y2H-188	Fibronectin type 3 and ankyrin repeat domains 1 (Fank1)	*Mus musculus*	0	fibronectin type 3 and ankyrin repeat domains protein 1 isoform X2	*Mus musculus*	1e-66	VEGLEPRTLYKFRLKV TSPSGEYEYSPVVSVA TTREPISSEHFHRAVS VNDEDLLLRILEGGHV MIDVPNKFGFTALMVA AQKGYTRLVKILVSNG TDVNLKNGSGKDRLAD SAGMPGWAATFREPFD STSTTSDE

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
Y2H-189	Cell division cycle associated 7 (Cdca7)	*Mus musculus*	0	ABC transporter ATP-binding protein	*D. youngiae*	0.5	EGSLVLDVVLMGTRVR ISAIRLP
Y2H-192	Paraoxonase 1 (Pon1) mRNA	*Mus musculus*	0	Paraoxonase 1	*Mus musculus*	1e-118	ESCQSHWAAGLRTDQC CCTCPSSACPSVSGHG EAASTHPRGTGVGTLQ EPSVFLSITLCTYWW
Y2H-193	PREDICTED: angiomotin (Amot) transcript variant X3	*Mus musculus*	0	hypothetical protein I79_007206	*Cricetulus griseus*	1.2	TYQLKNSVYGMNLELT LPSQSCGCNISGLFLD FPEEEGIFLVLTFTV
Y2H-194	Dynactin 4 (Dctn4), transcript variant 1	*Mus musculus*	0	Dynactin 4, isoform CRA_a	*Mus musculus*	2e-33	RKANKVGIFIKVTPQR EEGEVTVCFKMKHDFK NLAAPIRPVEEGDQGT EVIWLTQHVELSFGPL LP
Y2H-198	PREDICTED: DEAD (Asp-Glu-Ala-Asp) box polypeptide 19a (Ddx19a), transcript variant X1	*Mus musculus*	0	INO80 complex subunit D	*Gossypium arboreum*		DVFILQIGRKNGSQQI SDSRVIDNPIEDGGVF PPVHLVCAWLHCLFQS MGLHAT

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
Y2H-202	Secretory carrier membrane protein 1 (Scamp1), transcript variant 2	*Mus musculus*	0	No significant similarity found	NA	7e-45	RDTSEIHKMLIVFYFQ NWECSLSIMWRSPEVV LLNGKE
Y2H-203	DnaJ heat shock protein family (Hsp40) member B11 (Dnajb11) transcript variant 2	*Mus musculus*	0	Dnajb11 protein	*Mus musculus*	7e-57	KLVNEERTLEVEIEPG VRDGMEYPFIGEGEPH VDGEPGDLRFRIKVVK HRIFEKKGEGLPNFDS NNIKGSLIITFDVDFP KEQLTEEAKEGIKQLL KQGPVQKVYNGLQGY
Y2H-205	N-glycanase 1 (Ngly1) mRNA	*Mus musculus*	0	Peptide-N(4)-(N-acetyl-beta-glucosaminyl) asparagine amidase	*Mus musculus*	5e-101	DRYIRVSDNNINISGW ENGVWKMESIFRKVEK DWNMVYLARKEGSSFA YISWKFECGSAGLKVD TVSIRTSSQSFESGSV RWKLRSETAQVNLLGD KNLRSYNDFSGATEVT LEAELSRGDGDVAWQH TQLFRQSLNDSGENGL EIIITFNDL
Y2H-206	Zinc finger protein 431-like (LOC102639653	*Mus musculus*	0	PREDICTED: zinc finger protein 431-like isoform X2	*Mus musculus*	2e-48	VTTVVRPLHKAVLSES ISEHVQERDNMNVTSM VKPLQEAVVSNAIKDH VHERHHMNETKVVNPL

Prey #	Blastn	Organism	E Value	Blastp	Organism	E Value	Protein (continued)
							KEAVVLNIIIEHTQLK NPGM
Y2H-209	PREDICTED: heterogeneous nuclear ribonucleoprotein H1 (Hnrnph1) transcript variant X17	*Mus musculus*	0	PREDICTED: heterogeneous nuclear ribonucleoprotein H isoform X12	*Mus musculus*	6e-52	PVRVHIEIGPDGRVTG EADVEFATHEDAVAAM SKDKANMQHRYVELFL NSTAGASGGAYEHRYV ELFLNSTAGASGGAYG SQMMGGMGLSNQSSYG GPASQQLSGGYGGGYG GQSSMSGYDQVLQENS SDFQSNIA

Multiple Sequence Alignment

The parallel comparative genomics approach used multiple sequence alignments of five mammalian species to identify conserved sequences in myosin V introns. In all, 120 introns across the five species were analyzed. These represented three myosin V genes for each species and twenty introns within each gene. A region at the 3′ end of intron 18 of MYO5A demonstrated the most highly conserved sequence of any of the alignments, with 85% identity (Fig. 6). In fact, this region corresponds to the sequence of a predicted mouse MYO5A isoform from the NCBI database, myosin-Va isoform X7 (accession # XP_006510897.1).

To see if the region was expressed in mouse brain, RT-PCR and PCR amplification using primer pairs indicated in Figure 7 were used to generate a product for sequencing. Figure 8 shows amplimer bands resulting from various combinations of primers and nested PCR, with the final purified amplimers cloned into the pMiniT plasmid. Products using primers within exons 18 and 19 showed a perfect splice junction between the two exons (sequences from intron 18 were absent in these products). However, when forward primers within the conserved intron 18 sequence were used in combination with reverse primers in downstream exons, a product with the predicted size was produced. Sequencing of the resulting product showed the conserved intron 18 sequence spliced directly into exon 19 (Figs. 9, 10, 11)

Translation of the intron 18 region identified in the mRNA pool shows that downstream stop codons exist in all possible reading frames, either within intron 18 (reading frame 1) or in exon 19 (reading frames 2 and 3). Furthermore, alignment and

analysis of the conserved area of intron 18 shows that any potential protein product it may produce by splicing out a region with the stop codon would have no significant homology to the *C. elegans* HUM-2C unique lead peptide.

Further sequence analysis of the conserved intron 18 region showed that it contains putative promoter regions with transcription factor binding sites (TFBS) (Fig. 10). The Ebf1 TFBS is located at the 3′ end of intron 18 and was identified using the PAZAR database[52]. The Bhlhe40 regulatory element was found upstream from the conserved area, in the 5′ region of intron 18, using the UCSC Genome Browser[53] and Jaspar database[54]. In addition, an unnamed predicted promoter with a transcription factor binding site was found within the highly conserved intronic region via the Promoter 2.0 Prediction Server[55] and Promoter Prediction by Neural Network software[56].

A coiled-coil domain was found within the conserved region using COILS[57] by SIB ExPASy Bioformatics. Furthermore, a predicted TATA box was found 33 bp upstream from the putative predicted exon via GPMiner software[58] (data not shown).

```
Mus_I18       GTTTGTTTTTAACTTTTTATTTTGTGTGGTTTTCTTCCTGCCTTATTCTTCAT--CCTGG
Canis_I18     GTTTCTTGTTTAATTCTTACTTTGTGTGCGTGGCTTCTTGCCTGACTCTTCATGCTGGTG
Bos_I18       TTTAATCTTTTTGTTTTTATTTTGTATGTATTACCTCCTGCCTTTTTCTTCAC--ACTGG
Homo_I18      TTTTGGTTTTTGAATTTTATTTCATAAACGTTTCTTCCTGCCTTATTCTTCAC--ACTGG
Maca_I18      TTTTGGTTTTTGAATTTTATTTCATATGCATTTCTTCCTGCCTTATTCTTCAC--ACTGG
               **     **    * *** **  *     *  * ** *****   ******       *

Mus_I18       TGTTTTTGTTTCTTCAGC-TGCTCATGGAGTGTACAGCCTGATG---TAATCTAACAGCG
Canis_I18     TTTTTTTGTCCCTTCAGCTTTCTCATGGAACATACAGCCTGATGTAATGCTTACAGCCTG
Bos_I18       TGTTTTTGTCCCTTCAGCTTTCTCATGGAACTTACAGCCCGATGTAATCCTTACAGCCCG
Homo_I18      TGTTTTTGTCCCTTCAGC-TTCTCATGGAATATACAGCCCGCTGTAATCCTTACAACCTG
Maca_I18      TGTTTTTGTCCCTTCAGC-TTCTCATGGAATATACAGCCCGATGTAATCCTTACAACCTG
              * *******  ******* * ********   ******* * **   *  *   *    *

Mus_I18       GGTCAGTCAGAGGGAGTGTCAGCATCTCCATTTCCTTGAGGTGTTTAAAATCCAAACAAG
Canis_I18     AGTCAGTCAGAGGGAGTGTCAGCATCTCCATTTCCTTGAGATATTTAAAATGCAAACATG
Bos_I18       GGTCAGTCAGAGGGAGTGTCAGCATCTCCATTTCCTTGAGATATTTAAAATACAAACAAG
Homo_I18      GGTCAGTCAGAG--AGAGTCAGCATCTCCATTTCCTTGAGATGTTTAAAACACAAACAAG
Maca_I18      GGTCAGTCAGAG--AGAGTCAACATCTCCATTTCCTTGAGATGTTTAAAACACAAACAAG
               ***********  ** **** ****************** * *******  ****** *

Mus_I18       CAAATAGTAGCTGCGTAGAATCCACAACC-AGTTTTGGGCCTCAGATAATAATTCCAGGT
Canis_I18     TAAACAATAACTGCCTATAACCCACAACCAAGTTTTTGGCCTCAGATAATAATTTCAGGT
Bos_I18       CAAATAGTAGCTGCCTATAACCCACAACCAAGTTTTGGGCTTCAGATAATAATTCCAGGT
Homo_I18      CAAATAATAACTGCCTATAACCCACAACCAAGTTTTGGGCCTCAGATAATAATTCCAGGT
Maca_I18      CAAATAATAACTGCCTATAACCCACAACCAAGTTTTGGGCCTCAGATAATAATTCCAGGT
               *** * ** **** ** ** ******** ****** *** ************* *****

Mus_I18       AGAATAATTCTGGAAAATAACTCTTA------GGCTACCATTTTCTCTACAATCCTTTCT
Canis_I18     AGAATGATTCTGGAAAATAACTCTTAATTAAAAACTGCAGTTTTCCTTAAAATCCCTTCT
Bos_I18       AGACTAATTCTGGAAAATAACTCTTA------AACTGCAGTTTTCCTTAAAGTCCCTTCT
Homo_I18      AGAATAATTCTGGAAAATAACTCTTA------AACTGCAGTTTTTCTTAAAATCCCTTCT
Maca_I18      AGAATAATTCTGGAAAATAACTCTTA------AACTGCAGTTTTTCTTAAAATCCCTTCT
              *** * ********************        ** *  ****   ** * *** ****

Mus_I18       ATTTAAACAACTTATAAAGGATCAAC-AATCTGGCTCCATTAGTTCGTTGAAATGTATGT
Canis_I18     ATTTGCTCAACTTGCATAGGATTAAACAGTCTGACTTTATTAGTTAATTG----TGATGT
Bos_I18       ATTCGCTCAACTTGCCTAGGATCAAA-AAGCTGGCTCCATTAGATACTTG----TGATGT
Homo_I18      ATTTGCTCAACTTGCATAGGATTAAA-AATCTGGCTCCATTAGTTAATTG----TGATGT
Maca_I18      ATTTGCTCAACTTGCATAGGATTAAA-AATCTGGCTCCATTAGTTAATTG----TGATGT
              ***    ******    ***** **  *  *** **  ***** *  ***      ****

Mus_I18       AGCTCAGGTTTTAGATGATAGATTCATCTTTTTACTGATTCTACTTCTATATTTTTCCAT
Canis_I18     AGTCTAAGTTTTAGTTAATAGAGTCATCCCTTTACTGATTCTGCTTGTAGACATTTTTTT
Bos_I18       AGCCCAAGATTTAGCTAATAGTTTCATCCCTTTACTGATACTGCTTTTAGACTTTTTTTT
Homo_I18      AGCCCAAGTTTTAGCTAATAGAGTCATTCCTTTACTGATTCTGCTTTTAGACATTTTTTC
Maca_I18      AGCCCAAGTTTTAGCTAATAGATTCATTCCTTTACTGATTCTGCTTTTAGACATTTTTTC
              **   * * ***** * ****  ****   ********* ** *** ** *  ***
```

Figure 6. Alignment of the 3′ end of intron 18 of 5 mammals. Mouse (Mus_I18), dog (Canis_I18), cow (Bos_I18), human (Homo_I18), and macaque (Maca_I18). The highlighted area represents the most highly-conserved, semi-continuous nucleotide sequence identified in intron 18.

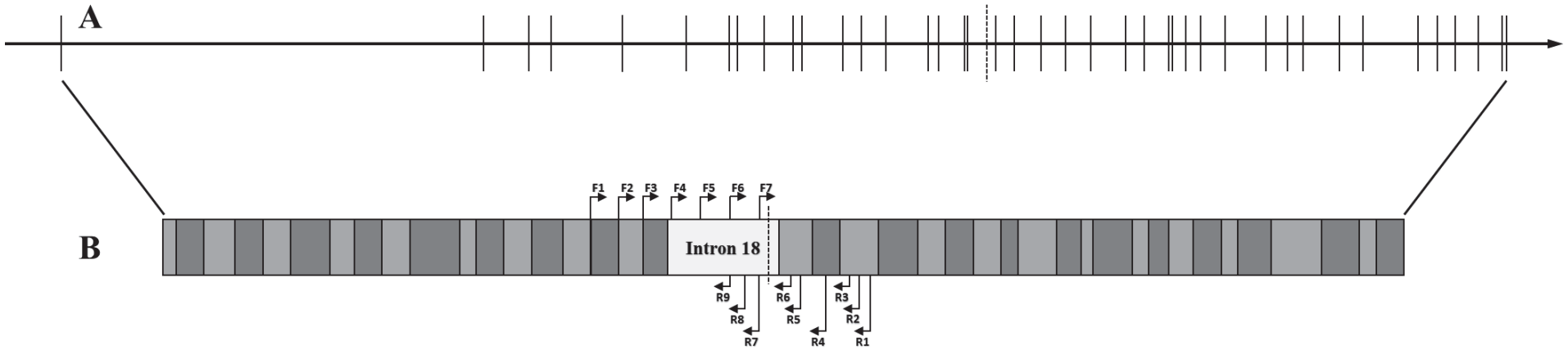

Figure 7. MYO5A model #1. Based on the Ensembl genome browser[59] predicted gene model. A) Genomic DNA model of full-length MYO5A gene showing all 41 exons. B) cDNA model of full-length MYO5A transcript with the putative exon intercalated on intron 18. Positions of the primers for the mouse cDNA PCR cycles are indicated by arrows and numbers*. The vertical dashed lines indicate the putative conserved intronic sequence.

* MMyo5A-F# and MMyo5A-R# as indicated in Appendix C with sequences

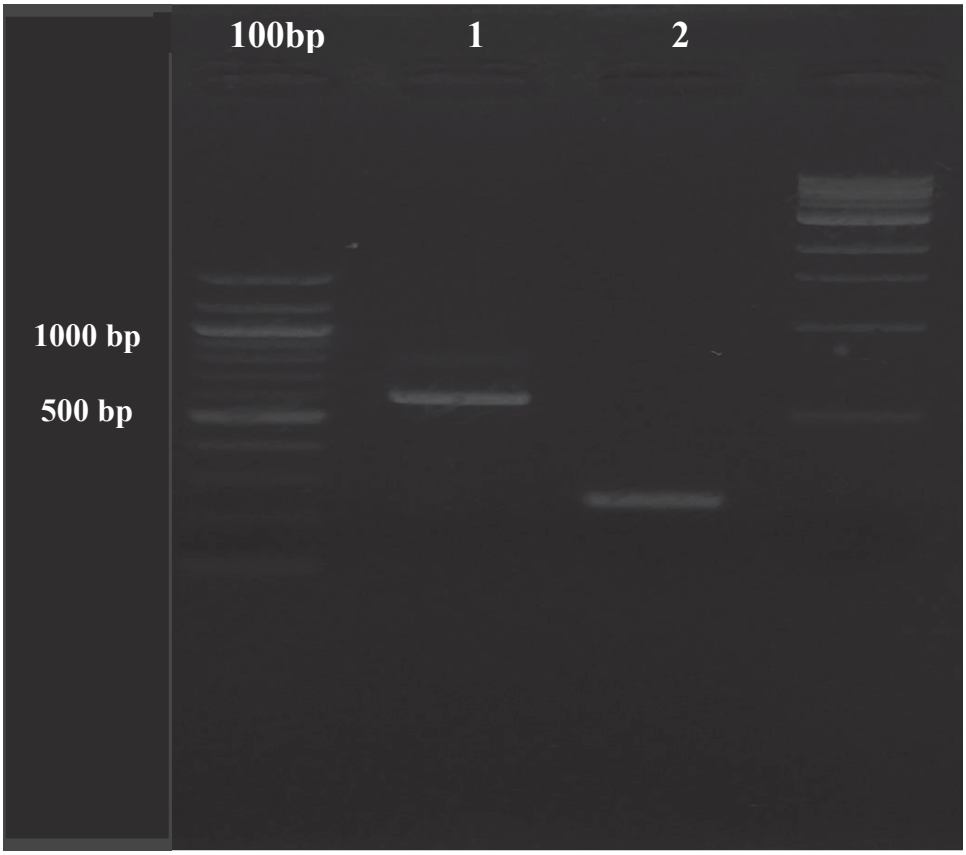

Figure 8. Agarose gel of PCR products from predicted exon in MYO5A intron 18.

NNNNNNATCGANNGATTGTTAGGTAATCGTCACCTGCAGGAAGGTTTAAACGCATTTAGGTGACA
CTATAGAAGTGTGTATCGCTCGAGGGATCCGAATTCAGGAGGTAAAAACCATGATcaaccagttt
tgggcctcagATAATAATTCCAGGTAGAATAATTCTGGAAAATAACTCTTAGGCTACCATTTTCT
CTACAATCCTTTCTATTTAAACAACTTATAAAGGATCAACAATCTGGCTCCATTAGTTCGTTGAA
ATGTATGTAGCTCAGGTTTTAGATGATAGATTCATCTTTTTACTGATTCTACTTCTATATTTTTC
CATCTTTAATTTCTTTTATACAGTCATTATTTCTCCTTATGTTTTTCCTCTAAAATAGGACAAGG
ATAAATACCAGTTTGGTAAGACAAAGATCTTTTTCCGTGCTGGTCAAGTGGCCTATCTTGAAAAA
TTGAGGGCTGACAAACTTCGGGCTGCCTGCATCCGGATCCAGAAGACCATTCGTGGGTGGCTTCT
AAGGAAGAGATACCTGTGT**ATG**CAGAGGGCAGCCATCACAGTGCAGCGATACGTGCGGGGCTATC
AGGCTCGATGCTATGCTAAGTTTCTGCGCAGAACCAAGGCAGCAACCACCATTCAAAAGTACTGG
CGCATGTATGTGGTCCGCAGGAGGTACAAGATTAGACGAGCTGCCACGATTGTTATTCAGTCTTA
CTTGAGAGGCTACTTGACGAGAAATAGGTATCGCAAGATACTCCGTGAATACAAAGCAGTCATCA
TTCAGAAACGTGTCCGTGGCTGGCTGGCCCGTACACATTATAAGAGGACCATGAAAGCCATCGTC
TACCTTTCAGTGCTGCTTCCCGGCGGATGATGGGNCAAGCNNGAGCTGANNAAACTCNAAANNNN
NTNNNNNTGNNNNNCTACTGATAATAANNNNNNANGTCNNANNCTCNAGNCNNNCANGNGCNTCN

Figure 9. Nucleotide sequence of the PCR product showing the splice site. Lower case letters indicate the forward primer sequence (MMyo5A-F7), highlighted letters indicates the sequence of exons 19 and 20, underlined letters represent the putative predicted exon, and the bolded letters are in-frame start codons.

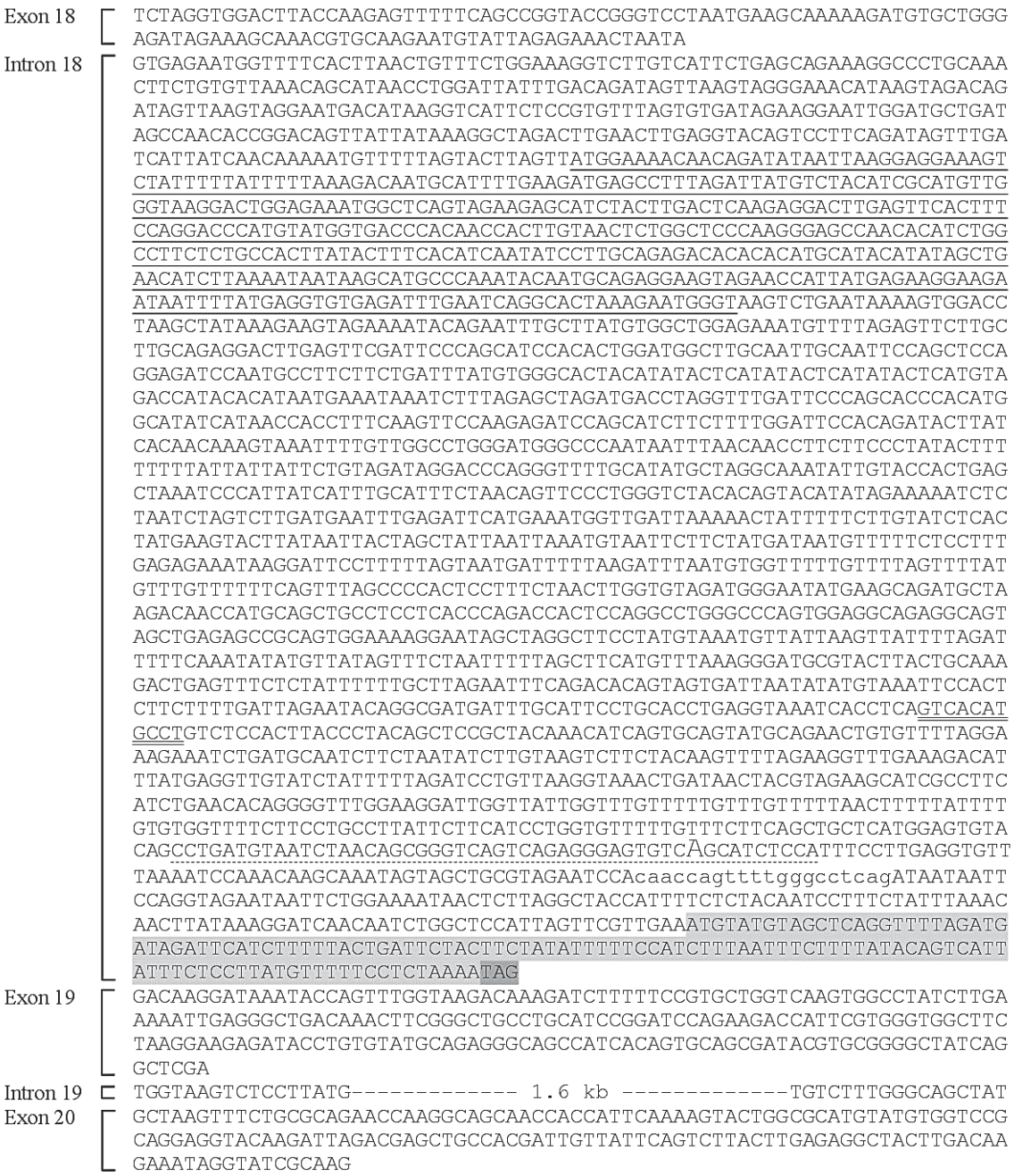

Figure 10. Continuous nucleotide sequences of the intronic region prior to exon 19. Lower case letters indicate the forward primer sequence (MMyo5A-F7), light highlight indicates the putative predicted exon and dark highlight is a stop codon. The single underline, double underline, and dotted underline letters represent Ebf1 TFBS, Bhlhe40 TFBS, and unnamed predicted promoter, respectively (see Appendix D, Table D1).

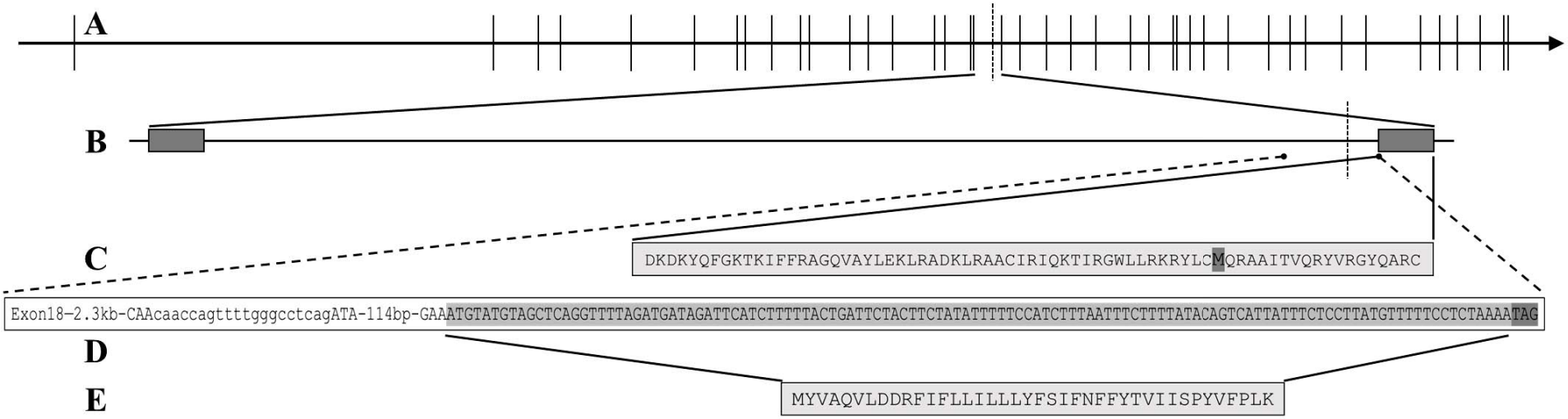

Figure 11. MYO5A model #2. Based on the Ensembl[59] and UCSC[53] genome browsers predicted gene models. A) Genomic DNA model of full-length MYO5A gene. B) enlarged region of genomic DNA showing intron 18 flanked by exons 18 and 19. C) Amino acid sequence corresponding to exon 19, highlighting the methionine encoded by a potential start codon. D) Nucleotide sequences of the intronic region prior to exon 19. Lower case letters indicate the forward primer sequence (MMyo5A-F7), light highlight indicates the putative predicted exon, and dark highlight is a stop codon. C) Putative predicted protein from the predicted exon within intron 18.

4. Discussion

Yeast Two-Hybrid System

The identification of a motorless myosin V that interacts with an RNA binding protein in mammalian brain tissue may provide a new pathway in studying neuronal development, plasticity, and synapse formation in the mammalian brain. Previous unpublished, experimental data showed that there is an interaction between the *C. elegans* truncated myosin V (HUM-2C) and VIG-1[41]. VIG-1 is an RNA-binding protein that has been shown to participate in post-transcriptional regulation of RNA via its association with the RNA-Induced Silencing Complex[60]. As shown in this study, closest mouse homologue to VIG-1 is the plasminogen activator inhibitor 1 RNA binding protein (PAI-RBP1).

PAI-RBP1 is an RNA-binding protein with relatively little data regarding its cellular activity. It has been shown to interact with the adenosine-rich 3′-UTR of plasminogen activator inhibitor mRNA[61], serving as a post-transcriptional regulatory protein[62]. Other biological roles appear to involve regulation of mRNA stability[63] and the apoptotic process[64]. PAI-RBP1 is expressed in cells of the vascular system and appears to regulate the expression of FOSB[65]. FOSB is a member of the FOS family of transcription factors and is essential for development and plasticity of neurons[66]. However, it is not known if PAI-RBP1 is co-expressed in neurons.

We used the carboxy-terminal portion of PAI-RBP1 as a yeast two-hybrid bait because this region is most similar to the region of VIG-1 that interacts with the *C.*

elegans HUM-2C lead peptide. In using this PAI-RBP1 region we hoped to identify a similar lead peptide of a truncated mouse MYO5 that interacts with PAI-RBP1. However, using this bait to screen all of the potential mouse brain proteins resulted in approximately 10^4 "positive" hits (yeast colonies carrying sequences coding for interacting proteins). Theoretically, one-sixth (16.7%) of the cloned mouse library cDNAs should be in the right orientation and reading frame[67]. However, 35 of the 65 sequences sampled from that large pool of hits (54%) were in-frame inserts, coding for known mouse proteins. This indicates a non-random selection for actual proteins in this screen. Unfortunately, none of the hits corresponded to any part of mouse myosin V and there did not appear to be any consensus protein, or protein family, identified as a target of PAI-RBP1.

Of the 65 sequenced hits, 23% were identical to, or partially overlapped with, other hits in the pool of sampled sequences. One set of five identical sequences, the Y2H-006 cluster (Table 1), is derived from a ribosomal protein pseudogene sequence and is inserted out-of-frame. The resulting peptide has no homology to any known protein. The overrepresentation of this sequence may have resulted from replication and transmission to daughter yeast cells during the mating step of the Y2H screen, but it is apparent that this sequence yields a peptide that is either "sticky" or mimics the shape of one of the natural ligands for PAI-RBP1.

Four mouse proteins were each represented twice in the sampled sequences. The Y2H-017/055 sequence codes for a portion of the 40S ribosomal protein S20; the Y2H-033/107 sequence codes for part of the arginine N-methyl transferase 9; Y2H-127/094

sequence codes for a peptide from an ATPase; and the Y2H-181/068 sequence codes for a portion of the mouse epididymal secretory protein E1 precursor. The primary sequences of these peptides bear no resemblance to each other, nor do they show any resemblance to any primary sequence derived from a mouse myosin V gene. Determining if any similarity in tertiary structures exists between these peptides is beyond the scope of this study. However, these interacting proteins may be of interest in future studies dedicated to defining functions of PAI-RBP1.

Despite the fact that the yeast two-hybrid system is a powerful method for screening interactions of a bait with other partners, the technique has potential artifacts that may have affected interactions between the PAI-RPB1 bait and potential prey proteins in the library. The generation of false positives due to non-specific interactions appears to be the most common problem associated with the yeast two hybrid screen in this study. False positives may arise through unnatural folding of the bait, resulting in abnormal shape and a dysfunctional protein that may aggregate or "stick" to random proteins[68, 69].

False negative results may also occur in yeast two hybrid screens. These types of artifacts may result from interactions of the bait (or prey) with an "unnatural environment," namely yeast cytoplasm[70]. Small molecules within the yeast cytosolic environment can indirectly prevent the bait from interacting with its partner through binding of the active site and/or altering the folding pattern of the protein so that normal interactions are inhibited[71, 72]. In addition, some proteins may only interact with the bait when they are membrane-anchored[73, 74]. Other false negative results may arise when the

expressed protein is not imported back to the nucleus[44,55]. The nucleus is not the natural environment for the fusion peptides, and their translocation into the nucleus is controlled by integral membrane proteins and many other soluble nuclear proteins that interact with the nuclear localization signal on the peptide[75–77].

This study showed that the yeast two-hybrid screening system was not an effective tool for identifying binding partners for the PAI-RBP1 bait, particularly when the bait protein is an artificial protein consisting of a short piece of PAI-RBP1 protein that has not been fully characterized. The Y2H screening may be revisited when a region of myosin V, identified as a potential start for a truncated product, can be used as a bait. Further work is required to use a proteomics approach in finding myosin V binding partners. A different mouse homologue to Vig-1 could be used; the protein structure resulting from cloned PAI-RBP1 fragment may not represent the native folding pattern that actually interacts with a myosin V peptide. Therefore, future Y2H experiments may employ longer segments of PAI-RBP1 as a bait. Alternatively, co-immunoprecipitation may also be employed to precipitate the motorless myosin V with its interacting proteins.

Multiple Sequence Alignment (MSA)

MSA is a powerful method that compares genomic regions of related organisms in order to identify conserved elements that include protein-coding, structural and regulatory sequences. Previous studies have effectively used MSAs to identify important regulatory sequences in the human genome[78]. In this study, alignments of mammalian intronic sequences from each of the myosin V genes (MYO5A, MYO5B, and MYO5C)

identified a region in intron 18 of Myo5A that is highly conserved between all five mammals included in the alignment. This sequence is transcribed and present in the pool of mouse brain RNA, as indicated by reverse transcription (RT) and nested PCR, but was only observed when the forward, amplifying primer occurred within the conserved sequence. Amplicons generated from forward primers placed upstream from intron 18 never contained the conserved intronic sequence, presumably because it was spliced out during mRNA processing. These results suggest that the conserved region is at the beginning of a transcribed product that occurs midway through the mouse MYO5A gene, similar to the *hum-2C* scenario in *C. elegans*.

The Genbank database contains a record of a similar mouse transcript, the X7 isoform of Myo5A (RefSeq# XM_006510834), that is predicted from EST and mRNA data. This isoform begins with part of the conserved intron 18 sequence identified in this study, and it is spliced directly to the exon 19 sequence. All of the subsequent downstream sequence in this predicted transcript is derived from known Myo5A exons. The predicted X7 isoform does not begin with a start codon, and not all of the intron 18 sequence identified in this study is contained in the X7 transcript. In fact, the predicted X7 transcript skips over the stop codons that are present in the conserved region. It is unclear if there is experimental evidence for this predicted splicing or if the Gnomon software utilized by NCBI and Genbank simply spliced these stop codons out in order to create a continuous reading frame in the predicted transcript.

This study used publicly available software to identify putative promoters and transcription factor binding sites in intron 18, both upstream from and within the

conserved region. Promoters and transcription factor binding sites play important roles in the regulation of gene expression, controlling such processes as cell-cycle progression[79], T cell development[80], and early mouse embryogenesis[81]. The Bhlhe40 transcription factor binding site identified in this study binds a member of the basic helix–loop–helix protein family that has a regulatory role in different biological processes, including apoptosis, cellular growth, cell proliferation, immune response, and the regulation of circadian rhythms[82–85]. A putative upstream promoter region was also identified in the intron 18 sequence, but it is not clear if it is associated with a downstream transcription initiation site.

Due to its biological significance, the bioinformatics analyses have focused on the transcription initiation site as a feature in developing advanced computational algorithms for predicting and distinguishing poorly defined consensus sequences that are often associated with transcription initiation sites. The computational identification of promoter sequences relies on the assumption that a transcription start site is located downstream from the regulatory region and that a TATA box (a general transcription factor-binding site) is located 25 to 35 base pairs upstream of the transcription start site[86]. A number of short AT-rich sequences within the conserved region of intron 18 could be interpreted as TATA boxes.

Any predicted peptide that could be derived from the intron 18 sequence would have no similarity to the *C. elegans* HUM-2C protein or any other protein in the database of known proteins (data not shown). However, the MYO5A protein sequence encoded by exons downstream from intron 18 includes coiled-coils motifs, which have been shown to

participate in both homo- and heterodimerization[87]. This is the region that leads to homodimerization of the full MYO5A protein[88], and its inclusion in a predicted, truncated product indicates that the truncated product would also exist as a homodimer.

The sequence features described above seem to indicate that most, if not all, of the conserved region in intron 18 plays a regulatory role, acting as a promoter and a 5′ untranslated region (UTR) involved in driving the expression of a downstream protein-coding sequence. Unfortunately, there is still no empirical evidence for the *in situ* existence of a truncated myosin V in mammals. However, there is anecdotal evidence for a truncated myosin VB product on gels, for which no explanation was provided[89]. The MSA analyses in the current study did not identify any non-coding Myo5b sequences with the level of conservation demonstrated in the Myo5A intron 18. Experimental evidence for a truncated Myo5A protein may best be approached in the future with western blots using monoclonal antibodies directed against the Myo5A tail region.

Finally, there is a large evolutionary gap between mammals and nematodes, the two groups represented by the model organisms in this research. Narrowing this evolutionary gap by examining taxonomic groups more closely related to *C. elegans* will be considered in future experimental approaches to determining how widespread motorless versions of myosin V may be in the Eukarya domain. Model organisms from the arthropods (insects), annelids (segmented worms), planaria (flat worms), and mollusks are examples of organisms that might be used to find a homologous version of the *C. elegans HUM-2C* protein.

5. References

1. Gibbons, I. R. & Rowe, A. J. Dynein: A protein with adenosine triphosphatase activity from cilia. *Science* **149,** 424–426 (1965).

2. Slep, K. C., Rogers, S. L., Elliott, S. L., Ohkura, H., Kolodziej, P. A. & Vale, R. D. Structural determinants for EB1-mediated recruitment of APC and spectraplakins to the microtubule plus end. *J. Cell Biol.* **168,** 587–598 (2005).

3. Kudryashov, D. S., Sawaya, M. R., Adisetiyo, H., Norcross, T., Hegyi, G., Reisler, E. & Yeates, T. O. The crystal structure of a cross-linked actin dimer suggests a detailed molecular interface in F-actin. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 13105–13110 (2005).

4. Oda, T., Iwasa, M., Aihara, T., Maéda, Y. & Narita, A. The nature of the globular-to fibrous-actin transition. *Nature* **457,** 441–445 (2009).

5. Mooseker, M. S. & Cheney, R. E. Unconventional myosins. *Annu. Rev. Cell Dev. Biol.* **11,** 633–675 (1995).

6. Richards, T. A. & Cavalier-Smith, T. Myosin domain evolution and the primary divergence of eukaryotes. *Nature* **436,** 1113–1118 (2005).

7. Berg, J. S., Powell, B. C. & Cheney, R. E. A millennial myosin census. *Mol. Biol. Cell* **12,** 780–794 (2001).

8. Krendel, M. & Mooseker, M. S. Myosins: tails (and heads) of functional diversity. *Physiology* **20,** 239–251 (2005).

9. Hodge, T., Jamie, M. & Cope, T. V. A myosin family tree. *J. Cell Sci.* **113,** 3353–3354 (2000).

10. Sellers, J. R. Myosins: a diverse superfamily. *Biochim. Biophys. Acta* **1496,** 3–22 (2000).

11. Reddy, A. S. & Day, I. S. Analysis of the myosins encoded in the recently completed *Arabidopsis thaliana* genome sequence. *Genome Biol.* **2,** research0024–1 (2001).

12. Volkmann, D., Mori, T., Tirlapur, U. K., König, K., Fujiwara, T., Kendrick-Jones, J. & Baluška, F. Unconventional myosins of the plant-specific class VIII: endocytosis, cytokinesis, plasmodesmata/pit-fields, and cell-to-cell coupling. *Cell Biol. Int.* **27,** 289–291 (2003).

13. Kipreos, E. T. *C. elegans* cell cycles: invariance and stem cell divisions. *Nat. Rev. Mol. Cell Biol.* **6,** 766–776 (2005).

14. Oiwa, K. & Kojima, H. Force-generating mechanisms of dynein revealed through single molecule studies. In: Knight, A. (ed.) *Single Molecule Biology*. pp 61–104. San Diego, CA, United States: Academic Press, (2009).

15. Pestic-Dragovich, L., Stojiljkovic, L., Philimonenko, A. A., Nowak, G., Ke, Y., Settlage, R. E., Shabanowitz, J., Hunt, D. F., Hozak, P. & de Lanerolle, P. A myosin I isoform in the nucleus. *Science* **290,** 337–341 (2000).

16. Vreugde, S., Ferrai, C., Miluzio, A., Hauben, E., Marchisio, P. C., Crippa, M. P., Bussi, M. & Biffo, S. Nuclear myosin VI enhances RNA polymerase II-dependent transcription. *Mol. Cell* **23,** 749–755 (2006).

17. Li, Q. & Sarna, S. K. Nuclear myosin II regulates the assembly of preinitiation complex for ICAM-1 gene transcription. *Gastroenterology* **137,** 1051–1060 (2009).

18. Lindsay, A. J. & McCaffrey, M. W. Myosin Vb localises to nucleoli and associates with the RNA polymerase I transcription complex. *Cell Motil. Cytoskeleton* **66,** 1057–1072 (2009).

19. Sellers, J. R. *Myosin*. Oxford, United Kingdom: Oxford University Press, (1999).

20. Wu, X., Wang, F., Rao, K., Sellers, J. R. & Hammer, J. A. Rab27a is an essential component of melanosome receptor for myosin Va. *Mol. Biol. Cell* **13,** 1735–1749 (2002).

21. Vale, R. D. The molecular motor toolbox for intracellular transport. *Cell* **112,** 467–480 (2003).

22. Wu, X., Xiang, X. & Hammer, J. A. Motor proteins at the microtubule plus-end. *Trends Cell Biol.* **16,** 135–143 (2006).

23. Hammer, J. A. & Sellers, J. R. Walking to work: roles for class V myosins as cargo transporters. *Nat. Rev. Mol. Cell Biol.* **13,** 13–26 (2012).

24. Li, X. D., Jung, H. S., Mabuchi, K., Craig, R. & Ikebe, M. The globular tail domain of myosin Va functions as an inhibitor of the myosin Va motor. *J. Biol. Chem.* **281,** 21789–21798 (2006).

25. Lu, H., Krementsova, E. B. & Trybus, K. M. Regulation of myosin V processivity by calcium at the single molecule level. *J. Biol. Chem.* **281,** 31987–31994 (2006).

26. Sellers, J. R., Thirumurugan, K., Sakamoto, T., Hammer, J. A. & Knight, P. J. Calcium and cargoes as regulators of myosin 5a activity. *Biochem. Biophys. Res. Commun.* **369,** 176–181 (2008).

27. Cheney, R. E., O'Shea, M. K., Heuser, J. E., Coelho, M. V., Wolenski, J. S., Espreafico, E. M., Forscher, P., Larson, R. E. & Mooseker, M. S. Brain myosin-V is a two-headed unconventional myosin with motor activity. *Cell* **75,** 13–23 (1993).

28. Sellers, J. R. & Veigel, C. Walking with myosin V. *Curr. Opin. Cell Biol.* **18,** 68–73 (2006).

29. Seperack, P. K., Mercer, J. A., Strobel, M. C., Copeland, N. G. & Jenkins, N. A. Retroviral sequences located within an intron of the dilute gene alter dilute expression in a tissue-specific manner. *Embo J.* **14,** 2326–2332 (1995).

30. Wagner, W., Fodor, E., Ginsburg, A. & Hammer, J. A. The binding of DYNLL2 to myosin Va requires alternatively spliced exon B and stabilizes a portion of the myosin's coiled-coil domain. *Biochemistry* **45,** 11564–11577 (2006).

31. Brown, J. R., Stafford, P. & Langford, G. M. Short-range axonal/dendritic transport by myosin-V: A model for vesicle delivery to the synapse. *J. Neurobiol.* **58,** 175–188 (2004).

32. Prekeris, R. & Terrian, D. M. Brain myosin V is a synaptic vesicle-associated motor protein: evidence for a Ca^{2+}-dependent interaction with the synaptobrevin-synaptophysin complex. *J. Cell Biol.* **137,** 1589–1601 (1997).

33. Rosé, S. D., Lejen, T., Casaletti, L., Larson, R. E., Pene, T. D. & Trifaró, J. M. Myosins II and V in chromaffin cells: myosin V is a chromaffin vesicle molecular motor involved in secretion. *J. Neurochem.* **85,** 287–298 (2003).

34. Dekker-Ohno, K., Hayasaka, S., Takagishi, Y., Oda, S. I., Wakasugi, N., Mikoshiba, K., Inouye, M. & Yamamura, H. Endoplasmic reticulum is missing in dendritic spines of Purkinje cells of the ataxic mutant rat. *Brain Res.* **714,** 226–230 (1996).

35. Takagishi, Y., Oda, S. I., Hayasaka, S., Dekker-Ohno, K., Shikata, T., Inouye, M. & Yamamura, H. The *dilute-lethal* (d^l) gene attacks a Ca^{2+} store in the dendritic spine of Purkinje cells in mice. *Neurosci. Lett.* **215,** 169–172 (1996).

36. Lapierre, L. A., Kumar, R., Hales, C. M., Navarre, J., Bhartur, S. G., Burnette, J. O., Provance, D. W., Mercer, J. A., Bähler, M. & Goldenring, J. R. Myosin Vb is associated with plasma membrane recycling systems. *Mol. Biol. Cell* **12,** 1843–1857 (2001).

37. Volpicelli, L. A., Lah, J. J., Fang, G., Goldenring, J. R. & Levey, A. I. Rab11a and Myosin Vb regulate recycling of the $m_4$muscarinic acetylcholine receptor. *J. Neurosci.* **22,** 9776–9784 (2002).

38. Rao, M. V., Engle, L. J., Mohan, P. S., Yuan, A., Qiu, D., Cataldo, A., Hassinger, L., Jacobsen, S., Lee, V. M., Andreadis, A. & Julien, J.P. Myosin Va binding to neurofilaments is essential for correct myosin Va distribution and transport and neurofilament density. *J. Cell Biol.* **159,** 279–290 (2002).

39. Ohashi, S., Koike, K., Omori, A., Ichinose, S., Ohara, S., Kobayashi, S., Sato, T. A. & Anzai, K. Identification of mRNA/protein (mRNP) complexes containing Purα, mStaufen, fragile X protein, and myosin Va and their association with rough endoplasmic reticulum equipped with a kinesin motor. *J. Biol. Chem.* **277,** 37804–37810 (2002).

40. Baker, J. P. & Titus, M. A. A family of unconventional myosins from the nematode *Caenorhabditis elegans*. *J. Mol. Biol.* **272,** 523–535 (1997).

41. Fields, S. An endogenous, motorless isoform of *C. elegans* myosin V has a distinctive expression pattern and acts as a neuronal cargo adaptor. Unpublished manuscript (2017) and personal communication.

42. Wormbase. (2017, May 5). Retrieved from http://www.wormbase.org

43. Shin, K. H., Choi, B., Park, Y. S. & Cho, N. J. Analysis of *C. elegans* VIG-1 expression. *Mol. Cells* **26,** 554–557 (2008).

44. Fields, S. & Song, O. K. A novel genetic system to detect protein-protein interactions. *Nature* **340,** 245–246 (1989).

45. Phizicky, E. M. & Fields, S. Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* **59,** 94–123 (1995).

46. Gavin, A. C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C.M. & Remor, M. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415,** 141–147 (2002).

47. Suter, B., Kittanakom, S. & Stagljar, I. Two-hybrid technologies in proteomics research. *Curr. Opin. Biotechnol.* **19,** 316–323 (2008).

48. Estojak, J., Brent, R. & Golemis, E. A. Correlation of two-hybrid affinity data with in vitro measurements. *Mol. Cell. Biol.* **15,** 5820–5829 (1995).

49. Van Criekinge, W. & Beyaert, R. Yeast two-hybrid: state of the art. *Biol. Proced. Online* **2,** 1–38 (1999).

50. Edgar, R. C. & Batzoglou, S. Multiple sequence alignment. *Curr. Opin. Struct. Biol.* **16,** 368–373 (2006).

51. Dubchak, I. & Frazer, K. Multi-species sequence comparison: the next frontier in genome annotation. *Genome Biol.* **4,** 122.1–122.6 (2003).

52. Portales-Casamar, E., Arenillas, D., Lim, J., Swanson, M. I., Jiang, S., McCallum, A., Kirov, S. & Wasserman, W. W. The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences. *Nucleic Acids Res.* **37,** D54–D60 (2008).

53. Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. & Haussler, D. The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).

54. Bryne, J. C., Valen, E., Tang, M. H. E., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B. & Sandelin, A. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.* **36,** D102–D106 (2007).

55. Knudsen, S. Promoter2.0: for the recognition of Pol III promoter sequences. *Bioinformatics* **15,** 356–361 (1999).

56. Reese, M. G. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput. Chem.* **26,** 51–56 (2001).

57. Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252,** 1162–1164 (1991).

58. Lee, T. Y., Chang, W. C., Hsu, J. B. K., Chang, T. H. & Shien, D. M. GPMiner: an integrated system for mining combinatorial cis-regulatory elements in mammalian gene group. *BMC Genom.* **13,** S3 (2012).

59. Aken, B. L., Achuthan P., Akanni, W., Amode, M. R., Bernsdorff, F., Bhai, J., Billis, K. Carvalho-Silva, D., Cummins, C., Clapham, P., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Juettemann, T., Keenan, S., Laird, M. R., Lavidas, I., Maurel, T., McLaren, W., Moore, B., Murphy, D. N., Nag, R., Newman, V., Nuhn, M., Ong, C. K., Parker, A., Patricio, M., Riat, H. S., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Wilder, S. P., Zadissa, A., Kostadima, M., Martin, F. J., Muffato, M., Perry, E., Ruffier, M., Staines, D. M., Trevanion, S. J., Cunningham, F., Yates, A., Zerbino, D. R. & Flicek, P. Ensembl 2017. *Nucleic Acids Res*. **45,** D635–D642 (2017).

60. Caudy, A. A., Myers, M., Hannon, G. J. & Hammond, S. M. Fragile X-related protein and VIG associate with the RNA interference machinery. *Gene Dev.* **16,** 2491–2496 (2002).

61. Heaton, J. H., Dlakic, W. M., Dlakic, M. & Gelehrter, T. D. Identification and cDNA cloning of a novel rna-binding protein that interacts with the cyclic nucleotide-responsive sequence in the type-1 plasminogen activator inhibitor mRNA. *J. Biol. Chem.* **276,** 3341–3347 (2001).

62. Heaton, J. H., Dlakic, W. M. & Gelehrter, T. D. Posttranscriptional regulation of PAI-1 gene expression. *Thromb. Haemost.* **89,** 959–966 (2003).

63. Tillmann-Bogush, M., Heaton, J. H. & Gelehrter, T. D. Cyclic nucleotide regulation of PAI-1 mRNA stability identification of cytosolic proteins that interact with an A-Rich sequence. *J. Biol. Chem.* **274,** 1172–1179 (1999).

64. Peluso, J. J., Yuan, A., Liu, X. & Lodde, V. Plasminogen activator inhibitor 1 RNA-binding protein interacts with progesterone receptor membrane component 1 to regulate progesterone's ability to maintain the viability of spontaneously immortalized granulosa cells and rat granulosa cells. *Biol. Reprod.* **88,** 1–10 (2013).

65. Walther, C., Tayebwa, J., Lilljebjörn, H., Magnusson, L., Nilsson, J., von Steyern, F. V., Øra, I., Domanski, H. A., Fioretos, T., Nord, K. H. & Fletcher, C. D. A novel SERPINE1-FOSB fusion gene results in transcriptional up-regulation of FOSB in pseudomyogenic haemangioendothelioma. *J. Pathol.* **232,** 534–540 (2014).

66. Ruffle, J. K. Molecular neurobiology of addiction: what's all the (Δ) FosB about? *Am. J. Drug Alcohol Abuse* **40,** 428–437 (2014).

67. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P. & Qureshi-Emili, A. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403,** 623–627 (2000).

68. Serebriiskii, I., Estojak, J., Berman, M. & Golemis, E. A. Approaches to detecting false positives in yeast two-hybrid systems. *BioTechniques* **28,** 328–337 (2000).

69. Bartel, P. L., Chien, C. T., Sternglanz, R. & Fields, S. Using the two-hybrid system to detect protein-protein interactions. In: Hartley, D. A. (ed.). *Cellular Interactions in Development: A Practical Approach*. pp. 153–179. Oxford, England: Oxford University Press, (1993).

70. Joshi, P. B., Hirst, M., Malcolm, T., Parent, J., Mitchell, D., Lund, K. & Sadowski, I. Identification of protein interaction antagonists using the repressed transactivator two-hybrid system. *BioTechniques* **42,** 635–644 (2007).

71. Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev. Drug Discovery* **3,** 301–317 (2004).

72. Archakov, A. I., Govorun, V. M., Dubanov, A. V., Ivanov, Y. D., Veselovsky, A. V., Lewi, P. & Janssen, P. Protein-protein interactions as a target for drugs in proteomics. *Proteomics* **3,** 380–391 (2003).

73. Hubsman, M., Yudkovsky, G. & Aronheim, A. A novel approach for the identification of protein-protein interaction with integral membrane proteins. *Nucleic Acids Res.* **29,** e18–e18 (2001).

74. Broder, Y. C., Katz, S. & Aronheim, A. The Ras recruitment system, a novel approach to the study of protein-protein interactions. *Curr. Biol.* **8,** 1121–1130 (1998).

75. Borg, J. P., Marchetto, S., Le Bivic, A., Ollendorff, V., Jaulin-Bastard, F., Saito, H., Fournier, E., Adélaïde, J., Margolis, B. & Birnbaum, D. ERBIN: a basolateral PDZ protein that interacts with the mammalian ERBB2/HER2 receptor. *Nat. Cell Biol.* **2,** 407–414 (2000).

76. Sugita, S., Hata, Y. & Südhof, T. C. Distinct Ca-dependent properties of the first and second C-domains of synaptotagmin I. *J. Biol. Chem*. **271,** 1262–1265 (1996).

77. Niethammer, M., Kim, E. & Sheng, M. Interaction between the C terminus of NMDA receptor subunits and multiple members of the PSD-95 family of membrane-associated guanylate kinases. *J. Neurosci.* **16,** 2157–2163 (1996).

78. McLean, C. Y., Reno, P. L., Pollen, A. A., Bassan, A. I., Capellini, T. D., Guenther, C., Indjeian, V. B., Lim, X., Menke, D. B., Schaar, B. T. & Wenger, A. M. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471,** 216–219 (2011).

79. Yang, Z. F., Mott, S. & Rosmarin, A. G. The Ets transcription factor GABP is required for cell-cycle progression. *Nat. Cell Biol.* **9,** 339–346 (2007).

80. Yu, S., Zhao, D. M., Jothi, R. & Xue, H. H. Critical requirement of GABPα for normal T cell development. *J. Biol. Chem.* **285,** 10179–10188 (2010).

81. Ristevski, S., O'Leary, D. A., Thornell, A. P., Owen, M. J., Kola, I. & Hertzog, P. J. The ETS transcription factor GABPα is essential for early embryogenesis. *Mol. Cell Biol.* **24,** 5844–5849 (2004).

82. Honma, S., Kawamoto, T., Takagi, Y., Fujimoto, K., Sato, F., Noshiro, M., Kato, Y. & Honma, K. I. Dec1 and Dec2 are regulators of the mammalian molecular clock. *Nature* **419,** 841–844 (2002).

83. Miyazaki, K., Miyazaki, M., Guo, Y., Yamasaki, N., Kanno, M., Honda, Z. I., Oda, H., Kawamoto, H. & Honda, H. The role of the basic helix-loop-helix transcription factor Dec1 in the regulatory T cells. *J. Immunol.* **185,** 7330–7339 (2010).

84. Seuter, S., Pehkonen, P., Heikkinen, S. & Carlberg, C. The gene for the transcription factor BHLHE40/DEC1/stra13 is a dynamically regulated primary target of the vitamin D receptor. *J. Steroid Biochem.* **136,** 62–67 (2013).

85. Sun, H., Lu, B., Li, R. Q., Flavell, R. A. & Taneja, R. Defective T cell activation and autoimmune disorder in Stra13-deficient mice. *Nat. Immunol.* **2,** 1040–1047 (2001).

86. Mahdi, R. N. & Rouchka, E. C. RBF-TSS: identification of transcription start site in human using radial basis functions network and oligonucleotide positional frequencies. *PloS one* **4,** e4878 (2009).

87. Li, C., Ching Han Chang, C., Nagel, J., Porebski, B. T., Hayashida, M., Akutsu, T., Song, J. & Buckle, A. M. Critical evaluation of in silico methods for prediction of coiled-coil domains in proteins. *Brief. Bioinform.* **17,** 270–282 (2015).

88. Espreafico, E. M., Cheney, R. E., Matteoli, M., Nascimento, A. A., De Camilli, P. V., Larson, R. E. & Mooseker, M. S. Primary structure and cellular localization of chicken brain myosin-V (p190), an unconventional myosin with calmodulin light chains. *J. Cell Biol.* **119,** 1541–1557 (1992).

89. Rodriguez, O. C. & Cheney, R. E. Human myosin-Vc is a novel class V myosin expressed in epithelial cells. *J. Cell Sci.* **115,** 991–1004 (2002).

Appendices

A. Reagents, supplies, and equipment

Table A1. Reagents used during thesis research with their supplemental information

Chemical and kits	Manufacturer	Lot Number	Notes
1 Kb DNA Ladder	New England BioLabs Inc., USA	1291506	DNA molecular weight marker
100 bp DNA Ladder	New England BioLabs Inc., USA	1091506	DNA molecular weight marker
Ambion® Nuclease-Free Water	Thermo Fisher, USA	1605210	Used for all PCR reactions
Ampicillin	Sigma-Aldrich, France	80K0784	
Bullseye Taq Plus Master Mix (2X)	MIDSCI, USA	010522	PCR reaction master mix
ChargeSwitch® PCR clean-up kit	Invitrogen, USA	1878522	Used to purify PCR products
HS Taq polymerase, 2X Master Mix Blue	MIDSCI, USA	160307E	PCR reaction master mix
In-Fusion® HD Cloning Kit	Clontech Laboratories, Inc., USA	1603367A	
Kanamycin	Sigma-Aldrich, China	SLBH9905V	
NEB® PCR Cloning Kit	New England BioLabs Inc., USA	12025	Utilizes the pMiniT vector
Ribonuclease A	Sigma, USA	SLBL6452V	
RNAlater RNA Stabilization Reagent	QIAGEN, Germany	151052596	
SimpleSeq™ Kit	Eurofins Genomics, USA	NA	DNA sequencing reaction mailer
Taq® 2X Master Mix	New England BioLabs Inc., USA	0291706	PCR reaction master mix

Table A2. Equipment used during thesis research with their supplemental information

Appliance	Manufacturer	Model	Notes
6-Tubes Magnetic Separation Rack	New England BioLabs Inc., USA	S1506S	Used to purify PCR product procedures
Applied Biosystems™ Thermocycler	Thermo Fisher, Singapore	2720 Thermal Cycler	PCR and RT-PCR protocols
Benchtop Centrifuge	Hettich, Germany	rotofix 32a	
Compound microscope	Olympus, Philippines	CX31RBSFA	
ENDURO™ Gel XL Electrophoresis System	Labnet, USA	E0160	Agarose gel electrophoresis processes during this research
FluorChem Imagers	ProteinSimple, USA	FluorChem M system	Agarose gel images documentation
Freezer	Kenmore, USA	22042	
Hemocytometer	Hausser Scientific, USA,	Bright-Line 3110	
Microbiological Incubator	Thermo Fisher, Germany	IGS 180	Used to incubate the yeast at 30°C and the bacteria at 37°C
Microcentrifuge	Thermo Fisher, Germany	accuSpin Micro 17R	
Precision™ Balance	Fisher Science Education, United Kingdom	AMF 1202	For weight determination
Rotatory Shaker	Benchmark Scientific, Inc., USA	INCU-SHAKER™ 10L	
Spectrophotometer	Thermo Scientific, Germany,	NanoDrop 2000	Determination of DNA/RNA concentration, $A_{260/280}$
Spectrophotometer	Thermo Scientific, Germany	NanoDrop 2000c	Determination of Yeast OD_{600}

Table A3. Organisms and microorganisms used during thesis research with their supplemental information

Organism / Strain	Source	Notes/Genotype
E. coli / DH5α	New England BioLabs Inc., USA	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
Mouse / C57BL/6J	Biological Sciences Department at Emporia State University	IACUC number 15-006 under Dr. Tim G. Burnett, Associate Professor and the chair of the department 2610307O08Rik; 2610307O08Rik; ERIS; MITA; MPYS; NET23; RGD1562552; RIKEN cDNA 2610307O08 gene; SAVI; STING; Sting; hMITA; hSTING; rSTING
S. cerevisiae / Y187	Clontech Laboratories, Inc., USA	*MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, gal80Δ, met–, URA3 : : GAL1$_{UAS}$–Gal1$_{TATA}$–LacZ, MEL1*
S. cerevisiae / Y2HGOLD	Clontech Laboratories, Inc., USA	*MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2 : : GAL1$_{UAS}$–Gal1$_{TATA}$–His3, GAL2$_{UAS}$–Gal2$_{TATA}$–Ade2 URA3 : : MEL1$_{UAS}$–Mel1$_{TATA}$ AUR1-C MEL1*

Table A4. Genomic sequences for class V myosin genes

Species	Gene	Ensembl number
Bos taurus	MYO5A	ENSBTAG00000006489
Bos taurus	MYO5B	ENSBTAG00000019455
Bos taurus	MYO5C	ENSBTAG00000003763
Canis lupus	MYO5A	ENSCAFG00000015800
Canis lupus	MYO5B	ENSCAFG00000018982
Canis lupus	MYO5C	ENSCAFG00000015680
Homo sapiens	MYO5A	ENSG00000197535
Homo sapiens	MYO5B	ENSG00000167306
Homo sapiens	MYO5C	ENSG00000128833
Macaca mulatta	MYO5A	ENSMMUG00000007509
Macaca mulatta	MYO5B	ENSMMUG00000003041
Macaca mulatta	MYO5C	ENSMMUG00000007504
Mus musculus	MYO5A	ENSMUSG00000034593
Mus musculus	MYO5B	ENSMUSG00000025885
Mus musculus	MYO5C	ENSMUSG00000033590

Table A5. Software and tools used during thesis research

Software and Webtools	Source	Version	Baseed
BLAST	NCBI	2.7.0	Online
Clustal Omega	EMBL-EBI	1.2.4	Online
Jaspar	Not applicable	7	Online
LALIGN	ExPASy	NA	Online
Microsoft Office 2016	Microsoft Corporation	1708	Offline
NANODROP 2000/2000c	Thermo Scientific	1.5	Offline
Notepad	Microsoft Corporation	1703	Offline
Oligo Analyzer 3.1	Integrated DNA Technologies	3.1	Online
Promoter 2.0 Prediction Server	The Center for Biological Sequence Analysis at the Technical University of Denmark	2	Online
Promoter Prediction by Neural Network	BDGP	2.2	Online
Translate tool	ExPASy	NA	Online

B. Microbial growth media and agarose gel recipes

B.1 Preparation of *E. coli* growth media

Table B1. Lysogeny broth (LB) Miller's

Component	Source	Lot#	w/v per L
LB media	Growcells, USA	G14-04	25 g

Table B2. LB-agar / kanamycin

Component	Source	Lot#	w/v per L
LB media	Growcells, USA	G14-04	25 g
Agar	Sigma, USA	SLBK5425V	20 g
Kan 50µg/ml	Sigma, China	SLBH9905V	1 ml

Table B3. LB-agar / ampicillin

Component	Source	Lot#	w/v per L
LB media	Growcells, USA	G14-04	25 g
Agar	Sigma, USA	SLBK5425V	20 g
Amp 50µg/ml	Sigma-Aldrich, France	80K0784	1 ml

B.2 Preparation of *S. cerevisiae* growth media

Table B4. Yeast minimal media: SD -Trp

Component	Source	Lot#	w/v per L
Minimal SD Base	Clontech Laboratories, Inc., USA	2926C152	26.7 g
-Trp DO Supplement	Clontech Laboratories, Inc., USA	2740C288	0.74 g
Agar	Sigma, USA	SLBK5425V	20 g

Table B5. Yeast minimal media: SD -Leu

Component	Source	Lot#	w/v per L
Minimal SD Base	Clontech Laboratories, Inc., USA	2926C152	26.7 g
-Leu DO Supplement	Clontech Laboratories, Inc., USA	0467B016	0.69 g
Agar	Sigma, USA	SLBK5425V	20 g

Table B6. Yeast minimal media: SD -Leu/-Trp

Component	Source	Lot#	w/v per L
Minimal SD Base	Clontech Laboratories, Inc., USA	2926C152	26.7 g
-Leu/-Trp DO Supplement	Clontech Laboratories, Inc., USA	1838B399	0.64 g
Agar	Sigma, USA	SLBK5425V	20 g

Table B7. Yeast minimal media SD: -Ade/-His/-Leu/-Trp

Component	Source	Lot#	w/v per L
Minimal SD Base	Clontech Laboratories, Inc., USA	2926C152	26.7 g
-Ade/-His/-Leu/-Trp DO Supplement	Clontech Laboratories, Inc., USA	1539B233	0.60 g
Agar	Sigma, USA	SLBK5425V	20 g

Table B8. YPDA yeast agar

Component	Source	Lot#	w/v per L
Bacto yeast extract	DIFCO, USA	0127-02	20 g
Bacto peptone	DIFCO, USA	0118-01-8	40 g
Glucose monohydrate	Fisher Science Education, USA	AD-8162-28	40 g
Adenine hemisulfate	MP Biomedicals, USA	2542KA	80 mg
Agar	Sigma, USA	SLBK5425V	20 g

Table B9. YPDA yeast broth

Component	Source	Lot#	w/v per L
Bacto yeast extract	DIFCO, USA	0127-02	20 g
Bacto peptone	DIFCO, USA	0118-01-8	40 g
Glucose monohydrate	Fisher Science Education, USA	AD-8162-28	40 g
Adenine hemisulfate	MP Biomedicals, USA	2542KA	80 mg

Table B10. 0.5x YPDA broth

Component	Source	Lot#	w/v per L
Bacto yeast extract	DIFCO, USA	0127-02	10 g
Bacto peptone	DIFCO, USA	0118-01-8	20 g
Glucose monohydrate	Fisher Science Education, USA	AD-8162-28	20 g
Adenine hemisulfate	MP Biomedicals, USA	2542KA	40 mg

Table B11. 2.0x YPDA broth

Component	Source	Lot#	w/v per L
Bacto yeast extract	DIFCO, USA	0127-02	40 g
Bacto peptone	DIFCO, USA	0118-01-8	80 g
Glucose monohydrate	Fisher Science Education, USA	AD-8162-28	80 g
Adenine hemisulfate	MP Biomedicals, USA	2542KA	160 mg

B.3 Preparation of Agarose Gel

Table B12. 1% agarose gel

Component	Source	Lot#	w/v per L of 0.5X TAE
Agarose	Sigma, USA	SLBN6401V	10 g
5 µg/ml EtBr	Sigma, USA	SLBF7130V	50 µl

Table B13. 2% agarose gel

Component	Source	Lot#	w/v per L of 0.5X TAE
Agarose	Sigma, USA	SLBN6401V	20 g
10 µg/ml EtBr	Sigma, USA	SLBF7130V	100 µl

C. Polymerase chain reaction supplemental information

C.1 Primers

Stock concentration of all primers used in this research were at a concentration of 25 µM unless otherwise stated. Primers were designed using the Oligo Analyzer 3.1 online tool and synthesized by Integrated DNA Technologies. The following parameters were considered during the design process:

- Melting temperature (MT) between 57°C and 63°C.
- (G+C) ratio between 50-60%.
- Length between 18–28 bases.

- Melting temperature between two primers are close.
- One or two Cs or Gs at the 3′ ends but not three or more.
- Primers were analyzed for primer dimers and secondary structure formation with Oligo Analyzer.

Table C1. Primer information

#	Name	DIR	Primer sequence 5′- 3′
1	PAI-RBP1-cDNA	R	TATGGCATCCAGTTAGGCCAG
2	PAI-RBP1-OR	R	TTAGGCCAGAGCTGGGAAGG
3	PAI-RBP1-OF	F	GTAGCGGCTCTCACAACTGG
4	PAI-RBP1-CF	F	CATGGAGGCCGAATTCGCGGCTCTCACAACTGGGGAACTG
5	PAI-RBP1-CR	R	GCAGGTCGACGGATCCAGCTGGGAAGGCCTCTGGGTCATC
6	T7	F	TAATACGACTCACTATAGGGC
7	pGBK-R	R	TTTTCGTTTTAAAACCTAAGAGTC
8	pGAD-R	R	AGATGGTGCACGATGCACAG
9	MMyo5A-F2	F	ATGAGAAGAGGGCAGTGCAGC
10	MMyo5A-F3	F	GGTCTAGGTGGACTTACCAAGAG
11	MMyo5A-F5	F	GGTCAGTCAGAGGGAGTGTC
12	MMyo5A-F6	F	CCTTGAGGTGTTTAAAATCC
13	MMyo5A-F7	F	CAACCAGTTTTGGGCCTCAG
14	MMyo5A-R1	R	ATTCTGCTCATCCACTTTGCGC
15	MMyo5A-R2	R	GTTCTCCATGCCAATATGGAGC
16	MMyo5A-R3	R	TAGCGTTCCACAGAGCGAG
17	pMiniT F	F	ACCTGCCAACCAAAGCGAGAAC
18	pMiniT R	R	TCAGGGTTATTGTCTCATGAGCG

C.2 PCR Mixtures

Table C2. PAI-RBP1 cycle mixture

Component	vol per 25µL	Notes
DNA polymerase	12.5	Bullseye Taq Plus Master Mix (2X)
H_2O	9.5	Nuclease-Free Water
Forward Primer	1	PAI-RBP1-OF primer
Reverse Primer	1	PAI-RBP1-OR primer
template	1	

Table C3. *E. coli* colony PCR mixture #1

Component	vol per 25µL	Notes
DNA polymerase	12.5	Bullseye Taq Plus Master Mix (2X)
H_2O	9.5	Nuclease-Free Water
Forward Primer	1	T7 primer
Reverse Primer	1	PGBK-R
template	Portion of *E. coli* colony.	

Table C4. *E. coli* colony PCR mixture #2

Component	vol per 25µL	Notes
DNA polymerase	12.5	HS Taq polymerase, 2X Master Mix Blue
H_2O	10.5	Nuclease-Free Water
Forward Primer	1	pMiniT F primer
Reverse Primer	1	pMiniT R primer
template	Portion of *E. coli* colony.	

Table C5. Yeast colony PCR mixture

Component	vol per 25µL	Notes
DNA polymerase	12.5	HS Taq polymerase, 2X Master Mix Blue
H_2O	10.5	Ambion® Nuclease-Free Water
Forward Primer	1	T7 primer
Reverse Primer	1	pGAD-R primer
template	Portion of yeast colony.	

Table C6. PCR mixture #1

Component	vol per 25µL	Notes
DNA polymerase	12.5	Taq® 2X Master Mix
H_2O	9.5	Nuclease-Free Water
Forward Primer	1	MMyo5A-F6
Reverse Primer	1	MMyo5A-R2
Template	1	cDNA from MMyo5A-R1

Table C7. PCR mixture #2

Component	vol per 25µL	Notes
DNA polymerase	12.5	Taq® 2X Master Mix
H_2O	9.5	Nuclease-Free Water
Forward Primer	1	MMyo5A-F7
Reverse Primer	1	MMyo5A-R3
Template	1	Mixture #1 product

Table C8. PCR mixture #3

Component	vol per 25µL	Notes
DNA polymerase	12.5	Taq® 2X Master Mix
H_2O	9.5	Nuclease-Free Water
Forward Primer	1	MMyo5A-F5
Reverse Primer	1	MMyo5A-R2
Template	1	cDNA from MMyo5A-R1

Table C9. PCR mixture #4

Component	vol per 25µL	Notes
DNA polymerase	12.5	Taq® 2X Master Mix
H_2O	9.5	Nuclease-Free Water
Forward Primer	1	MMyo5A-F6
Reverse Primer	1	MMyo5A-R3
Template	1	Mixture #3 product

Table C10. PCR mixture #5

Component	vol per 25µL	Notes
DNA polymerase	12.5	Taq® 2X Master Mix
H_2O	9.5	Nuclease-Free Water
Forward Primer	1	MMyo5A-F2
Reverse Primer	1	MMyo5A-R2
Template	1	cDNA from MMyo5A-R1

Table C11. PCR mixture #6

Component	vol per 25µL	Notes
DNA polymerase	12.5	Taq® 2X Master Mix
H_2O	9.5	Nuclease-Free Water
Forward Primer	1	MMyo5A-F3
Reverse Primer	1	MMyo5A-R3
Template	1	Mixture 5 product

C.3 PCR Cycles

Table C12. PAI-RBP1 PCR cycle conditions

Step	PCR phase	Temperature	Time
1	Initial denaturation	94°C	2 minutes
2	Denaturation	94°C	30 sec
3	Annealing	50°C	30 sec
4	Extension	72°C	1 minutes
5	Thermocycling	Repeat steps 2-4	35 times
6	Final extension	72°C	7 minutes
7	Hold	4°C	∞

Table C13. *E. coli* colony PCR cycle #1 conditions

Step	PCR phase	Temperature	Time
1	Initial denaturation	94°C	2 minutes
2	Denaturation	94°C	30 sec
3	Annealing	49°C	30 sec
4	Extension	72°C	3 minutes
5	Thermocycling	Repeat steps 2-4	30 times
6	Final extension	72°C	7 minutes
7	Hold	4°C	∞

Table C14. *E. coli* colony PCR cycle #2 conditions

Step	PCR phase	Temperature	Time
1	Initial denaturation	95°C	15 minutes
2	Denaturation	95°C	30 sec
3	Annealing	50°C	30 sec
4	Extension	72°C	1 minutes
5	Thermocycling	Repeat steps 2-4	35 times
6	Final extension	72°C	7 minutes
7	Hold	4°C	∞

Table C15. Yeast colony PCR cycle conditions

Step	PCR phase	Temperature	Time
1	Initial denaturation	95°C	15 minutes
2	Denaturation	95°C	30 secs
3	Annealing	49°C	30 secs
4	Extension	72°C	3 minutes
5	Thermocycling	Repeat steps 2-4	30
6	Final extension	72°C	7 minutes
7	Hold	4°C	∞

Table C16. PCR cycle #1 conditions

Step	PCR phase	Temperature	Time
1	Initial denaturation	95°C	30 sec
2	Denaturation	95°C	30 sec
3	Annealing	51°C	30 sec
4	Extension	68°C	50 Sec
5	Thermocycling	Repeat steps 2-4	35 times
6	Final extension	68°C	5 minutes
7	Hold	4°C	∞

Table C17. PCR cycle #2 conditions

Step	PCR phase	Temperature	Time
1	Initial denaturation	95°C	30 sec
2	Denaturation	95°C	30 sec
3	Annealing	52°C	30 sec
4	Extension	68°C	50 Sec
5	Thermocycling	Repeat steps 2-4	35 times
6	Final extension	68°C	5 minutes
7	Hold	4°C	∞

Table C18. PCR cycle #3 conditions

Step	PCR phase	Temperature	Time
1	Initial denaturation	95°C	30 sec
2	Denaturation	95°C	30 sec
3	Annealing	50°C	30 sec
4	Extension	68°C	50 Sec
5	Thermocycling	Repeat steps 2-4	35 times
6	Final extension	68°C	5 minutes
7	Hold	4°C	∞

Table C19. PCR cycle #4 conditions

Step	PCR phase	Temperature	Time
1	Initial denaturation	95°C	30 sec
2	Denaturation	95°C	30 sec
3	Annealing	49°C	30 sec
4	Extension	68°C	50 Sec
5	Thermocycling	Repeat steps 2-4	35 times
6	Final extension	68°C	5 minutes
7	Hold	4°C	∞

Table C20. PCR cycle #5 conditions

Step	PCR phase	Temperature	Time
1	Initial denaturation	95°C	30 sec
2	Denaturation	95°C	30 sec
3	Annealing	51°C	30 sec
4	Extension	68°C	50 Sec
5	Thermocycling	Repeat steps 2-4	35 times
6	Final extension	68°C	5 minutes
7	Hold	4°C	∞

Table C21. PCR cycle #6 conditions

Step	PCR phase	Temperature	Time
1	Initial denaturation	95°C	30 sec
2	Denaturation	95°C	30 sec
3	Annealing	52°C	30 sec
4	Extension	68°C	50 Sec
5	Thermocycling	Repeat steps 2-4	35 times
6	Final extension	68°C	5 minutes
7	Hold	4°C	∞

D. Multiple sequence alignments supplemental information

Table D1. Predicted regulatory regions with their sequences

Regulatory Region / Type	Sequence
Bhlhe40 / TFBS[§]	GTCACATGCCT
Ebf1 / TFBS[‡⁂]	ATGGAAAACAACAGATATAATTAAGGAGGAAAGTCTATTT TTATTTTTAAAGACAATGCATTTTGAAGATGAGCCTTTAG ATTATGTCTACATCGCATGTTGGGTAAGGACTGGAGAAAT GGCTCAGTAGAAGAGCATCTACTTGACTCAAGAGGACTTG AGTTCACTTTCCAGGACCCATGTATGGTGACCCACAACCA CTTGTAACTCTGGCTCCCAAGGGAGCCAACACATCTGGCC TTCTCTGCCACTTATACTTTCACATCAATATCCTTGCAGA GACACACACATGCATACATATAGCTGAACATCTTAAAATA ATAAGCATGCCCAAATACAATGCAGAGGAAGTAGAACCAT TATGAGAAGGAAGAATAATTTTATGAGGTGTGAGATTTGA ATCAGGCACTAAAGAATGGGT
Unnamed TFBS[⁂]	CCTGATGTAATCTAACAGCGGGTCAGTCAGAGGGAGTGT CAGCATCTCCA

§ PAZAR database

‡ UCSC Genome Browser & Jaspar database

⁂ Promoter 2.0 Prediction Server & Promoter Prediction by Neural Network database

E. Yeast two-hybrid library screen hits DNA sequences

DNA sequences of yeast two-hybrid library screen in FASTA format. The translation start codon and vector nucleotides are shown in bold font. Protein product sequences shown in table 1 correspond to translated nucleotide sequence in italics.

>Y2H-006
NNNNNNNNNNNNCNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGAAGTCCAGGAG
GACAGACAGGACTCTGTTACACAAAGCAACCCTGTCTCGAAAAACAAAACAAAACAAAA
CAAAAAACCCGGGTTCTAGGATTCAAATTCAGATCCTCATGCTTGAGGGAAAGCACCTC
GCCGCTCCGCTGAGCTGTGTCCTCAGCCTCCCTCCCTCTCCTGCTGGCTCTCCCTATTC
TCCATAAGCTCACATGAACCTTCTGGGCCAGCAGGGTGTGGGCAACAGCTGGTGAAAGA
AAAGGCAAGAAAGAAAGGCCAAGATCATAAAATTTGACAATGGAAACACAGTAATAAAT
TTTCATATTCCGAAAAAAAAAAAAAAAAAAGGGCCCCCCCGCCCTCAAAGGGGGGGCT
*TC*NATCGGGGNCCCTNNGNCCNNNACTTNAAAANAATTNNAAAATCCNAAAAACCCCCN
NNNTTTCNTTTNNNNNGNNNNNNANNNNNNNNNAAAAANNNNNNNNNNNNNNNNNNNN
NNNNGNNNNNNNGCCNAGCTCGGCCNNNGNNCAANANNNGNNCTNANNGANNNNNTTCN
NCGCNNTNNNCACNGAANANTNNNGNNTANNNTNNNANANNCGNNANTNTCTCATTNNN
CNTCNNNGNNGNTTNNNNACNGNNANGGTNNNNTCNCAGNACNNNNNNNNNGNCNGCNN
NNNNNNNNGNCNGNNNNANNNNNNNNNNCNCCNNTNNNNNNNNNNN

>Y2H-008
NNNNNNNNGNNNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GA*
ATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGAAGGAAGAAA
ACGCAACGATAGCAACTTATCCTGAATTTGGAGTTTTAGACTTAAAGTATTTTCCATAT
TATGGGAAAAAACGGCATGTTGGATATCGACAACCCCTAGTTGCCGTACAGGTCAAATT
TGACTCTGGTCTTAACAAGAAAGAAGTAACAGTTGAGTGCCATATTGCTGGAACCAGGA
ACCTAAAAAACAAGAATGAGCGTGACAAGTTCTTGGGACGTGTTTCGTTCAAAGTTACA
GCACGAGCCTAGGAATAGGATGTCTCCACAGAATTCATGTTGTGTTGTCGCCATTTTGT
ATCAGCTGGACCTGCCATTCTAGGATTATGAGGCCACCTTGGAGGAGGAAGTGGTGTGG
TACACACTTGGGTGACATCATAACATGCTTCCAGATCATAGTGTTCAGTGTCCTCTGAA
GTAACTGCCTGCTGCCTCTGCTGCCCTTGAACCCATGTACGGTCGCCAGACAGGGACCA
GTGAGCTCCCGATCCCAAACATGCAGAGTGGGGGCCCTGTCTTCTTCTGTGTGTAATTG
CCAAACGTCTAAAGCTTCATGTGCTGGGCTGTGTAAATATTTTATGGATTTAACACTGG
TAACTGCATATTTTGATGCCAACAAAGTTTTAGGGATACAATGGTACCTTGCCAACATC
GAGTGACTTTAGCTGGAAGAAATGTCTGAGAAGCTCTGTATGTGAGGAGGGAAAAGGAA
AATAAAGTGGATCTGAAAGTAGACTCTGCTTCTGCGCATGTGTGAGTGGTGCCTGTCCG
TATGCTCCAGACTGTTTCTTACAAGGTGAACTAGACCATATCTGATTTTGCTGAGAGGA
*CAA*NCTTTCTGTTGTAGTACATATTTAAAANTACTATTGGGCTGACTTTGAGTGTTGGA
ATCAAGTCTTCNNCATTCGGNTATCNANNNCCAGGNATGAANNNNNTTANNTTAGGTGA
TGGNCNNGATTTTNGTTGTNNGNTTTGGTTTNGATNATNAANTCCCANNNNAANTTNNA
NGTTTNCAANTNNNNNNANGGNNTNNNNNNNNNNNGGNNCNNANAANNNNTTNNNAANN
NNNAAAAAAANNNNNNNNNNNNNNNGGNNNNNTCCNGGCNNNNNNNNNNNNNNANNN
NNNNNNNNNNCNNNNNNNNGNNNNNNTNGNNNNNNNNNANNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNN

>Y2H-016

ATGGAGGCCAGT*GAATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCC*
GGGGATATATATACATTGCACACATTTTACCCTGTGCCTTAATGTTGAATCTACTGTTT
GCCTCACACTAAGTTACTACCAGAAGCCTTAGTCATGGGTGTGAGGTGGAAGGTTTCAT
TTTGTGTCAGAGTTGCTGAATGTTAATCTATTTCATAAAGAGTCACACTAATGATCTAG
CTGCTGGTAGCGGCCAGGCTTTTCCCTCTTGTCCCCTCCACCTCACATCTATCAGTTTG
CCACAGAAAGATGAGAGATGCCATTGGACAGAATGTGACCTTCCTAATGATGTCCTTTT
TTGTGACAGGAATTTTCTTAACTTGATTTTTTTGATTTAATTTTGATGAATTTATAGTC
*CTATACATTTTAAATAAAATTGTTTTTCAGTGCAAAAAAAAAAAAAA*NGGCCCCCCCNGC
CCTTAAGGGGGGGNCCTCAAACCGGGACCCCTCGGGCCTAAACTTCAAAAAAACCCAAA
AACCCAAAAAACCCCCAAATTTCCTTTCAACTTGNTNGGGANCCCCCNAAAAANGNNGG
GGGGNGNNNNGNNNNNNNNNNNNNNNNNNNNNNNNN

>Y2H-017

NNNNNNNNNNNNCNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGACCAGCCGCAA
CGTGAAGTCGCTGGAGAAGGTTTGTGCGGACTTGATCAGAGGCGCCAAGGAAAAGAATC
TGAAAGTGAAAGGACCGGTGCGCATGCCTACCAAGACTTTGAGAATCACTACCAGAAAA
ACACCTTGTGGTGAAGGTTCCAAGACCTGGGATCGATTCCAGATGAGGATCCACAAGCG
ACTCATTGATTTACATAGTCCTTCTGAGATTGTTAAGCAGATTACTTCCATCAGTATTG
AGCCGGGAGTTGAGGTTGAAGTCACCATTGCAGATGCCTAAGACAACTGAATAAATCGT
*CTTAATGGCCAAAAAAAAAAAAAAAAAGGGCCCCCCCGCCCTCAAGGGGGGGGG*NTTNN
ATNNGGGNCCCNTCNNNCNNNNNNNGNAAANAANTNNAAAANCCNAAAAACCCCCNNNN
NTTCNTTTCNNNNGNNNTCNNNCCCCANAANAA

>Y2H-020

NNNNNNNNNNNNCNTNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGAATCTCCCAT
GTCTGTGGTGAAGTCCATCGAGGTAGTGCTGCCCCAAGATGCGGTCTACCTGGCCGGCT
CTATCATAGACGGGCAGGTGGTCCTCACCCTCAACAGTACCCTCGTGGACCCCGTCGTG
AAGGTGGAGCTTGTCGGTAGGGGTTACGTTGAGTGGAACGAAGAAATTGGAGAGACCCG
TGATTATAGTAGAGATGTTATTTGCAATAATAAGGCAGATTACGTACATAAAACAAAAG
CATTCCCAATAAAAGAACTCCGTGTCTGTAGAAGCTGAGAAGAAGGTGTCATACAACTG
CTGCAGTCAGGGCTGGGTGAGCCTGCATGTGCAGATGAGCAAGAACACCTATGTGCCGG
GCGAGAAGGTGACATTCACATCAGAGATCCGCAATCACACGGGCAAGTACATCAAGACC
GTGGTGTTCGCCCTCTATGCCCATGTGCAGTACGAGGGCTTCACGCCAAGTGCCGAGAG
GCGCCGGCGGGCAGACAGCAGCGAGCTGCTGCGACAGATGGCCAACGCACGGATTCCAG
CCTTCAACAGTACCACAGTGGTTAGTGCCTTCAACCTGCCGCTGGTTTTGTCTGTGAGC
AGTGGCTCCCAGGAAAATGAGATCATGAGGACCAGCTATGAGCTTGTAGTCACGATTCA
CCTCCCCTGGTCTTTGTCGACAGTCAAAGCCCGACTCCCCATCATCATCACCAGCACCA
GGGAAGGCCAGGCCGACTGCCCCCGGCTGGATGAGCTGCCATATGAGGACCATATGGTT
*TGAGTGCCTGCATATTAAATATGAAAAGCTTATTGCAAAAAA*NGAAAAAAGAAAGGCCG
CNNNGGCCTCTAGAGGNTGGNCNNNNATACGGNATCNNNCGAGCTCAACTGCANNNAAT
NNNNNTACTGAAAACCCNNNNTNNNNGGNNNGGGNNNNNNNNNNNNAAAANNNAAAANN
NNNNNNNNNNNNN

>Y2H-021

GNNNNNNNNNNNNCNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GA*
ATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGCATGGAAATG
CCTTTAGCATTTCAATTACTGTATATTTTGTTTAAGGTGACCCTTCAGCATGCCGCTAA
TGTCTTTGTTAGTGACAGTGCATTTTGTAGTACTGTACAAGTGTTGTGCTAACGGTAAG
CCATTTCTTAAGTTTTTTGCCTTGATTAGGGTACCCTAATTTGAGGGTTAAAAAAAACA
AAACAAAACAAAGCTCTATTTTTGTTAATTACAAAACTGTAAAGAGCTACAGAGCTATC
CCCGTCTGGTTAGTGAAGGGGGTTTATCGCTAAATGTTTGGTGTAAAGTTGAGAAGACC
TTTTTCCATTTTGGTGGCAGGCTCCTTTGGGGGGAGGCAGCTTTCTGTTTTATAAATGC
ACACTTCTGTTTATTGAATGAAGCATATCTCAGTGTTTATCTGTCAGGTTTTGAAGCAT
TTCATATATGTCCAAATACTTGGCAGGATTTAAAAAAAAATAGTGAATTTGGTGTAAAG
TTGCTATTTTATGGAAATGCCTCTAACTTTACATTTTCATTCCATCTGTAGATTTTTCT
ATCTTTATAAAATATTGGAGTTATTTTTTAAGGAAAAATAGAAAAGTAGCTTGTGAATA
GCTCAAACTAAGCTTACAAATCGCATGTAAAAAAGCAGAAAAGTTATTTGTGTCTGTTT
CTATTGCGTCCTTTTTTAGCCTTTGTACCTGGACAGGTGACAGTAAGGGCCAAGCAGGA
GAGGCGGGCGCGGCCTTGTATAAAATAGGAGCCAGCGGGACTCTTGTATCTACCTGTTC
*TCTTTTTAGTCGGTTACTTCAAAAAAAGAAAA*NGCAAAAACAAAAAANCAAAAGCTGTA
CATTTNAACATAAAATAAATTATGATGAGCCAAAAAAAAAAAAAAANGCGNGNNNCNCC
NTNNNGGGGGGGNATNNNNNNCGNNNNNNNNNNNNGGNNNNNNNNNNNNNNNTNCNAGA
ACCNNAGTCTCCTTNNNNNNNNNNNNNNANNNNNNNAAANNNNNNNNNNN

>Y2H-022

NNNNNNNNNNGNNCNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GA*
ATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGACGGCGACC
TGGCCAACTGGATGGTGCCAGGCAAGAAGGTGAAGGGCATGGGTGGGGCCATGGACTTG
GTGTCTAGTAAAAAGACCAGAGTGGTAGTCACCATGGAACACTGTACCAAGACAAAGCA
GCCCAAAATCTTGAAGAAATGCACCATGCCGTTGACTGGCAAGCGCTGCGTGGACCTCA
TCATCACTGAGAAGGCGGTGTTTGAAGTGAACCACTCAAAGGGGCTGACGCTGGTGGAG
CTGTGGGAGGGCTCGTCGGTAGATGACATCAAGGCCACCACAGCCTGTTCATTTGCTGT
GTCCCCCAACCTCGAGCCCATGCAGCAGATTAAACTTGATGCTTGAGGAGCCCTCCAGG
GCTGATTTTCCAGTTAGTGACAACTGGACATCCTTAGCAGCACCTGACCACTGTGCCAC
ACTGGCTCCTCCGCCTTCCTTCTGCTAGATGGTGTCTAGTGAGAGCCATGTGACCTTAA
*TTAAAAACCCTAACCAAAAAAAAAAAAAAAAA*NGGCCCCCCCNGCCCTTAAGGGGGGGN
CATTCAATCGGGATCCCTTCAACTTCAACTTGAAAAAAATCCAAAAATCCGAAAAACCC
CCGAAGTTCCTTTCAATNNGNGGGGNNNCCNAAAAAAAANNNNNNNNNNNNNNNNNNG
NNNNCCNNGNNGGGNCCCNNGGGCNNGGGNCCCGCNNNGNNNCCNNNNNNNNNNNNNN
GGGNNNNNNNGNCCCCNNNGGGNNNCCCCNNNNCNNNNNTTNNNNNNNGCCGNNGNNNN
NCNNNNNNGGNNCNNCNNGCNNNNNNNGGGGNGGGGGGNNCCTNNNNNNNNCCCNNNNN
NCCCCCCCCNNNNNNCCCCNNNNNNNNNNGGNCNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNCCNNNNNNNNCCCNNNNNNNTTTNNGGNNNNNNCNNNNCCNNNNNNN
NNNNNNNNTNNGNNNNNNNNNNTNNNNTTTNNNTNNGNNNNNNNNNNNNNNGNNN

>Y2H-032
NNNNNNNNNNNNNNACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAAT*
TCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGACAGACAAGAAA
TACCTTCCGAAGTGCTCTGTAGATTTTTTTGATGAATGTTTTATGCTTGCATTTAAACT
TGAAATGTATGAGCAGAATGAGACAATCTTTACAATCAGAATTGAGAAGTGTTACAATT
GAATGGCCTTGTGCTGTAGCAATAAAATGACCAAGTGCAATGACTTTAATAAAATCATC
*TTTCAAACGTTAAAAAAAAAAAAAAAAAAAAGGCCCCCCCGGCCTC*NAAGGGGGGGGCA
TCAATCCGGGACCCATCGAGNTCAACCTGAAAAAAAATCNAAAAACCNNAAAANCCCCC
NNGNTTCNNTTNNNNNGNNNNNNNNNNANNNNNANAAN

>Y2H-033
NNNNCNNNNNNNNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAAT*
TCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGATAGCACCTTTT
ATTAACCAGTTTCAGGTGCCTATCCGGGTATGTCTGGATTTATCTTCTTTGCCATGTGT
CCCTTTAAGCCAACCAGTGGAGCTGTTGAGGTTGGATCTGATGACTCCATATTTGAACA
CTTCCAATAAAGAAGTAAAGGTCCGTGTGTGCAGGTCTGGTCGAGTGACCGCTGTTCCA
TTTTGGTTTCATCTGTGCCTGGATGACGAGGTGAGGCTGGACATCTCAGGTGAAGCCTC
CCACTGGAAACAAGCTGCAGTTGTTCTGGATAATCCCATCCAAGTACAAGCGGGAGAGG
AGCTTGTGCTCAGCGTTGAGCATCACAAAAGCAACGTCAGCATCGCAGTGAAGCCGTGA
GATTGTCATCGGACAATGAATGTGTGAAAATCCCTGGAGTCATCAAAAATGTTTAAGGG
*TTGCAATAATAAAGCCCATGTTAAAAGTCAAAAAAAAAAAAGGCCCCCCC*NGCCCTCAA
AGGGGGGGNATCCAATCGGGNTCCCTCCAACCCCAACTTCAAAAAAATCCAAAAATCCG
AAAAACCCCCCAAGTTCCATTCANNNNNNNGGGGNCNNNNNNAAAAAAANNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNN

>Y2H-045
NNNNNNNNNNNNNNNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GA*
ATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGTGTCTGAAC
AGGTGAAAAATGTGAAGCTTAGTGACGACAAACCCAAAGACTCCAAGTCTGAAGAGACT
CTTGATGAGGGTCCACCAAAATACACAAAATCTATTCTAAAAAAGGCTCGACTGGAGAT
CGAACCAGAATGGGCCTACGGAAAGAAAGGACAGCCCGATGCCAAAATTCCACCAAACA
CCAAGCTCATTTTTGAAGTGGAATTAGTAGACATTGACTGAAACCAGTACTTCAGATCG
GACATCAGCAAACATAAAAATGAGCTGTGGCTAATTAGAGCTTGTAAGGGAAGAGTCAA
CTGGAAAATTTAGAGTTAAAGCTTGTTTACTTGATCTCACATTTTGAGAAACATAATCC
TGTTACTCCCTTTAAGTCTCAGATATTTTCCTACAGCTATGTAAAGTATTCCAGAGAGC
TGCTTTCCTTTTACTTTGTAAACTAGAAGACTCAATAAAAGTGTCATGGATTGTCTTGA
*CAAAAAAAAAAAAAA*NGGGGCCCCCTGGCCTTTAGGGGGGGGGNATTCAATGGGGATCC
CTTGGGCCCGAGCTGGAAAAAAATNNAAAAATCCAAAAAACCCCCGAAGTTCCTTTCAN
NNGNAGGGGNNNNNNNAAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

>Y2H-047

NNNNNNNNNNNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGATCCGCTCCAT
TAATGACCCGGAGCATCCACTGACACTGGAGGAATTGAACGTAGTAGAGCAAGTTCGGA
TTCAGGTTAGCGACCCAGAGAGCACAGTAGCAGTGGCCTTCACACCCACCATCCCACAC
TGCAGCATGGCCACGCTTATTGGCCTCTCCATCAAAGTCAAGCTGCTGCGATCCCTTCC
CCAGCGGTTCAAGATGGATGTACACATTACACCCGGGACCCACGCCTCCGAGCATGCAG
TGAACAAACAACTTGCAGATAAGGAGCGAGTGGCAGCTGCCCTGGAGAATACACACCTG
CTTGAGGTTGTTAATCAGTGTTTATCCGCCCGCTCGTGAGCCTGGCTTTGGACCCCCAG
CCTGCACTCAGGTTGTCCTGGCCTCAGCTCCACCAAGGGCTCATGGCTTTGTTTTCTTG
AATCACTGACAAGGAGAAACTTAACACTTTTATTTTGTAATAAACCCTTAATTTATATT
*CAAGAAAAAAAAAAAAAAAAGGGCCCCCTGGCCCTTAAGGGGGGGG*NCTTCAATCGGGA
ACCCTTCAACTTCAACTTGAAAATAATTCGAAAATCCGAAAAACCCCCGAAGGTCCTTT
TCANNNTGNGGGNNNGCCCCAAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

>Y2H-055

NNNNNNNNNNNNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***G*
AATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGATCGATTC
CAGATGAGGATCCACAAGCGACTCATTGATTTACATAGTCCTTCTGAGATTGTTAAGCA
GATTACTTCCATCAGTATTGAGCCGGGAGTTGAGGTTGAAGTCACCATTGCAGATGCCT
AAGACAACTGAATAAATCGTCTTAATGGCCAAAAAAAAAAAAAAAAAAAAAGGCCCCCC
*CGGCCCC*NAGGGGGGGGGGCTGGAACCGGGGTCCCTCGGNNNNGGNCNGNAAAAAAANC
CNAAAANNNGAAAAACCCCCNNNNTTNNNTTNNNNNGGNNNNNNNNNNNNNNAAAANA
NNNNNTANCATANNNNANNNNTNNNCATCACAAANGCCNCCTCANNNNNNNNNGTGNGN
NNNNNAGATTGTCNTCNNNCNNTGANTGTGTGAAAATCCCNGGANTCATCANNNATGTT
NNNCGNTTGCNATNNNNNNNCNNNTGNNNAAANTCNNAAANANNNNNNNCCNNGGNNNN
NNNANNNNNNNGNNNTNNNNNNNNNNNNNNNNANNNNNNNNNNNNNNNNNNNTNNNG
NNNNNCCNNNGNNNNNNNNNNNNNNNNNNTNNNNNNNNNNNNNNNNNN

>Y2H-058

NNNNNNNNNGNNCNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGAAGTCCAGGAG
GACAGACAGGACTCTGTTACACAAAGCAACCCTGTCTCGAAAAACAAAACAAAACAAAA
CAAAAAACCCGGGTTCTAGGATTCAAATTCAGATCCTCATGCTTGAGGGAAAGCACCTC
GCCGCTCCGCTGAGCTGTGTCCTCAGCCTCCCTCCCTCTCCTGCTGGCTCTCCCTATTC
TCCATAAGCTCACATGAACCTTCTGGGCCAGCAGGGTGTGGGCAACAGCTGGTGAAAGA
AAAGGCAAGAAAGAAAGGCCAAGATCATAAAATTTGACAATGGAAACACAGTAATAAAT
TTTCATATTCCGAAAAAAAAAAAAAAAAAGGGCCCCCCCGCCCTTAAGGGGGGGGGCT
TCAATCGGGAACCCTCCGGCCCCAACTTGAAAAAAATTCAAAAATCCGAAAAACCCCCN
*AA*NTTTCTTTTANNTTGNNTNNNNNNANNNAAAAAAANNANNNNNNANAAGACNNNNTN
GGTAANAANNNNCNNNTANNTTTNCGANTGNNNAGAANGNANNNNNNNNANNNCNNCN

>Y2H-063

NNNNNNNNNNNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGCGGCTTATTTT
TGCAGATGGTTCTCGTATCATCTTCAGACTGAGTGGCACCGGGAGTGCAGGGGCTACCA
TCCGGCTGTATATTGATAGCTACGAGAAGGACGTTGCCAAGATCAACCAGGACCCCCAG
GTGATGCTGGCGCCCCTGATCTCCATTGCTCTGAAAGTGTCCCAGCTCCAGGAGAGGAC
GGGTCGCACTGCGCCCACTGTCATCACCTAGAAAGATGGCCAGATGTGGCACGTCCCTC
CCCCAGGACTATCCACGCCACCTGATTGAAGGGCACGGACAGAAACAGTGTATTTGCCT
GGGCCTTTTAGGACTCGGTCTTGATTTTTTACTTCCCACTTTGATAAACAGCCATTTGT
GAGGCACTGCCACTGTGTGGGGGAGAGGGAAAGCAGACCCGAGGGCCTGCACCTGTGTC
TGTTCCTTCTCTTGCCCTATTAAATTGCTGCTGTGTGAGCCCTTGTCTCCCTACATCAG
GTACAGTACACTACACTGTAAAGATTTGGAGGTGGAGGGTCAGCAACAAGGCCATCTCC
TGTCCATGGAGTGTGCCTGTAAATGATATCAAAGTTCCTGCTTCACTGTGAAAGTCTCC
CCATGTGTGTTTACATGTAATCCAACAGAATACGGTTCCTGGTGCCTTCTCTCTGAGTG
AGATGTGCTTGCCTTCAGACCTCTGTGTTATTTTGCATCTCCCATCTGCACAGATATTT
*GGGGTATTAAAGCAGAATGTGCTACGGTAAAAAAAA*NGAANNNAAAAAAAAAAANGGCC
CCCCNNNCCTCCAAGGGGGGGGNNGNAATCGGGNCTCTCGGNCTCAGCCGCAANTAATC
GAAANTCGGAAANCCCNGAGGTCCNTCNTGNGNNNNNGNNNNCCNTNTNTAANNNNNNN
NN

>Y2H-068

NNNNNNNNNGACNTNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAAT*
TCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGATAAAGGAGGTG
AATGTGAGCCCATGTCCCACCGATCCCTGTCAGCTGCACAAAGGCCAGTCCTACAGTGT
CAACATCACCTTTACCAGCGGCACTCAGTCCCAGAACAGCACGGCCTTGGTCCACGGCA
TCCTGGAAGGGATCCGGGTCCCCTTCCCTATTCCTGAGCCTGACGGTTGTAAGAGTGGA
ATCAACTGCCCCATCCAGAAAGACAAGGTCTACAGCTACCTGAATAAGCTTCCGGTGAA
GAATGAATACCCCTCTATAAAACTGGTGGTGGAATGGAAACTTGAAGATGACAAAAAGA
ATAATCTCTTCTGCTGGGAGATCCCAGTTCAGATCACAAGCTAGGCTCCTTGGCGCCTG
TGTCTGTGTGGGCTGAGAGGCCATGGACGGAGTGGGGGGGAAGAAACAGAAATCAGACC
CGAAATGGAATCAGTGCCATATGATGAACAGAATTTCAAGAATGCTGTTTTATGCCTTT
TAACCTCCAAAGCAGTAACTGCAAGCCTACTACTCTTGAGAGCGCGCTCAGAGCCATTG
TCCCCCGAGATAGCCTCTGGGGAGGCTTCGGGAGGGAAAGGGGAAACTGGAAAGATTAG
ATTGGTGTCCATGGCTGTTTGCTGTTGGATTACAGGCAGGTTCCACCAGACAGGATGGG
*ACGGGGTTGCTTA*NGGATGTCAGATAACTTGACCCAGGGCTATGGATCCGCTGTGAAGG
ATGACTTCNCAGAGTCTTCTGGCTGCCTGGGGGTGTTACCNNCCTGTTCTAAGTGNNTN
NTGNGTCNNCAGNCCNCNGGNTNATCAGTCNGNNAGTCNNNNGGGNANCNNNGNCNNNT
CGNNNANNNNNANCGACTANCNTNNNNNNNNNTNNNNNCNNNNTGNAGGNGNNNNNNNN
NNNNNNAAANNNN

>Y2H-075

GNNNNCNNNNNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAANCNGTGNNNTNAACGCAGAGTGGCCATTACGGCCGGGGGTCTCAGACAA
CAACATAAACATTTCTGGGTGGGAGAATGGTGTGTGGAAAATGGAATCCATATTCAGAA
AAGTCGAGAAAGACTGGAACATGGTTTATTTGGCCCGAAAAGAAGGATCATCTTTTGCT
TATATTTCCTGGAAGTTTGAATGTGGGTCAGCTGGTCTAAAAGTAGATACTGTTTCCAT
CCGCACAAGTAGCCAAAGCTTTGAGTCTGGATCAGTGAGGTGGAAACTGCGATCTGAGA
CAGCTCAAGTCAACCTGCTGGGAGATAAAAATCTACGTTCCTATAATGATTTTTCTGGT
GCCACAGAGGTTACATTAGAAGCAGAGTTAAGCAGAGGAGATGGAGACGTTGCTTGGCA
ACATACCCAACTGTTTAGACAAAGCTTAAATGACAGTGGAGAAAATGGCTTGGAAATAA
TTATAACGTTCAATGACCTCTGAAAGCCTGAACTTCAGGGAAATGCTGGCTGCAATCAC
CGACCCGCCATGGCCTCCAGTAGTGTATTAGCTCAGTGTGTATCTAGTTGGCAGTTCAC
CATCCTGTGCAACAGGTTTCTTTTGCTGGCTGTACATCATGTAATCCTCATTAAAGTAT
ATCTTCATAGTGTGGACCATAAGAAATCTTCAATTAAATGCTAACTTCTGCCTGTGATT
AGAATTTGGCATAAAATATTTGGCATAAAGTAACTGTATGAAATTCTAATCTAAACACT
GTTACTAAAGTTGTAAGTGAAAATGATGGTATTCATGATTAAATGCTGTTCATGAATGT
GATACAATCTAGGGCTACTAGTTACTTAACATTTTCTGATTAGATTTTTCCAAACATTC
*TCATTTT*NGAATAACTGTCAAAGTCAAGTATCTTATGTNNNATGATGCATGAAGGTTTN
ATCATTNANNNATCTGGACAGNAAANNNNNAANTGGATTTTTNNGCAAAANNNAAAAAA
NGGNCCCCNNNNCTCTNNGNGGNGGNNNNNNNNNNNNNCCTNNNNNNNNNNGNAGNNTN
NNNNNNNNGACCGNNNNNNNNNNNTTNGNGNGTGNNNNNNNNGGATNNNTANTAANNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNN

>Y2H-076

NNNNNNNNNNNNNNNNNTNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAG**
TGAATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGAAGTCCA
GGAGGACAGACAGGACTCTGTTACACAAAGCAACCCTGTCTCGAAAAACAAAACAAAAC
AAAACAAAAAACCCGGGTTCTAGGATTCAAATTCAGATCCTCATGCTTGAGGGAAAGCA
CCTCGCCGCTCCGCTGAGCTGTGTCCTCAGCCTCCCTCCCTCTCCTGCTGGCTCTCCCT
ATTCTCCATAAGCTCACATGAACCTTCTGGGCCAGCAGGGTGTGGGCAACAGCTGGTGA
AAGAAAAGGCAAGAAAGAAAGGCCAAGATCATAAAATTTGACAATGGAAACACAGTAAT
*AAATTTTCATATTCCGAAAAAAAAAAAAAAAAA*NGGGCCCCCTCGCCCTCTAGGGGGGG
GNAATCANTNNGGGACCCTTCGAGCTCAAGTTGCAAAGAATTCGAAAATNNGAAAAAAC
CCCCAAATTTCCTTTCAATTGGGCATCGGGCCCCATAAAGGGGGGGGGGGGNCGCCCTG
NNNNCTTNNNNNNNGTGGGGNACTCNGCNNTNGTNNNNNNNNNNTCCCCTNNNNGGCNN
NNNNNNTGANNNTCTCACGTTCTTTCTTNNNTTTNNNNNNTNNNNAGNNNNNNAGAGNN
NNCNNNATTCNNNTNNNNNNATCNNNNNNACNNNNNNNNNGGAGCANNNNNNNNNCNNN
NNNNGNCNNNNNNNNNCNNNNCANNNNNCGGTGNNN

>Y2H-077

GNNNNNNCTNNNNNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT**
GAATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGACAGTGGA
TGTAGAAGGAGATGGATGCTCATTTATCCAGGCCACCCTCAGGTACAACGTGCCACTTC
CTAAGGAGGCCTCAGGATTTTCCCTTTCCGTGAAAACTGGAAAAAGCAACTCTTCAGAT
GAATTCCAGACCAAGTTTGAACTCACGGTGACCCTCACCTACACTGGAGCTCGTGAAAG
CTCAGTTACAGTCCTTGTGGATGTGAAGATGCTCTCTGGCTTTACTCCTGTCGTCTCAT
CCACCGAGGAGCTTAAATTCAATAGTCAAGTAACGAAGACGGACATAAAGAATGGCCAC
GTTCTGTTCTACTTGGAAAATGTTCCCAAGGAAGCAACCAGCCTCACCTTCTCCATTGA
ACAAACTAACCATGTGGCCAATATCCAGCCAGCCCCAGTCACGGTCTACAGCTATGAAA
AAGGTGAATATGCTTTTGATTCTTACAACATCAATAGCATTTCGGATTCCCAATAAAAC
AAAGAAGCGCACAGAAAAAAAGAAAAAAAAAGTATGAGACTATTTCTTTGAATAAACTG
TTCTCCTGGTCACCATCCTACCAAGAGCCTGATCAAGCACCTCCCATCCTGAAGACTCT
*CCCATCACAATGACTTCTTGTCAGCTTCACTTCCTTCCTTCCTTC*NNNCCTTCCTTCCT
TACTTCCTNACTTCCTTCCTTCCTTCCATTNNATGGCATCATTTTAAAAAAAAAAANGA
AAGGNNNNCCCNNTNTCANGAGGGGCNNNNNNNAGGCGGGAACCATCGGGCTCCACCTG
NAAANNAATCGAAAATCCTGAAANCCCNNGGNNNTCCNNNNNNNNGNNNCCNGNGGACA
ANNTAGACNTATTNNNNNNNNGTNNNNNCNNNNNNNNNNNNNNCNNNNNNGTNNNNNNN
NNNTNGNTGNNNNNNNNNANNNNCCNNNN

>Y2H-080

NNNNNNNNNNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAAT*
TCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGATGGAAGCAAAT
GACCGAGAAGGAGACCCTCTTGACCTGTGGGTGACTCTGCATTCCATGGGCTATAATAA
GGCTCTGGAGCTGACAGAGGCATGCCCATTTGTCATCAACATCTACGCTGAAAGATGCA
AGCCAAGAATCAAAGTTGTCCATATGGAGGCATGCAGTGGGCAACTTGAAAAGGCCATT
TGCAAATCTGTCCTTGATAGAAGCGATGCTAAAGTCATGGATGGCTATGAAAATATAAT
CGTACATACTTGCAATTATGACACATGGATAACATCAATTATTGAAAATAAGTCAGATA
ATAAAGTGATTATTCACATCAATAACGAACTGAGTAAAAACTGTGTCAACAACAGAGGA
CTCAACATCTTTGCAGTAGAAGTGGCACCAAGATCTACAATGGTGTGTCAACATGTCAT
GCCCCTGAATGAACAAGAAGAATGGATATATTGTTGTGTGTATTCTCTTGTAGCTTAAA
TAATTCCATTTGGAGCTATCAAACTAAGAAGTATTTTAGATCGACTATTTTATTTATTA
ACTACATACTTTTCCATGTAACTGGTGTACTTAATAATCTGTTTCATTTCTATGCATTT
ATATTTTATGTCAGTGTTTTATATTTTATTATTAAAATACAAATAGTGTTTTAAGAAAA
*AAAAAAAAAAAAAA*NNNNCCNCCNNNNNCCNTNNNNNNGGGNGGGGNNNNNNNNNNGG
GAANNNNNNCNNNNTCNNNNNNNNNNAAANNNNTCNNNNGANNNNTGAAAAANCCCNNNN
GTTTCNNTNNNNNNNNNNNNNNNNNNNNNNAANAAA

>Y2H-086

NNNNNNNNNNNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGAAGTCCAGGAG
GACAGACAGGACTCTGTTACACAAAGCAACCCTGTCTCGAAAAACAAAACAAAACAAAA
CAAAAAACCCGGGTTCTAGGATTCAAATTCAGATCCTCATGCTTGAGGGAAAGCACCTC
GCCGCTCCGCTGAGCTGTGTCCTCAGCCTCCCTCCCTCTCCTGCTGGCTCTCCCTATTC
TCCATAAGCTCACATGAACCTTCTGGGCCAGCAGGGTGTGGGCAACAGCTGGTGAAAGA
AAAGGCAAGAAAGAAAGGCCAAGATCATAAAATTTGACAATGGAAACACAGTAATAAAT
*TTTCATATTCCGAAAAAAAAAAAAAAAAAAAGGCCCCC*NCGNNNTCNNNGGGGGGGGNN
TNNNNNNGGGATCCTTNNNNNNCNANCTGNAAANAAATCNNAAATACNGAAAAACCCCG
CAGGTTCCCTTCNNNNGNNCATCGGGANNCNNNNNANNNAACNNNNNNNNCCCNCCANG
NCCCCCNNGGNGCNGCCCCACTNANAANNNTCCNCNTCNNCNNNCNGANNNNNTCNCNA
NNNACNNCTCCCCCCCACTCNNCACTNNNNNNNCCNNNNCNANTCCNNNNNNANNCNNN
NNACCNNNTNGCNCN

>Y2H-088

NNNNNNNNNNNNNNNNNCGACGTACCNGATTACGCNCATATGAAC**ATGGAGGCCAGT***GA*
NTTCCACCCAAGCAGNNNTATCAACGCAGAGTGGCCATTACGGCCGGGGAGCCAATTCT
CCATCATCATCCGTTTTTACCTTAGTTAGAATTTTTTCTTATCATTTTTCTTCTTTCTT
TCCCTCTTCTCCTTTGGACATAACTTTCTGTTGACGCTGTTCTTTGGCTGGTTGGTTTT
AGTACTGTAAACTGCTTCTGAGCAAACACGGAAATTTAGCAAAATTATGTAAACTTGAT
CCTGAAGTTTTAGAATGGCAAATAAATGTACAATTGTTTACATAACAGAAAAGGCTAAG
CAGAAAGTAAATTTCAGTGTGTCAGTCTAGAGGCTCTACTTTATGCAGACTTAAATTCA
TGTGAGATATGTACCTTCATATTCAGAAATCTGGATGTCTCCTTCATACATTAAACTAT
TAATAAGCGCCACTTTTCTACTTGTATATTTTAAGCATAGAGTAAAATGATGGGCATTA
TCAACTGATGTTCCCTGCGTTAGCTCTGAAACACCCATCTTATTCCCTGTTGGGTTTGA
TTGTAGCAAGGCAGATAGAAGAAATTTCTGATTTTTCCGTGTCTTTATTAAATTTGTCA
CACTTTTCAAACTCATGCGCACACATCTCTTCAGTTTGGACTCTCTCTTTGCCACTATT
GCTACTTAGTAAGATTCTCTTGCTCAATAGTGAGAACCTTCTTGTAAAGTTTTTTCCCA
TAGTGGCTGTGACACAGATGTGGAGCATCTTCAGTGGCACCAGTGGTTTTTTCTAACCA
GAAAGCTTGAGACTTTTGAGCAACTTAGAAGCCACCACAGCTCTCAGTCAGCATATATT
GTTTATTTTTAAAGATTTCAGCATATTTTATATTGGATGATATTGATTCTTAACATTGG
*CC*NTTCCTGGTCAGCAAGAGTCAACATGCAGTTTTCCCTTTTCAGCATCTATGTCTNAA
NTTGAANGGTCCTCTCTCTGTAGTTGTGGNNNAANTNAAAGGATGANNNNTACTTGGGC
CACNNAAANGGANTGTNGGNNCNTTTNAANTGG

>Y2H-094

NNNNNNNNNNNNNNNNACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GA*
ATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGTACTACGGCT
TCCCTCTGCAGTATTATCCCTACTACGGCAAACTCCTGCAGCCCAAGTACCTGCAGCCC
CTGCTGGCCGTGCAGTTCACCAACCTCACCGTGGACACTGAAATCCGCGTCGAGTGTAA
GGCGTATGGTGAGAACATTGGGTACAGTGAGAAAGACCGTTTTCAGGGACGCTTTGATG
TAAAAATTGAAATTAAGAGCTGATCACAAGCACAAATCTTTCCCACTAGCCATTTAATA
AGTTAAAAAAAGATACACAAACCTACTAGTCTTGAACAAACTGTCATACGTATGGGACC
TACACTTAATCTCTATGCTTTACACTAGCTTCTGCATTTAATAGGTTAGAATGTAAATT
TAAAGTGTAGCAATAGCAACAAAATATTTATTCTACTGTAAATGACAAAAGAAAAATAA
*AAATTGAGCCTTGGGAAAAAAAAAAAAAAAAAGGGCCCCCTGGCCCTTAAGGGGGGGG*N
CTTGGTTGGGGATCCCTTGGGCCTGAACTTGAAAAAAATTCTAAAATCCGAAAAACCCC
CCAAGTTTCCTTTAAATTGGNNGGGNCCCCCCNTAAAAAGNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNN

>Y2H-097

NNNNNNNNNNNNNNNNNNCGACNNACCNGNTTACGCTCATATGAAC**ATGGAGGCCAGT***G*
AATTCCACCCAANCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGCAAAACAAA
ACAAAAAACCTAAACCAGAGACAAAAATCTAACGTGGTTATTAAAATTTTTCTCATTCT
TGACATTTGATTATTTTAATAAATGTCTATGAAATTGAGCATGTTACTTACTCCCTAGG
GATCTTGGGTTTTTCCGTTTGTTTTGTTGTGCTCATTCCTTGAGCCCCAAAAGACCACC
AAGGAGTCGATTCCAATCCAATTGCACTAGGGTCTTTATTCGAGTTCGAGCTGGAGCCT
TATTGTCATCACTGAAACAGCAGGATGGGAGGGGGAAGCCCCGAGCCCAGTTTCAAGCA
AGCATTTATAGAGGCAAGCAAACAAGCCAGGGGCCTTCCAGCTTGGCACACATCTGATT
AGGGGCTGTTATGGAATTTGTTGCTCATTAAAGCATGAGCAGGGACAGAGCCAGGACGT
AAAGAGAGAGAGCTACATGGTGGGACACACTGGCAGGGAAGTTGGCGTAAGTTTACCTG
TCTAACTGAAGGACTGCCACAGCAAAGAGACCAACAGCCTTACACTCTACAGATTTCTC
CATGTCTTTATTAGTAAACCTCAAGTAGGCCTCCTTAACCCCCACAATACAAACCAACC
CAAAACCACTCCCCACACACAAAATATGTCTTGAAATCCTTAATGTAAATTCTGTGACA
AACAGACAAACACGTCAGTCAGTCTTGGACATATTTTTATGACCACATACCTTTCATGA
GTGAAAGGAACTGTTTAGCACAGCGATGGAAAGGACAGCAAGCAGGCTATTTCTAGGTG
CCAAGAATATCACATTCTTTCCGAGACATTTTTTAATGTGATCGTAGAGTGAAGGAGAA
TGACATTAAGGTATTCCTAACTGTGTGGAACAAAGACTGTTTCTAAAAACCGTTAAAAG
*GCCAATTTCTA*NATTNAGTGAANTNNNTTTTGAANTTNANNTAAATTNANNCCTAAAAA
AAAAAAAAAAAGGNCCCCCNGNCNNNNNNG

>Y2H-098

NNNNNNNNNNNNCNTACGACGTACCAGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGATCAGAGGATG
GGATGAAGCGCTCTTGACTATGAGTAAAGGAGAAAAGGCTCGACTGGAGATCGAACCAG
AATGGGCCTACGGAAAGAAAGGACAGCCCGATGCCAAAATCCCACCAAACACCAAGCTC
ATTTTTGAAGTGGAATTAGTAGACATTGACTGAAACCAGTACTTCAGATCGGACATCAG
CAAACATAAAAATGAGCTGTGGCTAATTAGAGCTTGTAAGGGAAGAGTCAACTGGAAAA
TTTAGAGTTAAAGCTTGTTTACTTGATCTCACATTTTGAGAAACATAATCCTGTTACTC
CCTTTAAGTCTCAGATATTTTCCTACAGCTATGTAAAGTATTCCAGAGAGCTGCTTTCC
TTTTACCTCGTAAACTAGAAGACTCAATAAAAGTGTCATGGATTGTCAAAAAAAAAAAA
*AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA*NGGCCCC
CCCCCCCCCNGGGGGGGGGGNCCCCAAGGGGGGACCCCGGGGGCCCGGGGGAAAAAAAA
NTTAAAAATNAAAAAAACCCCCGGGTTTTTTTTTTNGGGGGGGNNCNGNAAGAAAAAAA
AAAAANCAAANNCNNNNNNNNNNNNNNNNGGCNNNNNNGNNCNGGNNNNNNCCNNNNGN
NCC

>Y2H-099

NNNNNNNNNNNCNNACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAAT*
TCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGACAAGGACACT
GACCAAACCAGTTACAGACCCGACTGGACAGAGCGCTTTTTGAAAGTGGAAGTAATATC
TAAGCCTGCTTCTGGACATGCTGTCAAAGTGGAGTTCTCGCTGAGAGTTTGACTTATGC
AGAAACAATGGTTCGACTGACGGTGGATTTAATTGCCAAAAACAGCAATCTTAAACCCA
GAAAAGAAGAGACCCTTGCACAATGCCTGAAGAAAATAACACACATAAATTTCTCAGAC
AGAAATATAGACTCGATTGATGACCTCTCTCTTTGCAGAAACCTGAGTGTTTTATATTT
ATATGATAATCGCATTAGTCAAGTCACTAACCTGAATTACACCACAAATCTGACCCACT
TGTACCTACAGAACAATTGCATTTCCTGCATTGAAAACCTCAGTTCACTAAAGAAATTG
GAAAAACTGTATCTGGGAGGCAATTACATTGCGGTCATAGAAGGCTTGGAAGGATTAGA
AGAACTAAGAGAGCTTCATGTTGAAAGTCAGAGGCTCCCTCTAGGCGAAAAGCTTCTGT
TTGATCCAAGAACTCTTCGTTCTCTAGCGAAATCCCTCTCTACTTTGAATATCAGCAAC
AATAACATTGATGACATAAAGGACTTAGAGATCTTGGAGAATCTTAATCACCAGATTGC
AGTTGACAACCAGCTGATGCACGTGAAGGTAACATGAAGTCGACGTAAAGTATTTACAT
*GAAATTGTGGCTTTACAATATACTTTGACATATTTATAAAAAA*NTGTGTCACCAAAAAA
AAAAAANGGNNNCCNNNNCNNTTAAGAGGGGNNNNNNNNNNGGGNACTCNGGGNCTCGA
GTGNAANGAATCGAANTCCGAAANCCCNNNTCTCGGNNNNNNGGGNNNNNANAATAATA
TNNNNTAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

>Y2H-107

NNNNNNNNNNNNNNNACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGATAGCACCTTT
TATTAACCAGTTTCAGGTGCCTATCCGGGTATGTCTGGATTTATCTTCTTTGCCATGTG
TCCCTTTAAGCCAACCAGTGGAGCTGTTGAGGTTGGATCTGATGACTCCATATTTGAAC
ACTTCCAATAAAGAAGTAAAGGTCCGTGTGTGCAGGTCTGGTCGAGTGACCGCTGTTCC
ATTTTGGTTTCATCTGTGCCTGGATGACGAGGTGAGGCTGGACATCTCAGGTGAAGCCT
CCCACTGGAAACAAGCTGCAGTTGTTCTGGATAATCCCATCCAAGTACAAGCGGGAGAG
GAGCTTGTGCTCAGCGTTGAGCATCACAAAAGCAACGTCAGCATCGCAGTGAAGCCGTG
AGATTGTCATCGGACAATGAATGTGTGAAAATCCCTGGAGTCATCAAAAATGTTTAAGG
*GTTGCAATAATAAAGCCCATGTTAAAAGTCAAAAAAAAAAAAAGGCCCCCCC*NGCCCTTA
AGGGGGGGGNATNNAATCGGGATCCCTTCAGGTTCAACTTCAAAATAATTNNAAAATCC
TAAAAACCCCCCAAGTTCCCTTTCACTGNNTGGGGGCCCCCNNTAAAAANNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

>Y2H-113

NNNNNNNNNNNNCNTNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGGAGAGTGTGG
GAAACCCAGAAGAACCACAAGCTGGGGTGGCCAGAGTGAGGAAGGAGCCCAGCAGGACG
AGCAGGTGGGAGAGTCACAGAAAGTGGCCCTGGCAGTTGGTGTGGATCAGTCCAAGAAT
GAAGTGAAGGCACAGAACTACTTTGGATCCATTAATATCTCTAATGCCAATGTGAAGCA
GTGTGTTTGGTTTGCCATGAAAGAATACAACAAGGAAAGTGAGGACAAATATGTCCTCC
TTGTGGACAAGATACTCCATGCCAAACTTCAGATCACAGACCGAATGGAATACCAAATT
GATGTTCAGATCTCCCGCAGTAATTGCAAAAAACCTTTAAACAATACTGAAAACTGCAT
CCCTCAAAAAAACCCGAACTGGAAAAAAAAATGAGTTGCAGCTTTCTGGTAGGAGCACT
TCCCTGGAATGGCGAGTTCAACCTGTTGAGCAAAGAATGTAAAGATGTCTAAGGGTGTC
TTGGAGCACCCCTCCCACTTCTGGGTCTGCTTTTTTCTGTAATAAAGAGCAACCTGTGG
*TTCAAGTCTGTCACCAACACAAAAAAAAAAAA*NGGCCCCCCCCCCCNTTAAGGGGGGGN
CTTCGATCGGGGATCCTTGGGGTTCAAGTTGCAAAAAANNNNAAAAACCTAAAAANCCC
CGGAAGTTCCTNNANNNNCNGGGCNNGCCNCNNAAAAAAAAAANNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

>Y2H-114

NNNNNNNNNNNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GA*
ATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGGGGCAGGCA
TTATTAGGTTAGTGTTACAGATCCAGTTTATAGACTGACAGCTTCAATTCACGGGACAC
CTTTTTTCCTTTGTGGTTTGTGTATTTCTGTGTTTTGTTTGCTTTTGTATCCGTTCGAA
AATTTAACCCACATTTTCACATAGTGAAAATTTCACATGGTCTGATTAGCCAAAAAAGA
ATAAGATCTAGAAGTAGAACTCACACCATTTTTTTTCTTAACTTTGATTTCTAAAACAA
CAAAAACTACCACATGAGCTGAATAAGAAAATTCACTAGCAAGTTCTCTGGATGATTTT
TGGTGCTGAACAATGACATGAGCCTCAGACTGTAAAATAGAGGTAGTTGGAACTAATGT
ACAGAACTAAATTTCTTAACCTTATTTGCGTTTAATTCTGTGAAGTTTCAGTTATCTAA
AATAAATGTGTAATGTTTCAGATTGCAAGGTGATAAGTAATGTAGCATTTGTAAGATAC
TCTTGTCAATATTAACTAGTAGGATTTTGATTTGTACAGTTTTAATTGGTTAAAATGAT
CTCATTTTAACATCCACTGCTATAGATGAATAATGTAAGTTCAGATTTAATGAATGGTG
GGGAGATGGTGCATGTAATTTTTTTGCAAGTATTGAGAGTTCTGTATGTTTTGAAAAGA
GTAATTTTAACGTTTGGGTGCCAAGAAGTGGGTTTTCTCAGAGTCCATTGCCGGCAATG
GGCAAGCCTGGCGGTACTGGCACGGAGCGTTAACCACACCTTACTAATAGCAAGGCCAA
*TAACTTTGAAATAAAGTTTTAGACAA*NGAAAAAAAAAAAAAAANGGCGNNNCCGCCTCC
AGAGGGGGNNATCNNCCGGNACCNTGGGNNNNANNGGNAANGNNCCGAANTCGGAAANC
CNNNNNTCCTCNNNNTTNGCCNGGNCNNACATAANNGNNNNN

>Y2H-121

NNNNNNNNNNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAAT*
TCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGATGGATGTGGTG
TCAGGACGCTCCGTTCCTGCTGACATCTTCCAGATGCAAGCAACAACCCGATACCCTGG
TGCCTATTACATTTTCCAGATCAAATCTGGCAACGAGGGTCGAGAGTTCTATATGCGGC
AAACAGGGCCTATCAGTGCCACCCTGGTGATGACACGCCCCATCAAAGGGCCTCGGGAC
ATCCAGCTGGACTTGGAGATGATCACTGTCAACACTGTCATCAACTTCAGAGGCAGCTC
CGTGATCCGACTGCGGATATATGTGTCGCAGTATCCGTTCTGAGCCTCTGGCTAAGGCC
TCTGACACTGCCTTTCACCAGCACCGAGGGACGGGAGGAGAAAGGAAACCAGCAAGAAT
GAGAGCGAGACAGACATTGCACCTTTCCTGCTGAATATCTCCTGGGGGCATCAGCCTAG
CATCTTGACCCATATCTGTACTATTGCAGATGGTCACTCTGAAGGACACCCTGCCCTCA
GTTCCTATGATGCAGTTATCCAAAAGTGTTCATCTTAGCCCCTGATATGAGGTTGCCAG
TGACTCTTCAAAGCCTTCCATTTATTTCCATCGTTTTATAAAAAAGAAAATAGATTAGA
TTTGCTGGGGTATGAGTCCTCGAAGGTTCAAAAGACTGAGTGGCTTGCTCTCACCTCTT
CCTCTCCTTCCTCCATCCCTTGCTGCATTGCCGCTTTGCAAAAGTCCTCATGGGCTCGT
GGGAAATGCTGGGAATAGCTAGTTTGCTTCTTGCATGTTCTGAGAAGGCTATGGGAACA
*CACCACAGCAGGATCGAAGGTTTTTATAGAGTCTATTTTAAA*NTCACATCTGGTATTTT
CAGCATAAAAGAAATTTTAGTTGTCTTTAAAATTTGTATGAGTGTTAACCTTTTCTTAT
TCATTTTGAGGCTTCTTAAAGTGGTANNNTTCCTTCNANNNNNNNNGATACATNNNNNN
TTCAGTCTTNCNNCNTNNTCCTTCNNNNNTNNNNCNNNGTTTNNNNANNNCCNNNNNNN
NNNNNNNNNNNNNNTTNNCNNNNNNNGNNNNANNNNGNGNT

>Y2H-124

NNNNNNNNNNGNNGNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT**
GAATTCCACCCAAGCAGTNGNATNAACGCAGAGTGGCCATTACGGCCGGGAGAGGGAGG
GTCTCCAGTACTGCCTGAAAGTCTAAACAGCTTGGGTTTAAGTTTTCACGGTCAGGTCA
GCCGTTTTGAAGGGAAAGTCAGACCAGTACTACCCCACCCCCAAAGTACCTGTAAAGTC
ATTGTATCATTGAGTACAATTCCATGCAAAAAGACAAAGCAACCCTCACCAGGAAGTTC
TGTCCCCCCCCACACACTCGATTTCTATATTTGTGGATTTTATAGTAATCAACTATGTC
AGATAAAGGAAACCTTTATTGAATGATTCAAAAAATTTAAAAAGTGATTTAAAACAGAG
GAATTTATTATGCACAGAAAAAAATGTGTCTTGTATTAAATATTTTTATTAAAAAGAAT
GTTTCATTAATGCATTCATAAGAAAACTAGGTTTGTTGTGCGGGATGTTTCTGGTTTCG
TTTCAAATGACTACATTCCTACTGCTACTTGCCCTGAGGTCCTGGCTCTGTCAGGGTGA
AGCCCTGATGCCCTTTGCCTCAGGAGCAGAACAGGCTAGTGCTTCGGGTGACTCAAAGT
GCTGACTCAGCTGTAGCCTGCATAAGTGGTTTATTTGGGGTTGTTTTGTTTTGTTTTGT
*TTAGCCTTTGGGGTGTTTTAAAGAACATGTTTAACAAA*NAATAAAAAANTAATAAAAGC
TAGGGCTCCTGCTGCCAAGATTTCTGTCAAGGAAACCTTGTTTGAGAACGATGTTAATA
AAATTCTATTGCAAAATTCTTTTCTAGAACNNAATAATAACACANAAAANGAAGAAATT
CCAAAGTAATAAAAACTTTTCTTGAAACAGATTTANAAANGTGTTTGCTTACTATTTGA
ATAAATAANTTNACTTACGTTTCTTANNNNNTTTAATTTTTTANNGTNTCTGGNNAGAN
TTGNNTGGACCATAATGTNNTTNCTGTAGAACAANTTATTNANNNGTTNCANTTCCNNN
NNAAANTNCNNTTTATGTTATNNNCTACCANNAATTGGNAACATTNGANGTNTTTAGGN
NNTTGAANTTGGNNNTGNTNATGTTTTTNANTAACAANNNNTTAACNNNANTCCNANCC
NNNNGGNN

>Y2H-126

NNNNNNNNNNNANNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTNNCAACGCAGAGTGGCCATTACGGCCGGGGGTGGCAACCGC
CGCTGGCTTTGGAGGAGCCTGGCACAAGGACACTGACCAAACCAGTTACAGACCCGACT
GGACAGAGCGCTTTTTGAAAGTGGAAGTAATATCTAAGCCTGCTTCTGGACATGCTGTC
AAAGTGGAGTTCTCGCTGAGAGTTTGACTTATGCAGAAACAATGGTTCGACTGACGGTG
GATCTAATTGCCAAAAACAGCAATCTTAAACCCAGAAAAGAAGAGACCCTTGCACAATG
CCTGAAGAAAATAACACACATAAATTTCTCAGACAGAAATATAGACTCGATTGATGACC
TCTCTCTTTGCAGAAACCTGAGTGTTTTATATTTATATGATAATCGCATTAGTCAAGTC
ACTAACCTGAATTACACCACAAATCTGACCCACTTGTACCTACAGAACAATTGCATTTC
CTGCATTGAAAACCTCAGTTCACTAAAGAAATTGGAAAAACTGTATCTGGGAGGCAATT
ACATTGCGGTCATAGAAGGCTTGGAAGGATTAGAAGAACTAAGAGAGCTTCATGTTGAA
AGTCAGAGGCTCCCTCTAGGCGAAAAGCTTCTGTTTGATCCAAGAACTCTTCGTTCTCT
AGCGAAATCCCTCTCTACTTTGAATATCAGCAACAATAACATTGATGACATAAAAGACT
TAGAGATCTTGGAGAATCTTAATCACCTGATTGCAGTTGACAACCAGCTGATGCACGTG
AAGGTAACATGAAGTCGACGTAAAGTATTTACATGAAATTGTGGCTTTACAATATACTT
*TGACATATTTATAAAAAAATGTGTCACCAAAAAAAAAAA*NGNNCCCNNNNNCNNTTAAG
AGGGGNCTTNNNNCNGGNANNNTNNNGCTCAACTTCNNNNANNNNNAAANNNNGAAANC
CNNNNTNCNNNTTNNNNNNNNCCTCCTCCNAGAGAGNGA

>Y2H-127

NNNNNNNNNNAGACCNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GA*
ATTCCACCCAAGCAGTGGTNNCAACGCAGAGTGGCCATTACGGCCGGGGGACCCACTGA
TGATGAAGTATAATCCAAATGTCCTGCCTGTTCAGTGCACTGGCAAGAGAGATGAAGAT
AAGGATAAAGTCGGGAACATAGAGTACTTTGGGATGGGCGGATACTACGGCTTCCCTCT
GCAGTATTATCCCTACTACGGCAAACTCCTGCAGCCCAAGTACCTGCAGCCCCTGCTGG
CCGTGCAGTTCACCAACCTCACCGTGGACACTGAAATCCGCGTCGAGTGTAAGGCGTAT
GGTGAGAACATTGGGTACAGTGAGAAAGACCGTTTTCAGGGACGCTTTGATGTAAAAAT
TGAAATTAAGAGCTGATCACAAGCACAAATCTTTCCCACTAGCCATTTAATAAGTTAAA
AAAAGATACACAAACCTACTAGTCTTGAACAAACTGTCATACGTATGGGACCTACACTT
AATCTCTATGCTTTACACTAGCTTCTGCATTTAATAGGTTAGAATGTAAATTTAAAGTG
TAGCAATAGCAACAAAATATTTATTCTACTGTAAATGACAAAAGAAAAATAAAAATTGA
GCCTTGGGACGTGCCCATTTTTACTGTAAATTAGATTCCGTAACTGACTTGTAGTGAGC
AGTGTTCTGGCCCCTAAGTATTGCCGCCTTGTCTATTTTATTTAGTGTACAGTACTATA
GGTGCGCACTCTGGTCATTTTTCAAGCCATGTTTTATCATATCTGTTTTCTACTTTACG
TGAGCAAGGTTTGCTGTCCAAGGTGTAAATACTCAACGGGAATAAAACTGGCATGGTAC
TTTCTCCTTTCTTTCTTATTTTCTTGGCTCTGAAATTTCAAAGGTAACGGCCCATCGAT
*GAGCATTNAACACACTCCATAGTCTCTTCCTGGG*NNTCAGGTCTTNNTANTNTTTTTTT
TNNTNNCTTATTNCCTGGGNNNGGGGGNNGGNCTNNNNNGGGGNAACTGCCNTTAANTT
TAAGNNNNNNNNGAAAATNNNNNNNNNGGNTT

>Y2H-128

NNNNNNNNNNNNNNNNNNNCCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCA**
GT*GAATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGCATCAA*
CATTCCTATGGATCCGAGCATCTTATCCACGCTTCCGTTACGATCAACTTATACATCTT
CTATGAAAAAAACTTTCTACCCCTAACACTAGCATTATGTATGTGACATATTTCTTTAC
CAATTTTTACAGCGGGAGTACCACCATACACATAAAAAAAAAAAAAAAAAAAAAAAAA
*AAAAAAGGGCCCCCCCGGCCT*NNANGGGGGGGGCTTCNATCNGGGNCCCTTNNNNNNNN
ANNTNNAAAAAAANCNAAAATCNNAAAAANCCCCNNNNTTTCNTTTNNNNGGNNNNNNN
NNNNNNNAAAANNANNNNNGNNATNNNNTNNCCGGNNNNNTANNNTNTTCCNGAGAGCT
GCGNTNGNNGTACCNNCGNNNGNNNNANACGGAANNAAAGCGACNNNNNNTTGNNNNCN
ACNNGNNGNNNNNNTGTGNNANNNNNNNGNCGNCNANNATNAATNNGCNCANNNGNNNA
NGGNGTGCTCGNCTNTCNNGAGNANCTCCNNNNNNNNNNNTGNNTANNATNNNNNNNNN
NNGNNNNNCNNNNNNAGNCNTCCNNNNNNNNNANNNNNNNNNNNNNNNCNNNNNNNGN
NNTGNNNNNNNGNANNNNNNNN

>Y2H-129

NNNNNNNNNNNNNNNNCGACGTACCNNATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTNTCAACGCAGAGTGGCCATTACGGCCGGGGATCAATGACCA
GTCCATCAACCTGCTCACTGCCGTCGGCACCACCACCCGAGACTCGCTGACAGAGGTCA
ACCTATCAGACTGTAATAAGGTAACTGACCTGTGCCTGTCCTTCTTCAAACGCTGTGGA
AATATCTGTCATATTGACCTGAGGTACTGCAAGCAAGCCACCAAGGAAGGCTGTGAGCA
ATTCATAGCTGAAATGTCTGTGAGTGTCCAATTTGGGCAAGTGGAAGAGAAACTCCTGC
AAAAACTAAGTTAGGCTCCAGGACAAGTATGTAAATATGGAGGGTTGGGGGGAATACAA
CCAGGAAGGGTGAAAGACTGCAAAATCAAGTGGGCCTATTATTTCCCTTTGGAATTTGA
AACAAGAGACTACAGGACACAGTTCTGTTTTGCAAATCTTGCCAAGGGGGAAGCCATTC
AACCACCCATTTTGGGATGAGTGAGCCAGATGTTTTCCTTTGCTCTTTCTGCATCGTAA
CTGCACTGCTTTCTGGACCATTTCTAAGGCGGCCTTTAGAAGACATTCCAATTGGAAAG
GAAGATGAGCTTTGGAGAAATCCTCTTAGAAGCCTAAGCTTGGTGTTTTTATCAGTAGT
TTTTGTTTGGGGCACTTGTTCCTTGCAATGCTGAGGCTTCTGGCGTCAACATGGAGTGG
ACCACACACACCATCTGCAGCCGTCGTTGATACATTTTGTTTAGTTGGTTTTGTTACAT
CTTACATTATGCAGAACTATTCTTGTATAAATTGTTTAAAAATTATTTATGGGAGATTT
GAATGCATACCAGTGTTTTTATTATTTAGATTGCCAACTGTCCTTCAGGTAGGAGAGTT
*AACCTGTACTATGTCAGCATAAAAAGCC*NATCTCCAGATGTGCCCAGGTGTGACAGGGG
NCNAGGANNTTCACTTNNNAGNNNGTTAACTGGATGCTGAANNNCTACTGTNNNNNNNA
GNNNNNACNNNNNNNTGTGGGGNNNNNGNNAANGTTANNCNGNNAANNNNTTTGCCAGN
TTNNNGNACTCCCNNNNNNNNNNNNNNNNTTNNNNNNNNNGNTTG

>Y2H-130

NNNNNNNNNNNNNNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGCACCTTCCTTG
GTTCCACCGCTACCCCTGATGACCCCTATGAAGTGAAAAGAGCCCGAGAGTTGAACCAG
CTGCTGGGTCCCAAGGGCACAGGCCAGGCTTTTGACTTTACCCTAGACCTGCACAACAC
CACAGCAAACACTGGAGTCTGCCTCATCTCTGAATCCAACATCTCCTTCAACTTGCACC
TGTGCCACTACCTACAGCGGCAGAACCCGGGGATGCCCTGCCGCCTCTTTCTGTATGAG
CCAGCTGGGACGGAGACCTTCAGCGTGGAATCTATATCCAAGAATGGAATCTGTCTGGA
GATGGGCCCACAGCCTCAGGGCGTGCTGCGGGCCGACCTGTTCTCCCGGATGCGAGCTC
TGGTGGCATCCATTCTGGACTTCATCGAGCTCTTCAACCAAGGCATGGACTTACCCGCC
TTTGAGATGGATATCTACAGGAACTTGGGCAGTGTGGACTTCCCACGCACTGCAGATGG
TGACCTGGCTGGCACTGTGCACCCTCAACTGCAGGACCATGACTTTGAGCCACTGAGGC
CTGGTGAACCCATCTTCAAGCTTTTCAGCGGAGAAGACGTACTGTATGAGGGGGACTCC
ATTGTGTACCCTGTGTTCATTAATGAGGCTGCCTATTATGAGAAGCACGTGGCATTCCT
GAAGTCTGAGAAGATCAGGGTCACAGTGCCTGCCCTGCTGAGGTTGACCCCCCGCTCCA
CCCAGACTCCCTAACCCAAATCATACCTGGCCTGATCTCAGTTTTCCTATCTGAACAAT
*GGGTCCCAAGCCCAGGGTCCCCTATCTCACTCTGGA*NNCNATATCCNTTTCNAGACANC
TGATCTCATTATATAAAATAGACTGTGCCCAAAAAAAAAAAAAAANNGGCCCNNNNNCC
NNAANGGGGGNNNNCNNNN

>Y2H-131
NNNNNNNNNNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAAT*
TCCACCCAAGCAGTGNTATCAACGCAGAGTGGCCATTACGGCCGGGGGTTTACTACCCA
GCTGACCCACCTCCATACATTGATGGGATTCGAATTAACAGCCCTCATTACTTGACTAA
GATAAAGCTGACCACGCCCGGGACTCACACCTTCACCCTAGTGGTTTCTCAGTATGAGA
AACAGAATACGATTCATTACACAGTCCGGGTGTATTCAGCCTGCAGCTTTACTTTTTCG
AAGATCCCTTCACCATACACCTTATCCAAACGGATAAATGGAAAGTGGAGTGGTCAAAG
TGCTGGAGGATGTGGAAATTTTCAGGAGACTCACAAAAATAACCCCATCTACCAATTCC
ATATAGATAAGACTGGACCGTTATTGATTGAGCTGAGAGGACCAAGGCAATACAGTGTC
GGATTTGAGGTTGTAGCAGTTTCCATAATGGGAGATCCTGGTCCTCATGGCTTTCAGAG
GAAATCTAGTGGTGACTATAGGTGTGGATTTTGCTACTTAGAACTGGAAAACATACCTG
CTGGGATCTTTAATATCATCCCTAGTACTTTTTTGCCTAAACAAGAAGGACCTTTTTTC
TTGGACTTTAATAGTACTGTCCCCATCAAGATAACACAACTCCAATGATTGAGAACTCT
CAGCTATTACTGGATTCCATATTTCCCAGTCTCAGCTCTTCAGACAAGGAGACAAATCA
CGATGGTGAACCATCAGAATGGAATTAAATCAAAAGCATGTTACGGAGTCTGGTGCTCG
TTGTGGTGTTTGGTAAGGATTTTAGATAGTGGAATTACCTTAATCACCTGAAATACTGT
TTACATGGTTTGTGCACACAGAGCTGGCGTGTACTCGGTGGCGGCTCTGTGTAAATCAC
*AGGCTAGATTCAGCC*NTCACTGGGAGATGCTTNCGNNNTAGGACACGAGGA

>Y2H-132
NNNNNNNNNNNNNNNNACGACGTACCAGATTACGCTCATATGAAC**ATGGAGGCCAGT***GA*
ATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGAGGCGTGCT
GTGCGGCTGGCTGTACGGTGAACTGTCTGGCCTTTAGCGCTTCATCCTTGGTTAAGGAA
ATGACCAACCAGTACAGTATTCTCTTCAAGCAAGAGCAAGCCCATGATGATGCCATATG
GTCAGTTGCCTGGGAGACAAACAAAAAGGAAAACATTGAAACAGTGGTCACAGGATCCC
TGGATGACCTGGTGAAGGTCTGGAAATGGCGTGATGAGAGGCTGGAGCTCCAGTGGAGC
CTGGAGGGACATCAGCTCGGGGTGGTGTCTGTGGATATCAGCCACACTCTTCCCATTGC
TGCATCCAGCTCTCTAGACGCTCATATTCGCCTCTGGGACTTGGAAAATGGCAAACAGA
TGAAGTCTATAGATGCAGGACCAGTGGATGCCTGGACTTTGGCATTTTCTCCTGACTCC
CAGTATCTGGCCACAGGAACTCACATGGGGAAAGTGAACATTTTTGGTGTGGAAAGTGG
AAAAAAGAATATTCTTTGGACACTAGAGGAAAATTCATCCTTAGTATTGCATATAGTCC
TGATGGGGAATACCTGGCCAGCGGAGCCATAGACGGAATCATCAATATTTTTGACATTG
CAACTGGAAAGCTTTTGCATACGCTGGAAGGCCATGCGATGCCCATTCGCTCCTTGACC
TTTTCCCCTGACTCCCAGCTCCTTGTCACGGCTTCAGATGATGGCTACATCAAGATCTA
TGATGTACAACATGCCAATTTGGCTGGCACACTGAGTGGCCATGCGTCCTGGGTGTTGA
ATGTTGCGTTCTGTCCTGATGACACTCACTTTGTCTCCAGTTCATCTGACAAAAGTGTG
*AAGGTT*NGGGATGTTGGAACAAGGACCTGTATNNNCACCTTCTTTGATNNCNAGGATCA
GGTTTGGGGAGTAAANNATAATGGAAATGGATCAAAATTGGATCTGTTGGAANTGACCA
GGAATNCATNNNNTGNNGNCNAATTANNNNNNGGATCCNNGGGNTAGGGCNNNNNNNNN
NGGGNNNNNNNNNANNCCCNAGNTTTTNGNNNNNNANGNNNNNNANN

>Y2H-133

NNNNNNNNNNNNNNNACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGCTAGTCAGTGG
CTTGATCCCATCAACACCAGTAGAATGTTCTGGAAGCAGTTATATCTGTTCCCTGTGCA
TGATATCTCCTTTCCCTTCTCTCTTCCCAGAGACACAAAGCATAGTCATAAAAATCCGA
TTAACAAAGCCAGGTGTGGTGGCCATGCCCTAATCACAGCACCGCAGAGGCAGGCAGAT
CTCTGCGGCATCAGGGCCAGCTTTGTCTACAGAGTGAAACACACCAGAACCCCAATGGA
TCAAACACTATAATTTTTTTGCTTCAATTTTTGGATGAATGGTGTATAATATATTGGAC
CTATTAGACTCTAGTTCATCAGCTTAGGGAGGGCATGGTTTTATGCCCACTTTACATAA
AATACAATAAGACAATACAATAGAGTTCTGCTAACTCTTCCACAATCACAGAGATGCAT
CAATACAAGGTAGACGGCGAGAAAGACCGACTTCCCTCACCAGAGGAGCACCCCTCGTC
TTCACCAAGCCACACTACCCCAAAACAGAACAAGTGAGCTATGGGCTTGCATCCTCACG
*GGGGGGGGGGGGGGGG*NNGAANNNNNNNGNNNGNCNNGGGGGGGNAANNNGTGGGNNNN
NNNACCGCNNNCNTTTGGGGAGGGAAACCGGCTGGGGGTCTTNNNNNNNNAANNGAAAA
TTTAGTCCCTTGGGNCGGGATTCCCGGGGGGGGNNNNGGGAATTTTCCAGNNCNGGNCT
GTTGTTGCTGAGGGGAAACTTGNCNNTGGGGATCTGGGCAAGNNNTNNANGGTCCTGCA
ANNNGGNNTTNNNNAANNNNCCGGAANANNCCNGGTNNNNCNTTGGCAAGGNNGGGGGG
GGGGG

>Y2H-137

NNNNNNCTGGAGNCNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GA*
ATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGAGCACACTC
TTGTGCAGGCCCCCCTGTTCAACACCTCACAGGGAGTTGTCATATACCTGCGACTGAGA
CTTTCCTTTAGTCTACAGGCTTCTGGGGACACTGTCTGTTCAAGAGTAAAGTGATTTGT
TTATGGAAGAAAAGAAAAAAAGTCAGATCTGTTCATCAGAAAAGACAAAAAAGAAGAAT
TAGCTTACCTGATGGAGTAAGGGTGATGAATACATCAGTCACTGAAAGGAGGTAGAGCG
CTGCCCTTTCTCTTGCTGCCTACTGTTTCTAACCCATCCCCCCGCCCCTGAAGCACAGT
CCTCCCTGCTTTTTCTATACCTTCAGATTGATTGTTGAAGGGAAAAGACAAATGGAGTT
TGTTCAACCAATTACTTTAACTTTATATACAACTCAAACCCATAGACGTTTGGATGAGG
ATCTCTCAAGGATTTCCCCGTAGAATGAAGAGCACAAGTTAACCTTTAACAACCATCAG
AAACTCTTGGGAGGAGTTAACAACAAAAACAAAAAGTCCTCATGCTTGTCTCCTGCTCC
CCATTCTGGAGACGCCATGCTATGGGGAGCTCTCCGGATAACTCTGCACCCTGCCAGAA
AGCACTCTATTATGCTGTGATTTTTTTGGGGGAGTTCTGGACAAGTGGACCTGTGTGAC
CCTGACCTAACTAAAAGAGAACAAAGAAATAGACTGTACTGAAACCCTGGCTGAGGGTG
CTGGGGAACTCTTCCCCTTTGGTGAGATTCTCCCTGGAGTATCATCAACATTTGTCCAG
ATGCTGCTGGAGTCTGCGCCTGGATACCAAGGATCTTTTCCTTCCCATTGCTTCCAAGC
TGCTGAGCCATGATTTCTTCCTTGTCTCTCCAGTTGGCAGCACATCTCTTTACAGTAGA
*TTGCTCTTCAGTTCACAGATATCATATCGTTCTCC*NANNGAATAAGTCAAGCAATCAGA
GNCTGTGGCTNNNACTGAAAACTNNNNNNNNNNNGCTTNNTTNNNNNNANTCANNCTTG
TNNTANCTTNCTGAGNCAAANTNTNNNNNNNNNNNGNATNNNNNNAGGNCAAGGNCCN
TTNNNNGGAANNNCTNNNCNANNNGATANNNNNAAAAANNGGN

>Y2H-138

NNNNNNNNNNNNNGNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAG**
T*GAATTCCACCCAAGCAGTGNTATNAACGCAGAGTGGCCATTACGGCCGGGGGGAGGAT*
GGCTGGAGCAGCCAGTGTCTACAGTTAAAGATTGCCAAGTGTATTTCAATATGGTTGTA
AAGATCACATTCAGATATGGTTTGCACAGGAGGCAGAGCAGCTCTAATGTGGATAGATG
ATAACCTTTACTGCTCTGCAGTAAGTTTTATAGGCTACATTTTCCTAATTTCATATTTA
ATTGTATTTTTACTACAAAAAACATTGGAGTACAAATCATATGAAATGAAAAAGGTGCC
TTCATGGAAAAATATTTTTTTGTATGTGTGTGATATAAAAATAATGTTGAGTAAATTTC
TGAGCAAGAGTTGTATACCCTTTGTCTAAATATGCTTGTCTGTTCACGGAAAGGACAGA
ACTGGAAATTTTAAGAAAGGAAAAAAGTACACTGAAGTACTTTGGATAACTTGAGTTAA
TGAAAAGTTCTAACTGGTTTGGTGCTGCCTAAGTGATTATGTCAAGAAATTGAAAAATG
ATAAACCCCTAACAACGGAAGAGTTAATTGCCTCTTAATGTCAGACAACAGTTGAGGCC
CCTGAGTATAAGTCAGTGGCTTCATTTCATTTTAAACATTCAGACAAGGTGGCCTAGAA
AAAAAGTCTTGCATCCTCTGTGTTTTTTAATTCTTAAAAAAAAAAAATTGAAATGTGTG
AAAAGCTTATGAGTAAAATATTAACAGATTGAAAATCTTACACTTTTTTCTANAAANTG
*ACATGTTGTC*NNNNAANAAAAAAAAAAAAAANGNNNNNCNNNNNNNCCNNNTANNGGGN
GGNNNNNNNNNNGGNNCNNNNCNANNNNNNNNNNNNANNNNNNNNNNCNGAANNNCNN
NNTNNNNNNNACTNNNCNNNNNNNGNNNNNNNNNA

>Y2H-142

NNNNNCNNNNNNNCNTNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GA*
ATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGAGTTCATTGA
GCTATAGTACACACTGTTTTCCTCTCTGTATAATGGTGACAAATATAGCAAGTGAACAT
GTCGTAAATATAAAGGTTCAAGTGATGTATTGAAAATAATTTTCTAACTGCTTTGTATG
TATCACATACTCTAAATTGCACAGTAGGCGTGTTTATTAAACAGATTGGCTGGGAAAGT
TTTGAAATAGTACTGAATTTACAGTCTCAGTGCTATTCCTGTTGCTGTGTATGTGGTAG
GTTATACCGAAGTAACTAATCTTCTAATGAATGCTGATATTTTATTAGAATGAAAGCAC
AGATTAGCATTAATATTTTTTATCCTGGAAATCCTTTGGCAAGTACTGGGCAGCCCAAA
TTTAGAAAACTATGAGATTTCAAACCACGTAAACATGGTTCTGTTGGACATTTTATTTT
TTAACTGTTAACTTAGGTGTTCGCAGTTCTGTTGAATGGGACTTCTGCAGGACTGTATT
TAGCATGCTGTAAGTACTTTGGGGTGAAATAGGCTCCGAGTCTAAATTCTCAGAAATGT
CTGTGATGTATTGACCATGAATAGCGTGTACCACATGACAGAAATGAAATTACACTCTT
AACCCCTTCTCAACGTGGACTGGTCTTTACAGGGGCACCAATAAATGACAGATCTGGAG
AGTCTCCTGGTGCTTATTTTAGAAGTCAAATCAACCCAATTAGTATCTTGTTAGTAGAT
ATAATGAGCCTATTTCTTTCTAAAAAGGCTTGTGGTCTGGACGTGCTTTTACAAGAAAT
*AGTGACCTTGGGGAATATGTAAACAAGAGATCTTTTTCT*NGAGTTNGGGGGNAAGGTTG
GGGGACCTGTATAATGACAAAAAAANTCGGCTTACACCATTGACACCANANANNNNNAA
TTCAAAAANNTNNCNANCCAAGNNTTNNNNCAACNNNNCNNGATGCTTTCAGNNNNTNN
GNCNTNNNNTTTTNGCAGATTTTTNCNNCCTTCCGTGGCCTCTTAATTNNNNGGNACCN
GNTTTNNNNNNNNNTAANNNNNNNNANNNGGNTNNCNNNNNNNCANCCCCNGGNAAAAA
ANCNNNNGNNNNN

>Y2H-144

NNNNNNNNNNNNNNNNNNNCGACGNACCNNATTACGCTCATATGAAC**ATGGAGGCCAG**
T*GAATTCCACCCACCACAGNGGTATCAGCGCAGAGGGGCCATTACGGCCGGGGCATCAA*
GGCAACTTAACCTCGCATTGCATGTCCAGTGCTCCCCTACAATTTTCTCACAGATTCTG
TACTGTGACACAAATGCATTTTATTTTTCTCTCCTTACCTTCTACTAATATCCTGATTT
CACTAAAGAGTGACTACTCTGTTCTCTGGAATCAAAAGAAACTTACTTTGATGATGAGT
TCAGACACTGATACTATCATAAAGATGCAAGAATGAATATTCTAAAATCTTGGGAATGC
TCCAGGAATCATGCTATAGCCCGGACAACCTCAGCATCTCTGCAGCACTTTACAGCAGG
ACATGCATTAGAACACCTAGTGATAACCATAAACGTTTCAACATGAACGTCTCACTATA
TGGAATTCTGGTATATAGCAAGCACTTTTAAAGTGACTGAACCAAAATTTTTTCCATGT
TTAAGAAGCTGAGAGTCGAGACTCTTAGCTGGTCTGCTCCAGTAAAGACATCAAATCTA
GAGGAACTCCTAAACTCAGAGAGCAGTTGGTAAGAACACCCCCCCTTTCTCCAGAATCC
CAGAGCTGCTGCTGGGAGTCCAGGAGTCCAGTGGACTTGGGACCCATCACCTACCAAAA
TCCCTGCCCTCCCGAATACCACTCCCCTTCCATCCTTTCATTCCCTTTAGCCTGATACT
*TT*NNGCTGGAGTAGACTCTTAGCTGGNNTNTTCCAGCGAACACATCGTATCCTGACCCA
TCCTAGAGNACNNACTNNNCTCTGANTANNNGNNNNCNTNNTNNNCNCNNNNNTTNNNA
GGATCNNNNNNNTTGNN

>Y2H-156

NNNNNNNNNGNNNNCNTACGACGTACCAGATTACGCTCATATGAAC**ATGGAGGCCAGT***G*
AATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGCAGACAAA
CGGGATGAAAACAAATGGATCAAGTTGCATGCTGATCAGGAGTATGTGCTTCAAGTCAG
CCTACAGAGAGTCCATTTTGGGCTTCCAAAGGGAAAACATGGGAATCACGCAGTTACTC
CTCGATTTCCAAAATTAAAAGATGAAGGCTGGTTTTTGATATTAGGAGAAGTGGATAAG
AGAGAGCTGATGGCTGTAAAACGAGTAGGATTTGTTCGAACACATCATGATGCTTCCAT
CTCGTTCTTTACTCCTGAAACACCTGGAAGATATATCTTTACCCTGTATCTCATGAGTG
ACTGTTACCTCGGCCTAGACCAGCGGTATGACATTTATCTCAATGTCATAAAAGCAAAC
ATTTCTACAAAGGACTCTGATGTCTTCACTGACTTGTCAGTATAGTATCCTGACCCAGA
CAGGCCATTTGACCAGAGGCTGAGAGTCTTGACTGTTCAGTCATGTGAACAAGTCGAAT
TACTTGGTGTCTACCTTGACAGAATCAACTTCAAACCTCAAACCAACTAGGAAATTGAC
AGCGGCTGCAGTGATGATTCCAGCAACGTTAGCCACAGAGGCCTTTTAGAAACTGCTGC
CTTTTGTAACAGATTTATGAGTTTCTTTACATGTAAGATAAGTAAGCATATAGTATCAT
*TTTTAAGTTGTGTATGTTGGAGGTGG*NNAGGTTATGGGTNNATCAAACTTTGCATCTCT
GGATGTACTGNTNACATGAATACCAGCTACNCCTTTTNNTAAGGAAAATTACATTGTGG
TGTGACGGGTTAAATGTTNAAGCAATCTAACATATATAANTCCATGNCGTGATACAAAG
TCGATCTTATTANNAAANNCACACCTTGGTTNGATTACATGTNGAAATCTANNNGANNN
NGNNNNNNNCCNNTATGAATTCCTTNGGAAGNCNNNNNNNANNATCTTTCAAGGACNNN
AANTGANTNNGCACATGTGAACCNGGNNNNGNNGNAAGTTANNNNTCNNNTGNTCNNNC
NAANCAANNNNNNNNNNCCNNNNNNNNNANGGGGGGGGGNAANCCGGNCCCCNNGNNCC
ANNNNCAAANNNNNNNNNNNNNNCTCCAGGNNNNNNNCNTNNNGGAN

>Y2H-158

NNNNNNNNNNNNNNCNTNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***G*
AATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGAAACGATA
AAGGTTTCATTCGTCTTGTCAGACCTAGTCAGTGTTGAAGGAGTGCCGGTAAATTATTT
GAAAATAACGACACGAACCATCATCTCCTAGCAGGTTTTAGCACTAATTGGCAACTTGC
TCTCCTCTGCGTTCCACGCCATTGCTCTAGTACAGCGGTTCTCAACCGGTGGGTGGTGA
CCGCCGTGGTGTCCACAGAATCTTTCACAGGAGTCCCCTAAGACCATCCTGCATATAGG
CGGTTTACATTATGATTCATAATGCAATTCATAATTCTCTTTTTTTTCTGTAGCTAATG
CAATCTTCGTGTAACCATTACAAAATTACAGTTATGAAGTAGCAACGAAAATAATGTTA
AGGTTGGGGGTCACCACAACGTGAGGAACTGTATTAAAGGGTGGCAGCATTAGGAAGGT
TGAGAACCACTGCTCTAGCCCTTGGCTAATGCGGCCCTCAAAAGTGAATGTATCAGAGA
AATTATTGGAAAATCAGAATTGCCAGTGCAGCTTTTAAAAATACAGTTCTGTCCTATTA
CTTTTACATAGATGGGTATCTTCTCTGAGGTAGAAGTAGATACACAGTCATTTGAAATC
TCCCTGGCCTTATGCAGTCAGAGGGCAAGGAGGAGGTGAACGAAGGCCAGCTGGGTAAT
TTGTTGTTTGCTATTTTACAGAAGTGACATATTCTCGAAGCACTGGGGAAAACGAACAA
GTTAGTCGCTGCCAGCAACGAGATCTCCGTCGAGTAAGAGACCTGTGTCGGATTCTGAA
GCTGCCGCTAACGTTTGAGGATACAGCAATCTCCTACTACCAGAAGGCATATCAGCTGT
*CTGGGATCCGAGCTGCCAGGCTGCAAAAAA*NGGAAGTGTTAGTCGGATGCTGTNTTTTA
ATCACCTGTCGGCNNCNAACTGGCCCNNCNNATGGNACCATCTNNNANTGCTNNNGCCG
ATCTGGACTTGTTTNNGGCCNTANTNGNNNNGGNAANNCNCNTTGGNNTGGANNNNCCT
CCNTGNCNTGGNANNNTNGNNNNNTNNACTGNAACAGGNNNGGNNNANNNGAANCNNGG
NNNNNNTGNNNCCNGGNNTTTTNCNNNNNNNTNTNNAN

>Y2H-160

NNNNNNNNNNNNNNNNNNACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT**
GAATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGAACAGCTA
CTAGTTTTCAACAATAGAGTGGAAGAAATGAGCAAAGGTAGAAATGTCAAGCAGGTCAC
AAGTCAGGGTGATTGGGGGAATCCTGTGCCAACAGCCTCACTTTGTAATTCCATCTGTC
ACATTTCAAAAGAACAGCAGCATAAGACAGGGATAAAAACCCACGTACCCTCCAAGGCT
TGAGTAAAAGTCCACACTCAGCACTTCAAAGACTAACGTCGTTGACTGCCCAAGGCTGC
CCTCTTAATACACCGCCTATGCATGTGCTGTGGAAGGCAACTCTGTGCATGTGCTGTGG
AGGAGATGGGCCTCATGGCTGTGCCTGGCTGCCCAGGAATCAGTATAGCTGTGGAAGGA
GACAGTATCCATAGACTCTGCTTTTCTGCAAGGAAAGCCCTTTTCCTTATACATGATTG
CCTATAATTCAGACAAATTTAAAATTGCTGCCTGCCTGAGCCCTCCACCTTTACTTTTG
CATTCCTCCAGTCATATTCTTTTGAGGCTAAGTGCCCTATCCGAGGAGATGGTTTCAAA
GGCTAACTAACTAATCTGCAGCTTTCCCAAGTGCCCAGAGGTATTTCTCAAAGTTGGTA
TGCTTAATAAGTGATGTAAATATTCCAGTTCTCTTAGGCAGCCTTACTCCTGTTGTCCC
*TGTAGTT*NCGAAAAAAAAAAAAAAAAGGNNCCNNNNNNNNTNNNNNNAANNGGNGGGG
NNNNNNNNNNNGNNNNNNNNTCNNNNNNNNNNNNNTNNNAANANAAANNNNNNNANANN
NNNNNAAAAACCCCNNNNNANGTTNNNNTNTNCANNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNN

>Y2H-161

NNNNNNNGNNNCNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAAT*
TCCACCCAANCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGATGGGAGGATGT
CACAGCAAGAAAGTTGTAACCCCAGATATTGAGACCTCAGCGAGATGCAGGTCGATGCT
TGGAAGTTATCAGTCATATATTCAGTCACAAAACAGAACATCGGGAAACTCACCAGACA
GTGGACAGACAGTATTTTCTCTGGAACACTTGGAGATTTGCCTGAAGGAAGCAGAGGAA
AAGGCTAGAGCTCTGTTGGAACAGCTTACAGCTTCTGAAGCAACAAAATCACAACTCCT
TGAGAAGGTCTCCATGCTGGAGGGAAGGCTAGAGGATGTGAACCGAAAGAACGTTGGTG
GAGAGTTGTATGAAAACATGGTTCTTGAGAAAGACAAGTGCATTGAGAAGTTGCAAGCT
GAAGTCAAAGCCTCCCAGGAGAAACTTAACATCCATTTGTCTCTTTCATTATTGAAGTG
GAGACGAGCTGCTGAGCCAAGGCTTTGACCATAGAAACGCCAGTAACAAGATGTGGTGC
TATGATTTAATAGCTGCCTACACATTTCAAATACAGTTTTAGACAGAGGTCCTGCGGAC
ATCCAACCAACATAATGAAGCAAAATGTGATTATTTACATGGCTTAGCATTCGGAGCCC
CACAACCCACATGAAGCTGATGTAGAACATGCATTTGTAATGTGGGCACTCTTATGGCA
AGATGGGAGACAGAGCCAAGAGAGTCCCTTCTGATCTCTTGGGACAACTAGCATAAGTA
GTGGCATCCGCAGAGGCCCTGTCCCAAGGTACAGGGAAATAACTGGCATGAGAGGTTGT
CCCCTAATCCTGACCTTCAGCTTGCCACATACGCATACACATCACACACATATTGAAAA
*TAATAGTAAATAAAATTTATAAGGGATACAAGCAAAAAAAAAAA*NGGNCCGCNNNNCTC
TNNNNGGGGNCGNNNNATTCGGGNACCTTNNNNNNNNCNTGGAANTAATCGNAANTCTG
AAAACCCNNNGNTCCTCATGGGGGNNNGGGGNNCNNNANAAANNNNNNNNAAANNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
N

>Y2H-163

NNNNNNNNNNNNCNNNCGACGTACCAGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGATATTGTATGT
GTCTATAACATTATTTATTACGTAAACTTTTTTTAAAAAAATTGAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGGGG
*GGGAAAAAAAAGGGGGGGGCCCCCCCCCCGAAAAAAAAAAAAAAAA*NNAAAAAAAAAA
AACCCAATTTTTTNANANCCCCCNNNNNTNNNNNNCCCAAAAAAAAAANNNNNNNNAAA
AAAAAAAANNNNNAAANAANNTNNNNNNNNNNNANANANNNNNANACNNNNNTCNNNNN
ANAANCNACNNNNNTNCNNNNGGCCCNNNNNNCTNNGANNNTGNNATGANNAANANNNG
CTGANCCNTACNNNNNNNNCTGNNNNNNTGNNNNNNNNNCNNTNNGNNNNNNGNNNNNN
NCTGCNNNNNNNNNN

>Y2H-178

NNNNNNNNNNNNCNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GA*
ATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGATGGCAACTC
TCTCTGTGTCCTTTTTTGTTCCTAAAGCTCTCAGTGTACAAGTGGACATTTGGATACAA
GACCTTATGTAACAAACAGAGGTGATATCTGAAGCATGTAAATTGGATATAAAATTCTA
CTCTTAAAGAGTTGATCTAGAGTATGGCTAAACATCTATATATGCAATCTATTAAAAGA
ACTTAATTCGGCAATTATGTCTTGATCTGATTGCAATTTATTCCTAATTAATCCTTACC
CCCATTAACCTGTTTTTATCTTGTACCAGATAGAGTGCAAAATGTATTTATAAGATTAA
TACCTGCCCATTCCCCTCTCCTCATTTCTCCTCCCCTGTCTTGATTATGGTATCTTATC
CTAACTCTGAAGACTGAAGGCACATGGTTCCCTCTAAAAACTACTATTGATACCACTGC
AGAAATAAGCCAGCAACAACAGGGCGGCGGGGTGGGCTTCTCCCCTCTCTTCCTAAATG
CATGTTCAAAGATGAGACCTTATTGATCTTAAACATCTGTCAGATGAGTCATACATTGG
ATTATTTTTTATATACATGTATACACAATATTTCAAATTGAAAGCAACATCTCAGTGGA
TTCAAAACTACTACATACTGTTGTCTAAAACAAAATTTATAATCGTACAAATGAAATGC
ACATATTGATATTTAAAATGCATAATTAAGAAAACCCAAAAAAAAAAAAAAAAANGGGCC
CCCCCGCCCTCAAAGGGGGGGNNNTCGAAACGGGACTCCTCGGCTCCAACCNNNAAANG
AATCGTAAATCCGAAANCCCCCNAAGTTCCTTAATNNGGNNNGGGGNCGNAAAAAAAAA
ANCNGCCNGNCG

>Y2H-181

NNNNNNNNNNNNCNTNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCANNCAGTGGTATNAACGCAGAGTGGCCATTACGGCCGGGGGTTATAAAGGA
GGTGAATGTGAGCCCATGTCCCACCGATCCCTGTCAGCTGCACAAAGGCCAGTCCTACA
GTGTCAACATCACCTTTACCAGCGGCACTCAGTCCCAGAACAGCACGGCCTTGGTCCAC
GGCATCCTGGAAGGGATCCGGGTCCCCTTCCCTATTCCTGAGCCTGACGGTTGTAAGAG
TGGAATCAACTGCCCCATCCAGAAAGACAAGGTCTACAGCTACCTGAATAAGCTTCCGG
TGAAGAATGAATACCCCTCTATAAAACTGGTGGTGGAATGGAAACTTGAAGATGGCAAA
AAGAATAATCTCTTCTGCTGGGAGATCCCAGTTCAGATCACAAGCTAGGCTCCTTGGCG
CCTGTGTCTGTGTGGGCTGAGAGGCCATGGACGGAGTGGGGGGGAAGAAACAGAAATCA
GACCCGAAATGGAATCAGTGCCATATGATGAACAGAATTTCAAGAATGCTGTTTTATGC
CTTTTAACCTCCAAAGCAGTAACTGCAAGCCTACTACTCTTGAGAGCGCGCTCAGAGCC
ATTGTCCCCCGAGATAGCCTCTGGGGAGGCTTCGGGAGGGAAAGGGGAAACTGGAAAGA
TTAGATTGGTGTCCATGGCTGTTTGCTGTTGGATTACAGGCAGGTTCCACCAGACAGGA
TGGGACGGGGTTGCTTAGGGATGTCAGATAACTTGACCCAGGGCTATGGATCCACTGTG
AAGGATGGCTTCCCAGAGTCTTCTGGCTGCCTGGGGGTGTTACCTCCCCTGTTTCTAAG
TGCCTCCTGAGTCCCCAGCCCCTGGCTTATCAGTCAGATGAGTCTCCTTGGTAGCCTCT
GCCCCATCGCTTCAGCAGTAGCGACTAGCTCTCCTCGGTATCCAGACTGGCTGANGGNC
*AGTCTGCC*NCAAAANTTAGGNNNNNNNNGNCTNNNNCTTNGGGGNTNNNCCNNNGTNNT
TGGTAGTTTTCATTAANGCNNNNNNTTGGATNNNNNNNNCTNNNNTTTNGNTTTNAACC
NNNNTNGNCCTGAACCGNCTNNNGGNAGGANNNNNNNACCNCCCNNNNNNNNNNNNNNN
NAAAAAANNANNNNNNNNNNNNNCCNTNNGNNNGGNNNNNNAANN

>Y2H-185

NNNNNNNNNNNNNCNTACGACGTACCNNATTANNCNCATATGAAC**ATGGAGGCCAGT***NA*
NTTCCACCNNNNCNNNGGTATNAACGCAGAGTGGCCATTACGGCCGGGGNGTTCATTGA
GCTATAGTACACACTGTTTTCCTCTCTGTATAATGGTGACAAATATAGCAAGTGAACAT
GTCGTAAATATAAAGGTTCAAGTGATGTATTGAAAATAATTTTCTAACTGCTTTGTATG
TATCACATACTCTAAATTGCACAGTAGGCGTGTTTATTAAACAGATTGGCTGGGAAAGT
TTTGAAATAGTACTGAATTTACAGTCTCAGTGCTATTCCTGTTGCTGTGTATGTGGTAG
GTTATACCGAAGTAACTAATCTTCTAATGAATGCTGATATTTTATTAGAATGAAAGCAC
AGATTAGCATTAATATTTTTTATCCTGGAAATCCTTTGGCAAGTACTGGGCAGCCCAAA
TTTAGAAAACTATGAGATTTCAAACCACGTAAACATGGTTCTGTTGGACATTTTATTTT
TTAACTGTTAACTTAGGTGTTCGCAGTTCTGTTGAATGGGACTTCTGCAGGACTGTATT
TAGCATGCTGTAAGTACTTTGGGGTGAAATAGGCTCCGAGTCTAAATTCTCAGAAATGT
CTGTGATGTATTGACCATGAATAGCGTGTACCACATGACAGAAATGAAATTACACTCTT
AACCCCTTCTCAACGTGGACTGGTCTTTACAGGGGCACCAATAAATGACAGATCTGGAG
AGTCTCCTGGTGCTTATTTTAGAAGTCAAATCAACCCAATTAGTATCTTGTTAGTAGAT
ATAATGAGCCTATTTCTTTCTAAAAAGGCTTGTGGTCTGGACGTGCTTTTACAAGAAAT
*AGTGACCTTGGGGAATATGTAAACAAGAGATCTTTTTCTAAGAGT*NNGGGGGNAANNNG
GGGGANCTGTATANTNACAANAAANTCGGGTTNCACNGTNNNNGNNNNNNANNNANATT
NAANGAATNNGNNTNNNNNGTTTTNAAGCAANGCCNNNNGAGAGCTTCCAGGANGNNAG
NCNNNAANTTTNNGNNNNNTTTTNNNNNCNTNNCNNNNNNCTTTNANTNNNNNCGGNCN
NGNNTTTNNNNNCNTCTNNACTANNNNNNAANNNGGGNTTCCAANNNNNCNNCCCCCNG
GNAAANNNNNNNNCGGNNNACCCNN

>Y2H-186

NNNNNNNNNNNNCNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGGACCTGCAGC
CGCTGCAGGGATGAAGGTCTGTCAGTTTGTCGTCTCTGTCAATGGCCTCAACGTGCTGA
ATGTAGACTACCGCACCGTGAGCAATCTGATTCTGACAGGCCCAAGGACAATTGTCATG
GAAGTCATGGAGGAGTTGGACTGCTGAGCACAGAGACATCCCTGCTTGTCCCCTATCCC
AGAACAACACAACAGTGGCGAGGCAAGGCTGCCAGGCACATGGATAATTGGGACATTGC
AGATCTTTTCTCTGATCATATGCACCTTGTCTTTCATACGCTGTTTGGATGTGGAAGAA
CCAGGTTACATATCTTTAAAAACCAAAACACAGCAGGGGCTTGGTGGCTCAAGTCTATA
ATCCCAGCATTCAGGAGGGCAAGGCAGAAGTGCTGCAAATTCAAGGCAAGCGTCTACTA
ACAGTGTGAGTGCCTGTCTCAAAAACCACACTCGGGATCTTTCCACACAAGTGCACATA
CCACACACATGCACACACGCACACATGCACACCACACACACACGCACCAAAGCCCAGTT
TAGAGTTCTCAATCATTGTTCTTCATGTTTTCGGGAAGTTGTTGTGAAACCAAATTAGC
AACATGGAAAAATGCCATTTTTAGAAGGTCATAGAATAGTTACTACATATTTTACGTCA
TGCCTACTGTGCTTGTGTGGTAAGAATAAGGAACCAGCACCTGCAATCTGGGCTACTTA
GGATGCTGAGGCAGGAGGATGATCACTTGGGTTCTGGTAGTAATGTGAGCCTGGGCAAC
*ATAGTGAAACCCCACATTAATTTCAAAAA*NAAAAAAAAAAANGGCCCCNNGCCTCTTAA
GGGGGGGNCTCCAACCGGNACCCCCNGNCTCCGCTCGAAAAAANTCGNNNNCCGNNNCC
NNNNNGTCCNNNNNNNNNNNNGNNNNNNNTNTNTAAAAANNNNNNNNNNNNNNN

>Y2H-187

NNNNNNNNNNNCNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAAT*
TCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGTGGAGGAGGCT
ACACAATCCGGAATGTTGCCCGATGTTGGACATACGAGACTGCAGTTGCCCTTGATTGT
GAAATTCCCAATGAGTTGCCATATAATGATTACTTTGAGTATTTTGGACCAGACTTCAA
ACTGCATATTAGTCCTTCAAACATGACAAACCAGAACACTCCAGAATATATGGAAAAGA
TAAAACAGCGTTTATTTGAAAATCTACGTATGTTACCACATGCACCTGGTGTTCAAATG
CAAGCTATTCCAGAGGATGCTGTTCATGAAGACAGTGGAGATGAGGATGGAGAAGACCC
GGACAAAAGAATTTCCATTCGAGCATCAGACAAACGGATAGCTTGCGATGAAGAGTTTT
CAGATTCTGAGGATGAAGGTGAAGGAGGTCGTAGGAATGTTGCTGATCATAAGAAAGGA
GCAAAGAAGGCTAGGATTGAAGAAGACAAGAAGGAGACAGAGGACAAGAAGACAGATGC
TAAGGAAGAAGACAAATCCAAGGACAATAGTGGTGAGAAAACAGACCCCAAAGGAGCCA
AGTCAGAACAACTCAGCAACCCTTGAATTTGACTCTCTAACTTTAGGAACCTCGAAAAG
TGAGACGATTCTGGGATAAGAAACCTTCCCTGTTTGAGGACATTGGCTTCATTTTATAC
TGTTTTGGCATGGACTGTATTTATTTTCAAAATGGCTTGTTTTTGTTTTTCTTGGCAAG
TTTTATTGTGAGTTTTTCTAATTATGAAGCAAATTTTTTTTTCCACCATGCTTTATGTG
ATTGTATTTAAATTGATGTGTTATTATGTCAAAAGCCGGATCTATTAAAGAAACAATTG
*GCCTTTCTGAGCTGATTTTTCCATCTTT*NGGAATTATCTTTANNAAAAANTTGTACTTG
CAAANNNAAAAANGGCGCCNNNNNNNTTAAGNGGGGGNNTNGGAAGGGNTCCNTNNNNN
NCNNNTNNAAAANNNTNNAGAACNNAAANCCNNCCNCCTCGCTNNNNNNNNNCCNNNNN
NAAAAANNNNNAAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNN

>Y2H-188

NNNNNNNNNGNNCNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGTTGAAGGCCT
AGAACCAAGGACACTATACAAATTTCGACTGAAGGTCACCAGCCCCTCTGGAGAATATG
AGTACAGCCCAGTCGTCTCGGTGGCTACGACCCGGGAGCCCATAAGTAGCGAACACTTT
CATCGTGCTGTCAGTGTGAACGATGAAGATCTCCTGCTTCGAATTTTGGAAGGAGGCCA
TGTCATGATTGACGTGCCCAACAAATTTGGCTTTACCGCTCTCATGGTTGCTGCTCAGA
AAGGTTACACCAGGCTTGTGAAGATCCTAGTTTCCAATGGCACAGATGTGAATCTGAAG
AACGGAAGTGGCAAGGACAGGCTTGCTGATTCAGCTGGGATGCCTGGTTGGGCTGCGAC
TTTCAGGGAGCCTTTTGACTCTACCTCCACCACTAGTGATGAATAGGTACCTGCAGTCA
CACATGGCTTTCTACATGCCTGTCAGAGATCTGAACTCAGGCTCTTCAGCTTGCACAGA
AAACACGCTCCTCACCAAGCCATCTCCATAACTCAAAATGTGGCTTTCTCCAGCCAGAA
AGCCTCTCTGCTCCATTCCAGTGGAAGATAACATTTAGACAACAATTTTAAATTATTTT
CTGAATGAGTCTTAAACCGATTTGGGACTGAGTGATCAACTGGGCCAAGATCTTGGGAA
GAATAATGCCAGTTTATAAAAGAAAATATAAGGAAAGTAGAAAACAGGCAAGAATCAAA
AGAGACCTGGGAACTGATGAGTGAGGATATTTTCATGTTGCAAAAATATAAGTCATCTC
*TTAAGAAAAATCTTAATGATAAACCAATGTCCTGTACATCAAAGAAAAGAGAA*NTAAAA
TTGGATATTGGGNANGGGAAANNANAAAAAAAAANGGCCCNNNNNCCNTTAAGGGGGGN
NNTNGAANGGNATCCTCGGCTCCAGTTCAAAAANTNTAAANNNGAAANCCNGAGTCTCG
TCCNNNNNNNCCCNNNCTNNTATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

>Y2H-189

NNNNNNNNGNNNNNCNTNCGACGTACCNNATTACGCTCATATGAAC**ATGGAGGCCAGT***G*
AATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGAAGGCTCT
CTTGTTCTCGACGTTGTGTTGATGGGAACCCGAGTAAGAATCTCGGCCATCCGTCTGCC
TTAGAACTTGGTCGCGTTGATCTCGTATCTCACACTTGTTTCTGAAGCATCATGAAAGG
CATGCTACTCCTTACATTCCAGCGATATGTAGTGTCTCAGAGCTTGCCCCGGCCCCACC
TTGTAGCCTCCCTGGTCTTTTGCCACCACTTAGATTCTGAGCTATCTTTAACTGGCAGC
TTGAGAATTTGTGTGATCAGGACATACCCAAGCACTTTGGGGTACAGTAGTATTGTGAA
GCAACGGTTGTGATGAATCACAGAGAAGAGCCCCTTAACTTGTTTACACAGAATCAACG
CACATAGTTTAGTGTCTAGTGTTTATATTGGCTAATGTTTTAAATGGAAAGGGCCAAGG
CACAAAGGGCCGGAAACCATGAGGAAGAATTAGGTAGTTGTTGTAAACTGGAGTATGGC
CACATTGGATGGCTTTCTGGAGGAATACATCTCTGGGGTGCTGTGCTCAGTGTACGTCT
TGTCACAGCTGACACTGTTGCTGCTGTCATCTGGCCTACCCAGTCACCCAAAGGTTTGT
GGGAACTTGGTACCATAATTACTAAAAATTGCTGTAGAGAAGGTGTTCTGGGAAATTAA
AGATGATAGCTCAGCAGCGGGTTGATCTGAAACCTTGGTGTTCTGCTTCAGGAACTTAA
CCTTAATAATGTCACTTTAGATCCTGGCAGTTTCCTAAAGAGGTAACCCTGGGGGTTGT
GTCCAGGGTTTCTATATATTTAGGTTTTAAACACTTTTGGAATGATGGGAAGTGGCCTT
*TTAGAGATGGGTAGCTTTGAACCTCTCTGC*NATGTAGTTNNNGGNTGGNTNCCTGANNN
ACCNTCCNTCCNTTTNCCTACGGGGGNTCATTNNNACTGGNNACAANTGANNNNCCNNN
ANNGGNTAGCNNNNNNNGGNNCCAANNN

>Y2H-192

NNNNNNNNNNNNCNTNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGAGAGTTGTCA
AAGCCACTGGGCTGCAGGTCTCAGGACTGACCAGTGTTGCTGCACTTGTCCATCCTCAG
CCTGTCCATCTGTCTCAGGCCATGGCGAAGCTGCTAGCACTCACCCTCGTGGGACTGGT
GTTGGCACTTTACAAGAACCATCGGTCTTCCTATCAATAACACTGTGTACCTACTGGTG
GTAAACCACCCAGACTCCTCGTCCACTGTGGAGGTATTTAAATTTCAAGAAGAGGAAAG
ATCACTCTTGCATCTGAAAACCATCACACATGAACTTCTGCCTAGGTACTGAGCATCAA
CGATATAGCTGCTATTGGACCTGAGAGCTTTTATGCCACAAATGATCACTATTTTGCTG
ACCCATACTTACGGTCCTGGGAAATGTACTTGGGTCTGTCGTGGTCCAATGTTGTTTAC
TGCAGTCCAGATAAAGTCCAGGTGGTAGCAGAAGGGTTTGATTTCGCGAATGGCATTGG
CATTTCCCTTGATGGTATGTCTATATAGCTGAATTGCTGGCTCACAAGATTCATGTGTA
TGAAAAGCATGCTAATTGGACTTTAACACCATTGAAGGTCCTCAACTTTGACACCCTTG
TGGATAACATCTCTGTGGATCCTGTGACAGGGGACCTCTGGGTGGGATGCCATCCCAAC
GGAATGAGGATCTTTTTCTATGACGCAGAGAATCCTCCCGGCTCAGAGGTGCTTCGAAT
CCAGAACATTTTATCCGAAGACCCCAAAATAACTGTAGTTTATGCAGAGAATGGTACCG
TCCTGCAAGGCACCACGGTCGCCTCTGTGTACAAAGGGAAACTGCTGATTGGCACTGTG
*TTCCACAAAGCTCTTTACTGTGATCTGTGACAAG*NAACTGCNCNACGTCATCGAAAGTG
AGGCNNTNATNNCATCCTTTGCCATACTGTAAGNNCNTANNNNAGTGANNTNTGCTANT
AATGANNGNNNNNNNNNNNNNNGNANNNNNNAANGNNTNGAGNNTTNCCTGGCNNNNNN
TNNNAACNNNNNNGGNTNNNNTNNNNTTNNCTNNNNNNNNNNTNAAGNNCCCCNTNNNN
NTNAAAANTTTTTTNCCCCNNAAAAAAAAANNN

>Y2H-193

NNNNNNNNNNNNNNNNTNCGACGTACCAGATTACGCTCATATGAAC**ATGGAGGCCAGT***G*
AATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGACATACCAG
TTAAAAAATTCAGTATATGGGATGAATCTCGAATTAACCTTGCCTTCACAGTCGTGTGG
ATGTAACATTTCAGGCCTTTTCCTGGACTTCCCAGAAGAGGAGGGAATATTTCTTGTCT
TAACTTTTACTGTCTAGTTGTCTATGAGGATCTTATAATTCATTTTGTTATCCCCAATG
TCCCATGTCCAGCAAAATTTCATTTGAGGGAACTATACTTGATACAAAACTATTAACAT
TGTTGAGGAAGTAATTTAACTTCCTTTGAAAGTCATGTATTGTGACTGTTGTCTTAATG
AAGGTTTTAATTTCAGTGTTGCTTTTGAGCTGCAGGGTGACGGGGCAGTAGAAACTTGA
AACATCTGCCATCATCCATGGCAGGAAACATAGCAAGATGTTCTCTTTTTACTTTTGCA
CTGTTGAGTGGGTCTAACAAACTGATTTCCCAGCTAGTTTTTAAAATGAAGAATCTAAA
TAATGTGGCAGCTTAAAGTATAGATGTATGTGCTAACTTGAATAAATGCTCATGTCTGT
TAGCCATACAGTCTGTTATCTTACATCCTCATTCATACATGGATTATGTTACTTAAGAT
ATGTTCCTTCTTACAGAAAATATTTTCAAAGGGAACTATTCACTGTGGGTTCATTGTCA
CTGATTTCTGCCTGCCACAATCTTCAGAAATTCTTCTCTCATGTGCCACCCATTGACAG
TAGGACAAGGGTTGGAAATACATTTTCTTCTGGTGGTAAGGGTGGGTGACAAGACGTTT
TCAAATCCTTCTGGCCTCCTTTGCGATCGATGGCATGTGTTTAATAAGATGGCTAGATG
*GGACTGGCAGATTGTTC*NNNANAATTACTTTTGGCTNAATGTACATGGTNNGGNAGGGA
TGGNANCNNNTTCNNNNNCNNGAGNGTAGATCAATAANNNNAAGCAGGTAAATCCNNNN
NGGNCNNTCTTTCATAAGGTAGCCNNNNANNNGGAAAAAAAAAANGGNCCCNNNNNNNN
NNNNGGGGGNNNNNNNNCNGGNTNNNNGNNNNNNNNNNAANNNNGGG

>Y2H-194

NNNNNNNNNGTNCNTNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGAGAAAAGCCAA
CAAAGTGGGCATCTTCATCAAAGTCACCCCACAGCGTGAGGAAGGGGAAGTGACCGTGT
GCTTCAAGATGAAACACGATTTTAAAAACCTGGCTGCCCCCATCCGCCCAGTGGAAGAA
GGTGACCAAGGCACAGAGGTCATCTGGCTCACGCAGCACGTGGAGCTGAGCTTTGGCCC
TCTTCTTCCTTAAAACACAGTGACCCCACCCACAGCTGTGTCAACACAACTCTAACAGT
GGAAGCCGCTGCTTCACCAGTCGTGTTTATAAAGATGTACCCATCACTACTGAATCCTG
TGGAGGGGGGAACAGGCCAGGGTTGGGAGCACAGAGTAACTGCTGTTGACACCAGGAAG
CCTGTGTCTCTGACACTTGCACATTGAGTAACAGCATGGCTCCATCCCAGTGAAGGAGT
GGCTGTTCCTCAGAGGGGCGTTAGATTTACTCCCTAAGACATGTGCTGTGTGCCTGTCA
CAGTCTGACAGAGAACCTTAGGACGGTAATCCAGAGCACACCTGGTCTTCATGCCGCAG
CATTGGCCTCGGTGTTGAAAGGCTTCCTGCATTCCACTAAGGATAACCCTGTAATGATA
ACCCTTCATAGCCTTGGCTCTCACTTTGCCACTGCTCTCCATTTGACAGAGCAGTAGGA
CTGTCAGTGTGTCTTTAACTTTGTGATGACAGGGCAACCTGTCACTTTCCTAAGGCACT
TTGTTTCTTCCAAAAGAACACAGTATACCTGTCACTTGCTTGTCATTTTCCTCACTAGT
ATATGGGACTTTTAAGAATTTGAGTGGCAAGCAATTACATGTCATGAAATTCTATTTGT
TTCAATATTACAATACCCAGGGGCTAACGCTGCCAAGGGGAGACAGTTATGTATCTCAC
*TATCTGCC*NTCCNNNTTCTGANNNCNGGTGGNTTTGTTTTCAGATAGGNTCGCTCCATG
TAACTCTTNCNGGNCTGGANNNNGGCCNNNNNGGATTAANGAACTTGCCCCNTNAAAAA
AAAAAANNNCNNNNNNNNNNNTNNNGGNTGNNNNNNNNNCCGGNCTCANNNNCCTTNNN
NNAANNNNNNTNNNAGAACNNNNCNNNNNNCCCCNTCCTNNNNNNNNNCNNNNNNNAAN
NCAAAAAAAAAANNANNNN

>Y2H-198

NNNNNNNNGNGTNCNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GA*
ATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGATGTCTTTA
TTCTTCAGATTGGCAGAAAAAATGGAAGTCAGCAAATCTCAGACTCGAGGGTCATAGAC
AATCCCATTGAAGATGGAGGTGTCTTTCCTCCTGTACATCTAGTCTGTGCCTGGCTCCA
TTGTTTGTTCCAATCCATGGGACTTCATGCCACGTAACAACTTAGAAAGGAAGTTATGA
CCAAGCAAAAACTCTGAATTATGCTATGAGCAGACTTATGGGTTACTTCATACATTGTA
TGAATAAGGTTTATAAATCAGTGGTAACACGTCATTCTTAGAAACTTTAATTATGAGAA
ATTTTGCTAATCATGATTTGTTATTTGTATTTTTCTGCTTACTAGACTAATATATGCTC
ATTAAAAATGTGTATAAGAATGTTTGTGTACACTTGAGGCACCATAGCTACCAGCCTAC
TCGGCATGTCCCTCTGTCTTCCACAAGTGTGAACACTTCACACTAGTTTGTAAGCTGAT
GGTTCACTTGTATGCCCATCTCCCTCTGGAAAGTAGCTGTCACCCTCATGTGTAGCCAC
CCTGGCCCAGGTTTCCTCAGTGATAGGTGTAGAGGGTTTTTAGCATGTTGTTACCTACA
GCGACAGTGACTGGCTCAGTCCCTGTCTTGGACCACAGGGGGTAGAATTCTTGGGTCAG
AGTTGGAACATTTGAAGTCATAATTCACCAACTCTGTAGTATTCTCCTACTTTTTGTTG
TTTGAGATAAGTTCTCACTGTGTAGCCCTAATTGTCTTTGAACTCGGAAATCTGCCTGC
*CTCTGCCTCCTAAATGCTTGGATTAAATGGGTTTATATTATCAAATTTAC*NNNAAAAAA
ANNNAAAANGGCGNNNNNNCCCNTNGGGGGGGGNATCAANCGGNACCNNNNNNCNCAGC
TCAGAAGAACCTAAANCGGAAANCCCNNAGTCTNNGNNNNGGGGGGGGNNNNAAAANNC
ATAAANNNNNNTNNNNNNNN

>Y2H-199

NNNNNNNNNGTNCNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGAAGTCCAGGAG
GACAGACAGGACTCTGTTACACAAAGCAACCCTGTCTCGAAAAACAAAACAAAACAAAA
CAAAAAACCCGGGTTCTAGGATTCAAATTCAGATCCTCATGCTTGAGGGAAAGCACCTC
GCCGCTCCGCTGAGCTGTGTCCTCAGCCTCCCTCCCTCTCCTGCTGGCTCTCCCTATTC
TCCATAAGCTCACATGAACCTTCTGGGCCAGCAGGGTGTGGGCAACAGCTGGTGAAAGA
AAAGGCAAGAAAGAAAGGCCAAGATCATAAAATTTGACAATGGAAACACAGTAATAAAT
*TTTCATATTCCGAAAAAAAAAAAAAAAAAAGGGCCCCCCCGCCCTCAAAGGGGGGG*NCT
TCANNCGGGGACCCTCCGACCTCAACCTGAAAAAAANCCGAAAATCCGAAAAACCCCCC
AAGTTCCCTTTAACTGGNNNGGNAACCCCCNTTAAAAANNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNN

>Y2H-202

NNNNNNNNNNNNCNTACGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGAGAGACACATC
AGAGATTCACAAAATGCTAATAGTTTTTTATTTTCAAAATTGGGAGTGTTCCCTGTCCA
TTATGTGGAGGTCCCCAGAGGTGGTATTATTAAATGGAAAAGAATAGTTATAACTCACA
TTATGTTTTTGCGTATGAACAAGGAAAATGCTGTTTTCAATCCAATTGTTTGGATTCAT
ACACGTTTGAAGGAGATTATGTTGTTTGAAAACACAAAGGAGCTTTACAGCCATGGTGG
TGTCTGTGGCTGGGTCAGGTGGGAGCGGCTCGAAGCCCTTACCTGCTGTCCTCTCAGCA
GCAGTCTGTCAGCAGTGCACTGTGATTTGGTGAAAGCCTGTTGACGTCTTCTATTCAAT
CCCAATGTTACTGCAGCCATGAGACGAGCTGCACGAGGACAGGCATCCATGTCACGTCT
TAGTTGGCGATGTTAAATGGTGACTGTTTTATTTAAAACTCGGGTCTGATCACTGGAAA
CGATTGCTTGGTGATTTGTGATTTGGGGCTCGGTTCATTGCTCTAGGGATGTTTGTGTA
*TTGTGTA*NGGTGCGTGTACTACTGCGCTGCCCCTGTGGTATGTCACGGTTCTGTGTAAA
CGCCCTCCATGTGGCCCTCTCATGTTGCATGTTTTCAGTGGTTTGGAAGTTTTGTATAT
TTATTGTATTAATGCAGAGTGTCATAAAATAAAATAATGTTTACTGGAAAAAAAAAAAA
NGGCCCNCCCCNNTCTAAGGGGGGGNATTCNNTCGCGGACTNNNNATCTCCGCCTCCAG
NGAACCGAAAACCNNNCNNNGCCNNNNNCCTTAGTGNTNTTNAANGTCACCTNATAATA
ACACACNNNNNACGNANNCNCANGNNNNNNNNNGNNCCNANNNNNNN

>Y2H-203

NNNNNNNNGNNNNNNNCGACGTACCNGATTACGCTCATATGAAC**ATGGAGGCCAGT***GAA*
TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGAAATTAGTGAA
TGAAGAACGAACACTAGAAGTGGAAATAGAGCCTGGGGTGCGAGATGGCATGGAGTACC
CCTTTATTGGAGAAGGTGAGCCGCATGTGGATGGGGAACCCGGAGACTTACGGTTCCGA
ATCAAAGTTGTCAAGCACCGGATATTTGAGAAGAAAGGGGAAGGGCTGCCCAACTTTGA
TAGCAATAACATCAAGGGCTCTTTGATAATCACTTTTGATGTGGACTTTCCAAAAGAAC
AGCTGACAGAGGAAGCAAAAGAAGGTATCAAGCAGCTTCTCAAACAAGGCCCCGTGCAG
AAGGTGTACAACGGGCTGCAGGGCTATTAGGAGCGAAGAGAACTGGACTTTGAAAGAGG
TGGATCAGCGATATTTATTATCTGCAGTGGGTTTTTGTGTGTGTGTTTCATTATTTTCA
ATATGCAAGTTTGGCTTAGTCCTTTTTTCCCTGATGACCGTCCTGACAAGAACTAGAGG
TTTGAGGCTCTGTCTATTGCATTCAGACAACAATGGCCAGCAGAAGGTTACCTAATACC
TCCCCCTCTGGGGATTTAACGTCTGGTGCTGCCCGAGGTTCAAGGAGTAACGTCTCAGA
GGATGCCCGAGCAAAGGAAACCAGATGAGGGTTAGAGACAATTGTTAGCTATTTCAAAA
TGCCAACTGGAGAAGTCTGTTTTTAAATACATCTTGTTGTTATTTTTTTTTTCTTGACT
CCTGTTTCTTTTTCCTGTGAGCCGATGGGAAGTCTTCCCATCAGGCAGGATGGCCACAC
AGCATGCAGAGGAACGATGAGAGCCCGGCTCCCAGGCCCACTAAACACATGGTGTCTAT
*AGCAACGAGCTCGAANAAACCGAGTAGACTAGAGACTGAGCTACTTGAATTAA*NGTTAN
NNAANGANNNCACTGTCTTTNNGAGANCTATTTNCCNGTTTNCTTNGGTANNNNTAANT
CTGATTNGGNCAAAAAAANAAAAAAANGGCCCCCNNNCCCNNAAGGGGGGGNNNAANNC
GGNACNTNNNNNNNTCTCNNNNGAANNNNNNNTTANNCNGNNNNGANNNNNNTTTTNNN
GGNNNGGGNNNGNCACAANTATAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNN

>Y2H-205

NNNNNNNNNNNNNNNNCGACGTACCAGATTACGCTCATATGAACATGGAGGCCAGTGA
ATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGGACCGTTACA
TTCGTGTCTCAGACAACAACATAAACATTTCTGGATGGGAGAATGGTGTGTGGAAAATG
GAATCCATATTCAGAAAAGTCGAGAAAGACTGGAACATGGTTTATTTGGCCCGAAAAGA
AGGATCATCTTTTGCTTATATTTCCTGGAAGTTTGAATGTGGGTCAGCTGGTCTAAAAG
TAGATACTGTTTCCATCCGCACAAGTAGCCAAAGCTTTGAGTCTGGATCAGTGAGGTGG
AAACTGCGATCTGAGACAGCTCAAGTCAACCTGCTGGGAGATAAAAATCTACGTTCCTA
TAATGATTTTTCTGGTGCCACAGAGGTTACATTAGAAGCAGAGTTAAGCAGAGGAGATG
GAGACGTTGCTTGGCAACATACCCAACTGTTTAGACAAAGCTTAAATGACAGTGGAGAA
AATGGCTTGGAAATAATTATAACGTTCAATGACCTCTGAAAGCCTGAACTTCAGGGAAA
TGCTGGCTACAATCACTGACCCGCCATGGCCTCCAGTAGTGTATTAGCTCAGTGTGTAT
CTAGTTGGCAGTTCACCATCCTGTGCAACAGGTTTCTTTTGCTGGCTGTACATCATGTA
ATCCTCATTAAAGTATATCTTCATAGTGTGGACCATAAGAAATCTTCAATTAAATGCTA
ACTTCTGCCTGTGATTAGAATTTGGCATAAAATATTTGGCATAAAGTAACTGTATGAAA
TTCTAATCTAAACACTGTTACTAAAGTTGTAAGTGAAAATGATGGTATTCACGATTAAA
TGCTGTTCATGACTGTGATACAATCTAGGGCTACTAGTTACTTAACATTTTCTGATTAG
ATTTTTCCAAACATTCTCAATTTTGAATAACTGTCAAAGTCAAGTATCTTATGTCATAT
*GATGCATGAAGGTTTNATCATTGTAATATCTGGACAGTTAAAA*NNAAANTGGANANTTT
TTNCNNNNAAAANNNNAAAAAAANNNNNCNNTCNNNNTCTTAANGGGGGGGNNNNNANN
NGGNACCNTCNNNNNCGATNNNNANNANNNNGANTNNNNAAANCGCNNNNNNNNNNNTN
CNNCCCNNNNNNNNNGCNNNNAAANTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNN

>Y2H-206

NNNNNNNNNNNNNNNNNCGACGTTCCNCATTACGCTCATATGAAC**ATGGAGGCCAGT***G*
AATTCCACCCAAGCAGTGGTATCAGCGCAGAGTGGCTTTTACGGCCGGGGGTAACCACT
GTGGTAAGGCCTTTGCACAAAGCAGTTCTCTCCGAATCCATAAGCGAACATGTACAGGA
GAGAGACAATATGAATGTAACTAGTATGGTAAAGCCTTTGCAGGAAGCAGTGGTCTCCA
ATGCCATAAAAGATCACGTACACGAAAGACACCATATGAATGAAACCAAGGTGGTAAAT
CCTTTGAAGGAAGCAGTGGTCTTGAATATCATAATCGAACACACACAGCTGAAAAACCC
AGGAATGTAACCAATGTGGTAAAGTCTTTGCAGAAAGGAGTGACCTTTAAAGAGAAAAG
ACCATACAGGAGAGAAACTCTATGTAGGTATCTGATGTTATAAAGCCTTTGCAGCAAGC
ATCGATCTCCAAGTACATAAATGAACATATACCAGAGAGGAACCTTTTAAACGTAACCA
ATGTGTTTAACCTTTTTCGAAAACCAATAGTCTTTACACACATAAAAGCACATATAGTT
GAGAGAAATGCTTTGATCTTAACCAATGTGGTTAAACTTTTTCAAAAAGCAGTCATCAA
CAAAATCATGAAAGAGCACGTATAGGAGAGAGATTTTTGAATGTTAGCAATGGTGCTTT
TATAGCTTTTTCACAAAGCAGTGGCCTCCAAATATATAAAGGAACACATACAGGAGAGA
AACCCTATGAATATTGGCAATGTTTTAAAACACTTCTACATCACAATCATCTTCAAATC
TGTAAGGAACATGAGGTTTTGAGAAATCCTCTGAATAACACCAATTAGGGAAAGCAATT
*GTACAAAAGAGTCATCTCCCAACACAAA*NGTATACACACTAGATAGAAACCCTATGNNN
ATAGGCAATGTGGTAANNNTTTNATAAGTCNNNGTAGTCTCTANGTAANTAAANGANNN
ACAGTTGAGAAANCCNTATGAAATGANNAGTGNNNGNGCCTNGNNNTTAAATCTTCAAN
NNTAANNNNNNNNNANNGNNNGAATCCNNNNNNNAAGCATGGGNATTNNNNNNTNNNNN
NNTTNCAGGACCANNGNNNNNNNNNNNAANGGNTAAGTGGTTNNNNGGGNNNTTTNNNA
NTNNNNNNNNCCNNGGCANNNNNNCCGGANNNTTNGNNNNNNNNNNNNAAANNNNN

>Y2H-209

NNNNNNNNNNNNNNNNTNCGACGTACCNNATTACGCTCATATGAAC**ATGGAGGCCAGT***G*
AATTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGGGCCTGTGAGA
GTACATATTGAGATTGGACCTGATGGCAGAGTAACTGGTGAAGCAGATGTTGAGTTCGC
AACCCATGAAGATGCTGTTGCAGCTATGTCAAAAGATAAAGCAAATATGCAACACAGAT
ATGTAGAGCTCTTCTTGAATTCTACAGCAGGAGCAAGCGGTGGTGCTTACGAACACAGA
TATGTAGAACTCTTCTTGAATTCTACAGCAGGAGCAAGCGGTGGTGCTTATGGTAGCCA
AATGATGGGAGGCATGGGCTTGTCAAACCAGTCCAGTTACGGTGGCCCAGCCAGCCAGC
AGCTGAGTGGTGGTTATGGAGGTGGTTACGGTGGCCAGAGCAGCATGAGCGGATATGAC
CAAGTTTTACAGGAAAACTCCAGTGATTTTCAATCAAACATTGCATAGGTAGCCAAGGA
GCAGTGAACAGCAGCTACTACAGTAATGGAAGCCGAGCATCTATGGGCGTGAATGGAAT
GGGAGGGATGTCTAGCATGTCTAGTATGAGTGGTGGATGGGGAATGTAATTGATTTTGA
*TCACTGACTCTTGG*NCAAGAAGAAATTTTTTCGAAAACCCCTCGGTTTTTCAGAAANNN
NTTTTTTNCCAGCCTNGTGGTTTTANNNNNGCTCTTAAGTGCCCATTNNCNNCTCTCAA
AGANGNCTCANNGNNNNCCATNNNNTNCATCTANNNNGATAGAGCACAGCNNNCTGAAC
TGNNNNAGCNAGACTTAACTGCNNCCNTGGGNNCNNAGTANCAGTGTGATGTAAAAANN
CCCGNNNNTNNNNNTANANNNNNGGNNNNAANCCNNNNNNNNGGNTTGGGGNTCNNNAT
AACNNNNNACCNNNNNNNCNNNNNNTTGGNNNNNNNGNNTTGNCCCCNAAAANCTTGGN
NGGNNNNNNNNNCNNCCNACTGNCCCCNNATNNTTTNGGGCNNNNTANNNNNCCCCNNN
TNNNCNNTCGNNAAAAAAAANTAAANNAAANNNNATCTTAAAANNNNNNNTNNNNAAAN
NCCNNNNNCCNTNNNNNNNNNNGGTTNAANNNNNNAANNNNNCCNTNNNNNNNNNNG
NNNNNNAANTNNNNNNNNNNTTTTNNNAANNNNNNNNNNNNCCCNNNNNNNNANGG
GGNNNNNNNAANCCNNNNNNNNNNNNNNNNNNNNNN

I, Abdulkareem A. AlShaheeb, hereby submit this thesis to Emporia State University as partial fulfillment of the requirements for an advanced degree. I agree that the Library of the University may make it available to use in accordance with its regulations governing materials of this type. I further agree that quoting, photocopying, digitizing or other reproduction of this document is allowed for private study, scholarship (including teaching) and research purposes of a nonprofit nature. No copying which involves potential financial gain will be allowed without written permission of the author. I also agree to permit the Graduate School at Emporia State University to digitize and place this thesis in the ESU institutional repository.

_____

Signature of Author

_____

Date

Identification of Unique Regulatory and Coding Regions in Mouse Myosin V Introns: A Comparison of Proteomics- and Genomics-Based Methods

Title of Thesis

_____

Signature of Graduate School Staff

_____

Date Received