**Functional variants within the 5' UTR support reduced exon usage as a molecular mechanism underlying the quantitative trait gene *Hnrnph1* for reduced methamphetamine sensitivity**

**Running title:** Functional variants within the 5' UTR of *Hnrnph1*

Qiu T. Ruan[1,2,3], Neema Yazdani[1,2,3], Eric R. Reed[4], Jacob A. Beierle[1,2,3], Lucy P. Peterson[2], Kimberly P. Luttik[1], Karen K. Szumlinski[5], William E. Johnson[6], Peter E. A. Ash[7], Benjamin Wolozin[7], Camron D. Bryant[1,2,3]*

1. Laboratory of Addiction Genetics, Department of Pharmacology and Experimental Therapeutics and Psychiatry, Boston University School of Medicine, 02118
2. Biomolecular Pharmacology Training Program, Department of Pharmacology and Experimental Therapeutics, Boston University School of Medicine, 02118
3. Transformative Training Program in Addiction Science, Boston University School of Medicine, 02118
4. Ph.D. Program in Bioinformatics, Boston University, 02118
5. Department of Psychological and Brain Sciences, Molecular, Cellular and Developmental Biology, Neuroscience Research Institute, University of California, Santa Barbara, 93106-9660
6. Department of Medicine, Computational Biomedicine, Boston University School of Medicine, 02118
7. Laboratory of Neurodegeneration, Department of Pharmacology and Experimental Therapeutics and Neurology, Boston University School of Medicine, 02118

***Corresponding Author**

Camron D. Bryant, Ph.D.

Department of Pharmacology and Experimental Therapeutics and Department of Psychiatry

Boston University School of Medicine

72 E. Concord St. L-616

Boston, MA 02118 USA

E: camron@bu.edu

P: (617) 358-9581

## NONSTANDARD ABBREVIATIONS

psychostimulant use disorders (PUDs)

methamphetamine (MA)

quantitative trait locus (QTL)

substance use disorders (SUDs)

DBA/2J (D2J)

C57BL/6J (B6J)

*Hnrnph1* mutants ($H1^{+/-}$)

quantitative trait variant (QTV)

real-time quantitative PCR (qPCR)

**ABSTRACT**

We previously identified a 210 kb region on chromosome 11 (50.37-50.58 Mb, mm10) containing two protein-coding genes (*Hnrnph1*, *Rufy1*) that was necessary for reduced methamphetamine-induced locomotor activity in C57BL/6J congenic mice harboring DBA/2J polymorphisms. Gene editing of a deletion in the first coding exon of each gene supported *Hnrnph1* as a quantitative trait gene. We since showed that *Hnrnph1* mutants also exhibit a reduction in methamphetamine-induced reward, reinforcement, and dopamine release. However, the quantitative trait variants (QTVs) that modulate *Hnrnph1* at the molecular level are not known. There are nine SNPs and seven indels that distinguish C57BL/6J from DBA/2J within *Hnrnph1*, including four variants within the 5' UTR. Here, we show that a 114 kb introgressed region containing *Hnrnph1* and *Rufy1* was sufficient to induce a decrease in MA-induced locomotor activity. Transcriptome analysis of 114 kb congenics versus *Hnrnph1* mutants identified a nearly perfect correlation of fold-change in expression in differentially expressed genes common to both mouse lines, indicating functionally similar effects on the transcriptome and behavior. Ten overlapping genes showed differential exon usage, including *Hnrnph1* and *Ppp3ca*. Differential 5' UTR exon usage of *Hnrnph1* and *Ppp3ca* were validated using real-time quantitative PCR. Cloning of the *Hnrnph1* 5' UTR containing all four variants together (but none of them individually) induced a functional decrease in reporter expression in both HEK293 and N2a cells, thus identifying a set of QTVs underlying molecular regulation of *Hnrnph1*.

**KEY WORDS**

RNA Binding Protein, Functional Variants, Positional Cloning, Alternative Splicing, Calcineurin

## INTRODUCTION

Psychostimulant use disorders (**PUDs**), including methamphetamine (**MA**) and cocaine dependence, are a serious public health concern in the United States. While the opioid epidemic crisis continues to garner warranted attention, there has been much less focus on the recent steep surge in PUDs, as evidenced by the steep increase in PUD-related deaths, especially MA-related deaths (1). This public health concern is particularly problematic, given that there are no FDA-approved treatments for PUDs. Both genetic and environmental factors contribute to PUDs (2), yet genome-wide association studies to date have identified very few genetic factors (3). One notable example was the identification of a genome-wide association between *FAM53B* and cocaine dependence in humans (4). This finding was particularly interesting in the context of a mouse forward genetic study of psychostimulant addiction traits that identified a *trans*-expression quantitative trait locus (**QTL**) originating from a locus containing *Cyfip2* and *Hnrnph1* that influenced *Fam53b* expression and was associated with cocaine intravenous self-administration (5). This is one of the first examples demonstrating a direct correspondence between genome-wide genetic association, functional changes in gene expression, and trait-relevant behaviors (cocaine self-administration) between rodents and humans. Thus, it highlights the power and translational relevance of systems genetics in mice in the study of substance use disorders (**SUDs**).

There are several advantages to using rodent forward genetics in studying the genetic basis of molecular and behavioral traits relevant to SUDs, including the ability to control allelic frequency, genetic background, the environment, the amount and length of drug exposure, the sample size, the collection of the appropriate tissue at the appropriate time points, and the ability to both map and validate causal functional variants *in vivo* within the same animal species. While there are clear limitations to modeling the genetic basis of psychiatric disease traits in rodents, it is important to point out studies in which there have been promising successes. In addition to the translational link of *Fam53b* with cocaine addiction traits in mice and humans (4, 5), there are multiple other recent examples of forward genetic discoveries in rodents that have yielded a high likelihood for translation to humans. A missense SNP in *Taar1* (trace amine-associated receptor 1) was mapped in mice for the aversive properties of MA self-administration, body temperature (6) and toxicity (7–9) and genetic variants in *TAAR1* in humans was associated with drug craving in methamphetamine dependence (10, 11). *CNIH3* (Cornichon Family AMPA Receptor Auxiliary Protein 3) was identified as a GWAS hit for opioid dependence in humans that was corroborated with a forward genetic study of morphine physical dependence in mice (12). Whole genome sequencing plus QTL analysis and functional validation identified *Grm2* (mGlu2 receptor) in escalation of alcohol consumption in rats (13). A mouse strain survey correlating gene expression with behavior found a link between glyoxylase 1 (*Glo1*) and anxiety-like behavior that was confirmed via viral knockdown and overexpression (14). A *cis*-eQTL and copy number variant containing *Glo1* was subsequently discovered to segregate in outbred mice and is fixed in several classical inbred strains, thus revealing a genetic and molecular mechanism linking *Glo1* expression with anxiety-like behavior (15). *Glo1* metabolizes methygloxyal which, in light of mouse preclinical studies, was found to act as an agonist for the GABA-A receptor and exert anxiolytic-like effects (16, 17).

Increasing methylglyoxal levels preclinically reduced alcohol drinking (18, 19), induced a rapid antidepressant-like response (20), and reduced seizures (21). To summarize, mouse forward genetic studies have led to the identification and/or corroboration of several promising therapeutic targets for SUDs and other psychiatric disorders.

To fine map a QTL in which the DBA/2J (**D2J**) allele was associated with reduced MA-induced locomotor activity (22), we previously employed an interval-specific congenic approach in C57BL/6J (**B6J**) mice carrying various introgressed regions on chromosome 11 from the DBA/2J strain (23) and detected a fortuitous recombination event that revealed a 210 kb region on chromosome 11 (50.37 - 50.58 Mb) that was necessary for reduced sensitivity in the locomotor stimulant response to MA (24). Specifically, replacement of this polymorphic region with the background C57BL/6J allele completely eliminated the MA-induced behavioral phenotype, thus demonstrating that this region was necessary for reduced MA-induced locomotor activity (24). The 210 kb region contains two protein-coding genes – *Hnrnph1* and *Rufy1*. Introduction of a heterozygous deletion within the first coding exon of each gene provided strong support for *Hnrnph1* (and not *Rufy1*) as a quantitative trait gene (**QTG**) underlying reduced MA sensitivity (24) and we subsequently expanded the phenotypic repertoire of the *Hnrnph1* mutants to include a reduction MA reinforcement, MA reward and MA-induced dopamine release (25).

Thus far, we have only demonstrated that inheritance of the polymorphic region containing *Hnrnph1* and *Rufy1* was *necessary* for the reduction in MA-induced behavior in response to inheritance of the D2J allele within the 210 kb region. In order to demonstrate that this region is also *sufficient*, in the present study, we backcrossed and screened for congenic mice capturing only *Hnrnph1* and *Rufy1* and we identified a founder containing a 114 kb introgressed region. Following the observation of reduced MA-induced locomotor activity in 114 kb mice, we assessed their transcriptome at both the gene- and exon-level and compared these results to the transcriptome of *Hnrnph1* mutants (**H1**$^{+/-}$) in order to provide further support for *Hnrnph1* as a QTG and to identify the quantitative trait variant(s) (**QTV**)s. After discovering differential exon usage of the 5' UTR of *Hnrnph1* in 114 kb mice as a potential molecular mechanism underlying the QTL and QTG, we validated this result via real-time quantitative PCR (**qPCR**) by targeting the exon junction and the specific exons. Finally, to identify the QTV(s), we cloned the 5' untranslated region (**5' UTR**) of *Hnrnph1* containing either the individual *Hnrnph1* 5' UTR variants, or the combined set of the four 5' UTR variants, fused to a luciferase reporter gene and tested the effect on reporter expression in two different cell lines. The results identify a set of 5' UTR variants within *Hnrnph1* that likely represent the QTVs underlying gene regulation and likely behavior.

## MATERIALS AND METHODS

### Generation of B6J.D2J 114 kb congenic mice

All procedures involving mice were conducted in accordance with the Guidelines for Ethical Conduct in the Care and Use of Animals and were approved by the Institutional Animal Care and Use Committee at Boston University (#AN-15326). The founder 114 kb congenic mouse was identified by monitoring the distal single nucleotide polymorphism (**SNP**) sites of offspring generated via backcrossing heterozygous congenic mice from Line 4a carrying one copy of a ~11 Mb introgressed interval from the D2J strain to the background B6J strain (24). We monitored for retention of the most proximal D2J SNP that was just upstream of *Hnrnph1* (rs29383600; 50,373,006 bp; mm10) and this SNP defined the proximal end of the 114 kb interval. We previously genotyped several purported SNP markers proximal to rs29833600 that were monomorphic until 49,873,463 bp (mm10), which was catalogued by Sanger as a B6J/D2J polymorphic marker and that we genotyped as homozygous for B6J (24). As we were unable to identify any polymorphisms between 49,873,463 bp (mm10) and rs29383600, we defined the proximal interval of this smaller congenic with the marker rs29383600 (50,373,006 bp). The remainder of the genome is isogenic for B6J as determined by several other markers on chromosome 11 and by a medium-density genotyping array containing 882 informative SNP markers from our previous study (24).

We also simultaneously monitored for the detection of a recombination event (i.e., the observation of a homozygous B6J genotype) at a second marker located just distal to *Rufy1* (rs254771403; 50,486,998 bp; mm10). Following detection of a recombination event at rs254771403, we then genotyped at an additional upstream proximal marker at rs29459915 (50,484,260 bp; mm10) that was also downstream of *Rufy1* and found retention of the D2J allele at this locus. Thus, the new remarkably smaller congenic interval was conservatively defined by a region spanning proximal rs29383600 (50,373,006 bp) and distal rs254771403 (50,486,498 bp), yielding a 114 kb interval. Thus, we named this new congenic **"114 kb".**

PCR primers were designed to amplify a ~200 bp amplicon that contained and flanked each SNP. The PCR products were run on a 1.5% agarose gel and visualized for band specificity. Single bands were excised according to their predicted fragment size and gel-purified (Promega Wizard SV Gel and PCR Clean-Up System, Cat A9281), and prepared for Sanger sequencing (Genewiz, Cambridge, MA, USA). Mice homozygous for the 114 kb region (referred to as 114 kb) and B6J littermates were generated via heterozygous-heterozygous 114 kb breeders. To avoid genetic drift, heterozygous 114 kb breeders were maintained by mating heterozygous male offspring with C57BL/6J females purchased from The Jackson Laboratory in Bar Harbor, ME, USA. Homozygous 114 kb mice were generated by breeding the heterozygous breeders. Mice were SNP-genotyped using genomic DNA extracted from tail biopsies and two Taqman SNP markers: rs29383600 (50.37 Mb) and rs29459915 (50.48 Mb) (ThermoFisher Scientific, Waltham, MA, USA).

### Methamphetamine (MA)-induced locomotor activity in 114 kb mice

Both female and male littermates (56-100 days old at the start of the experiment), were phenotyped for MA-induced locomotor activity. Mice were housed in same-sex groups of 2 to 4 mice per cage in standard

shoebox cages and housed within ventilated racks under standard housing conditions. Colony rooms were maintained on a 12:12 h light–dark cycle. The estimated sample size required to detect a significant effect (Cohen's d = 0.72), with 80% power ($p < 0.05$) was n = 25 per genotype based on the previously published phenotype in the larger congenics capturing a QTL for reduced MA-induced locomotor activity (24). Mice were tested for baseline locomotor activity on Days 1 and 2 over 60 min and then administered MA (2 mg/kg, i.p.) on Days 3, 4, and 5 and were video-recorded for distance traveled over 60 min using Anymaze (Stoelting Co., Wood Dale, IL, USA) as previously described (24). Data are presented in six, five-min bins or as the summed distance traveled over 60 min for each of the five days of injections.

**Transcriptome analysis**

Striatum punches were harvested bilaterally from 114 kb mice homozygous for the congenic region (n = 8) and B6J wild-type littermates (n = 8) and processed for total RNA extractions (24). RNA samples were bioanalyzed (RIN > 8) and cDNA library was prepared using the Illumina TruSeq Standard mRNA LT (100 bp paired-end reads). The 16 samples were multiplexed and sequenced over three lanes on Illumina HiSeq 2500. The data is available on the NCBI GEO database under accession number GSE76929. FastQ files were aligned to the mouse mm9 reference genome using TopHat (26). HTSeq Python framework (27) was used to compute the read counts per gene followed by limma (28) to integrate counts and detect differentially expressed genes. For differential analysis, a linear model was used to compare gene expression between genotypes with "Cage" included as a covariate. Furthermore, between-technical replicate correlation was accounted for using the duplicateCorrelation(), limma function (29).

**Differential exon usage analysis**

For differential usage of exons using limma (28), reads were aligned to the Ensembl-annotated genome that contains extensive annotation of coding and non-coding exons and quantified using DEXSeq (30) with a requirement of one read count per exon bin and a minimum of at least 10 reads across all replicates. Similar to differential analysis, for each gene, a linear model was used to detect differential exon usage between genotypes with "Cage" included as a covariate and between-technical replicate correlation was accounted for using the duplicateCorrelation(), limma function (29). Differential exon usage was defined as the proportion of total normalized reads per gene that are counted within an exon bin for that gene. Statistical significance was evaluated using gene-level tests, including an F test that reflects a consensus signal across a gene and is highly powered to detect differential exon usage when more than one signal is observed across multiple exons per gene. We also report the results form a Simes multiple testing procedure that assess all exons within a gene which is more powered to detect differential exon usage when a single strong signal is observed.

**Real-time quantitative PCR (qPCR) validation**

For qPCR validation of exon usage, oligo-dT primers were used to synthesize cDNA from total RNA using the same samples that were used in RNA-seq analysis using SYBR Green (ThermoFisher Scientific, Waltham, MA, USA; Cat# 4309155). The qPCR primers used for exon usage in *Hnrnph1* and *Ppp3ca* (two of

the top genes with differential exon usage) are shown in **Supplementary Table 1.** Each sample was run in triplicate, averaged, and normalized to its own expression level using *GAPDH* as a housekeeping gene. Differential exon usage was reported as the fold-change in 114 kb mice relative to B6J littermates using the $2^{-(\Delta\Delta C_T)}$ method (31).

***Hnrnph1:luc2* and B6J/D2J reporter lines construction**

We cloned the B6J *Hnrnph1* promoter sequence (Chr11: 50,375,375 to 50,378,330; mm10) starting at 2956 bp upstream of the annotated transcription start site for *Hnrnph1*. This sequence was amplified from B6J genomic DNA extracted from spleen tissue using iProof High-Fidelity PCR kit (BIO-RAD, Hercules, CA, USA; Cat# 172-5330) with the following primers containing *XhoI* and *HindIII* restriction enzyme sites in the sense and antisense primers respectively: sense (5'-GATTCTCGAGGCTCCCGTGATCAGATACAG-3') and anti-sense (5'-GTAAAGCTTCGTCCCTTCGGTGGTCCTGGC- 3'). The sequence was subsequently cloned into the multiple cloning site of pGL4.17[luc2/Neo] (Promega, Madison, WI, USA; Cat# E6721) with restriction enzymes *XhoI* and *HindIII,* placing the *Hnrnph1* promoter was inserted upstream of firefly luciferase luc2 coding sequence.

To replace B6J variants with D2J variants within the 5' UTR of *Hnrnph1*, we conducted site-directed mutagenesis using the B6J *Hnrnph1:luc2* construct as a template (Agilent QuikChange II, Santa Clara, CA, USA; Cat# 200521) to generate four luciferase reporter lines, each containing a single D2J variant. To introduce the three SNPs and the single indel all together into the 5' UTR of *Hnrnph1*, multi-site directed mutagenesis was performed by using the *GA7546G:luc2* construct with rs221962608 SNP variant as the template (Agilent, Santa Clara, CA, USA; Cat# 200514). The mutagenic primers are provided in **Supplementary Table 2.**

**Cell culture**

Human embryonic kidney (**HEK**)293T cells were grown in DMEM, high glucose with L-glutamine (Gibco, Waltham, MA, USA; Cat# 11965-092), supplemented with 10% fetal bovine serum (Gibco, Waltham, MA, USA; Cat# 26140-079), and 1% penicillin-streptomycin (Gibco, Waltham, MA, USA; Cat# 15140-122). Mouse neuro(N)2a (**N2a**) neuron-like cells were grown in 1:1 ratio of DMEM, high glucose with L-glutamine (Gibco, Waltham, MA, USA; Cat# 11965-092) to Opti-MEM (GibCO, Waltham, MA, USA; Cat# 31985-070), supplemented with 5% fetal bovine serum (Gibco, Waltham, MA, USA; Cat# 26140-079), and 1% penicillin-streptomycin (Gibco, Waltham, MA, USA; Cat# 15140-122). Cells were split every 3 to 4 days and were grown at 37°C in an atmosphere of 5% $CO_2$.

**Cell transfection**

For HEK293T, cells were seeded in 6-well plate at a density of 2.5 x $10^5$ cells per well 24 hours prior to transfection and would be about 90% confluent at the time of transfection. For N2A (given they grow faster than HEK293T cells), cells were seeded in 6-well plate at a density of 2 x $10^5$ cells per well. A suspension of plasmids, Lipofectamine 3000 and P3000 reagents (ThermoFisher Scientific, Waltham, MA, USA; Cat# L3000008) in Opti-MEM medium was pre-incubated at room temperature for 15 min before being added to each well of the 6-well plates. The cells were co-transfected with 1 μg of the experimental firefly luciferase construct and 0.2 μg internal

pRL control plasmid expressing *Renilla* luciferase driven by the CMV promoter (Promega, Madison, WI, USA; Cat# E2261). 24 hours after transfection, HEK293T cells are detached with 0.25% of Trypsin-EDTA (ThermoFisher Scientific, Waltham, MA, USA; Cat# 25200056) and counted and re-seeded at a density of 2.5 x $10^4$ cells per well in 96-well plate. N2a cells were re-seeded at a density of 2 x $10^4$ cells per well in 96-well plate. Cells were allowed to grow for approximately 48 h before measurement of luciferase activity.

**Dual luciferase reporter assay**

To account for background fluorescence signal, untransfected cells were used to subtract out background signal from the 96-well plate. Negative control cells were transfected with the promoter-less plasmid pGL4.17[luc2/Neo]. The Dual-Glo luciferase assay system (Promega, Madison, WI, USA; Cat# E2920) was used to measure luciferase activity. The growth medium was removed first and cells were then washed with PBS. Cell lysates were prepared by adding 50 µl of 1X passive lysis buffer to each well to lyse the cells by shaking the plate on rocker for 15 minutes. 20 µl of lysate from each well were transferred to wells of white opaque 96-well microplate (Corning, Corning, NY, USA; Cat# 3610). Dual-Glo LARII reagents (100 µl) were added into each well and the firefly luminescence was measured with SpectraMax i3x microplate reader after 15 min. An equal amount (100 µl) of Stop & Glo reagent was then added to the wells and Renilla luminescence was measured after 15 min.

Firefly and Renilla luciferase signal values were subtracted from the average background signal first. The background-adjusted firefly luciferase activity (FLA) for each sample was normalized to the background-adjusted internal control Renilla luciferase activity (RLA) to correct for differences in transfection efficiency and cell death. Relative luciferase activity for each experimental construct is expressed as the fold increase over the B6J *Hnrnph1:luc2* control:

$$Normalized\ luciferase\ activity = \frac{\left(\frac{FLA}{RLA}\right)_{Experimental}}{\left(\frac{FLA}{RLA}\right)_{\text{B6J}\ Hnrnph1:luc2\ control}}$$

**Experimental design and statistical analyses**

All data are presented as means of replicates from each experiment ± SEM. In summary, for experiments in which two conditions were compared, a two-tailed unpaired t-test was used to analyze the data unless otherwise specified. For the behavioral experiments in which multiple factors were evaluated, ANOVA was used and significant interaction was deconstructed along the relevant factor to examine main effects and group differences followed-up using t-tests. For the qPCR validation and luciferase report assay analyses in which multiple factors were evaluated, ANOVA was used and significant interaction was deconstructed with post hoc multiple-comparisons. Statistical p value threshold for t-test, ANOVA, and post-hoc tests was set to 0.05. The data comprising MA locomotor activity were analyzed in R (http://www.R-project.org/). All other data were performed in GraphPad Prism (Version 8.3.0).

## RESULTS

### Positional cloning of a 114 kb interval that is necessary for reduced MA sensitivity

In backcrossing the congenic line "Line 4a" that captured a QTL for reduced MA-induced locomotor activity and contained a heterozygous introgressed region from the D2J strain that spanned 50 Mb to 60 Mb, we monitored for recombination events at rs254771403 (50,486,998 bp, mm10). We identified a crossover event that defined 114 kb congenic region that influenced MA sensitivity. There are two protein coding genes within this 114 kb region - *Hnrnph1* and *Rufy1* **(Fig.1A)**. 114 kb mice homozygous for the D2J allele within this region on an otherwise isogenic B6J background displayed no difference in saline-induced locomotor activity, compared to their wild-type homozygous B6J littermates **(Fig.1B,C)**. In response to an acute dose of MA on Day 3, 114 kb mice showed reduced locomotor activity relative to B6J **(Fig.1D)**. Again, in response to a second dose of MA on Day 4, the 114 kb mice showed a decrease in MA-induced locomotion, compared to wild-type B6J littermates **(Fig.1E)**. No genotypic difference in MA-induced locomotor activity were observed on Day 5 after the third MA injection **(Fig.1F)**. Both 114 kb and B6J mice showed habituation to the testing apparatus as indicated by a reduction in saline-induced locomotor activity from Day 1 to Day 2 **(Fig.1G)**. Relative to the B6J mice, the 114 kb mice showed greater sensitization to repeated doses of MA from Day 3 to Day 5 **(Fig.1G)**. In comparing summed locomotor activity over 60 min across the five days, the 114 kb mice showed a significant reduction in MA-induced locomotor activity compared to B6J wild-type littermates on Day 4 **(Fig.1G)**. Taken together, these results indicate that 114 kb mice are less sensitive to the locomotor stimulant response to MA.

### Normalized MA-induced locomotor activity, as well as MA-induced locomotor sensitization, in 114 kb mice

Next, we examined MA-induced locomotor activity while accounting for individual differences in baseline locomotion in response to saline on Day 2, as well as MA-induced locomotor sensitization, via differences in MA-induced locomotor activity between MA treatment days. When accounting for non-drug locomotor activity, we again found that 114 kb mice showed less acute MA-induced locomotor activity relative to B6J mice **(Fig.2A)**, as well as after repeated exposure to MA **(Fig.2B-C)**. After two MA injections (comparing Day 4 versus Day 3 activity), 114 kb mice showed a significant decrease in locomotor sensitization early post-injection (**Fig.2D**). After three MA injections (comparing between Day 5 and Day 3), 114 kb mice showed a significant increase in locomotor sensitization, starting at 20 min post-MA **(Fig.2E)**. Again, when comparing Day 5 versus Day 4, 114 kb mice also showed significantly greater MA locomotor sensitization relative to B6J mice **(Fig.2F)**. The greater sensitization in 114 kb mice is likely due to their initially lower MA response.

### Transcriptome analysis identifies differential 5' UTR exon usage of *Ppp3ca* as well as *Hnrnph1* in 114 kb mice

To further understand the molecular mechanism of the 114 kb congenic region in reducing MA sensitivity, we performed both differential gene expression and differential exon usage analysis on RNA-seq data collected from the striatum of naïve 114 kb and their B6J wild-type littermates. We identified 69 differentially expressed genes ($p < 0.001$; **Table 1**) and 35 genes exhibiting differential exon usage **(Table 2A, B)**, which is defined as

the proportion of total normalized reads per gene that are counted within an exon bin for that gene. Notably, *Hnrnph1* (the first protein-coding gene within the 114 kb congenic interval) was one of the top genes exhibiting differential exon usage, providing direct evidence at the exon level for one or more functional variants within the gene that regulates exon usage and ultimately underlies differences in behavior. It should also be noted that we did not identify any significant difference in gene expression or in exon usage of *Rufy1* (the second protein-coding gene within the interval), thus limiting functional effects of D2J variants at the transcript and exon level to *Hnrnph1* and further supporting the candidacy of *Hnrnph1* as a quantitative trait gene underlying MA behavior (24).

We previously showed that mice heterozygous for a 16-bp deletion within the first coding exon of *Hnrnph1* (refer to as the $H1^{+/-}$ mice) were less sensitive to the stimulant, rewarding, and reinforcing properties of MA and showed a reduction in MA-induced dopamine release (24, 25). To directly compare the functional effects of inheriting the deletion versus the 114 kb congenic region on gene expression and exon usage, we identified 21 overlapping genes that were differentially expressed in the striatum between 114 kb mice and $H1^{+/-}$ mice relative to their B6J wild-type littermates **(Fig.3A)**, which was significantly greater than what would be expected by chance (Fisher's exact test: $p = 9.31 \times 10^{-23}$). In addition, there were 10 overlapping genes between $H1^{+/-}$ and 114 kb mice that showed differential exon usage, relative to their B6J littermates **(Fig.3B)**. This, again, was significantly greater than what would be expected by chance (Fisher's exact test: $p = 2.96 \times 10^{-07}$). Finally, in correlating differential expression for the 21 shared differentially expressed genes between 114 kb mice and $H1^{+/-}$ mice, we found a nearly perfect relationship between the magnitude and direction of change in gene expression **(Table 3 and Fig.4)**. These results suggest functionally similar effects of the 114 kb region and the $H1^{+/-}$ allele on their overlapping, convergent transcriptome.

*Ppp3ca* (or calcineurin A subunit gene) was identified as the top gene with differential exon usage in the 114 kb mice, relative to B6J wild-type littermates **(Table 2)** and it was also one of the 10 overlapping genes with differential exon usage **(Fig.3B)**. Fine mapping of *PPP3CA* reveals novel alternative splicing patterns in human subjects with high addiction vulnerability (32). In order to validate differential exon usage in *Ppp3ca* in the 114 kb mice, we designed primers flanking exons to detect the splicing isoforms of *PPP3CA* identified by Chiocco et al. (32) **(Fig.5A; Supplemental Table 1)**. 114 kb mice showed greater usage of exons 1 to 2 but no difference in the other 6 exon junctions **(Fig.5A)**. To further investigate the specific exon, we designed primers to target exons 1 and 2 separately. Relative to B6J, increased exon 1 usage was shown in the 114 kb mice **(Fig.5B)**, with no difference in exon 2 usage **(Fig.5C)**. Exon 1 is part of the 5' UTR of *Ppp3ca*. Thus, we conclude that the 114 kb mice show an increased number of transcripts that include this 5' UTR exon.

The physical positions and exon numbers comprising differential exon usage of *Hnrnph1* in 114 kb relative to B6J mice are shown in **Table 4**. Note that fewer normalized reads were mapped to the junction comprising exons 3 and exon 4 in 114 kb mice versus B6J. Additionally, a greater number of normalized reads mapped to the junction comprising exons 6 and 7, as well as the junction comprising exons 7 and 8 in the 114 kb

mice relative to B6J **(Table 4)**. It is important to note that *Hnrnph2* (homolog of *Hnrnph1*) did not show differential exon usage in our analysis. To validate differential exon usage of *Hnrnph1*, we performed qPCR with primers flanking those exons with significant differential exon usage. Out of the three exon junctions, 114 kb mice showed less usage of exons 3 to 4, but no difference in the other two exon junctions **(Fig.6A)**. To further validate the specific exons exhibiting differential usage, we designed primers to target exon 3 and exon 4 separately. Less usage of exon 3 was detected in the 114 kb mice **(Fig.6B)**, with no change in exon 4 **(Fig.6C)**. Exons 1 to 3 are part of the 5' UTR of *Hnrnph1* **(Fig.7)**. Thus, we conclude that 114 kb mice show a reduced number of *Hnrnph1* transcripts that include this 5' UTR exon.

**Identification of a set of 5' UTR functional variants in *Hnrnph1* that decrease luciferase reporter expression**

To identify potential functional variants in the 5' UTR of *Hnrnph1* underlying MA-induced behavior, we used the whole genome sequence dataset (33, 34) and the online Sanger variant query tool (https://www.sanger.ac.uk/sanger/Mouse_SnpViewer/rel-1505) to identify three SNPs and one indel (insertion/deletion) that were located within the 5' UTR exons and introns **(Table 5;** highlighted in orange**)**. There are 12 additional *Hnrnph1* variants downstream of the 5' UTR **(Table 5 and Fig.7)**. Given that the 5' UTR contains a promoter element for transcriptional regulation, we employed a luciferase reporter system to assay the function of the *Hnrnph1* promoter in both Human Embryonic Kidney (**HEK**) 293T and Neuro2a (**N2a**) cells. Using this system combined with site-directed mutagenesis, we examined the effect of each of the four 5' UTR *Hnrnph1* variants on *Hnrnph1* promoter activity. To examine *Hnrnph1* promoter activity, we developed a *Hnrnph1:luc2* construct by cloning 2956 bp of the B6J *Hnrnph1* promoter into the pGL4.17[luc2/Neo] promoter-less vector **(Supplemental Fig.1 and 2A, and Fig.8A)**. We next tested whether this cloned promoter could drive expression of the firefly luciferase gene, *luc2*. The 2956-bp promoter increased firefly luminescence compared to the pGL4.17[luc2/Neo] vector, and this increase in signal showed a dose-dependency, depending on the number of plated cells **(Supplemental Fig.2B).**

Using site-directed mutagenesis, we constructed 4 *Hnrnph1* promoters each possessing its own D2J variant within the 5' UTR **(Supplemental Table 2)**. In HEK293T cells, none of the four promoters with a single D2J variant differed significantly from the B6J *Hnrnph1* promoter with regard to driving luciferase expression **(Fig.8B)**. However, when all four D2J variants were introduced into the promoter, there was a significant reduction in luciferase activity compared to the B6J promoter **(Fig.8C)**. The findings were subsequently replicated in N2a cells **(Fig.8D-E)**. Taken together, these results identify a set of functional 5' UTR variants within *Hnrnph1* that decrease exon 3 usage of *Hnrnph1* and reduce *Hnrnph1*-promoter driven transcription.

## DISCUSSION

The 114 kb region containing two protein genes (*Hnrnph1*, *Rufy1*) was not only necessary (24), but was also sufficient, to induce a decrease in sensitivity to the locomotor stimulant response to MA **(Fig.1)**. Our prior work strongly implicated *Hnrnph1* and not *Rufy1* as the QTG responsible for the reduction in MA-induced locomotor activity (24) and we subsequently expanded the behavioral repertoire of $H1^{+/-}$ mice to include reduced MA-induced reinforcement, reward, and dopamine release (25). Our exon-level transcriptome analysis and comparative differential gene expression analysis between the 114 kb congenics and $H1^{+/-}$ mice further supports *Hnrnph1* as the QTG underlying the reduction in MA-induced behavior **(Fig.3)**. Specifically, we found differential exon usage of the 5' UTR of *Hnrnph1* in 114 kb mice that was independently validated via qPCR **(Fig.6)** and was associated with the inheritance of four SNP/indel variants near the 5' UTR as demonstrated by a reporter assay to measure *Hnrnph1* promoter strength **(Fig.8)**.

Identification of QTGs and QTVs is extremely rare in mammalian forward genetic studies, in particular for behavior (35), and especially when considering the identification of noncoding functional variants. In the recent identification of a noncoding functional variant underlying an eQTL, Mulligan and colleagues took advantage of the reduced genetic complexity of closely related C57BL/6 substrains (36) to deduce and validate an intronic variant in *Gabra2* (alpha-2 subunit of the GABA-A receptor) near a splice site that drastically reduced transcript and protein expression and modulated behavior (37). Correction of the single base deletion in C57BL/6J to the wild-type C57BL/6N allele via CRISPR/Cas9-mediated gene editing restored *Gabra2* transcript and protein levels to a normal level, thus identifying the QTN (37). The vast majority of other examples that successfully identified both the QTG and the QTV involve functional coding mutations, including, e.g., *Cyfip2* (38), *Grm2* (13), and *Taar1* (39). With regard to *Hnrnph1*, a human candidate gene association study identified a functional intronic variant in *OPRM1* (mRNA target of hnRNP H) affects binding to hnRNP H to regulate splicing that was associated with the severity of heroin addiction (40).

We identified a set of four D2J variants within the 5' UTR of *Hnrnph1* and its introns that were associated with differential 5' UTR exon usage **(Fig.6)** and decreased transcription **(Fig.8)**. A reasonable assumption is that inheritance of these 5' UTR variants is also sufficient to decrease hnRNP H1 protein expression, however, direct demonstration would require cloning of the full-length gene plus a reporter tag since there are currently no selective antibodies that can distinguish hnRNP H1 from hnRNP H2. At the exon-level, we validated a decrease in *Hnrnph1* transcripts containing exon 3 **(Fig.6B)**. The decrease in reporter expression was observed in our *in vitro* assays comprising two different cell lines, despite the fact that we did not observe any significant gene-level change in *Hnrnph1* transcription in 114 kb mice carrying these 4 variants [plus the 12 other additional variants; $\log_2 FC = 0.016$; $t(14) = 0.70$; $p = 0.49$]. Together, these results indicate that D2J variants within and surrounding the 5' UTR of *Hnrnph1* leads to exclusion of exon 3, which in turn leads to a functional decrease in protein translation. These variants may also cause other functional changes in the 5' UTR leading to a decrease in protein

translation by modifying the regulatory elements within the promoter to reduce efficacy of translation (40), or changing the binding sites for other RNA-binding proteins **(RBPs)** to repress initiation of translation (41, 42).

The top gene demonstrating differential exon usage in 114 kb mice was *Ppp3ca* **(Table 2, Fig.3**), a gene coding for the alpha catalytic subunit of calcineurin (a.k.a. PP2B), a $Ca^{2+}$-responsive serine/threonine protein phosphatase that regulates dopamine and excitatory signaling, as well as the behavioral response to psychostimulants (43). We independently validated our result via qPCR by showing an increase in transcription of exon 1 of *Ppp3ca*, which is a 5' UTR exon **(Fig.5)**. A human genetic fine mapping study identified a trinucleotide repeat in the 5' UTR of *PPP3CA* that showed allele-specific expression and altered expression of novel, alternatively spliced isoforms that were associated with addiction vulnerability and Alzheimer's disease (32). These 5' UTR variants were associated with the production of splice variants with predicted functional alterations as the coding sequences lack portions of the catalytic and/or regulatory subunits (32). Another study using convergent functional genomic analysis of transcriptome and GWAS datasets of cocaine addiction traits identified *PPP3CA* as one of the top genes (44). Other genetic studies found *de novo* mutations in *PPP3CA* that cause severe neurodevelopmental disorders and seizures (45). We hypothesized that more subtle genetic mutations in RNA-binding proteins could alter the neurodevelopment of the reward circuitry and processing that, in turn, influences addiction vulnerability (46). This hypothesis is supported by our current observation of an association of 5' UTR mutations in *Hnrnph1* with 5' UTR usage in *Ppp3ca*. Future experiments will determine whether hnRNP H protein directly binds to *Ppp3ca* mRNA to regulate alternative splicing of the *Ppp3ca* 5' UTR, and whether decreased hnRNP H1 protein leads to increased exon 1 usage within the 5' UTR of *Ppp3ca* to affect PPP3CA protein expression and calcineurin function.

The functional impact of increased 5' UTR exon usage in *Ppp3ca* is unclear. A recent study demonstrated in an amoeba species that *Ppp31* (protein phosphatase 3 regulatory subunit B, alpha), a gene coding for one of the $Ca^{2+}$-binding regulatory subunits of calcineurin, shows expression of two different isoforms at the mRNA and protein level and their mass sizes correspond to two different alternative start codons. The smaller sized transcript lacked the 5' UTR and the more upstream start codon, thus demonstrating differential 5' UTR usage of this subunit (47). Calcineurin dephosphorylates residues of various dopamine signaling molecules that are phosphorylated by kinase cascades in response to D1 dopamine receptor activation (e.g., PKA, casein kinase I), including DARPP-32 (48, 49) and has been implicated in a variety of MA effects, including MA-induced locomotor sensitization (50–52), destabilization of contextual memory (53), apoptosis (54), and neurotoxicity (55). In addition to dephosphorylation in dopamine receptor signaling, mitochondrial dysfunction can initiate a stress response that activates calcineurin and the RNA-binding protein hnRNPA2 that acts as a coactivator to regulate transcription of nuclear target genes (56, 57). hnRNP H binds exon 2 of the mitochondrial gene *Ndufb11* and regulates alternative splicing (58). We previously conducted synaptosomal proteomic analysis and found indirect evidence for mitochondrial dysfunction in $H1^{+/-}$ mice at baseline and in response to MA, including changes in complex I and V proteins (25). We propose that *Hnrnph1* D2J polymorphisms decrease transcription

of *Hnrnph1* **(Fig.6)**, preventing the binding and subsequently exon 1 exclusion of *Ppp3ca* **(Fig.5)**. A change in post-transcriptional processing of *Ppp3ca* could contribute to multiple downstream neurobiological events (a change in dephosphorylation in dopamine signaling and/or the mitochondrial stress response) underlying the reduction in MA behavior. Support for this hypothesis comes from other examples where hnRNP H promotes exon exclusion (59–61).

To summarize, we provide further causal behavioral and molecular evidence for *Hnrnph1* as a QTG for MA sensitivity by demonstrating that inheritance of a very small, 114 kb chromosomal region was not only *necessary* (24) but also *sufficient* to induce the behavioral phenotype. We provide functional evidence for a set of four QTVs within the 5' UTR that plausibly regulate differential exon usage and gene expression. The functional significance of these variants at the behavioral level is unknown but our comparative transcriptome analysis of the 114 kb mice with the $H1^{+/-}$ mice (24, 25) demonstrates functionally similar genotypic effects on the degree and direction of fold-change in gene expression and behavioral response to MA. We predict that inheritance of these 5' UTR variants in *Hnrnph1* alone would also be sufficient to induce a functional effect at the behavioral level. However, the subtle effect of the 114 kb region on behavior raises the question of whether modeling these variants *in vivo* is worth the effort –convergent evidence provides strong evidence that we have identified the QTVs underlying functional molecular and behavioral changes. Finally, we identified *Ppp3ca* as a candidate RNA-binding target of hnRNP H1 that could bridge *Hnrnph1* polymorphisms and function with methamphetamine-induced dopaminergic signaling mechanism(s) underlying addiction-relevant behaviors (25). Future experiments using CLIP-seq and proteomics will address whether *Ppp3ca* is a binding target of hnRNP H protein and identify additional RNA-binding targets within dopaminergic and striatal neurons of the reward circuitry.

## AUTHOR CONTRIBUTIONS

Q. T. Ruan, N. Yazdani, K.K. Szumlinski, P. E. A. Ash, B. Wolozin, and C. D. Bryant designed research; Q. T. Ruan, N. Yazdani, and E. R. Reed analyzed data; Q. T. Ruan, N. Yazdani, J. A. Beierle, L. P. Peterson, and K. P. Luttik performed research; P. E. A. Ash and B. Wolozin contributed new reagents or analytical tools; Q.T. Ruan and C. D. Bryant wrote the paper.

## FIGURE LEGENDS

**Figure 1. MA-induced locomotor activity in the 114 kb B6J.D2J congenic mouse line. (A):** Congenic mice containing an introgressed 114 kb region on chromosome 11 from the D2J strain on an isogenic B6J background were identified via backcrossing and screening for a recombination event at rs254771403 (50,486,998 bp; mm10) while monitoring for the retention of polymorphic rs29383600 (50,373,006 bp; mm10) and polymorphic rs29459915 (50,484,260 bp; mm10), and. Mice were then backcrossed again to B6J to generate additional heterozygotes for intercrossing (heterozygous-heterozygous breeding) which yielded heterozygous offspring as well as offspring homozygous for the 114 kb region (homozygous D2J) and wild-type homozygous B6J littermates. Behavioral analysis and results are shown for the homozygous genotypes. The 114 kb region contains two protein-coding genes: *Hnrnph1* and *Rufy1*. Mice were injected on Days 1 to 2 with saline (i.p.; panels B and C) and on Days 3, 4, and 5 with 2 mg/kg MA (panels D, E, and F). Locomotor activity was video-recorded for 60 min on each day. Locomotor activity is presented in 5-min bins over 60 min. **(B,C):** No genotypic differences were observed in distance traveled in response to a saline injection on Day 1 [mixed ANOVA: $F(1,46)_{Genotype}$ = 1.48, p = 0.229; $F(11,506)_{Genotype\ x\ Time}$ = 1.48, p = 0.133] or Day 2 [mixed ANOVA: $F(1,46)_{Genotype}$ = 0.003, p = 0.958; $F(11,506)_{Genotype\ x\ Time}$ = 1,29, p = 0.227]. A main effect of Sex was detected for both Days 1 and 2 [$F(1,46)_{Day1}$ = 10.52, p = 0.002; $F(1,46)_{Day2}$ = 9.83, p = 0.003]. However, there was no significant Genotype x Sex interaction for Day 1 or Day 2 [$F(1,46)_{Day1}$ = 0.32, p = 0.576; $F(1,46)_{Day2}$ = 2.32, p = 0.134]. **(D):** Locomotor activity on Day 3 in response to MA is shown in 5-min bins over 60 min. 114 kb mice showed reduced acute MA-induced locomotor activity relative to B6J wild-type littermates [mixed ANOVA: $F(1,46)_{Genotype}$ = 3.41, p = 0.071; $F(11,506)_{Genotype\ x\ Time}$ = 2.80, p = 0.002; unpaired t-test for each 5-min bin: $t(48)_{30min}$ = 2.17, *p = 0.035; $t(48)_{50min}$ = 2.05, *p = 0.045]. A main effect of Sex was also detected [F(1,46) = 4.32, p = 0.043], however, there was no significant Genotype x Sex interaction [F(1,46) = 2.74, p = 0.104]. **(E):** MA-induced locomotor activity on Day 4 is shown in 5-min bins over 60 min. 114 kb mice showed reduced MA-induced locomotor activity relative to B6J mice [mixed ANOVA: $F(1,46)_{Genotype}$ = 5.40, p = 0.025; $F(11,506)_{Genotype\ x\ Time}$ = 1.96, p = 0.031; unpaired t-test for each 5-min bin: $t(48)_{15min}$ = 2.43, *p = 0.019; $t(48)_{20min}$ = 3.10, *p = 0.003; $t(48)_{25min}$ = 2.90, *p = 0.006; $t(48)_{30min}$ = 2.22, *p = 0.031]. A main effect of Sex was also detected [F(1,46) = 7.33, p = 0.009] but no significant Genotype x Sex interaction [F(1,46) = 2.05, p = 0.159]. **(F):** MA-induced locomotor activity is shown for Day 5 in 5-min bins over 60 min. 114 kb mice showed no difference in MA-induced locomotor activity compared to B6J [mixed ANOVA: $F(11,506)_{Genotype\ x\ Time}$ = 0.59, p = 0.841]. A main effect of Sex was detected [F(1,46) = 6.09, p = 0.017], however, there was no significant Genotype x Sex interaction [F(1,46) = 0.39, p = 0.535]. **(G):** Locomotor activity summed over 60 min is shown for Days 1 and 2 (saline) and on Days 3, 4, and 5 (MA). In examining habituation via changes in locomotor activity in response to saline injections on Day 1 versus Day 2, both B6J and 114 kb showed similar degree of habituation [mixed ANOVA: $F(1,46)_{Genotype}$ = 0.47, p = 0.496; $F(1,46)_{Day}$ = 225.51, p < 2E-16; $F(1,46)_{Genotype\ x\ Day}$ = 1.35, p = 0.252; Bonferroni's multiple comparison test: $t(50)_{Day\ 1}$ = 6.34, #p = 3.26E-04; $t(46)_{Day\ 2}$ = 6.64, #p = 1.55E-04]. In examining sensitization of MA-induced

locomotor activity across Day 3, Day 4, and Day 5, 114 kb mice showed increased locomotor sensitization relative to B6J [mixed ANOVA: $F(1,46)_{Genotype} = 2.74$, $p = 0.105$; $F(2,92)_{Day} = 65.71$, $p < 2E-16$; $F(2,92)_{Genotype\ x\ Day} = 5.03$, $p = 0.009$]. Only the 114 kb mice showed an increase in MA-induced locomotor activity across the three MA treatment days [Bonferroni's multiple comparison tests: Day 3 versus D4: $t(50)_{B6J} = 2.40$, $p = 0.120$; $t(46)_{114\ kb} = 2.76$, $^{*}p = 0.050$; Day 3 versus D5: $t(50)_{B6J} = 2.45$, $p = 0.108$; $t(46)_{114\ kb} = 4.52$, $^{*}p = 2.57E-04$; Day 4 versus D5: $t(50)_{B6J} = 0.12$, $p = 0.999$; $t(46)_{114\ kb} = 1.92$, $p = 0.368$]. In addition, 114 kb mice showed reduced MA-induced locomotor activity compared to B6J mice on Day 4 when compared across all five days [mixed ANOVA: $F(4,184)_{Genotype\ x\ Day} = 3.96$, $p = 0.004$; unpaired t-test for each day: $t(48)_{Day4} = 2.16$, $^{\%}p = 0.036$]. A main effect of Sex was detected [$F(1,46) = 9.17$, $p = 0.004$] but there was no significant Genotype x Sex interaction [$F(1,46) = 2.14$, $p = 0.150$]. Data are represented as mean ± SEM.

**Figure 2. MA-induced locomotor sensitization in 114 kb mice.** Mice were injected (i.p.) with saline on Days 1 to 2 and with 2 mg/kg MA on Days 3, 4, and 5. Locomotor activity was recorded for 60 min. **(A):** Day 3 – Day 2 distance traveled represents the acute locomotor response to MA while accounting for individual differences in saline locomotor activity. 114 kb mice showed less acute MA-induced locomotor activity relative to B6J wild-type littermates [mixed ANOVA: $F(1,46)_{Genotype} = 3.43$, $p = 0.070$; $F(11,506)_{Genotype\ x\ Time} = 2.36$, $p = 0.008$; unpaired t-test for each 5-min bin: $t(48)_{25min} = 2.30$, $^{*}p = 0.033$; $t(48)_{30min} = 2.18$, $^{*}p = 0.035$]. **(B):** Day 4 – Day 2 distance traveled represents the locomotor response to second repeated dose of MA while accounting for individual differences in saline locomotor activity. 114 kb mice showed less MA-induced locomotor activity relative to B6J [mixed ANOVA: $F(1,46)_{Genotype} = 5.25$, $p = 0.027$; $F(11,506)_{Genotype\ x\ Time} = 1.76$, $p = 0.058$; unpaired t-test for each 5-min bin: $t(48)_{10min} = 2.15$, $^{*}p = 0.037$; $t(48)_{15min} = 2.47$, $^{*}p = 0.017$; $t(48)_{20min} = 2.80$, $^{*}p = 0.007$; $t(48)_{25min} = 3.34$, $^{*}p = 0.002$; $t(48)_{30min} = 2.24$, $^{*}p = 0.030$]. **(C):** Day 5 – Day 2 distance traveled represents locomotor response to third repeated dose of MA while accounting for individual differences in saline locomotor activity. No genotypic difference in locomotor activity in response to third dose of MA [mixed ANOVA: $F(1,46)_{Genotype} = 0.17$, $p = 0.680$; $F(11,506)_{Genotype\ x\ Time} = 0.92$, $p = 0.523$]. **(D):** Day 4 – Day 3 distance traveled is shown to represent locomotor sensitization from the first to the second MA injection. 114 kb mice showed a decrease in locomotor sensitization during the first 15 min [mixed ANOVA: $F(1,46)_{Genotype} = 0.47$, $p = 0.495$; $F(11,506)_{Genotype\ x\ Time} = 3.09$, $p = 4.92e-5$; unpaired t-test for each 5-min bin: $t(48)_{5min} = 2.47$, $^{*}p = 0.017$; $t(48)_{15min} = 2.47$, $^{*}p = 0.017$]. **(E):** Day 5 – Day 3 distance traveled is shown to indicate the degree of sensitization relative to the first MA exposure following two subsequent injections. 114 kb mice showed an increase in locomotor sensitization relative to B6J [mixed ANOVA: $F(1,46)_{Genotype} = 5.35$, $p = 0.025$; $F(11,506)_{Genotype\ x\ Time} = 1.95$, $p = 0.032$; unpaired t-test for each 5-min bin: $t(48)_{20min} = 2.30$, $p = {}^{*}0.026$; $t(48)_{25min} = 2.07$, $^{*}p = 0.044$; $t(48)_{30min} = 2.11$, $^{*}p = 0.040$; $t(48)_{45min} = 2.27$, $^{*}p = 0.028$; $t(48)_{50min} = 2.80$, $p = 0.007$]. **(F):** D5 – D4 distance traveled is shown to indicate which genotype continued to sensitize, even after two MA injections. Again, 114 kb mice showed increased locomotor sensitization relative to B6J [mixed ANOVA: $F(1,46)_{Genotype} = 8.50$, $p = 0.006$; $F(11,506)_{Genotype\ x\ Time} = 1.75$, $p = 0.061$; unpaired t-test for each 5-min bin: t(48)5min = 2.67, $^{*}p = 0.010$;

t(48)15min = 3.28, *p = 0.002; t(48)20min = 4.59, *p = 3.17E-05; t(48)25min = 3.43, *p = 0.001; t(48)35min = 2.24, *p = 0.030]. No main effect of Sex was detected for any locomotor sensitization measures (panels A-F) [$F(1,46)_{Sex} < 0.8$, all p's > 0.3]. Data are represented as mean ± SEM.

**Table 1. Differentially expressed genes between 114 kb and B6J.** The table shows the 69 genes that are differentially expressed in the striatum of 114 kb relative to the B6J wild-type littermates ($p < 0.001$).

**Table 2. Genes exhibiting differential exon usage between 114 kb and B6J. (A - B):** The table shows genes exhibiting differential exon usage in striatal tissue from 114 kb mice vs. their B6J wild-type littermates. Differential exon usage was detected using either an F-test (**A**) or a Simes test (**B**) in limma.

**Figure 3. Overlap in differentially expressed genes and genes exhibiting differential exon usage between 114 kb mice and $H1^{+/-}$ mice. (A):** Venn diagram showing the number of differentially expressed genes (**DEGs)** in 114 kb ($p < 0.001$) versus $H1^{+/-}$ mice ($p < 0.001$). The 21 DEGs that are shared between the two are significantly greater than what would be expected by chance (Fisher's exact test: p = 9.31E-23) as it represents nearly ½ of the 48 total DEGs identified in 114 kb mice. **(B):** Venn diagram shows the number of non-overlapping and overlapping genes with differential exon usage in 114 kb mice ($p < 0.001$) versus $H1^{+/-}$ mice ($p < 0.001$). Ten genes showed differential exon usage in both the 114 kb mice and the H1+/- mice which is significantly greater than chance (Fisher's exact test: p = 2.96E-07) and include *Ppp3ca, Cdhr4, Hnrnph1, Mapk9, Hnrnpm, Ptpn5, Ip6k1, Uba7, Slc8a3,* and *Mtap4*.

**Table 3. Correlation of differential expression between 114 kb mice and $H1^{+/-}$ mice.** Table showing the $log_2FC$ values for the 21 overlapping genes in the 114 kb and $H1^{+/-}$. The change in gene expression (up/down) is relative to B6J wild-type littermates.

**Figure 4. Correlation of differential expression between 114 kb mice and $H1^{+/-}$ mice.** The direction and the magnitude of change in gene expression relative to B6J wildtype are nearly identical between the 114 kb and $H1^{+/-}$ for every gene [Pearson correlation: r = 0.960, $R^2 = 0.922$, p = 5.41E-12].

**Figure 5. Increased 5' UTR exon usage of *Ppp3ca* in striatal tissue of 114 kb mice.** *Ppp3ca* was the most significant gene showing differential exon usage as indicated via the proportion of total normalized reads counted within an exon bin between the 114 kb versus B6J wild-type littermates. Exon usage was defined as the proportion of total normalized reads per gene that are counted within an exon bin for that gene**.** To validate differential exon usage, we designed a set of qPCR primers based on differential *PPP3CA* isoform expression identified in substance-dependent human subjects (32). The primer sequences are provided in Supplemental Table 1. **(A):** Out of the seven exon junctions tested, 114 kb mice showed increased usage of the first 5' UTR exon junction, namely exon 1-2 [unpaired t-test for each exon junction: $t(13)_{exon\ 1\text{-}2} = 2.67$, *p = 0.003; $t(13)_{exon\ 1\text{-}3} = 0.23$, p = 0.825; $t(6)_{exon\ 1\text{-}11} = 0.12$, p = 0.813; $t(6)_{exon\ 2\text{-}9} = 1.31$, p = 0.248; $t(6)_{exon\ 8\text{-}10} = 1.05$, p = 0.340; $t(13)_{exon\ 12\text{-}13} = 0.47$, p = 0.645; $t(13)_{exon\ 12\text{-}14} = 1.77$, p = 0.100]. **(B-C):** To further test specific exons, primers targeting either exon 1 or 2 of *Ppp3ca* were used to demonstrate that 114 kb mice showed increased usage of exon 1 [unpaired t-test: t(13) =

2.73, *p = 0.017] but no difference in exon 2 [unpaired t-test: t(13) = 1.16, p = 0.268]. Bars and error bars are represented as mean ± SEM.

**Table 4. Exon usage of *Hnrnph1* in 114 kb mice.** Table showing the four exon positions of *Hnrnph1* are shown that display differential exon usage in 114 kb mice relative to B6J wild-type littermates with $p < 0.05$ using the Simes test. Exon usage was defined as the proportion of total normalized reads per gene that are counted within an exon bin for that gene.

**Figure 6. Decreased 5' UTR exon usage of *Hnrnph1* in striatal tissue of 114 kb mice.** *Hnrnph1* was one of the top-ranked genes showing differential exon usage as indicated via the proportion of total normalized reads counted within an exon bin between 114 kb mice versus B6J wild-type littermates. To validate differential exon usage, we conducted a set of qPCR experiments targeting the exon junction or the individual exons. Primer sequences, as well as the exon junctions or individual exons they target are provided in Supplemental Table 1. **(A)** Decreased usage of the junction comprising exons 3 and 4 was detected in 114 kb mice [unpaired t-test for each exon junction: $t(13)_{\text{exon 3-4}} = 5.65$, *p = 7.9e-5; $t(6)_{\text{exon 6-7}} = 0.58$, p = 0.585; $t(6)_{\text{exon 7-8}} = 0.20$, p = 0.846]. **(B-C)** To further test specific exons, primers targeting either exon 3 or 4 of *Hnrnph1* were used to demonstrate that 114 kb showed decreased usage of exon 3 [unpaired t-test: t(13) = 3.25, *p = 0.006] but no difference in exon 4 [unpaired t-test: t(13) = 1.53, p = 0.150]. Bars and error bars are represented as mean ± SEM.

**Table 5. SNPs and indels in *Hnrnph1* between B6J and D2J.** A query for *Hnrnph1* genetic variants (SNPs plus indels) between the B6J and D2J parental strains was conducted using the Sanger database query tool (mm10, REL-1505). The variants highlighted in orange are located in close proximity to exon 3 and were cloned and tested functionally in the luciferase reporter assay.

**Figure 7. *Hnrnph1* variants between the B6J and D2J parental strains.** Schematic representation of the position of SNPs and indels in *Hnrnph1* with the exon numbers listed. The exon with decreased usage in 114 kb mice is boxed in red.

**Figure 8. Luciferase reporter assay for the *Hnrnph1* promoter in HEK293T and N2a cells to test functional 5' UTR variants.** Firefly luciferase was used as a reporter to assess the transcriptional activity of B6J-*Hnrnph1:luc2* versus promoters containing individual 5' UTR variants from the D2J strain. *Renilla* luciferase was used as an internal control reporter. **(A):** Schematic representation of the *Hnrnph1:luc2* reporter construct that was transfected to HEK293T **(B and C)** or N2a **(D and E)** cells. **(B):** Single D2J variants were engineered within the 5' UTR of the *Hnrnph1* promoter and had no effect on luciferase reporter signal in HEK293T cells [one-way ANOVA comparing B6J-*Hnnrph1:luc2* against the single D2J variants: F(4,50) = 3.03, p = 0.025; Dunnett's multiple comparison test with comparing each of the four variants with the *B6J-Hnnrph1:luc2* control group: $D_{\text{Dunnett(T7288C)}} = 0.36$, p = 0.988; $D_{\text{Dunnett(GA7546G)}} = 1.88$, p = 0.199; $_{\text{DDunnett(T7795G)}} = 2.11$, p = 0.125; $D_{\text{Dunnett(G7937C)}} = 2.14$, p = 0.117]. Data points represent the normalized luciferase signal from 2 independent replicates. **(C):** Engineering of all four D2J variants (D2J-*Hnrnph1:luc2*) in the promoter of *Hnrnph1* decreased luciferase

luminescence relative to B6J-*Hnrnph1:luc2* in HEK293T cells [unpaired t-test, $t(34) = 9.39$, *$p = 5.68E-11$]. Data points are normalized luciferase signal from 3 independent replicates. **(D):** Singly engineered D2J variants within the promoter had no effect on luciferase reporter signal in N2a cells [one-way ANOVA comparing B6J-*Hnnrph1:luc2* against the single D2J variants: $F(4,75) = 4.42$, $p = 0.003$; Dunnett's multiple comparison test with comparing each of the four variants with the *B6J-Hnnrph1:luc2* control group: $D_{Dunnett(T7288C)} = 1.18$, $p = 0.584$; $D_{Dunnett(GA7546G)} = 2.20$, $p = 0.100$; $DDunnett(T7795G) = 0.18$, $p = 0.999$; $D_{Dunnett(G7937C)} = 1.79$, $p = 0.229$]. Data points are normalized luciferase signal from 2 independent replicates. **(E):** Engineering all four D2J variants (D2J-*Hnrnph1:luc2*) within the *Hnrnph1* promoter decreased luciferase luminescence relative to the B6J-*Hnrnph1:luc2* promoter in N2a cells [unpaired t-test, $t(46) = 5.59$, *$p = 1.19E-6$]. Data points are normalized luciferase signal from 3 independent replicates. Error bars are represented as mean ± SEM.

**REFERENCES**


1. Fogger, S. A. (2019) Methamphetamine Use: A New Wave in the Opioid Crisis? *J Addict Nurs* **30**, 219–223
2. Pierce, R. C., Fant, B., Swinford-Jackson, S. E., Heller, E. A., Berrettini, W. H., and Wimmer, M. E. (2018) Environmental, genetic and epigenetic contributions to cocaine addiction. *Neuropsychopharmacology* **43**, 1471–1480
3. Jensen, K. P. (2016) A Review of Genome-Wide Association Studies of Stimulant and Opioid Use Disorders. *Mol.Neuropsychiatry* **2**, 37–45
4. Gelernter, J., Sherva, R., Koesterer, R., Almasy, L., Zhao, H., Kranzler, H. R., and Farrer, L. (2014) Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Mol.Psychiatry* **19**, 717–723
5. Dickson, P. E., Miller, M. M., Calton, M. A., Bubier, J. A., Cook, M. N., Goldowitz, D., Chesler, E. J., and Mittleman, G. (2015) Systems genetics of intravenous cocaine self-administration in the BXD recombinant inbred mouse panel. *Psychopharmacology (Berl)*
6. Harkness, J. H., Shi, X., Janowsky, A., and Phillips, T. J. (2015) Trace Amine-Associated Receptor 1 Regulation of Methamphetamine Intake and Related Traits. *Neuropsychopharmacology* **40**, 2175–2184
7. Miner, N. B., Elmore, J. S., Baumann, M. H., Phillips, T. J., and Janowsky, A. (2017) Trace amine-associated receptor 1 regulation of methamphetamine-induced neurotoxicity. *Neurotoxicology* **63**, 57–69
8. Reed, C., Baba, H., Zhu, Z., Erk, J., Mootz, J. R., Varra, N. M., Williams, R. W., and Phillips, T. J. (2017) A Spontaneous Mutation in Taar1 Impacts Methamphetamine-Related Traits Exclusively in DBA/2 Mice from a Single Vendor. *Front Pharmacol* **8**, 993
9. Shi, X., Walter, N. A. R., Harkness, J. H., Neve, K. A., Williams, R. W., Lu, L., Belknap, J. K., Eshleman, A. J., Phillips, T. J., and Janowsky, A. (2016) Genetic Polymorphisms Affect Mouse and Human Trace Amine-Associated Receptor 1 Function. *PLoS ONE* **11**, e0152581
10. Shi, X., Walter, N. A., Harkness, J. H., Neve, K. A., Williams, R. W., Lu, L., Belknap, J. K., Eshleman, A. J., Phillips, T. J., and Janowsky, A. (2016) Genetic Polymorphisms Affect Mouse and Human Trace Amine-Associated Receptor 1 Function. *PLoS One* **11**, e0152581
11. Loftis, J. M., Lasarev, M., Shi, X., Lapidus, J., Janowsky, A., Hoffman, W. F., and Huckans, M. (2019) Trace amine-associated receptor gene polymorphism increases drug craving in individuals with methamphetamine dependence. *PLoS ONE* **14**, e0220270
12. Nelson, E. C., Agrawal, A., Heath, A. C., Bogdan, R., Sherva, R., Zhang, B., Al-Hasani, R., Bruchas, M. R., Chou, Y. L., Demers, C. H., Carey, C. E., Conley, E. D., Fakira, A. K., Farrer, L. A., Goate, A., Gordon, S., Henders, A. K., Hesselbrock, V., Kapoor, M., Lynskey, M. T., Madden, P. A., Moron, J. A., Rice, J. P., Saccone, N. L., Schwab, S. G., Shand, F. L., Todorov, A. A., Wallace, L., Wang, T., Wray, N. R., Zhou, X., Degenhardt, L., Martin, N. G., Hariri, A. R., Kranzler, H. R., Gelernter, J., Bierut, L. J., Clark, D. J., and

Montgomery, G. W. (2016) Evidence of CNIH3 involvement in opioid dependence. *Mol.Psychiatry* **21**, 608–614

13. Zhou, Z., Karlsson, C., Liang, T., Xiong, W., Kimura, M., Tapocik, J. D., Yuan, Q., Barbier, E., Feng, A., Flanigan, M., Augier, E., Enoch, M. A., Hodgkinson, C. A., Shen, P. H., Lovinger, D. M., Edenberg, H. J., Heilig, M., and Goldman, D. (2013) Loss of metabotropic glutamate receptor 2 escalates alcohol consumption. *Proc.Natl.Acad.Sci.U.S.A.* **110**, 16963–16968
14. Hovatta, I., Tennant, R. S., Helton, R., Marr, R. A., Singer, O., Redwine, J. M., Ellison, J. A., Schadt, E. E., Verma, I. M., Lockhart, D. J., and Barlow, C. (2005) Glyoxalase 1 and glutathione reductase 1 regulate anxiety in mice. *Nature* **438**, 662–666
15. Williams, R., Lim, J. E., Harr, B., Wing, C., Walters, R., Distler, M. G., Teschke, M., Wu, C., Wiltshire, T., Su, A. I., Sokoloff, G., Tarantino, L. M., Borevitz, J. O., and Palmer, A. A. (2009) A common and unstable copy number variant is associated with differences in Glo1 expression and anxiety-like behavior. *PLoS ONE* **4**, e4649
16. Distler, M. G., Plant, L. D., Sokoloff, G., Hawk, A. J., Aneas, I., Wuenschell, G. E., Termini, J., Meredith, S. C., Nobrega, M. A., and Palmer, A. A. (2012) Glyoxalase 1 increases anxiety by reducing GABAA receptor agonist methylglyoxal. *J. Clin. Invest.* **122**, 2306–2315
17. McMurray, K. M. J., Du, X., Brownlee, M., and Palmer, A. A. (2016) Neuronal overexpression of Glo1 or amygdalar microinjection of methylglyoxal is sufficient to regulate anxiety-like behavior in mice. *Behav. Brain Res.* **301**, 119–123
18. de Guglielmo, G., Conlisk, D. E., Barkley-Levenson, A. M., Palmer, A. A., and George, O. (2018) Inhibition of Glyoxalase 1 reduces alcohol self-administration in dependent and nondependent rats. *Pharmacol. Biochem. Behav.* **167**, 36–41
19. McMurray, K. M. J., Sidhu, P. S., Cook, J. M., Arnold, L. A., and Palmer, A. A. (2017) Genetic and pharmacological manipulation of glyoxalase 1 regulates voluntary ethanol consumption in mice. *Addict Biol* **22**, 381–389
20. McMurray, K. M. J., Ramaker, M. J., Barkley-Levenson, A. M., Sidhu, P. S., Elkin, P. K., Reddy, M. K., Guthrie, M. L., Cook, J. M., Rawal, V. H., Arnold, L. A., Dulawa, S. C., and Palmer, A. A. (2018) Identification of a novel, fast-acting GABAergic antidepressant. *Mol. Psychiatry* **23**, 384–391
21. Distler, M. G., Gorfinkle, N., Papale, L. A., Wuenschell, G. E., Termini, J., Escayg, A., Winawer, M. R., and Palmer, A. A. (2013) Glyoxalase 1 and its substrate methylglyoxal are novel regulators of seizure susceptibility. *Epilepsia* **54**, 649–657
22. Parker, C. C., Cheng, R., Sokoloff, G., and Palmer, A. A. (2012) Genome-wide association for methamphetamine sensitivity in an advanced intercross mouse line. *Genes Brain Behav.* **11**, 52–61
23. Iakoubova, O. A., Olsson, C. L., Dains, K. M., Ross, D. A., Andalibi, A., Lau, K., Choi, J., Kalcheva, I., Cunanan, M., Louie, J., Nimon, V., Machrus, M., Bentley, L. G., Beauheim, C., Silvey, S., Cavalcoli, J.,

Lusis, A. J., and West, D. B. (2001) Genome-tagged mice (GTM): two sets of genome-wide congenic strains. *Genomics* **74**, 89–104

24. Yazdani, N., Parker, C. C., Shen, Y., Reed, E. R., Guido, M. A., Kole, L. A., Kirkpatrick, S. L., Lim, J. E., Sokoloff, G., Cheng, R., Johnson, W. E., Palmer, A. A., and Bryant, C. D. (2015) Hnrnph1 Is A Quantitative Trait Gene for Methamphetamine Sensitivity. *PLoS Genet.* **11**, e1005713
25. Ruan, Q. T., Yazdani, N., Blum, B. C., Beierle, J. A., Lin, W., Coelho, M. A., Fultz, E. K., Healy, A. F., Shahin, J. R., Kandola, A. K., Luttik, K. P., Zheng, K., Smith, N. J., Cheung, J., Mortazavi, F., Apicco, D. J., Ragu Varman, D., Ramamoorthy, S., Ash, P. E. A., Rosene, D. L., Emili, A., Wolozin, B., Szumlinski, K. K., and Bryant, C. D. (2019) A mutation in Hnrnph1 that decreases methamphetamine-induced reinforcement, reward, and dopamine release and increases synaptosomal hnRNP H and mitochondrial proteins. *J. Neurosci.*
26. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat.Protoc.* **7**, 562–578
27. Anders, S., Pyl, P. T., and Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169
28. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47
29. Smyth, G. K., Michaud, J., and Scott, H. S. (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**, 2067–2075
30. Anders, S., Reyes, A., and Huber, W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–2017
31. Schmittgen, T. D. and Livak, K. J. (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nature protocols* **3**, 1101–1108
32. Chiocco, M. J., Zhu, X., Walther, D., Pletnikova, O., Troncoso, J. C., Uhl, G. R., and Liu, Q. R. (2010) Fine mapping of calcineurin (PPP3CA) gene reveals novel alternative splicing patterns, association of 5' UTR trinucleotide repeat with addiction vulnerability, and differential isoform expression in Alzheimer's disease. *Subst.Use Misuse* **45**, 1809–1826
33. Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., Furlotte, N. A., Eskin, E., Nellaker, C., Whitley, H., Cleak, J., Janowitz, D., Hernandez-Pliego, P., Edwards, A., Belgard, T. G., Oliver, P. L., McIntyre, R. E., Bhomra, A., Nicod, J., Gan, X., Yuan, W., van der Weyden, L., Steward, C. A., Bala, S., Stalker, J., Mott, R., Durbin, R., Jackson, I. J., Czechanski, A., Guerra-Assuncao, J. A., Donahue, L. R., Reinholdt, L. G., Payseur, B. A., Ponting, C. P., Birney, E., Flint, J., and Adams, D. J. (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294

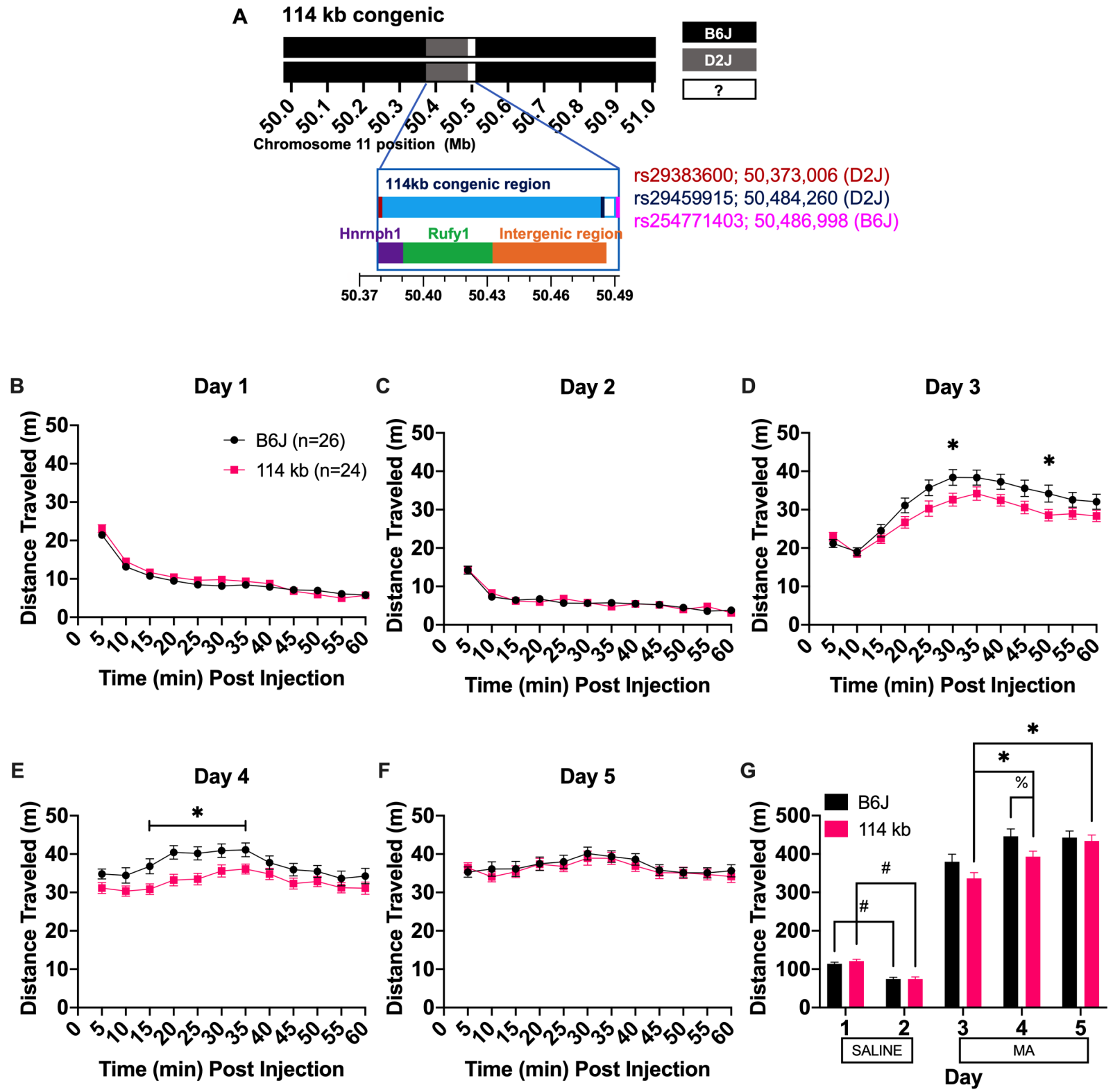
34. Yalcin, B., Wong, K., Agam, A., Goodson, M., Keane, T. M., Gan, X., Nellaker, C., Goodstadt, L., Nicod, J., Bhomra, A., Hernandez-Pliego, P., Whitley, H., Cleak, J., Dutton, R., Janowitz, D., Mott, R., Adams, D. J., and Flint, J. (2011) Sequence-based characterization of structural variation in the mouse genome. *Nature* **477**, 326–329
35. Yalcin, B. and Flint, J. (2012) Association studies in outbred mice in a new era of full-genome sequencing. *Mamm.Genome*
36. Bryant, C. D., Ferris, M. T., De Villena,F. P. M., Damaj, M. I., Kumar, V., and Mulligan, M. K. (2018) Reduced complexity cross design for behavioral genetics. In *Molecular-Genetic and Statistical Techniques for Behavioral and Neural Research* (Gerlai, R. T., ed) pp. 165–190,
37. Mulligan, M. K., Abreo, T., Neuner, S. M., Parks, C., Watkins, C. E., Houseal, M. T., Shapaker, T. M., Hook, M., Tan, H., Wang, X., Ingels, J., Peng, J., Lu, L., Kaczorowski, C. C., Bryant, C. D., Homanics, G. E., and Williams, R. W. (2019) Identification of a functional non-coding variant in the GABAA receptor α2 subunit of the C57BL/6J mouse reference genome: Major implications for neuroscience research. *Frontiers in Genetics* 540211
38. Kumar, V., Kim, K., Joseph, C., Kourrich, S., Yoo, S. H., Huang, H. C., Vitaterna, M. H., de Villena, F. P., Churchill, G., Bonci, A., and Takahashi, J. S. (2013) C57BL/6N mutation in Cytoplasmic FMRP interacting protein 2 regulates cocaine response. *Science* **342**, 1508–1512
39. Stafford, A. M., Reed, C., Baba, H., Walter, N. A., Mootz, J. R., Williams, R. W., Neve, K. A., Fedorov, L. M., Janowsky, A. J., and Phillips, T. J. (2019) Taar1 gene variants have a causal role in methamphetamine intake and response and interact with Oprm1. *Elife* **8**
40. Xu, J., Lu, Z., Xu, M., Pan, L., Deng, Y., Xie, X., Liu, H., Ding, S., Hurd, Y. L., Pasternak, G. W., Klein, R. J., Cartegni, L., Zhou, W., and Pan, Y. X. (2014) A heroin addiction severity-associated intronic single nucleotide polymorphism modulates alternative pre-mRNA splicing of the mu opioid receptor gene OPRM1 via hnRNPH interactions. *J.Neurosci.* **34**, 11048–11066
41. Steri, M., Idda, M. L., Whalen, M. B., and Orrù, V. (2018) Genetic variants in mRNA untranslated regions. *Wiley Interdiscip Rev RNA* **9**, e1474
42. Dassi, E. (2017) Handshakes and Fights: The Regulatory Interplay of RNA-Binding Proteins. *Front Mol Biosci* **4**, 67
43. Oliver, C. J. and Shenolikar, S. (1998) Physiologic importance of protein phosphatase inhibitors. *Front. Biosci.* **3**, D961-972
44. Forero, D. A. and González-Giraldo, Y. (2019) Convergent functional genomics of cocaine misuse in humans and animal models. *Am J Drug Alcohol Abuse* 1–9
45. Myers, C. T., Stong, N., Mountier, E. I., Helbig, K. L., Freytag, S., Sullivan, J. E., Ben Zeev, B., Nissenkorn, A., Tzadok, M., Heimer, G., Shinde, D. N., Rezazadeh, A., Regan, B. M., Oliver, K. L., Ernst, M. E., Lippa, N. C., Mulhern, M. S., Ren, Z., Poduri, A., Andrade, D. M., Bird, L. M., Bahlo, M., Berkovic, S. F.,

Lowenstein, D. H., Scheffer, I. E., Sadleir, L. G., Goldstein, D. B., Mefford, H. C., and Heinzen, E. L. (2017) De Novo Mutations in PPP3CA Cause Severe Neurodevelopmental Disease with Seizures. *Am. J. Hum. Genet.* **101**, 516–524

46. Bryant, C. D. and Yazdani, N. (2016) RNA-binding proteins, neural development and the addictions. *Genes Brain Behav.* **15**, 169–186
47. Aichem, A. and Mutzel, R. (2001) Unconventional mRNA processing in the expression of two calcineurin B isoforms in Dictyostelium. *J. Mol. Biol.* **308**, 873–882
48. Nishi, A., Snyder, G. L., and Greengard, P. (1997) Bidirectional regulation of DARPP-32 phosphorylation by dopamine. *J Neurosci* **17**, 8147–8155
49. Nishi, A., Bibb, J. A., Snyder, G. L., Higashi, H., Nairn, A. C., and Greengard, P. (2000) Amplification of dopaminergic signaling by a positive feedback loop. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12840–12845
50. Lin, X.-H., Hashimoto, T., Kitamura, N., Murakami, N., Shirakawa, O., and Maeda, K. (2002) Decreased calcineurin and increased phosphothreonine-DARPP-32 in the striatum of rats behaviorally sensitized to methamphetamine. *Synapse* **44**, 181–187
51. Ide, S., Kobayashi, H., Tanaka, K., Ujike, H., Sekine, Y., Ozaki, N., Inada, T., Harano, M., Komiyama, T., Yamada, M., Iyo, M., Ikeda, K., and Sora, I. (2004) Gene polymorphisms of the mu opioid receptor in methamphetamine abusers. *Ann.N.Y.Acad.Sci.* **1025**, 316–324
52. Wang, J., Sun, L.-L., Zhu, W.-L., Sun, Y., Liu, J.-F., Lu, L., and Shi, J. (2012) Role of calcineurin in the VTA in rats behaviorally sensitized to methamphetamine. *Psychopharmacology (Berl.)* **220**, 117–128
53. Yu, Y.-J., Huang, C.-H., Chang, C.-H., and Gean, P.-W. (2016) Involvement of protein phosphatases in the destabilization of methamphetamine-associated contextual memory. *Learn. Mem.* **23**, 486–493
54. Jayanthi, S., Deng, X., Ladenheim, B., McCoy, M. T., Cluster, A., Cai, N.-S., and Cadet, J. L. (2005) Calcineurin/NFAT-induced up-regulation of the Fas ligand/Fas death pathway is involved in methamphetamine-induced neuronal apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 868–873
55. Koike, K., Hashimoto, K., Fukami, G., Okamura, N., Zhang, L., Ohgake, S., Koizumi, H., Matsuzawa, D., Kawamura, N., Shimizu, E., and Iyo, M. (2005) The immunophilin ligand FK506 protects against methamphetamine-induced dopaminergic neurotoxicity in mouse striatum. *Neuropharmacology* **48**, 391–397
56. Guha, M., Pan, H., Fang, J.-K., and Avadhani, N. G. (2009) Heterogeneous nuclear ribonucleoprotein A2 is a common transcriptional coactivator in the nuclear transcription response to mitochondrial respiratory stress. *Mol. Biol. Cell* **20**, 4107–4119
57. Guha, M., Tang, W., Sondheimer, N., and Avadhani, N. G. (2010) Role of calcineurin, hnRNPA2 and Akt in mitochondrial respiratory stress-mediated transcription activation of nuclear gene targets. *Biochim. Biophys. Acta* **1797**, 1055–1065

58. Panelli, D., Lorusso, F. P., Papa, F., Panelli, P., Stella, A., Caputi, M., Sardanelli, A. M., and Papa, S. (2013) The mechanism of alternative splicing of the X-linked NDUFB11 gene of the respiratory chain complex I, impact of rotenone treatment in neuroblastoma cells. *Biochim. Biophys. Acta* **1829**, 211–218
59. Rahman, M. A., Azuma, Y., Nasrin, F., Takeda, J., Nazim, M., Bin Ahsan, K., Masuda, A., Engel, A. G., and Ohno, K. (2015) SRSF1 and hnRNP H antagonistically regulate splicing of COLQ exon 16 in a congenital myasthenic syndrome. *Sci Rep* **5**, 13208
60. Lefave, C. V., Squatrito, M., Vorlova, S., Rocco, G. L., Brennan, C. W., Holland, E. C., Pan, Y.-X., and Cartegni, L. (2011) Splicing factor hnRNPH drives an oncogenic splicing switch in gliomas. *EMBO J.* **30**, 4084–4097
61. Huelga, S. C., Vu, A. Q., Arnold, J. D., Liang, T. Y., Liu, P. P., Yan, B. Y., Donohue, J. P., Shiue, L., Hoon, S., Brenner, S., Ares, M., Jr, and Yeo, G. W. (2012) Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell.Rep.* **1**, 167–178

# Figure 1


A
114 kb congenic
B6J
D2J
?
50.0 50.1 50.2 50.3 50.4 50.5 50.6 50.7 50.8 50.9 51.0
Chromosome 11 position (Mb)
114kb congenic region
Hnrnph1
Rufy1
Intergenic region
50.37 50.40 50.43 50.46 50.49
rs29383600; 50,373,006 (D2J)
rs29459915; 50,484,260 (D2J)
rs254771403; 50,486,998 (B6J)
B
Day 1
C
Day 2
D
Day 3
E
Day 4
F
Day 5
G
Distance Traveled (m)
Time (min) Post Injection
B6J (n=26)
114 kb (n=24)
B6J
114 kb
SALINE
MA
Day

# Figure 2

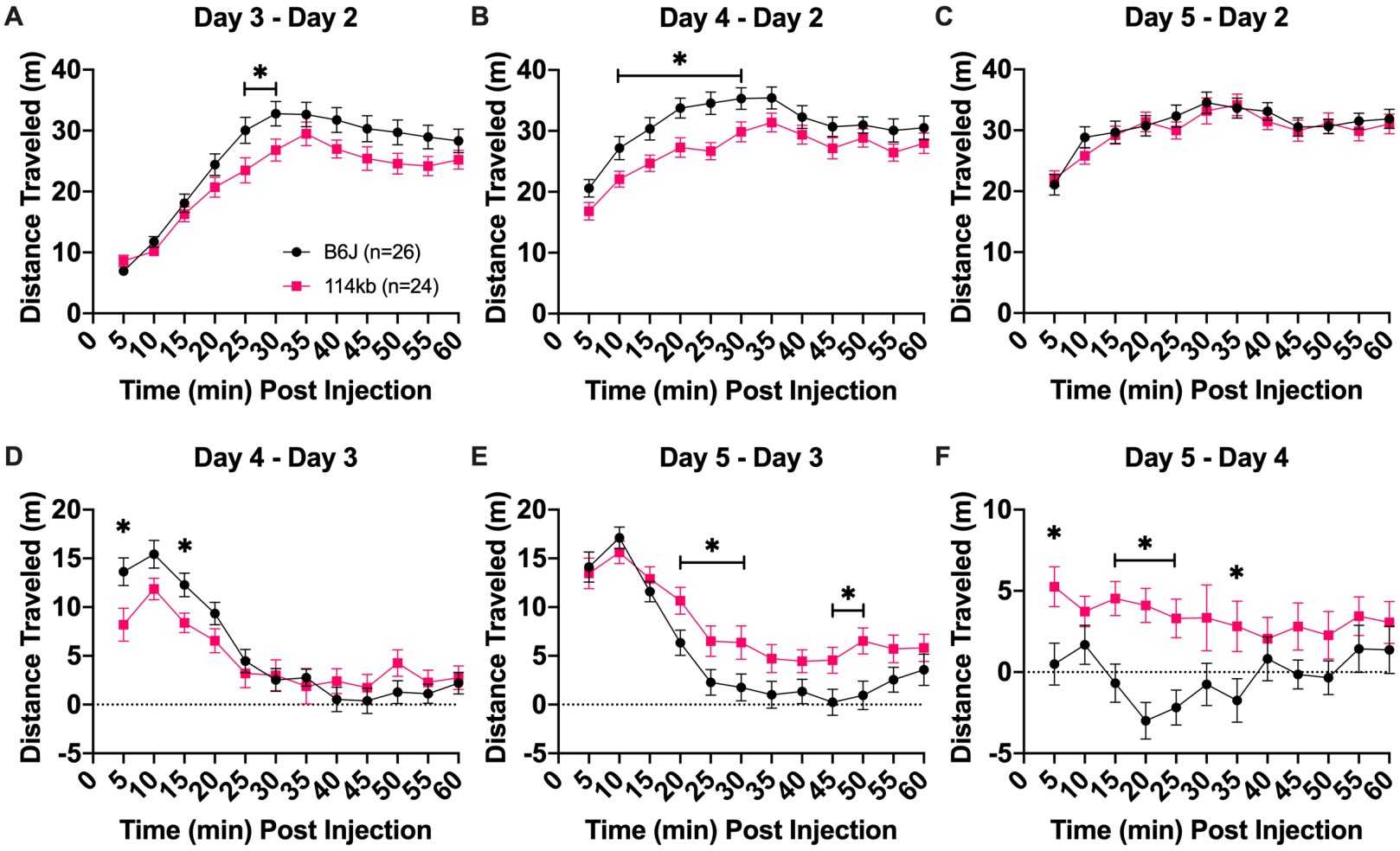

A
Day 3 - Day 2
B
Day 4 - Day 2
C
Day 5 - Day 2
D
Day 4 - Day 3
E
Day 5 - Day 3
F
Day 5 - Day 4
Distance Traveled (m)
Time (min) Post Injection
B6J (n=26)
114kb (n=24)
*

**Table 1. Differentially expressed genes between 114 kb and B6J.** The table shows the 69 genes that are differentially expressed in the striatum of 114 kb relative to the B6J wild-type littermates ($p < 0.001$).

| Gene | logFC | AveExpr | t | p value | Adjusted p value | B |
|---|---|---|---|---|---|---|
| Baiap3 | -0.534 | 4.730 | -5.500 | 1.14E-06 | 0.0162 | 5.180 |
| Hap1 | -0.302 | 7.192 | -5.065 | 5.36E-06 | 0.0271 | 3.919 |
| Wdr6 | -0.348 | 5.830 | -5.047 | 5.72E-06 | 0.0271 | 3.856 |
| Zcchc12 | -0.225 | 7.100 | -4.726 | 1.75E-05 | 0.0623 | 2.815 |
| Trpc4 | -0.562 | 3.444 | -4.556 | 3.13E-05 | 0.0713 | 1.838 |
| Tmem130 | -0.270 | 7.087 | -4.545 | 3.26E-05 | 0.0713 | 2.237 |
| Prkcd | -0.524 | 4.363 | -4.523 | 3.51E-05 | 0.0713 | 2.076 |
| Trp53inp2 | 0.183 | 8.099 | 4.399 | 5.33E-05 | 0.0857 | 1.752 |
| Wnk1 | 0.123 | 8.451 | 4.347 | 6.35E-05 | 0.0857 | 1.578 |
| Scd1 | 0.184 | 7.767 | 4.338 | 6.55E-05 | 0.0857 | 1.568 |
| Zic1 | -0.630 | 4.393 | -4.317 | 7.03E-05 | 0.0857 | 1.482 |
| Cacng5 | -0.454 | 3.203 | -4.291 | 7.65E-05 | 0.0857 | 1.005 |
| Col6a3 | -0.892 | 2.255 | -4.259 | 8.51E-05 | 0.0857 | 0.249 |
| Kcnj2 | 0.201 | 6.042 | 4.259 | 8.53E-05 | 0.0857 | 1.378 |
| Ecel1 | -0.187 | 5.713 | -4.203 | 0.000102 | 0.0857 | 1.220 |
| Fos | 0.994 | 2.801 | 4.201 | 0.000103 | 0.0857 | 0.773 |
| Lbhd2 | -2.694 | -0.765 | -4.196 | 0.000105 | 0.0857 | -1.978 |
| Magel2 | -0.704 | 1.965 | -4.186 | 0.000108 | 0.0857 | 0.001 |
| Zic2 | -0.692 | 2.885 | -4.118 | 0.000136 | 0.0986 | 0.386 |
| Osbpl8 | 0.115 | 8.486 | 4.099 | 0.000144 | 0.0986 | 0.808 |
| Qk | 0.152 | 8.422 | 4.097 | 0.000145 | 0.0986 | 0.805 |
| Gatm | 0.193 | 7.062 | 4.056 | 0.000166 | 0.1074 | 0.730 |
| Nrp2 | -0.391 | 3.477 | -4.005 | 0.000196 | 0.1127 | 0.423 |
| Itm2c | -0.089 | 10.023 | -4.001 | 0.000198 | 0.1127 | 0.461 |
| Ezr | -0.170 | 6.308 | -3.982 | 0.000211 | 0.1127 | 0.537 |
| Zic4 | -0.971 | 1.787 | -3.980 | 0.000213 | 0.1127 | -0.702 |
| Ermn | 0.240 | 6.924 | 3.978 | 0.000214 | 0.1127 | 0.502 |
| Zcchc18 | -0.103 | 8.156 | -3.954 | 0.000231 | 0.1175 | 0.381 |
| Pnmal2 | -0.226 | 7.963 | -3.925 | 0.000253 | 0.1216 | 0.300 |
| Fth1 | 0.169 | 10.805 | 3.922 | 0.000256 | 0.1216 | 0.196 |
| Dlx1 | -0.701 | 3.098 | -3.895 | 0.000279 | 0.1252 | -0.051 |
| Icam5 | -0.104 | 8.147 | -3.891 | 0.000283 | 0.1252 | 0.195 |
| Sept7 | 0.107 | 8.420 | 3.883 | 0.000290 | 0.1252 | 0.162 |
| Pcdh8 | -0.391 | 3.388 | -3.853 | 0.000319 | 0.1337 | 0.027 |
| Ttll7 | 0.104 | 7.896 | 3.840 | 0.000332 | 0.1346 | 0.055 |
| Rasd1 | -0.842 | 1.393 | -3.830 | 0.000344 | 0.1346 | -1.180 |
| Tspan2 | 0.175 | 7.668 | 3.824 | 0.000350 | 0.1346 | 0.019 |
| Slc22a17 | -0.120 | 9.192 | -3.810 | 0.000366 | 0.1371 | -0.082 |
| Gpm6b | 0.091 | 10.728 | 3.777 | 0.000406 | 0.1378 | -0.233 |
| Tppp | 0.091 | 9.176 | 3.775 | 0.000409 | 0.1378 | -0.184 |
| Cadm2 | 0.134 | 7.765 | 3.764 | 0.000423 | 0.1378 | -0.161 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Gprasp1 | -0.102 | 9.355 | -3.762 | 0.000426 | 0.1378 | -0.230 |
| Adcy1 | 0.111 | 8.611 | 3.761 | 0.000426 | 0.1378 | -0.201 |
| Npas4 | 0.694 | 4.629 | 3.758 | 0.000431 | 0.1378 | -0.052 |
| Rapgef4 | 0.146 | 9.876 | 3.744 | 0.000451 | 0.1378 | -0.299 |
| Tro | -0.103 | 7.983 | -3.744 | 0.000451 | 0.1378 | -0.233 |
| Grin2a | 0.232 | 4.844 | 3.736 | 0.000462 | 0.1378 | -0.111 |
| Scn5a | -0.522 | 2.614 | -3.734 | 0.000465 | 0.1378 | -0.572 |
| Maged1 | -0.152 | 8.935 | -3.694 | 0.000526 | 0.1528 | -0.410 |
| Kcna2 | 0.109 | 8.165 | 3.684 | 0.000543 | 0.1547 | -0.410 |
| Ano3 | 0.105 | 9.169 | 3.668 | 0.000571 | 0.1594 | -0.493 |
| Slit1 | -0.361 | 4.198 | -3.652 | 0.000601 | 0.1645 | -0.361 |
| Oip5os1 | 0.146 | 6.009 | 3.621 | 0.000661 | 0.1747 | -0.483 |
| Lpgat1 | 0.090 | 8.070 | 3.620 | 0.000663 | 0.1747 | -0.589 |
| Ptprd | 0.143 | 7.600 | 3.602 | 0.000700 | 0.1812 | -0.619 |
| Ngb | -0.792 | 1.543 | -3.596 | 0.000714 | 0.1815 | -1.419 |
| Rgs7bp | 0.118 | 9.071 | 3.589 | 0.000729 | 0.1821 | -0.716 |
| Camk4 | 0.101 | 9.316 | 3.583 | 0.000744 | 0.1825 | -0.743 |
| Thsd7a | 0.117 | 7.467 | 3.561 | 0.000793 | 0.1892 | -0.728 |
| Hipk2 | 0.191 | 5.988 | 3.560 | 0.000798 | 0.1892 | -0.651 |
| Pnck | -0.209 | 5.758 | -3.542 | 0.000843 | 0.1968 | -0.686 |
| Camk2n1 | 0.106 | 10.768 | 3.519 | 0.000904 | 0.2022 | -0.979 |
| Ret | -0.589 | 2.081 | -3.516 | 0.000911 | 0.2022 | -1.307 |
| Zeb2 | 0.142 | 6.285 | 3.512 | 0.000922 | 0.2022 | -0.798 |
| Bcas1 | 0.185 | 8.413 | 3.507 | 0.000937 | 0.2022 | -0.920 |
| Samd5 | 0.260 | 3.992 | 3.507 | 0.000937 | 0.2022 | -0.747 |
| S100b | 0.206 | 7.036 | 3.498 | 0.000964 | 0.2033 | -0.884 |
| Gpc3 | -0.594 | 1.672 | -3.495 | 0.000971 | 0.2033 | -1.579 |
| Gpi1 | -0.136 | 9.102 | -3.488 | 0.000991 | 0.2045 | -1.003 |

**Table 2. Genes exhibiting differential exon usage between 114 kb and B6J. (A - B):** The table shows genes exhibiting differential exon usage in striatal tissue from 114 kb mice vs. their B6J wild-type littermates. Differential exon usage was detected using either an F-test (**A**) or a Simes test (**B**) in limma.

| **(A): F-test (p < 0.001)** | | | | | |
|---|---|---|---|---|---|
| **Gene** | **Fstat** | **Fstat_P** | **Fstat_FDR** | **Simes_P** | **Simes_FDR** |
| Ppp3ca | 2.87 | 5.12E-07 | 0.0092 | 1.30E-09 | 2.30E-05 |
| Cdhr4, Ip6k1, Uba7, Gm20661, Gm20662 | 1.65 | 1.36E-06 | 0.012 | 0.0015 | 1 |
| Hnrnph1 | 2.19 | 7.50E-06 | 0.045 | 0.0055 | 1 |
| Mrps14 | 4.92 | 2.67E-05 | 0.1 | 4.10E-05 | 0.19 |
| Pfdn5, Myg1 | 2.1 | 2.89E-05 | 0.1 | 0.028 | 1 |
| Scn3a | 2.14 | 4.26E-05 | 0.13 | 0.002 | 1 |
| Chn2 | 2.53 | 5.58E-05 | 0.14 | 0.026 | 1 |
| Myo5b | 1.89 | 7.13E-05 | 0.16 | 0.036 | 1 |
| Mapk9 | 2.33 | 1.20E-04 | 0.21 | 0.022 | 1 |
| Calm2 | 3.6 | 1.30E-04 | 0.21 | 0.0024 | 1 |
| Hnrnpm, March2 | 1.83 | 1.40E-04 | 0.21 | 0.014 | 1 |
| Cdc42bpa | 1.75 | 1.40E-04 | 0.21 | 0.18 | 1 |
| Arrdc1 | 2.11 | 1.90E-04 | 0.26 | 0.21 | 1 |
| Igbp1 | 3.98 | 3.50E-04 | 0.43 | 0.014 | 1 |
| Sptbn1 | 1.84 | 3.60E-04 | 0.43 | 0.017 | 1 |
| Tm4sf1 | 3.64 | 4.30E-04 | 0.49 | 0.018 | 1 |
| Ptpn5 | 1.94 | 5.60E-04 | 0.6 | 0.2 | 1 |
| **(B): Simes (p < 0.001)** | | | | | |
| **Gene** | **Fstat** | **Fstat_P** | **Fstat_FDR** | **Simes_P** | **Simes_FDR** |
| Ppp3ca | 2.87 | 5.12E-07 | 0.0092 | 1.30E-09 | 2.30E-05 |
| Tspan32 | 2.96 | 0.0013 | 1 | 2.27E-05 | 0.19 |
| Mrps14 | 4.92 | 2.67E-05 | 0.1 | 4.09E-05 | 0.19 |
| Txn2 | 1.59 | 0.048 | 1 | 4.13E-05 | 0.19 |
| Map2 | 1.6 | 0.0018 | 1 | 1.50E-04 | 0.53 |
| Map4, Mtap4 | 1.57 | 0.007 | 1 | 1.80E-04 | 0.54 |
| Slc8a3 | 2.88 | 0.0011 | 1 | 2.20E-04 | 0.57 |
| Alpk3 | 1.76 | 0.052 | 1 | 3.80E-04 | 0.81 |
| Jaml | 2.74 | 0.006 | 1 | 4.10E-04 | 0.81 |
| Dph1, Ovca2 | 1.29 | 0.15 | 1 | 4.90E-04 | 0.87 |
| Rab21 | 2.69 | 0.15 | 1 | 5.40E-04 | 0.89 |
| Usp24 | 0.83 | 0.86 | 1 | 9.70E-04 | 1 |

Figure 3

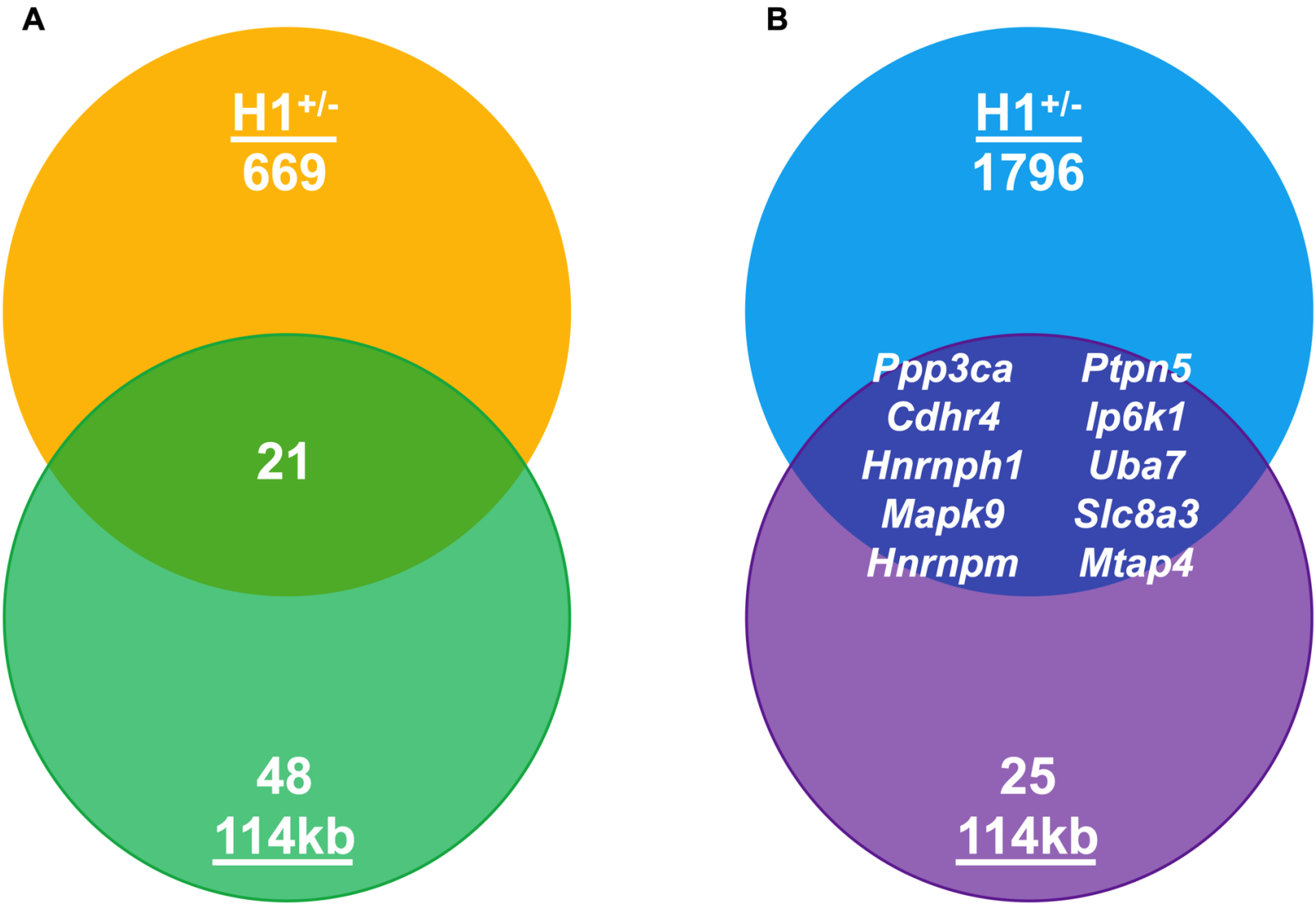

A
H1+/-
669
21
48
114kb
B
H1+/-
1796
Ppp3ca
Cdhr4
Hnrnph1
Mapk9
Hnrnpm
Ptpn5
Ip6k1
Uba7
Slc8a3
Mtap4
25
114kb

**Table 3. Correlation of differential expression between 114 kb mice and H1$^{+/-}$ mice.** Table showing the $log_2$FC values for the 21 overlapping genes in the 114 kb and H1$^{+/-}$. The change in gene expression (up/down) is relative to B6J wild-type littermates.

| Gene | 114 kb $log_2$FC | H1$^{+/-}$ $log_2$FC |
|---|---|---|
| Magel2 | -0.704 | -0.857 |
| Dlx1 | -0.701 | -0.686 |
| Zic2 | -0.692 | -0.811 |
| Zic1 | -0.630 | -0.897 |
| Trpc4 | -0.562 | -0.786 |
| Prkcd | -0.524 | -0.528 |
| Wdr6 | -0.348 | -0.420 |
| Hap1 | -0.302 | -0.380 |
| Zcchc12 | -0.225 | -0.297 |
| Pnck | -0.209 | -0.353 |
| Ecel1 | -0.187 | -0.596 |
| Zcchc18 | -0.103 | -0.154 |
| Camk4 | 0.101 | 0.311 |
| Ano3 | 0.105 | 0.212 |
| Kcna2 | 0.109 | 0.247 |
| Osbpl8 | 0.115 | 0.129 |
| Thsd7a | 0.117 | 0.303 |
| Rgs7bp | 0.118 | 0.335 |
| Wnk1 | 0.123 | 0.220 |
| Cadm2 | 0.134 | 0.192 |
| Oip5os1 | 0.146 | 0.369 |

Figure 4

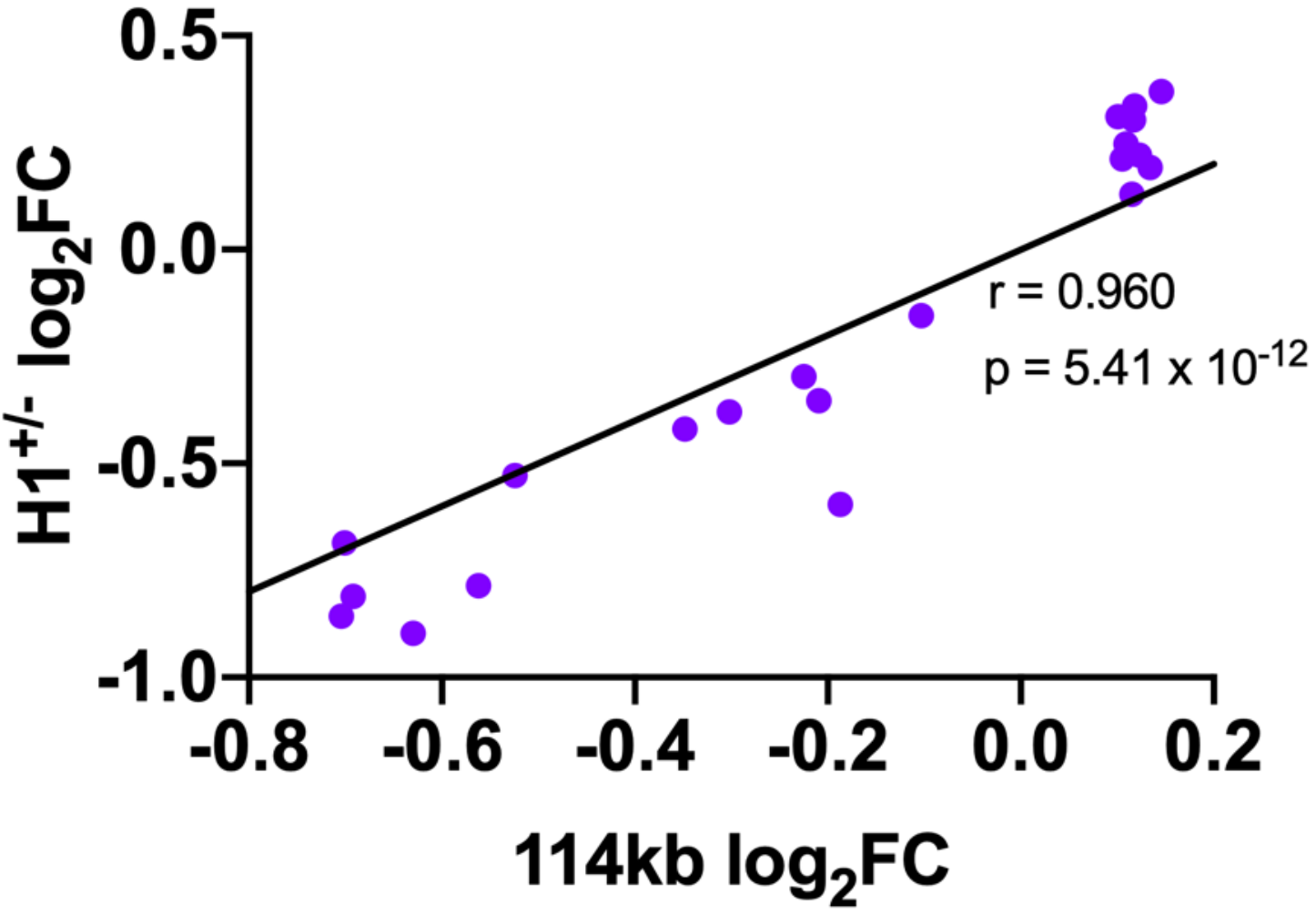

0.5
0.0
-0.5
-1.0
H1+/- log2FC
-0.8
-0.6
-0.4
-0.2
0.0
0.2
114kb log2FC
r = 0.960
p = 5.41 x 10-12

Figure 5

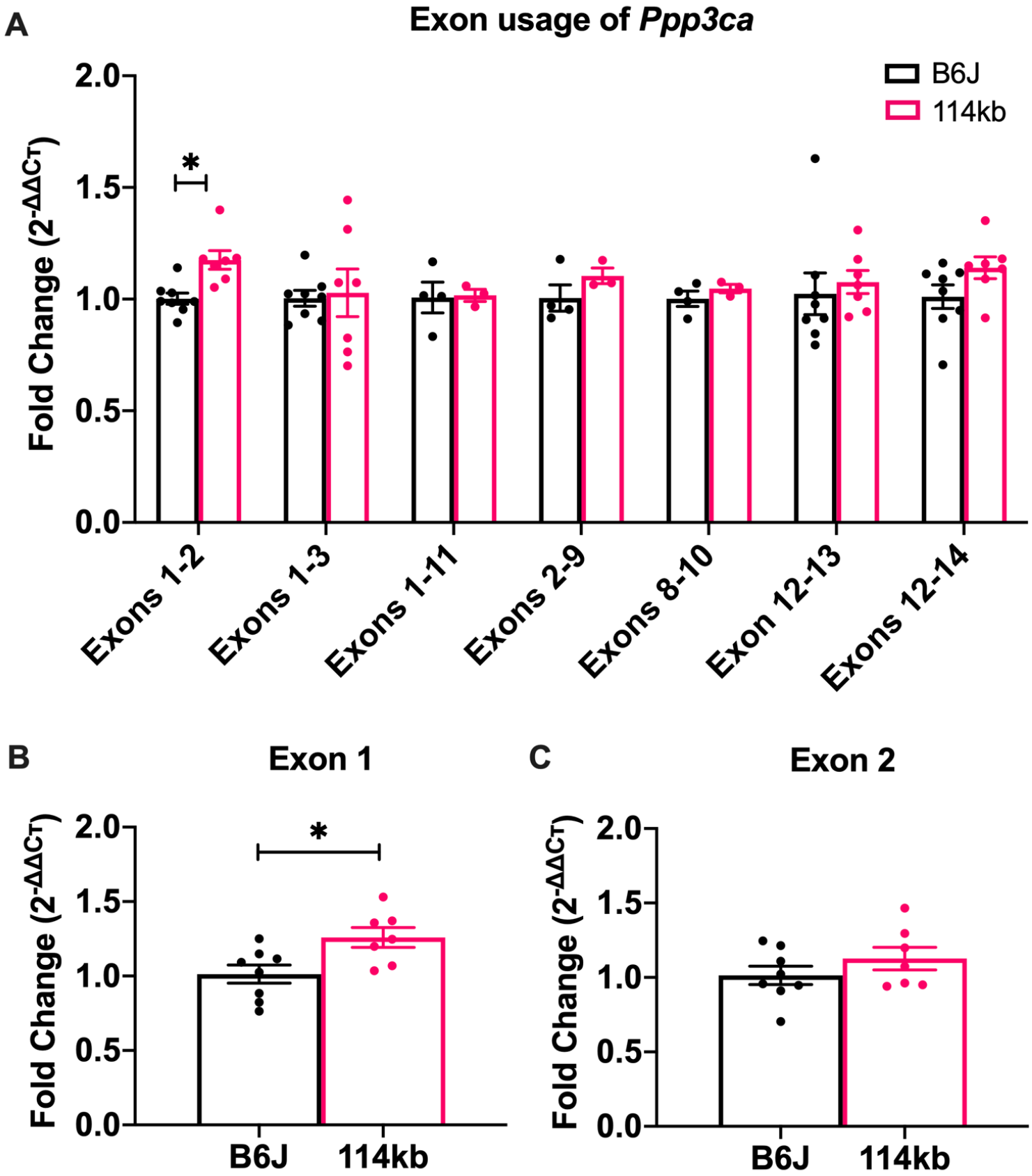

A
Exon usage of *Ppp3ca*
B6J
114kb
Fold Change ($2^{-\Delta\Delta C_T}$)
2.0
1.5
1.0
0.5
0.0
*
Exons 1-2
Exons 1-3
Exons 1-11
Exons 2-9
Exons 8-10
Exon 12-13
Exons 12-14
B
Exon 1
C
Exon 2
B6J
114kb

**Table 4. Exon usage of *Hnrnph1* in 114 kb mice.** Table showing the four exon positions of *Hnrnph1* are shown that display differential exon usage in 114 kb mice relative to B6J wild-type littermates with $p < 0.05$ using the Simes test. Exon usage was defined as the proportion of total normalized reads per gene that are counted within an exon bin for that gene.

| Position (mm10) | Corresponding exon junction | $log_2FC$ | p value | FDR |
|---|---|---|---|---|
| chr11:50,377,719-50,377,750 | 3-4 | -0.484 | 0.018 | 1 |
| chr11:50,380,191-50,380,825 | 6-7 | 0.137 | 0.001 | 1 |
| chr11:50,381,598-50,382,137 | 7-8 | 0.094 | 0.017 | 1 |
| chr11:50,382,138-50,382,536 | 7-8 | 0.167 | 3.80E-04 | 1 |

# Figure 6

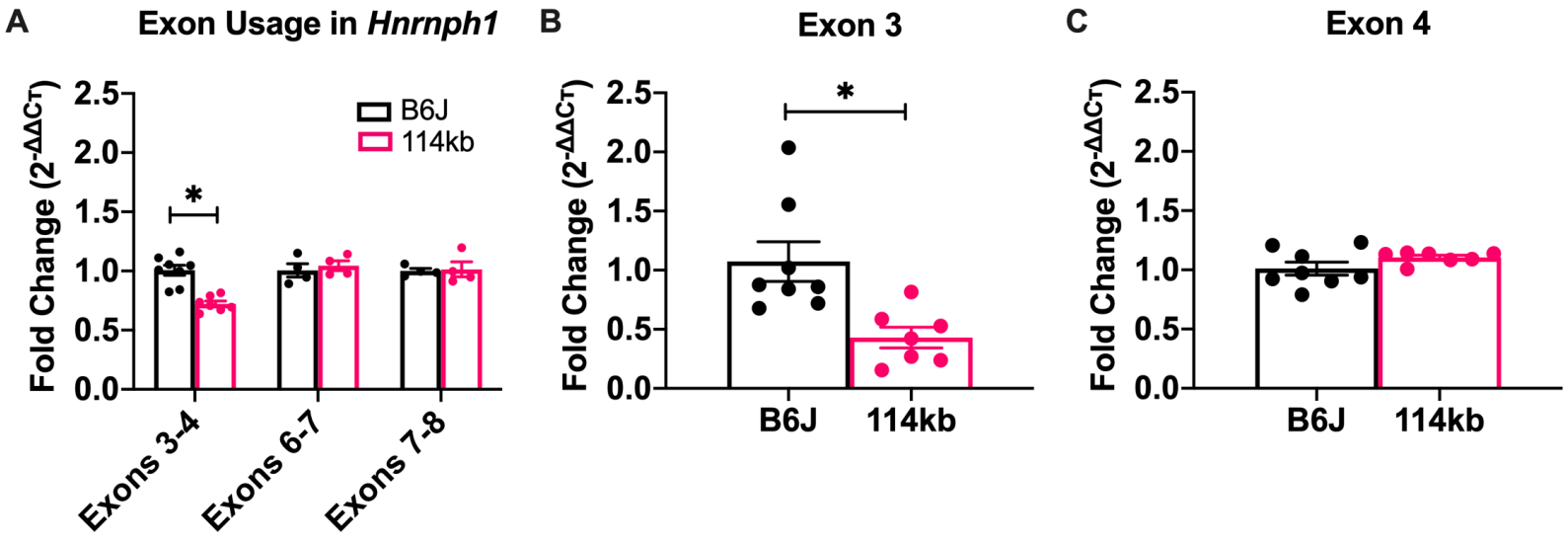
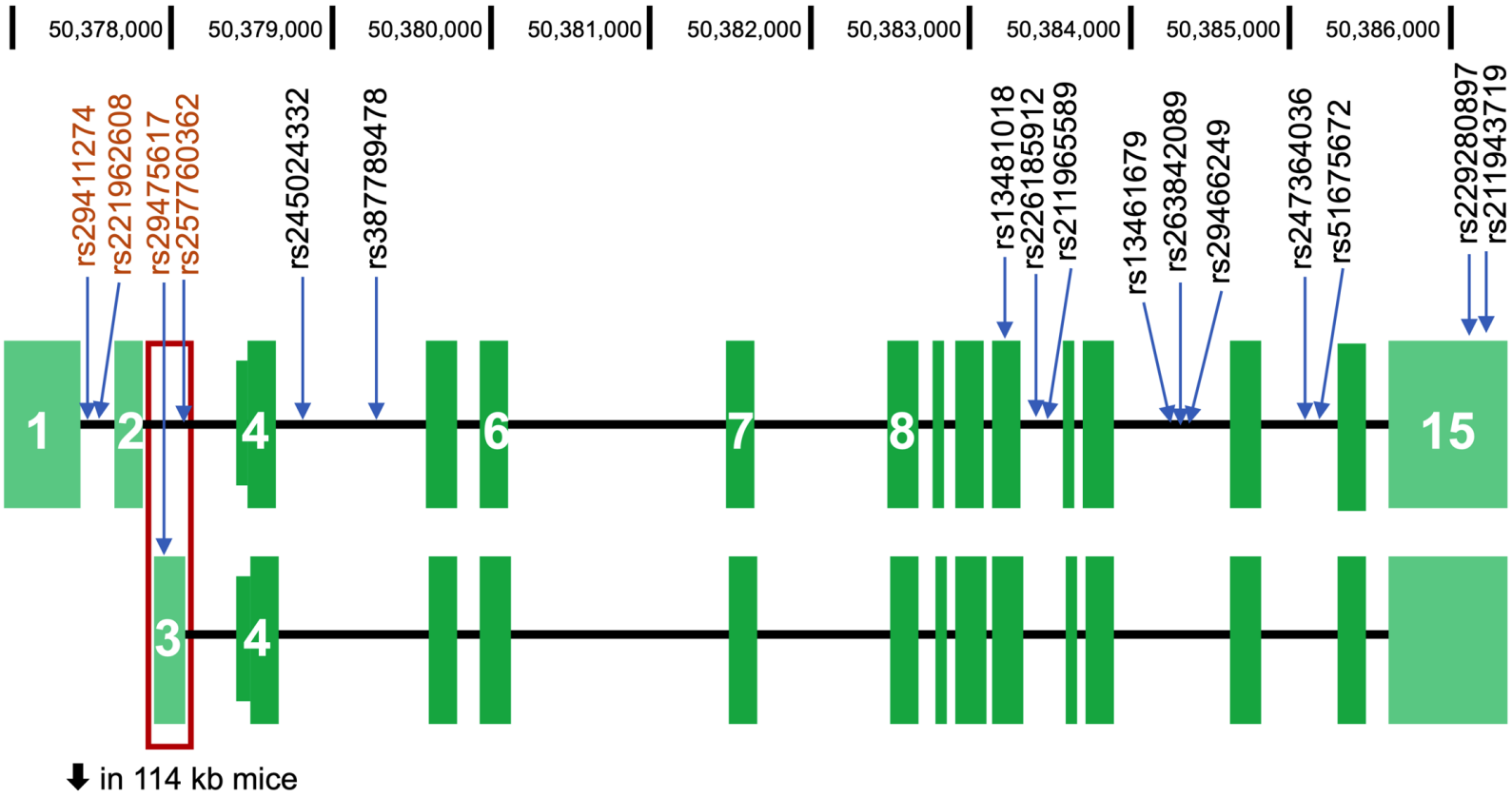

A
Exon Usage in *Hnrnph1*
B6J
114kb
Fold Change ($2^{-\Delta\Delta C_T}$)
2.5
2.0
1.5
1.0
0.5
0.0
*
Exons 3-4
Exons 6-7
Exons 7-8
B
Exon 3
*
B6J
114kb
C
Exon 4
B6J
114kb

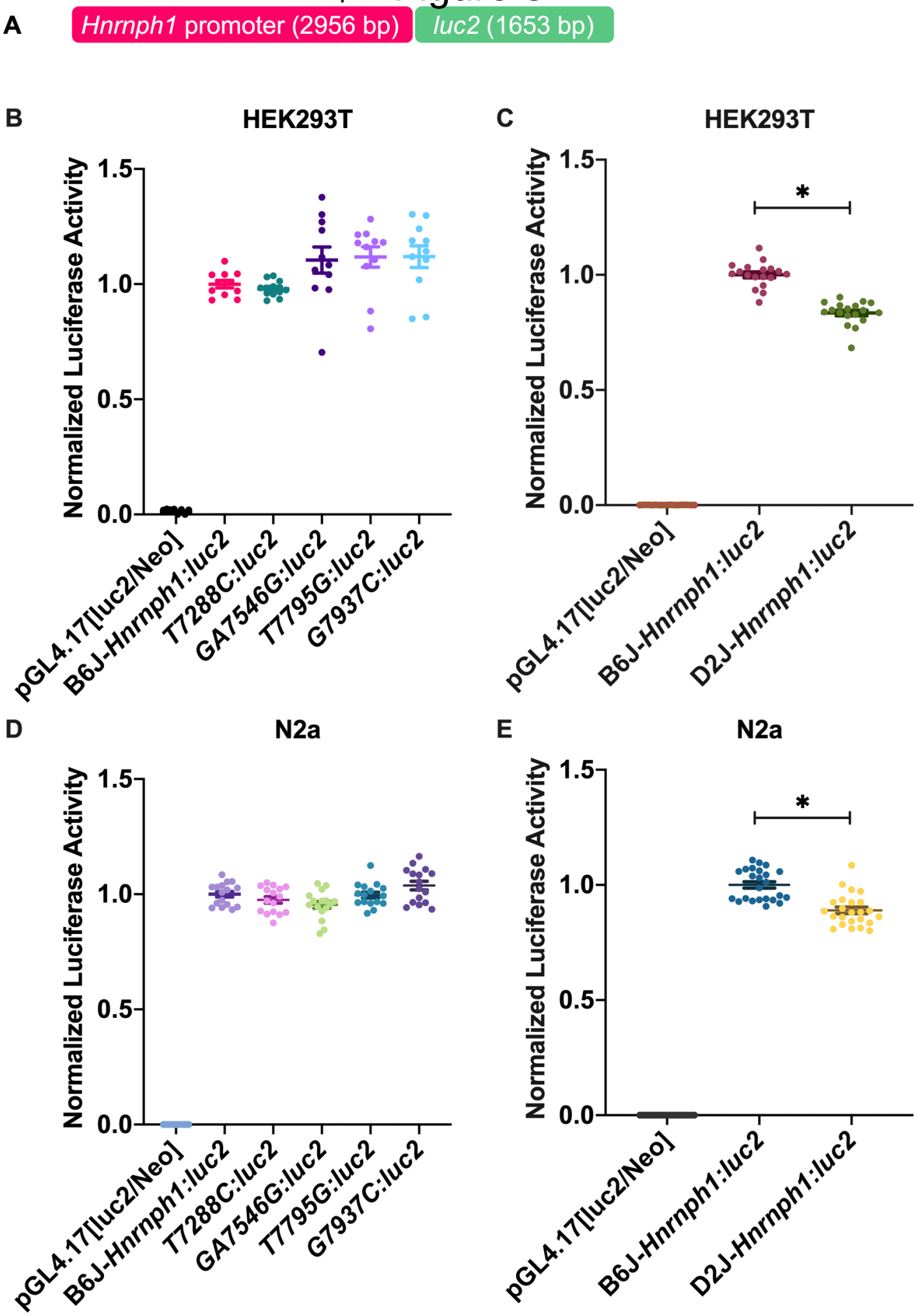
**Table 5. SNPs and indels in *Hnrnph1* between B6J and D2J.** A query for *Hnrnph1* genetic variants (SNPs plus indels) between the B6J and D2J parental strains was conducted using the Sanger database query tool (mm10, REL-1505). The variants highlighted in orange are located in close proximity to exon 3 and were cloned and tested functionally in the luciferase reporter assay.

| Position (mm10) | dbSNP | C57BL/6J | DBA/2J | Type |
|---|---|---|---|---|
| 50,377,288 | rs29411274 | T | C* | SNP; intron variant |
| 50,377,546 | rs221962608;rs224352813 | GA | G* | Indel; intron variant |
| 50,377,795 | rs29475617 | T | G* | SNP; 5'UTR variant |
| 50,377,937 | rs257760362 | G | C* | SNP; intron variant |
| 50,378,499 | rs245024332 | T | C* | SNP; intron variant |
| 50,379,377 | rs387789478;rs233140941 | GTT | GT* | Indel; intron variant |
| 50,383,262 | rs13481018 | C | T* | SNP; synonymous variant |
| 50,383,443 | rs226185912 | ATG | A* | Indel; intron variant |
| 50,383,487 | rs211965589 | T | TA* | Indel; intron variant |
| 50,384,116 | rs13461679 | G | A* | SNP; intron variant |
| 50,384,175 | rs263842089 | T | TAGATG* | Indel; intron variant |
| 50,384,378 | rs29466249 | A | T* | SNP; intron variant |
| 50,385,052 | rs247364036 | GTT | G* | Indel; intron variant |
| 50,385,066 | rs51675672 | G | T* | SNP; intron variant |
| 50,386,305 | rs229280897 | T | t/c* | SNP; 3'UTR variant |
| 50,386,341 | rs211943719;rs240064076 | CTT | CTTT* | Indel; 3'UTR variant |

# Figure 7


50,378,000
50,379,000
50,380,000
50,381,000
50,382,000
50,383,000
50,384,000
50,385,000
50,386,000
rs29411274
rs221962608
rs29475617
rs257760362
rs245024332
rs387789478
rs13481018
rs226185912
rs211965589
rs13461679
rs263842089
rs29466249
rs247364036
rs51675672
rs229280897
rs211943719
1
2
3
4
4
6
7
8
15
in 114 kb mice

# Figure 8


A
B6J-Hnrnph1:luc2
Hnrnph1 promoter (2956 bp)
luc2 (1653 bp)
B
HEK293T
Normalized Luciferase Activity
1.5
1.0
0.5
0.0
pGL4.17[luc2/Neo]
B6J-Hnrnph1:luc2
T7288C:luc2
GA7546G:luc2
T7795G:luc2
G7937C:luc2
C
HEK293T
Normalized Luciferase Activity
*
pGL4.17[luc2/Neo]
B6J-Hnrnph1:luc2
D2J-Hnrnph1:luc2
D
N2a
Normalized Luciferase Activity
pGL4.17[luc2/Neo]
B6J-Hnrnph1:luc2
T7288C:luc2
GA7546G:luc2
T7795G:luc2
G7937C:luc2
E
N2a
Normalized Luciferase Activity
*
pGL4.17[luc2/Neo]
B6J-Hnrnph1:luc2
D2J-Hnrnph1:luc2

**Supplemental Figure 1. DNA sequence of the cloned *Hnrnph1* promoter.** The promoter contains DNA sequence 2956 bp upstream of transcription start site which corresponds to chromosome 11 nucleotide position 50,375,375 to 50,378,330 (mm10). The nucleotides indicated in red are the positions of the SNPs, and the nucleotides indicated in blue are the positions of the indel.

**Supplemental Figure 2. *Hnrnph1:luc2* reporter assay in HEK293T cells. (A):** Schematic representation of the procedure for cloning the *Hnrnph1* promoter into pGL4.17[luc2/Neo]. The promoter region of *Hnrnph1* was PCR-amplified followed by restriction enzyme digest by *XhoI* and *HindIII* for insertion into pGL4.17[luc2/Neo]. *The Hnrnph1* promoter was defined as 2956 bp upstream of the transcriptional start site of *Hnrnph1* and was fused to the firefly luciferase (luc2) to make the *Hnrnph1:luc2* reporter. **(B):** Different numbers of HEK293T cells were seeded to determine whether the cloned *Hnrnph1* promoter could drive expression of firefly luciferase, *luc2*. Firefly luminescence increased as the cell number increased [two-way ANOVA: $F(3,16) = 64.19$, $p = 3.86E-09$; Bonferroni's multiple comparison test: $t(16)_{10,000} = 1.15$, $p > 1$; $t(16)_{20,000} = 2.94$, $*p = 0.005$; $t(16)_{30,000} = 8.35$, $*p = 1.25e-6$; $t(16)_{50,000} = 19.36$, $*p = 6.30e-12$]. Error bars are represented as mean ± SEM.

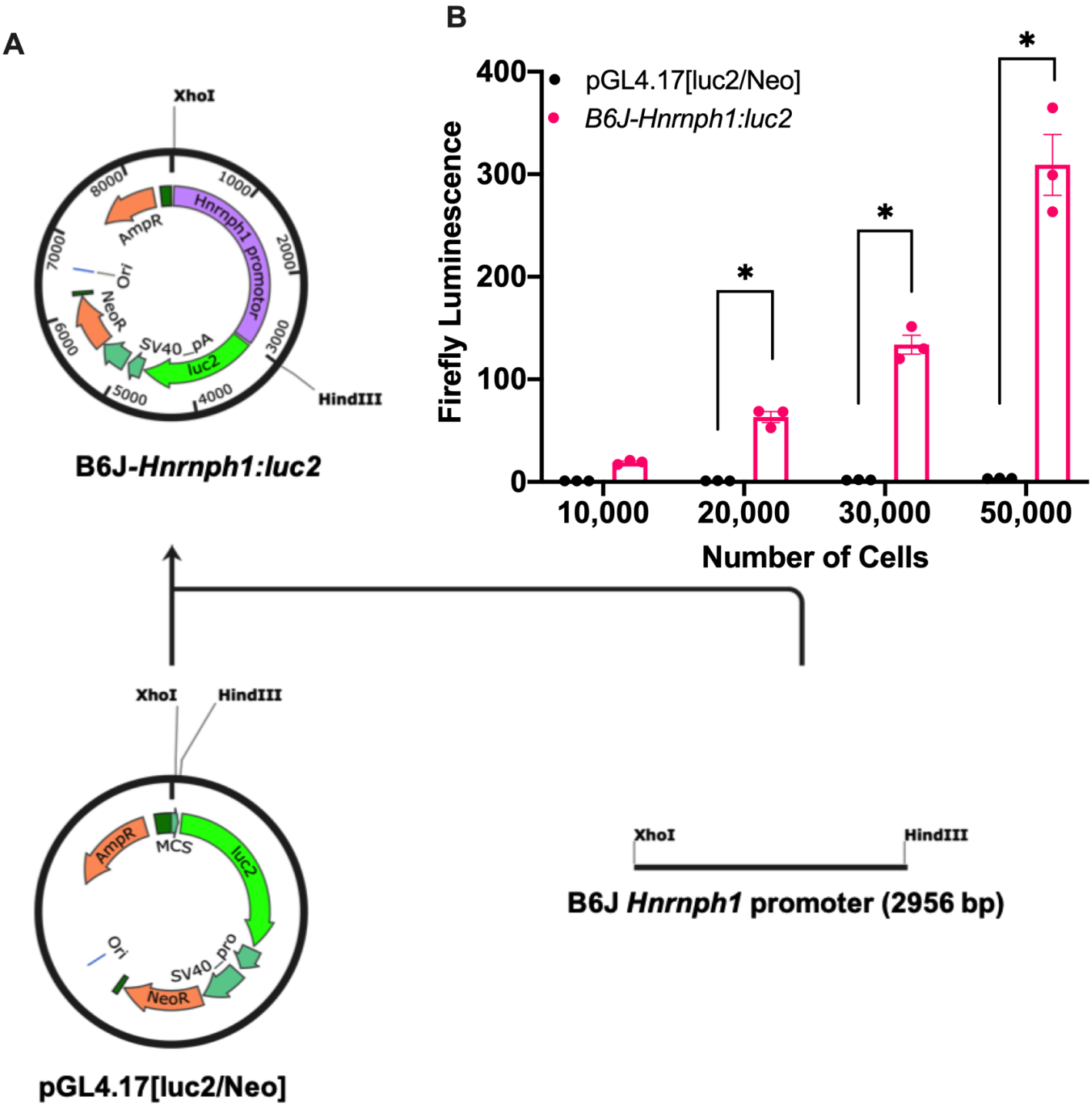
**Supplemental Table 1. RT-qPCR primers for exon usage.**

| Gene | Targeted exon(s) | Sequences |
|---|---|---|
| *Ppp3ca* | 1-2 | Forward: 5'-AGGCGATTGATCCCAAGTTGTC-3' |
| | | Reverse: 5'-CCGGTGACTTGGTGGAAATG-3' |
| *Ppp3ca* | 1-3 | Forward: 5'-AGGCGATTGATCCCAAGTTGTC-3' |
| | | Reverse: 5'-AAGTCAAAGAATTGTCCATGGATGTC-3' |
| *Ppp3ca* | 1-11 | Forward: 5'-AGGCGATTGATCCCAAGTTGTC-3' |
| | | Reverse: 5'-GATCTTGTTTCTGATGACTTCCTTCC-3' |
| *Ppp3ca* | 12-13 | Forward: 5'-CTCTCTGGCGGGAAACAGACT-3' |
| | | Reverse: 5'-TTGTGGTGAAAATCCTTTGATGGCT-3' |
| *Ppp3ca* | 12-14 | Forward: 5'-CTCTCTGGCGGGAAACAGACT-3' |
| | | Reverse: 5'-GTGATCTTATGTTGTGGTGAAAATCC-3' |
| *Ppp3ca* | 8-10 | Forward: 5'-ACAACAGGCTTCCCGTCTCTAATTAC-3' |
| | | Reverse: 5'-AGCAGATGTTGAGAACATTGACCA-3' |
| *Ppp3ca* | 2-9 | Forward: 5'-GCTTCGATTCTCCGACAGGAAAAA-3' |
| | | Reverse: 5'-GCTTCGATTCTCCGACAGGAAAAA-3' |
| *Ppp3ca* | 1 | Forward: 5'-CAGTAACTTTCGAGCCAGCC-3' |
| | | Reverse: 5'-TCGTCGACAACTTGGGATCA-3' |
| *Ppp3ca* | 2 | Forward: 5'-CATTTCCACCAAGTCACCGG-3' |
| | | Reverse: 5'-GTGCGTCGATATCCAGCAAG-3' |
| *Hnrnph1* | 3-4 | Forward: 5'-CGCCCGCTTATCATTTCTCC-3' |
| | | Reverse: 5'-CTTCACCACGAAGCCCTCT-3' |
| *Hnrnph1* | 6-7 | Forward: 5'-ACGGCTTAGAGGACTCCCTTT-3' |
| | | Reverse: 5'-CGTACTCCTCCCCTGGAAGT-3' |
| *Hnrnph1* | 7-8 | Forward: 5'-CAGGAAATAGCTGAAAAGGCTCTA-3' |
| | | Reverse: 5'-GCCATAAGTTTTCGTGGTGG-3' |
| *Hnrnph1* | 3 | Forward: 5'-AGCTTGCTGGAAGGGCTT-3' |
| | | Reverse: 5'-ACCTCCACAAAAAGCCTCG-3' |
| *Hnrnph1* | 4 | Forward: 5'-GACCACCGAAGGGACGAT-3' |
| | | Reverse: 5'-AAAAAGCGCTGCACTTCATC-3' |
| *GAPDH* | 5-6 | Forward: 5'-GCCTTCCGTGTTCCTACC-3' |
| | | Reverse: 5'-CCTCAGTGTAGCCCAAGATG-3' |

**Supplemental Figure 1**

```
>Chr11: 50,375,375-50,378,330
GCTCCCGTGATCAGATACAGATAGTTACATAGTCCCAGTATTAAGTTAGTAAGGATTAAA  60
AGAAAGGGAAAGCCCATGTAGTGAAAGTCAGGCGTTTGCAGCCCTTGTACACAAAAGGAT  120
CTTCAAATCCTTCCATTGTTAGAAATTACACAAATACTCAGAGCCTTCTGCCAAACAATT  180
GTCCACCAGAAAATCCTCATTGCGTGTGGTGGCACACACACTTTAATCCTGTCATTCTGG  240
AGGCGGATGATCAAACACTTGAGACCATGCATGACTGCAAGGGGAGTTTGAGGCCGGCCT  300
GGGATGTGTGCAATCAGGACTCAAAACAATAAGGAAAGGACACAATAAGGAAAGATAAAA  360
GTTTTTATCTATGATTATACACTGCATTTTTCTCTTACGATGGAAAAATAACCTCCCATA  420
CAACTAGAATAGAATTTAGAGAAACATCGTAGGAATGTGGTTGTGTTGTCAGAAGGGAAA  480
CATTTCCCCCCCTTCAGGTGAGAGACCTGGATTGCTAAGTTGCAGTGACTCTCCCATCTC  540
AGCATCTCAAGTAGATGGGACTACAGGCACATTACTGTGCAAAGTGAAAAAAGATGATTC  600
AAACAAAAAACGGTGTGTGTGTGTGAAAATATAGTTGAGTGGTAGGACTGGTTTGCTCAA  660
AGCCTTTCTAGTATGCATTGCTCTTCCCATACATCTAACCAACGATCGCAGGTCATCTAA  720
TAAGTATAGGGCTCTCAAATGGTGTGAATTAGCTAAAGAAGAGTAATCTTAGTTTCTTGG  780
TAATCAATCTCATGTCCCAAGACCATGTAGGGGGAAATATACATGTTAAGCAAGCTCTCT  840
ACCACTACTATTTCTCTTGTCTTCTTTTTGAGACAGGCTTTCAGTAAATTTCCCAGGCTG  900
CCTTGAACGCACTCTGTTGCTCAGGCAAGTCTTGAAGTTGCAGTTCTCCTCGGCTGGGCT  960
TGTGTTCTTGCCTTTCACCCACCTCCTGGCCTGGAGACTTGGGAGTTCTCACTTTTCGAT  1020
GATCTAAACTAGTGTTTTCTGTCTCCTCTAGATAGATCCTTTGCAGGTGAGGTCATGTGC  1080
AGTTATTCTCCCAGCCTGTCTCATCCGAGTTTCCCTTACACCTGAAGGTTCACGTTACAC  1140
AGCTCCTGTTGCTCTCTTGAAGTCTTGTGATGCCCTTCTACTTAACACCCTTGCGCAGAC  1200
TTAAAAGGTCTGAGTTCTGCACGCAGTCTCTGCACCTTTCCTTTGATGACACCCCCCCCT  1260
TTTAAAAAACAAACTGTTCAAACTGGATTTAGCCTCCTCAATGTTGCACTTGAAAGAAAA  1320
GCGTATTTTTGCCTGTAACCTCTTGTGATGGCTTTCAGAGGCTTACCAATTGTGGTACCT  1380
ACCGTGGCACCGCCCAAACAATAACTCTTATTGTTAATTTCTAGGGTAGGTCAGTCAGAT  1440
TCGGAGAAACTCGGGACATCATCCGTATTCTGCTATGTTGCAGGAATTCTCTTAAACGTC  1500
ACCCATTTACCTCTGTAAGGGCACACGGGATGACCCGGGCGTGTCTCCACCTTGGGAGAG  1560
GCCCGTTGGTCGGTCGGGCGCAGCGCTCTGGGGCTGCGTCCTTCGCGCCCCTGGCAGTCT  1620
CAGCCTTCCCTCGCGAGGCGCCGCACACCGAGCGCCTCTCGTGCAAGGGCGAGGGGTTCG  1680
TAGAGTGCGAGTCCGTTAGGCAAAGTGGGGTACAGGACTAGCTCGGGCACTGGCAGCCCA  1740
TGTGGGTCGGGACTGACCCCCTAAAGCGGGCAGGTAAGGGGACTGGTCCCAGGCGCCCCA  1800
CGGAGATCGGGGATTTGGGCGCAGGGGGCGGGGCGGTACCCAGAGGAGGTTTGGCACAAC  1860
CCGGCCCTCCAGCTCCTACGCGTTCTGAGGAGAAAGGGCATGGCTTCCTTTTCTTCGTGG  1920
TGTCTTTCAAAACCGACAGGAGCGCGGATGAAAAGCTGCGGATTAAGGGTAGAACTACCG  1980
CCCCTGCAAACCTGAACGAACTCTGGATTACTACCCAGTCCCTGCGGGGAGCTGATTAGC  2040
GCCATTCCCCTCCCCCACTCCGGCGGGTAGGTCTAGAGAACGGTTTGAAGGTAGCCGCAT  2100
GCGCATTAAGTAAAACAAGAAACCGCTGAGACGCCAAGCGCAATAATTGCAAAAAGGCGC  2160
AGGCGCAGAAAGAAAAAAAAAAAACAACCAAAAACACCTCGGTCAGTTTATGAAGAGGGC  2220
TGAGCCCAAACGCATGCGTGTAATAATAGGACGGCGGGAGAGAGCTCGGGGTTTCTTAAT  2280
ACGCATGCGTCCAATCTCTCTGTGACGCAAGTGGGGCGGGGCCGGCAACAGGGTGCGCGT  2340
GCGCAGAGCTTGCTGGAAGGGCTTTATTTAGCCTGCGCAGGCGCCCGCTTATCATTTCTC  2400
CTCAGCCACGCCGAGGCTTTTTGTGGAGGTGAGTGACTTTCCGTCTTCGACTGCTGCCCG  2460
AGCTCTCCACGGACACCTCTTTGGGTGACGATTGCTTGTCGCTGAGCTTAGGTTCTAACC  2520
TGAGGGGCCGTGGCAGGCGCAGGCTGCCCCTGGGAGCCGCGAGTGTCTCGCCGGCGTGGC  2580
AGCGCCGCCTAGGCCTGCAGGCCGCGCCGCTTTCTATTGTGTGAGATGTCTGAGGAGGCG  2640
GAGCGGAAGCGGCGGCCGCCATTTCCTCGTCTCAACCGTGGCTGTGGGCCAGGGTCTTAT  2700
TAGCTTGGCGCCTCGGCCCTGGGCGCACTGGGCGGCTTTGGGCATCGGGTCGCTAGCCGA  2760
GGCCGGGTTGCATGCTTGAGTCTCCGGGGCCCAGGCATCCTGACGGGCGGGACTCGCATG  2820
GCGGCGCCCGGTGCCCGCTGTGCGCCGTGAAGTGGGAGGGCCATTTAAATGCCTGTTTTC  2880
TCAGACGCGTTCCTGACCCCCACCCCCACCCCGCAGCCCCGCAATCTGTGTGGACGCCAG  2940
GACCACCGAAGGGACG  2956
```

**Supplemental Figure 2**



**Supplemental Table 2. Primers used for site directed mutagenesis.**

| Mutation | Position<br>SNP/Indel | Mutagenesis primer Sequences |
|---|---|---|
| T7288C | 50,377,288 | Forward: 5'-tgaaagacaccacgaggaaaaggaagccatgccc-3' |
| | rs29411274 | Reverse: 5'-gggcatggcttccttttcctcgtggtgtctttca-3' |
| GA7546G | 50,377,546 | Forward: 5'-gtttttggttgtttttttttttctttctgcgcctgcgcctttt-3' |
| | rs221962608 | Reverse: 5'-aaaaggcgcaggcgcagaaagaaaaaaaaaaacaaccaaaaac-3' |
| T7795G | 50,377,795 | Forward: 5'-cacctccacacaaagcctcggcgtggctga-3' |
| | rs29475617 | Reverse: 5'-tcagccacgccgaggctttgtgtggaggtg-3' |
| G7937C | 50,377,937 | Forward: 5'-cggcgagacagtcgcggctccca-3' |
| | rs257760362 | Reverse: 5'-tgggagccgcgactgtctcgccg-3' |
| All four mutations (GA7546G:luc2 mutant used as template for multi-site mutagenesis) | | Forward1: 5'-gggcatggcttccttttcctcgtggtgtctttca-3' |
| | | Reverse1: 5'-tgaaagacaccacgaggaaaaggaagccatgccc-3' |
| | | Forward2: 5'-tcagccacgccgaggctttgtgtggaggtg-3' |
| | | Reverse2: 5'-cacctccacacaaagcctcggcgtggctga-3' |
| | | Forward3: 5'-tgggagccgcgactgtctcgccg-3' |
| | | Reverse3: 5'-cggcgagacagtcgcggctccca-3' |