UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Accelerated Reverse and Comparative Genetics: Results from Nxf1 and Ankfn1

Permalink
https://escholarship.org/uc/item/4mz3834k

Author
Ross, Kevin

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Accelerated Reverse and Comparative Genetics:
Results from *Nxf1* and *Ankfn1*

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Kevin Donald Ross

Committee in charge:

    Professor Bruce Hamilton, Chair
    Professor Xiang-Dong Fu
    Professor Pamela Mellon
    Professor Miles Wilkinson
    Professor Gene Yeo

2020

The Dissertation of Kevin Donald Ross is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2020

iii

DEDICATION

I dedicate this work to my family,
and especially my parents,
Anne Webb Ross and Donald Ross.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Ankfn1   Ankyrin Repeat and Fibronectin Type III Domain Containing I

CRISPR   Clustered Regularly Interspaced Palindromic Repeat

Cas9   CRISPR-associated protein 9

IAP   Intracisternal A Particle

LRR   Leucine rich repeat

Nmf9   Neurological mutagenesis facility 9

NXF1   Nuclear Export Factor 1

PAM   Protospacer-adjacent motif

RA domain   Ras association domain

sgRNA   single guide RNA

TAP   Tip-associated Protein

UBA domain Ubiquitin association domain

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

"Deleting an *Nr4a1* Super-Enhancer Subdomain Ablates Ly6C[low]

Monocytes while Preserving Macrophage Gene Function". Thomas,

Graham D.; Hanna, Richard N.; Vasudevan, Neelakatan T.; Hamers,

Anouk A.; Romanoski, Casey E.; McArdle, Sara; Ross, Kevin D.;

Blatchley, Amy; Yoakum, Deborah; Hamilton, Bruce A.; Mikulski,

Zbigniew; Jain, Mukesh K.; Glass, Christopher K.; Hedrick, Catherine C.,

Immunity, 2016. The dissertation author was a co-author of this paper and

provided technical insight towards the generation of CRISPR/Cas9 edited

mice.

Chapter 2, also in part is a reprint of the material as it appears in

"The lipid elongation enzyme ELOVL2 is a molecular regulator of aging in

the retina". Chen, Daniel; Chao, Daniel L.; Rocha, Lorena; Kolar, Matthew;

Anh Nguyen Huu,Viet; Krawczyk, Michal; Dasyani, Manish; Wang, Tina;

Jafari, Maryam; Jabari, Mary; Ross, Kevin D.; Saghatelian, Alan;

Hamilton, Bruce A.; Zhang, Kang; Skowronska-Krawczyk, Dorota, Aging

Cell, 2020. The dissertation author was a co-author and co-investigator of

this paper, and designed and generated the Elovl2[C234W] mutant mice.

Chapter 3, "*Ankfn1* vestibular defects in zebrafish require mutations

in both ancestral and derived paralogs" is being prepared for publication

and is coauthored with Ren, Jie; Zhang, Ruilin; Chi, Neil and Hamilton,

Bruce. The dissertation author was the primary author of this chapter.

VITA

2009 Bachelor of Science, University of Illinois, Urbana-Champaign

2020 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Ross KD, Ren J, Zhang R, Chi NC, Hamilton BA. (2020). Ankfn1 vestibular defects in zebrafish require mutations in both ancestral and derived paralogs. (In preparation).

Chen D, Chao DL, Rocha L, Kolar M, Nguyen Huu VA, Krawczyk M, Dasyani M, Wang , Jafari M, Jabari M, Ross KD, Saghatelian A, Hamilton BA, Zhang K, Skowronska-Krawczyk D. (2020). The lipid elongation enzyme ELOVL2 is a molecular regulator of aging in the retina. Aging Cell. 00:e13100.

Thomas GD, Hanna RN, Vasudevan NT, Hamers AA, Romanoski CE, McArdle S, Ross KD, Blatchley A, Yoakum D, Hamilton BA, Mikulski Z, Jain MK, Glass CK, Hedrick CC. (2016). Deleting an Nr4a1 Super-Enhancer Subdomain Ablates Ly6Clow Monocytes while Preserving Macrophage Gene Function. Immunity. 45(5): 975-987.

Zhang S, Ross KD, Seidner GA, Gorman MR, Poon TH, Wang X, Keithley EM, Lee PN, Martindale MQ, Joiner WJ, Hamilton BA. (2015). Nmf9 encodes a highly conserved protein important to neurological function in mice and flies. PLoS Genet. 11(7): e1005344.

Concepcion D, Ross KD, Hutt KR, Yeo GW, Hamilton BA. (2015). Nxf1 natural variant E610G is a semi-dominant suppressor of IAP-induced RNA processing defects. PLoS Genet. 11(4): e1005123.

Urrichio LH, Chong JX, Ross KD, Ober C, Nicolae DL. (2012). Accurate imputation of rare and common variants in a founder population from a small number of sequenced individuals. Genet Epidemiol. 36, 312-319.

Huang L. et al. (2011). TMEM is mutated in patients with a Joubert syndrome related disorder and expands the role of the TMEM family at the ciliary transition zone. Am J Hum Genet. 89, 713- 730.

Torgerson D.G. et al. (2011). Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat Genet. 43, 887-892.

FIELDS OF STUDY

Major Field: Biomedical Sciences

     Studies in Molecular Genetics
     Professor Bruce Hamilton

     Studies in Human Population Genetics
     Professor Carole Ober

     Studies in Behavioral Genetics
     Professor Brent Roberts

ABSTRACT OF THE DISSERTATION

Accelerated Reverse and Comparative Genetics:
Results from *Nxf1 and Ankfn1*

by

Kevin Donald Ross

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2020

Professor Bruce Hamilton, Chair

Here I report my collaborative work employing genome editing to
generate mutations in diverse model organisms to accelerate reverse
genetic studies. We confirmed the mapping of a modifier of endogenous
retroviral mutations with nucleotide-level precision by creating a point
mutation in *Nxf1* which mimics a naturally occurring, suppressive allele. I
also applied this technology to a novel, highly conserved protein coding
gene, *nmf9/Ankfn1*. I confirmed the positional cloning of this gene,

identified in a forward genetic screen for vestibular behaviors in mice, by

creating alleles which phenocopied the behavioral outcomes of the ENU-

induced mutation. We validated functional conservation with mutant

alleles of the *Drosophila* homolog, disrupting two annotated functional

domains and an additional conserved, unannotated domain. Mutant flies

had locomotor defects, abnormal sleep patterns, and early lethality. The

quick adoption, refinement, and distribution of methods for making

transgenic animals facilitated other collaborative projects. In one, I

created an amino acid substitution in ELOVL2 to selectively inactivate its

ability to process certain polyunsaturated fats important for retinal and

brain function. This mutation recapitulated molecular signatures of aging in

the murine retina predicted by progressive loss of function associated with

methylation of the gene promoter. In another, I provided methodological

insight to dissect an enhancer element involved in monocyte development.

Enhancer deletions in the *Nr4a1* gene identified sequences which alter

Ly6Clow monocyte development. Chapter 3 reports my study of *Ankfn1*

zebrafish, where I saw the opportunity to test functional conservation of

Ankfn1 paralogs. Zebrafish have two copies of this gene, an ancestral

copy which contains a Ras-association (RA) domain, and a derived copy

which does not. I analyzed the physicochemical constraint of 115

homologs of Ankfn1 representing 800 million years of evolutionary divergence and edited the most constrained peptide sequence to create disrupting alleles. Interestingly, I discovered an overt swim bladder phenotype which was, with two exceptions, observed only in fish with mutations in both alleles of each paralogous copy. These studies expand our understanding of a genetic modifier of IAP insertions, characterize a gene important in neurological function, and demonstrate the power of gene editing to accelerate reverse genetics.

CHAPTER 1: INTRODUCTION

CRISPR/Cas9 Genome Editing Accelerates Genetic Studies

CRISPR/Cas9 editing is a method for RNA-guided DNA editing which is derived from a bacterial system for targeted DNA cleavage (Doudna and Charpentier 2014). Clustered regularly interspaced short palindromic repeats (CRISPR) were first reported as an oddity within bacterial genomes, where short stretches of 21-37 bp were interrupted by non-repetitive sequences, some of which could be traced to mobile genetic elements (Haft et al. 2005). It was hypothesized that CRISPR-associated (Cas) genes in the vicinity of these repeats may encode enzymes to target and defend against invading phage and plasmid DNA (Makarova et al. 2006). This bacterial adaptive immune system could specifically cleave double-stranded DNA, and then integrate sequences as DNA spacers for future defense (Garneau et al. 2010).

In short order, groups sought to demonstrate that this system could be harnessed in other contexts, specifically as a gene editing reagent that could circumvent the difficulties of contemporary technologies like TAL effector nucleases (Christian et al. 2010). The cleavage dynamics of CRISPR/Cas9 were characterized in vivo using a chimeric RNA which

included the CRISPR RNA (crRNA) and trans-activating RNA (tracrRNA) sequences necessary to guide double-stranded DNA cleavage (Jinek et al. 2012). Further advances quickly followed, including simultaneous reports that CRISPR/Cas9 editing could be reconstituted and multiplexed in mammalian cells and facilitate homology directed repair (HDR) by including a donor DNA template (Cong et al. 2013), and that an alternative short guide RNA (sgRNA) derived from the tracr-/cr-RNA sequences could direct cleavage (Mali et al. 2013). These experiments showed that it was feasible to direct precise gene editing using a three-part system of Cas9 mRNA, synthetic sgRNA, and a DNA repair template to produce HDR in other model systems. Genome editing was demonstrated in multicellular organisms before, however the design and implementation of these technologies was time and cost-intensive (Tesson et al. 2011; Sung et al. 2013). While we were undertaking the design and production of CRISPR/Cas9 reagents to edit *Nxf1*, breakthrough reports of CRISPR/Cas9 editing in common model organisms including mice (Wang et al. 2013), flies (Bassett et al. 2013; Gratz et al. 2013), and fish (Hwang et al. 2013) were published.

The range of potential Cas9 cleavage targets is only constrained by the requirement for a specific sequence motif adjacent to the cut site,

called the protospacer-adjacent motif (PAM), which for Cas9 is a three nucleotide NGG or NAG sequence. This flexibility comes with a concern over off-target effects, since single mismatches between the sgRNA and the DNA target can be tolerated, depending on where in the target sequence they occur (Pattanayak et al. 2013). This could complicate the interpretation of editing experiments and temper expectations for the therapeutic value of Cas9. However, Cas9 has been the subject of extensive improvements and iterations to address these and other aspects of its use.

An early solution to the off-target concern was to design dual sgRNAs that would direct a variant Cas9 protein which could only produce a nick in dsDNA for cooperative editing (Ran et al. 2013; Mali et al. 2013; Shen et al. 2014). Genome wide sgRNA libraries targeting mice and humans were developed as a powerful tool for both forward and reverse genetics, and have been used determine optimal design rules to maximize on-target and minimize off-target effects (Y. Zhou et al. 2014; Shalem et al. 2014; Doench et al. 2014; 2016). Methods to direct the timing and activity and repair efficiency of genome editing include compounds targeting the HDR mechanism of the cell, photoactivatable control of CRISPR/Cas9, doxycycline-inducible constructs (Dow et al. 2015;

Nihongaki et al. 2015; Chu et al. 2015). Beyond its nuclease activity,

CRISPR interference and activation systems modulate gene activity via

fusions of nuclease-dead Cas9 with effector proteins for highly specific

regulation (Perez-Pinera et al. 2013; Mali et al. 2013; Chavez et al. 2015;

Hilton et al. 2015).

Cas9 is not the only CRISPR-associated protein, and variants were

discover or engineered with different PAM requirements, sizes, and

substrate targeting (O'Connell et al. 2014). CRISPR/Cas has led to the

development of a suite of tools with transformative value in genetic studies

and potentially in the treatment of certain genetic diseases (Schwank et al.

2013; Wu et al. 2013; Long et al. 2014). However, at the time of our

studies of *Nxf1* and *nmf9/Ankfn1* described in this dissertation,

commercially available editing reagents were not available, and the

methods, while published, had not yet been adopted at UCSD and

elsewhere.

NXF1 Is The Primary Eukaryotic mRNA Export Factor

Nxf1 is a highly conserved, essential nuclear export receptor

responsible for the majority of mRNA export in the eukaryotic cell (Gatfield

and Izaurralde 2002; Andrea Herold, Teixeira, and Izaurralde 2003;

4

Natalizio and Wente 2013). It was first characterized as Tip-associated

protein (TAP) in the context of herpesvirus saimiri (Yoon et al. 1997),

followed by studies showing that it could export certain retroviral RNAs out

of the nucleus through its recognition of specific sequence elements,

namely the constitutive transport element (CTE), as well as the RNA

transport element (RTE), found in retroviral elements (Grüter et al. 1998;

Lindtner et al. 2006) It was later shown that the sequence-dependent

export of retroviral RNAs represents a subversion of the normal mRNA

export pathways, of which Nxf1 is an indispensable component in

eukaryotic organisms ranging from yeast to humans (A. Herold et al. 2000;

A. Herold, Klymenko, and Izaurralde 2001; Wilkie et al. 2001). Nxf1

consists of four domains that have been structurally and functionally

examined, primarily as they relate to the role of Nxf1 in nuclear export.

The N-terminus contains a non-canonical ribonucleoprotein (RNP)-type

RNA binding domain followed by four leucine-rich repeats (LRRs) that

allow Nxf1 to bind directly to target mRNAs and to interact indirectly with

transcripts via accessory proteins that bind to RNAs. The C-terminal half

of the protein includes a nuclear transport factor 2-like (NTF2L) domain

that is required for heterodimerization with NTF2-related export protein-1

(NXT1/p15), an interaction that is necessary for nuclear export of mRNAs

(Braun et al. 2001). The C-terminus of Nxf1 ends in a ubiquitin-associated

domain that allows Nxf1 to interact with unstructured phenylalanine-

glycine (FG) repeats of the nuclear pore complex (NPC). These functional

domains of Nxf1 confer the abilities to bind directly to mRNA, recognize

and interact with the exon-exon junction complex (EJC) (Z. Zhou et al.

2000), be recruited to the mRNP export complex (Sträßer and Hurt 2001;

Luo et al. 2001), interact and pass through the NPC, and eventually

deliver its mRNA cargo before being recycled to the nucleus.

IAPs Are Abundant Sequence Elements Shaping Murine Evolution

In addition to its role in mRNA export, NXF1 can suppress

mutations caused by intronic insertions of a class of retroviral elements

called intracisternal A particles (IAPs), which are a type of highly active

endogenous retrovirus (ERV) in mice, responsible for up to 15% of

spontaneous mutations in lab strains (Hamilton et al. 1997; Waterston et

al. 2002). Transposition of ERVs is responsible for a much larger

proportion of variation in mice than is observed for human de novo

mutations (Makarova et al. 2006). This is because while most human

ERVs are extinct, the murine genome contains full length ERVs which are

not only retrotransposition competent, but autonomous (Dewannieux et al. 2004).

ERV sequences are highly polymorphic, and can impact alternative splicing choices, expression level, polyadenylation, or can be benign. (Y. Zhang et al. 2008; Hayward and Katzourakis 2015). IAP insertions are the underlying cause of a range of mouse mutant lines (Hamilton et al. 1997; Floyd et al. 2003; Concepcion, Flores-García, and Hamilton 2009). The intronic insertion of related elements resulted in premature termination and aberrant splicing (Adachi, Watanabe-Fukunaga, and Nagata 1993). In some cases, ERVs generate alternative transcripts which confer useful biological functions. The LTR sequence of an HERV-K endogenous retrovirus creates an alternative short splice isoform of the human leptin receptor gene OBRa/b (Kapitonov and Jurka 1999). Retention of endogenous retroviral sequence between the terminal two exons in the human vascular endothelial growth factor receptor 3 (VEGFR-3) generates two transcripts which differ in their function (D. C. Hughes 2001).

Nxf1 can interact with IAP inserted transcript via protein intermediates. RBM15 binds to the RNA transport element (RTE) of IAPs, tethering them to Nxf1 for nucleocytoplasmic transport as unspliced

mRNAs (Lindtner et al. 2006). The IAP RTE sequence is required for retrotransposition of these elements (Andrei S. Zolotukhin et al. 2008), and is functionally similar to Nxf1-binding sequences in other retroviral elements (A. S. Zolotukhin et al. 2001). IAP insertions can alter local patterns of methylation, a likely defense mechanism to suppress these mobile elements (Deniz et al. 2018).

IAPs are abundant sequence elements in the mouse genome, and active components of their evolutionary history, and can have significant effects on the regulation of inserted genes, contributing a great deal to function variation. Because Nxf1 is able to suppress the effects of at least some of these mutations, it as a central component of a regulatory node based on the modification of genes with IAP mutations.

Nxf1 Is A Genetic Modifier of Endogenous Retroviral Mutations

Our interest in *Nxf1* began with its discovery as a genetic suppressor of neurodegeneration in a mouse disease model. *Vibrator* (*vb*) mouse mutants exhibit phenotypes including degeneration of the brain stem and spinal cord, severe action tremor, and uniform juvenile lethality (Welmar et al 1928). The *vb* mutation is caused by a sense-oriented intracisternal A particle (IAP) insertion into the fourth intron of the

phosphatidylinositol transfer protein alpha (*Pitpna*) gene, encoding a protein involved in phospholipid signaling pathways in diverse cell types (Hamilton et al. 1997). The *vb* IAP insertion reduces the amount of appropriately spliced *Pitpna* mRNA to 18% of wild-type levels. Interestingly, when *vb* mice were bred to the CAST/Ei laboratory strain in a mapping cross, the progeny had reduced tremor compared to *vb* mutants on a B6 background as well as increased life span (Hamilton et al. 1997). A single locus capable of ameliorating these phenotypes was mapped to a region on proximal chromosome 19 and named *Moderator of vibrator 1* (*Mvb1*).

Positional complementation cloning revealed that *Nxf1* is the gene within the *Mvb1* locus the controls the level of appropriately processed mRNA from genes that were mutated by IAP insertions (Floyd et al. 2003). *Nxf1* alleles modify the more severe phenotypes of the *vb* mutation through a mechanism involving mRNA biogenesis. Correctly processed mRNA and protein from *Pitpna* was increased two-fold in the CAST/Ei *Mvb1* homozygotes over B6 *Mvb1* homozygotes (from 18% to 36% of wild-type levels). Additionally, a low-abundance, aberrant transcript containing 5' exons of *Pitpna* was reduced in favor of appropriately spliced transcript. Importantly, opposing effects on normal and mutant-specific

9

transcripts has been replicated for several additional IAP mutations, together suggesting that Nxf1 has a role in mRNA transcript processing in addition to its function as an export factor. Specifically, it suggests that Nxf1 contributes, in some cases, to alternative splicing decisions. Our lab has shown that suppression of the *vb* mutation is just the first example of a range of different IAP mutations and strain polymorphisms that can be suppressed (Concepcion, Flores-García, and Hamilton 2009).

The idea that Nxf1 is involved in gene regulation by means other than licensing spliced mRNAs for export is surprising. Nxf1 associates with transcription/export (TREX) complexes localized to nuclear speckles, which are active sites of transcription, splicing, and exon-exon junction complex (EJC) formation (Teng and Wilson 2013). Serine-arginine (SR) splicing factors interact with mRNA transcripts in a manner that can be disrupted by Nxf1 binding in a phosphorylation-dependent manner (Tintaru et al. 2007). Further, U2 auxiliary factor (U2AF) has binding sites for Nxf1 (Andrei S. Zolotukhin et al. 2002). Proximity ligation-based methods for measuring protein-protein interactions (PPIs) have revealed additional partners of Nxf1, including RNA PolII, which suggests a possible explanation for the suppression of IAP insertion mutations (Youn et al. 2018).

Alleles of *Nxf1* differ between B6 and *castaneus* at many synonymous and non-coding sites (~1/300bp) and two amino acid positions, a serine-to-proline substitution (S48P), and a glutamate-to-glycine substitution (E610G). The E610G polymorphism occurs within the UBA domain at the C-terminus of the protein, which is known to interact with components of the NPC and other binding partners, making it a strong functional candidate. In chapter 2 of this dissertation, we explore the range of IAP isforms that are suppressed by Nxf1, and report gene editing experiments which attribute the modifier effect of the suppressive Nxf1 variant to the E610G polymorphism.

Nmf9/Ankfn1 Is Highly Conserved Gene Involved in Neurological Function

Despite having sequenced the human genome, there are still a significant proportion of genes which are un- or understudied (Wood et al. 2019). *Nmf9/Ankfn1* is a highly conserved protein coding gene with an interesting evolutionary history. Our characterization of *Nmf9/Ankfn1* represents one third of the published reports which comprise the entire literation of this gene, with the unpublished work in chapter 3 of this dissertation extending what we know to an additional species which has a

duplication of this conserved gene. Notably, the two reports that preceded

ours offered differing interpretations of the structure and function of the

gene, which we were able to resolve using genome-edited alleles.

Each of the published studies began with a forward genetic screen.

One was an ENU-mutagenesis screen for neurological mutants in mice (S.

Zhang et al. 2015). Another was a genome-wide RNAi screen for genes

involved in asymmetrical cell division (Mauri et al. 2014). The third was a

screen of transposon lines in drosophila for genes involved in sleep (Liu et

al. 2014).

Ankfn1 Homolog *Banderoula* (*Bnd*) Regulates Asymmetric Cell Division

Asymmetric cell division (ACD) is important for the development of

the fruit fly nervous system. Development is partly regulated by the Notch

inhibitor Numb, which differentially segregates to generate the cell types of

the central and peripheral nervous system. Mauri et al. performed an

RNAi screen to identify other genes involved in ACD by GFP-tagging the

Numb-binding partner Pon, expressed in sensory organ precursor cells,

from which the other cell types of the fly nervous system are derived

(Mauri et al. 2014). Among the three newly identified ACD genes,

CG45058 was the only one involved in establishing, rather than maintaining, polarity. The CG45058 sequence encodes two ankyrin repeats and a fibronectin type III domain, as well as a conserved non-motif domain, followed by a terminal Ras-association (RA) domain, consistent with the conserved features of *Ankfn1*. The group named this gene *banderoula* (*bnd*) for the random crescent like-shape in mutant animals. Immunohistochemistry revealed specific defects stemming from Bnd loss-of-function, namely mis-localization of Numb protein, resulting in spindle alignment and cell polarity defects. Interestingly, deletion of *bnd* resulted in pupal-lethality, and loss of function of both Bnd and its interaction partner discs-large (Dlg) results in neuroblast overproliferation, suggesting a possible role for Bnd in tumor suppression.

Ankfn1-Containing Gene *Wide awake* (*wake*) Regulates *Drosophila* Sleep

Liu et al. reported sleep phenotypes in fruit flies with putative null P-element alleles and one deletion allele of an ORF which includes the highly conserved, *Ankfn1*-homologous CG6954 sequence and 5' unstructured CG42686 sequence (now reclassified as CG45058/WAKE). They observed decreased sleep amount and increased sleep latency

phenotypes in *wake* mutants, but did not report early lethality as observed

by Mauri et al. Immunostaining of whole-mount fly brain shows

expression of WAKE in Clock neurons. They offer evidence that WAKE

expression level varies during the day, but that this variation is dampened

by mutations in canonical circadian clock genes, placing WAKE as a

downstream regulator of sleep. They also offer genetic and

immunoprecipitation data for an interaction between WAKE and $GABA_A$

receptor *Resistant to dieldrin* (RDL). The interaction data, along with

patch clamp recordings from ventrolateral neurons (LVns) suggest that the

mode of action of WAKE is to upregulate RDL in LNvs at dusk to inhibit

arousal.

Chapter 2 of this dissertation includes the positional cloning of a

vestibular mutation in mice to the murine homolog of *Ankfn1*, and the

behavioral characterization of mutants, including animals with genome

edited alleles of a highly conserved peptide sequence. We also created

precise mutations in domains of *Ankfn1* in *Drosophila,* and observed

sexually dimorphic sleep phenotypes which differed from those in *wake*

mutants, in additional to pupal lethality of homozygous animals, consistent

with *bnd* homozygotes.

14

Investigating Functional Conservation and Gene Fate After

Duplication in zebrafish *Ankfn1* Paralogs

Ankfn1/Nmf9 is a highly conserved protein coding gene with an interesting evolutionary history. In our study of an ENU-induced mutation in the mouse homolog of this gene and CRISPR alleles in mice and flies, we've demonstrated that it has a clear role in neurological function (S. Zhang et al. 2015). Zebrafish have two paralogs of *Ankfn1*, possibly resulting from a genome-wide duplication event 300-400 million years ago (Kumar et al. 2017). These paralogs differ in their sequence composition, and by the inclusion of the C-terminal RA domain, which is also missing in mammals orthologs. Studying these paralogs lets us address an ongoing debate over the fate of duplicate genes and their role in shaping the evolution.

Gene Duplications and Evolution

Gene duplication events are a significant source of evolution (Ohno 1970), but the path towards retention or loss is not well-understood. Gene duplications vary in scale from whole genome to smaller segments, and can arise from unequal crossing over, retroposition, and aneu/poly-ploidization (J. Zhang 2003; Ganko, Meyers, and Vision 2007). Duplicates

represent a significant portion of eukaryotic genomic sequence, with estimates ranging from 30-50% (W. H. Li et al. 2001; Rubin et al. 2000). However, the subject of when and how many duplications may have occurred in the evolution of different lineages is a topic of sustained debate (McLysaght, Hokamp, and Wolfe 2002; Holland and Ocampo Daza 2018; Smith and Keinath 2015; Smith et al. 2018; Pascual-Anaya et al. 2018) Polyploidization in particular has played a role in the evolution of arthropods (Z. Li et al. 2018). Multiple rounds of gene duplication, especially in plants, have generated gene families whose members have been subject to variation (Guo et al. 2016). Such clusters also exist in animals, where for instance the *Hox* genes have garnered significant attention (Lemons and McGinnis 2006; Brooke, Garcia-Fernàndez, and Holland 1998; Ferrier and Holland 2001; Garcia-Fernàndez and Holland 1994; Holland and Ocampo Daza 2018).

Duplication events are readily identified, but there is not a clear consensus on what happens when a gene is duplicated. The fate of duplicated genes is constrained by the function of the gene, and its interaction partners (Wapinski et al. 2007). The scale of the duplication event matters, as selection may act differently when a gene is duplicated as part of a whole genome or chromosomal event rather than a smaller

scale event (Scannell et al. 2007; Conant and Wolfe 2008; James A.

Birchler and Veitia 2012). Tandem duplication may have different

outcomes stemming from coregulation by shared genomic elements (Lan

and Pritchard 2016; Naseeb et al. 2017) relative to gene duplicate that are

not physically linked, as in the case of zebrafish *Ankfn1* paralogs.

Models of Gene Duplicate Fate

Models of gene duplicate fate vary by the type of selection applied

to duplicates, the time to fixation, and the time to and type of fate-

determining mutation that eventually arise (Innan and Kondrashov 2010).

Most duplicate genes will be lost due to genetic drift, but if they become

fixed, they will subsequently acquire fate-determining mutations. These

mutations may result in loss of function (pseudogenization), acquisition of

new function (neofunctionalization), or partitioning of existing functions

(subfunctionalization). An estimate of the frequency of gene duplication

and retention suggests that 50% of genes may duplicate and be expected

to rise to high frequency on a long enough time scale (M. Lynch and

Conery 2000).

Several models assume neutral selection on duplicates, and it has

been demonstrated that young gene duplicates have a signature of

relaxed selection compared to duplicate pairs with a higher divergence

time (M. Lynch and Conery 2000; Lan and Pritchard 2016). Under the

classical neofunctionalization model proposed by Susumu Ohno, the

ancestral copy is maintained while the derived copy is under reduced

selective pressure, and is therefore able to acquire mutations which yield

new function (Ohno 1970). While Ohno's work is considered the starting

point for discussions of gene duplicate fate, it is not clear that ancestral

gene copies are protected from variation as he suggests.

In duplication-degeneration-complementation (DDC), duplicate

copies are both subject to degenerative mutations which damage function,

and which can become fixed as the two copies compartmentalize the

original functions of the ancestral copy (Force et al. 1999; Michael Lynch

and Force 2000). If a gene has multiple functions which cannot be

independently improved, a duplication may offer the opportunity for

selection to act on each function independently, leaving to functional

partitioning (subfunctionalization) (A. L. Hughes 1994). The rate of

subfunctionalizing mutations would need to fall within a certain range

relative to null mutations, around 0.3-0.7 relative to nulls for 10-50%

retention, in order for DDC to be an appreciable explainer of gene

duplication retention (Force et al. 1999). *In silico* analysis of protein

evolution suggests that subfunctionalization can lead to changes in a binding partner to exclude interactions with one of the duplicates (Teufel et al. 2019). Alterations in tissue-specific patterns of expression after duplication may represent a form of sub/neofunctionalization, and can contribute to the retention of gene duplicates (Sémon and Wolfe 2007). For example, differences in the structure and composition of the heart between Chondrostei/Holostei and Teleostei fish are explained by the duplication of the elastin gene and subsequent mutation which led to changes in expression pattern and protein properties (Moriyama et al. 2016). Escape from adaptive conflict (EAC) is a type of subfunctionalization where a gene encoded two functions which could not be independently improved without a decrement in the alternative function (Des Marais and Rausher 2008). Following a gene duplication, these two functions could be subject to independent selection and evolution.

Several models are based on the effect of dosage on gene duplicate fate, which may be subject to positive or purifying selection (Kondrashov and Koonin 2004). A prevalent model for the retention or loss of gene duplicates is the dosage balance hypothesis, which suggests that there is selection for gene expression level, especially for genes which encode proteins that participate in protein complexes (J A Birchler and

Newton 1981; Papp, Pál, and Hurst 2003; Edger and Pires 2009).

Following from this, it has been shown that ohnologs, gene duplicates

arising from a whole genome duplication event, may be more resistant to

later small scale duplication and copy number variation, and that dosage

imbalance may underlie diseases like Down Syndrome (Makino and

McLysaght 2010; Makino, McLysaght, and Kawata 2013).

The fate of genes which participate in the formation of protein

complexes increases the number of possible outcomes for duplicates and

their interaction partners. This may result in either compensation, where

one paralog is able to compensate for loss of function of the other, or

dependency, especially in the case of paralogs whose products physically

interact (Diss et al. 2017). Competitive interference between paralogs

may be a factor. If a duplicate loses some of the function of the ancestral

copy, it may be detrimental if it still has the ability to interact with

complexes. This was modeled in a reconstructed ancestral transcription

factor in yeast, where it was also shown that compensatory mutations may

counteract interference (Baker, Hanson-Smith, and Johnson 2013). In

some cases, for genes that produce products that regulate responses to

the environment, increased dosage may be positively

selected. Duplication can also buffer against deleterious mutations. In

paramecium, when ohnologs have similar expression levels, it is less likely

that one of the duplicates was lost in a related species but when onhologs

had divergent expression, it was twice as likely that the duplicate was lost

in related species, and that the lower expression copy was lost (Gout and

Lynch 2015).

Chapter 3 of this dissertation reports our experiments of *Ankfn1*

functional conservation in zebrafish. When both paralogs are disrupted,

the swim bladders of mutant fish do not inflate normally by five days post-

fertilization, and these animals do not survive past two weeks. This result

has implications for both the biological role of *Ankfn1*, and offers an

opportunity to address the unique evolution of this gene in a model

organism.

REFERENCES

Adachi, Masashi, Rie Watanabe-Fukunaga, and Shigekazu Nagata. 1993. "Aberrant Transcription Caused by the Insertion of an Early Transposable Element in an Intron of the Fas Antigen Gene of Lpr Mice." *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.90.5.1756.

Baker, Christopher R., Victor Hanson-Smith, and Alexander D. Johnson. 2013. "Following Gene Duplication, Paralog Interference Constrains Transcriptional Circuit Evolution." *Science*. https://doi.org/10.1126/science.1240810.

Bassett, Andrew R., Charlotte Tibbit, Chris P. Ponting, and Ji Long Liu. 2013. "Highly Efficient Targeted Mutagenesis of Drosophila with the

CRISPR/Cas9 System." *Cell Reports*.
https://doi.org/10.1016/j.celrep.2013.06.020.

Birchler, J A, and K J Newton. 1981. "Modulation of Protein Levels in Chromosomal Dosage Series of Maize: The Biochemical Basis of Aneuploid Syndromes." *Genetics*.

Birchler, James A., and Reiner A. Veitia. 2012. "Gene Balance Hypothesis: Connecting Issues of Dosage Sensitivity across Biological Disciplines." *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1207726109.

Braun, Isabelle C., Andrea Herold, Michaela Rode, Elena Conti, and Elisa Izaurralde. 2001. "Overexpression of TAP/P15 Heterodimers Bypasses Nuclear Retention and Stimulates Nuclear MRNA Export." *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M100400200.

Brooke, Nina M., Jordi Garcia-Fernàndez, and Peter W.H. Holland. 1998. "The ParaHox Gene Cluster Is an Evolutionary Sister of the Hox Gene Cluster." *Nature*. https://doi.org/10.1038/31933.

Chavez, Alejandro, Jonathan Scheiman, Suhani Vora, Benjamin W. Pruitt, Marcelle Tuttle, Eswar P R Iyer, Shuailiang Lin, et al. 2015. "Highly Efficient Cas9-Mediated Transcriptional Programming." *Nature Methods*. https://doi.org/10.1038/nmeth.3312.

Christian, Michelle, Tomas Cermak, Erin L. Doyle, Clarice Schmidt, Feng Zhang, Aaron Hummel, Adam J. Bogdanove, and Daniel F. Voytas. 2010. "Targeting DNA Double-Strand Breaks with TAL Effector Nucleases." *Genetics*. https://doi.org/10.1534/genetics.110.120717.

Chu, Van Trung, Timm Weber, Benedikt Wefers, Wolfgang Wurst, Sandrine Sander, Klaus Rajewsky, and Ralf Kühn. 2015. "Increasing the Efficiency of Homology-Directed Repair for CRISPR-Cas9-Induced Precise Gene Editing in Mammalian Cells." *Nature Biotechnology*. https://doi.org/10.1038/nbt.3198.

Conant, Gavin C., and Kenneth H. Wolfe. 2008. "Turning a Hobby into a Job: How Duplicated Genes Find New Functions." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg2482.

Concepcion, Dorothy, Lisbeth Flores-García, and Bruce A. Hamilton. 2009. "Multipotent Genetic Suppression of Retrotransposon-Induced Mutations by Nxf1 through Fine-Tuning of Alternative Splicing." *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1000484.

Cong, Le, F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, et al. 2013. "Multiplex Genome Engineering Using CRISPR/Cas Systems." *Science* 339 (6121): 819–23. https://doi.org/10.1126/science.1231143.

Deniz, Özgen, Lorenzo de la Rica, Kevin C.L. Cheng, Dominik Spensberger, and Miguel R. Branco. 2018. "SETDB1 Prevents TET2-Dependent Activation of IAP Retroelements in Naïve Embryonic Stem Cells." *Genome Biology*. https://doi.org/10.1186/s13059-017-1376-y.

Dewannieux, Marie, Anne Dupressoir, Francis Harper, Gérard Pierron, and Thierry Heidmann. 2004. "Identification of Autonomous IAP LTR Retrotransposons Mobile in Mammalian Cells." *Nature Genetics* 36 (5): 534–39. https://doi.org/10.1038/ng1353.

Diss, Guillaume, Isabelle Gagnon-Arsenault, Anne Marie Dion-Coté, Hélène Vignaud, Diana I. Ascencio, Caroline M. Berger, and Christian R. Landry. 2017. "Gene Duplication Can Impart Fragility, Not Robustness, in the Yeast Protein Interaction Network." *Science*. https://doi.org/10.1126/science.aai7685.

Doench, John G., Nicolo Fusi, Meagan Sullender, Mudra Hegde, Emma W. Vaimberg, Katherine F. Donovan, Ian Smith, et al. 2016. "Optimized SgRNA Design to Maximize Activity and Minimize Off-Target Effects of CRISPR-Cas9." *Nature Biotechnology*. https://doi.org/10.1038/nbt.3437.

Doench, John G., Ella Hartenian, Daniel B. Graham, Zuzana Tothova, Mudra Hegde, Ian Smith, Meagan Sullender, Benjamin L. Ebert, Ramnik J. Xavier, and David E. Root. 2014. "Rational Design of

Highly Active SgRNAs for CRISPR-Cas9-Mediated Gene Inactivation." *Nature Biotechnology*. https://doi.org/10.1038/nbt.3026.

Doudna, Jennifer A., and Emmanuelle Charpentier. 2014. "The New Frontier of Genome Engineering with CRISPR-Cas9." *Science* 346 (6213): 1258096. https://doi.org/10.1126/science.1258096.

Dow, Lukas E., Jonathan Fisher, Kevin P. O'Rourke, Ashlesha Muley, Edward R. Kastenhuber, Geulah Livshits, Darjus F. Tschaharganeh, Nicholas D. Socci, and Scott W. Lowe. 2015. "Inducible in Vivo Genome Editing with CRISPR-Cas9." *Nature Biotechnology*. https://doi.org/10.1038/nbt.3155.

Edger, Patrick P., and J. Chris Pires. 2009. "Gene and Genome Duplications: The Impact of Dosage-Sensitivity on the Fate of Nuclear Genes." *Chromosome Research*. https://doi.org/10.1007/s10577-009-9055-9.

Ferrier, David E.K., and Peter W.H. Holland. 2001. "Ancient Origin of the Hox Gene Cluster." *Nature Reviews Genetics*. https://doi.org/10.1038/35047605.

Floyd, Jennifer A., David A. Gold, Dorothy Concepcion, Tiffany H. Poon, Xiaobo Wang, Elizabeth Keithley, Dan Chen, et al. 2003. "A Natural Allele of Nxf1 Suppresses Retrovirus Insertional Mutations." *Nature Genetics*. https://doi.org/10.1038/ng1247.

Force, Allan, Michael Lynch, F. Bryan Pickett, Angel Amores, Yi Lin Yan, and John Postlethwait. 1999. "Preservation of Duplicate Genes by Complementary, Degenerative Mutations." *Genetics*.

Ganko, Eric W., Blake C. Meyers, and Todd J. Vision. 2007. "Divergence in Expression between Duplicated Genes in Arabidopsis." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msm158.

Garcia-Fernàndez, Jordi, and Peter W.H. Holland. 1994. "Archetypal Organization of the Amphioxus Hox Gene Cluster." *Nature*. https://doi.org/10.1038/370563a0.

Garneau, Josiane E., Marie Ève Dupuis, Manuela Villion, Dennis A. Romero, Rodolphe Barrangou, Patrick Boyaval, Christophe Fremaux, Philippe Horvath, Alfonso H. Magadán, and Sylvain Moineau. 2010. "The CRISPR/Cas Bacterial Immune System Cleaves Bacteriophage and Plasmid DNA." *Nature*. https://doi.org/10.1038/nature09523.

Gatfield, David, and Elisa Izaurralde. 2002. "REF1/Aly and the Additional Exon Junction Complex Proteins Are Dispensable for Nuclear MRNA Export." *Journal of Cell Biology*. https://doi.org/10.1083/jcb.200207128.

Gout, Jean Francois, and Michael Lynch. 2015. "Maintenance and Loss of Duplicated Genes by Dosage Subfunctionalization." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msv095.

Gratz, Scott J., Alexander M. Cummings, Jennifer N. Nguyen, Danielle C. Hamm, Laura K. Donohue, Melissa M. Harrison, Jill Wildonger, and Kate M. O'connor-Giles. 2013. "Genome Engineering of Drosophila with the CRISPR RNA-Guided Cas9 Nuclease." *Genetics*. https://doi.org/10.1534/genetics.113.152710.

Grüter, Patric, Carlos Tabernero, Cayetano Von Kobbe, Christel Schmitt, Claudio Saavedra, Angela Bachi, Matthias Wilm, Barbara K. Felber, and Elisa Izaurralde. 1998. "TAP, the Human Homolog of Mex67p, Mediates CTE-Dependent RNA Export from the Nucleus." *Molecular Cell*. https://doi.org/10.1016/S1097-2765(00)80065-9.

Guo, Hai Song, Yan Mei Zhang, Xiao Qin Sun, Mi Mi Li, Yue Yu Hang, and Jia Yu Xue. 2016. "Evolution of the KCS Gene Family in Plants: The History of Gene Duplication, Sub/Neofunctionalization and Redundancy." *Molecular Genetics and Genomics*. https://doi.org/10.1007/s00438-015-1142-3.

Haft, Daniel H., Jeremy Selengut, Emmanuel F. Mongodin, and Karen E. Nelson. 2005. "A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes." *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.0010060.

Hamilton, Bruce A., Desmond J. Smith, Kenneth L. Mueller, Anne W. Kerrebrock, Roderick T. Bronson, Victor Van Berkel, Mark J. Daly, et al. 1997. "The Vibrator Mutation Causes Neurodegeneration via Reduced Expression of PITPα: Positional Complementation Cloning and Extragenic Suppression." *Neuron.* https://doi.org/10.1016/S0896-6273(00)80312-8.

Hayward, Alexander, and Aris Katzourakis. 2015. "Endogenous Retroviruses." *Current Biology* 25 (15): R644–46. https://doi.org/10.1016/j.cub.2015.05.041.

Herold, A., T. Klymenko, and E. Izaurralde. 2001. "NXF1/P15 Heterodimers Are Essential for MRNA Nuclear Export in Drosophila." *RNA.*

Herold, A., M. Suyama, J. P. Rodrigues, I. C. Braun, U. Kutay, M. Carmo-Fonseca, P. Bork, and E. Izaurralde. 2000. "TAP (NXF1) Belongs to a Multigene Family of Putative RNA Export Factors with a Conserved Modular Architecture." *Molecular and Cellular Biology.* https://doi.org/10.1128/mcb.20.23.8996-9008.2000.

Herold, Andrea, Luis Teixeira, and Elisa Izaurralde. 2003. "Genome-Wide Analysis of Nuclear MRNA Export Pathways in Drosophila." *EMBO Journal.* https://doi.org/10.1093/emboj/cdg233.

Hilton, Isaac B., Anthony M. D'Ippolito, Christopher M. Vockley, Pratiksha I. Thakore, Gregory E. Crawford, Timothy E. Reddy, and Charles A. Gersbach. 2015. "Epigenome Editing by a CRISPR-Cas9-Based Acetyltransferase Activates Genes from Promoters and Enhancers." *Nature Biotechnology.* https://doi.org/10.1038/nbt.3199.

Holland, Linda Z., and Daniel Ocampo Daza. 2018. "A New Look at an Old Question: When Did the Second Whole Genome Duplication Occur in Vertebrate Evolution?" *Genome Biology.* https://doi.org/10.1186/s13059-018-1592-0.

Hughes, A. L. 1994. "The Evolution of Functionally Novel Proteins after Gene Duplication." *Proceedings of the Royal Society B: Biological Sciences.* https://doi.org/10.1098/rspb.1994.0058.

Hughes, David C. 2001. "Alternative Splicing of the Human VEGFGR-3/FLT4 Gene as a Consequence of an Integrated Human Endogenous Retrovirus." *Journal of Molecular Evolution*. https://doi.org/10.1007/s002390010195.

Hwang, Woong Y., Yanfang Fu, Deepak Reyon, Morgan L. Maeder, Shengdar Q. Tsai, Jeffry D. Sander, Randall T. Peterson, J. R.Joanna Yeh, and J. Keith Joung. 2013. "Efficient Genome Editing in Zebrafish Using a CRISPR-Cas System." *Nature Biotechnology*. https://doi.org/10.1038/nbt.2501.

Innan, Hideki, and Fyodor Kondrashov. 2010. "The Evolution of Gene Duplications: Classifying and Distinguishing between Models." *Nature Reviews Genetics* 11 (2): 97–108. https://doi.org/10.1038/nrg2689.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Kapitonov, Vladimir V., and Jerzy Jurka. 1999. "The Long Terminal Repeat of an Endogenous Retrovirus Induces Alternative Splicing and Encodes an Additional Carboxy-Terminal Sequence in the Human Leptin Receptor." *Journal of Molecular Evolution*. https://doi.org/10.1007/PL00013153.

Kondrashov, Fyodor A., and Eugene V. Koonin. 2004. "A Common Framework for Understanding the Origin of Genetic Dominance and Evolutionary Fates of Gene Duplications." *Trends in Genetics*. https://doi.org/10.1016/j.tig.2004.05.001.

Kumar, Sudhir, Glen Stecher, Michael Suleski, and S. Blair Hedges. 2017. "TimeTree: A Resource for Timelines, Timetrees, and Divergence Times." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msx116.

Lan, Xun, and Jonathan K. Pritchard. 2016. "Coregulation of Tandem

Duplicate Genes Slows Evolution of Subfunctionalization in Mammals." *Science*. https://doi.org/10.1126/science.aad8411.

Lemons, Derek, and William McGinnis. 2006. "Genomic Evolution of Hox Gene Clusters." *Science*. https://doi.org/10.1126/science.1132040.

Li, W. H., Z. Gu, H. Wang, and A. Nekrutenko. 2001. "Evolutionary Analyses of the Human Genome." *Nature*. https://doi.org/10.1038/35057039.

Li, Zheng, George P. Tiley, Sally R. Galuska, Chris R. Reardon, Thomas I. Kidder, Rebecca J. Rundell, and Michael S. Barker. 2018. "Multiple Large-Scale Gene and Genome Duplications during the Evolution of Hexapods." *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1710791115.

Lindtner, Susan, Andrei S. Zolotukhin, Hiroaki Uranishi, Jenifer Bear, Viraj Kulkarni, Sergey Smulevitch, Martina Samiotaki, George Panayotou, Barbara K. Felber, and George N. Pavlakis. 2006. "RNA-Binding Motif Protein 15 Binds to the RNA Transport Element RTE and Provides a Direct Link to the NXF1 Export Pathway." *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M608745200.

Liu, Sha, Angelique Lamaze, Qili Liu, Masashi Tabuchi, Yong Yang, Melissa Fowler, Rajnish Bharadwaj, et al. 2014. "WIDE AWAKE Mediates the Circadian Timing of Sleep Onset." *Neuron*. https://doi.org/10.1016/j.neuron.2014.01.040.

Long, Chengzu, John R. McAnally, John M. Shelton, Alex A. Mireault, Rhonda Bassel-Duby, and Eric N. Olson. 2014. "Prevention of Muscular Dystrophy in Mice by CRISPR/Cas9-Mediated Editing of Germline DNA." *Science*. https://doi.org/10.1126/science.1254445.

Luo, Ming Juan, Zhaolan Zhou, Keith Magni, Claudine Christoforides, Juri Rappsilber, Matthias Mann, and Robin Reed. 2001. "Pre-Mrna Splicing and MRNA Export Linked by Direct Interactions between UAP56 and Aly." *Nature*. https://doi.org/10.1038/35098106.

Lynch, M., and J. S. Conery. 2000. "The Evolutionary Fate and

Consequences of Duplicate Genes." *Science*. https://doi.org/10.1126/science.290.5494.1151.

Lynch, Michael, and Allan Force. 2000. "The Probability of Duplicate Gene Preservation by Subfunctionalization." *Genetics*.

Lynch, Michael, and Vaishali Katju. 2004. "The Altered Evolutionary Trajectories of Gene Duplicates." *Trends in Genetics*. https://doi.org/10.1016/j.tig.2004.09.001.

Makarova, Kira S., Nick V. Grishin, Svetlana A. Shabalina, Yuri I. Wolf, and Eugene V. Koonin. 2006. "A Putative RNA-Interference-Based Immune System in Prokaryotes: Computational Analysis of the Predicted Enzymatic Machinery, Functional Analogies with Eukaryotic RNAi, and Hypothetical Mechanisms of Action." *Biology Direct*. https://doi.org/10.1186/1745-6150-1-7.

Makino, Takashi, and Aoife McLysaght. 2010. "Ohnologs in the Human Genome Are Dosage Balanced and Frequently Associated with Disease." *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0914697107.

Makino, Takashi, Aoife McLysaght, and Masakado Kawata. 2013. "Genome-Wide Deserts for Copy Number Variation in Vertebrates." *Nature Communications*. https://doi.org/10.1038/ncomms3283.

Mali, Prashant, John Aach, P. Benjamin Stranges, Kevin M. Esvelt, Mark Moosburner, Sriram Kosuri, Luhan Yang, and George M. Church. 2013. "CAS9 Transcriptional Activators for Target Specificity Screening and Paired Nickases for Cooperative Genome Engineering." *Nature Biotechnology*. https://doi.org/10.1038/nbt.2675.

Marais, David L. Des, and Mark D. Rausher. 2008. "Escape from Adaptive Conflict after Duplication in an Anthocyanin Pathway Gene." *Nature*. https://doi.org/10.1038/nature07092.

Mauri, Federico, Ilka Reichardt, Jennifer L. Mummery-Widmer, Masakazu Yamazaki, and Juergen A. Knoblich. 2014. "The Conserved Discs-Large Binding Partner Banderuola Regulates Asymmetric Cell

Division in Drosophila." *Current Biology*.
https://doi.org/10.1016/j.cub.2014.06.059.

McLysaght, Aoife, Karsten Hokamp, and Kenneth H. Wolfe. 2002.
"Extensive Genomic Duplication during Early Chordate Evolution."
Nature Genetics. https://doi.org/10.1038/ng884.

Moriyama, Yuuta, Fumihiro Ito, Hiroyuki Takeda, Tohru Yano, Masataka
Okabe, Shigehiro Kuraku, Fred W. Keeley, and Kazuko Koshiba-
Takeuchi. 2016. "Evolution of the Fish Heart by
Sub/Neofunctionalization of an Elastin Gene." *Nature
Communications*. https://doi.org/10.1038/ncomms10397.

Naseeb, Samina, Ryan M. Ames, Daniela Delneri, and Simon C. Lovell.
2017. "Rapid Functional and Evolutionary Changes Follow Gene
Duplication in Yeast." *Proceedings of the Royal Society B: Biological
Sciences*. https://doi.org/10.1098/rspb.2017.1393.

Natalizio, Barbara J., and Susan R. Wente. 2013. "Postage for the
Messenger: Designating Routes for Nuclear MRNA Export." *Trends in
Cell Biology*. https://doi.org/10.1016/j.tcb.2013.03.006.

Nihongaki, Yuta, Fuun Kawano, Takahiro Nakajima, and Moritoshi Sato.
2015. "Photoactivatable CRISPR-Cas9 for Optogenetic Genome
Editing." *Nature Biotechnology*. https://doi.org/10.1038/nbt.3245.

O'Connell, Mitchell R., Benjamin L. Oakes, Samuel H. Sternberg,
Alexandra East-Seletsky, Matias Kaplan, and Jennifer A. Doudna.
2014. "Programmable RNA Recognition and Cleavage by
CRISPR/Cas9." *Nature*. https://doi.org/10.1038/nature13769.

Ohno, Susumu. 1970. *Evolution by Gene Duplication. Evolution by Gene
Duplication*. https://doi.org/10.1007/978-3-642-86659-3.

Papp, Balázs, Csaba Pál, and Laurence D. Hurst. 2003. "Dosage
Sensitivity and the Evolution of Gene Families in Yeast." *Nature*.
https://doi.org/10.1038/nature01771.

Pascual-Anaya, Juan, Iori Sato, Fumiaki Sugahara, Shinnosuke Higuchi, Jordi Paps, Yandong Ren, Wataru Takagi, et al. 2018. "Hagfish and Lamprey Hox Genes Reveal Conservation of Temporal Colinearity in Vertebrates." *Nature Ecology and Evolution*. https://doi.org/10.1038/s41559-018-0526-2.

Pattanayak, Vikram, Steven Lin, John P. Guilinger, Enbo Ma, Jennifer A. Doudna, and David R. Liu. 2013. "High-Throughput Profiling of off-Target DNA Cleavage Reveals RNA-Programmed Cas9 Nuclease Specificity." *Nature Biotechnology*. https://doi.org/10.1038/nbt.2673.

Perez-Pinera, Pablo, D. Dewran Kocak, Christopher M. Vockley, Andrew F. Adler, Ami M. Kabadi, Lauren R. Polstein, Pratiksha I. Thakore, et al. 2013. "RNA-Guided Gene Activation by CRISPR-Cas9–Based Transcription Factors." *Nature Methods* 10 (10): 973–76. https://doi.org/10.1038/nmeth.2600.

Ran, F. Ann, Patrick D. Hsu, Chie Yu Lin, Jonathan S. Gootenberg, Silvana Konermann, Alexandro E. Trevino, David A. Scott, et al. 2013. "Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity." *Cell*. https://doi.org/10.1016/j.cell.2013.08.021.

Rubin, Gerald M., Mark D. Yandell, Jennifer R. Wortman, George L. Gabor Miklos, Catherine R. Nelson, Iswar K. Hariharan, Mark E. Fortini, et al. 2000. "Comparative Genomics of the Eukaryotes." *Science*. https://doi.org/10.1126/science.287.5461.2204.

Scannell, Devin R., A. Carolin Frank, Gavin C. Conant, Kevin P. Byrne, Megan Woolfit, and Kenneth H. Wolfe. 2007. "Independent Sorting-out of Thousands of Duplicated Gene Pairs in Two Yeast Species Descended from a Whole-Genome Duplication." *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0608218104.

Schwank, Gerald, Bon Kyoung Koo, Valentina Sasselli, Johanna F. Dekkers, Inha Heo, Turan Demircan, Nobuo Sasaki, et al. 2013. "Functional Repair of CFTR by CRISPR/Cas9 in Intestinal Stem Cell Organoids of Cystic Fibrosis Patients." *Cell Stem Cell*.

https://doi.org/10.1016/j.stem.2013.11.002.

Sémon, Marie, and Kenneth H. Wolfe. 2007. "Consequences of Genome Duplication." *Current Opinion in Genetics and Development*. https://doi.org/10.1016/j.gde.2007.09.007.

Shalem, Ophir, Neville E. Sanjana, Ella Hartenian, Xi Shi, David A. Scott, Tarjei S. Mikkelsen, Dirk Heckl, et al. 2014. "Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells." *Science*. https://doi.org/10.1126/science.1247005.

Shen, Bin, Wensheng Zhang, Jun Zhang, Jiankui Zhou, Jianying Wang, Li Chen, Lu Wang, et al. 2014. "Efficient Genome Modification by CRISPR-Cas9 Nickase with Minimal off-Target Effects." *Nature Methods*. https://doi.org/10.1038/nmeth.2857.

Smith, Jeramiah J., and Melissa C. Keinath. 2015. "The Sea Lamprey Meiotic Map Improves Resolution of Ancient Vertebrate Genome Duplications." *Genome Research*. https://doi.org/10.1101/gr.184135.114.

Smith, Jeramiah J., Nataliya Timoshevskaya, Chengxi Ye, Carson Holt, Melissa C. Keinath, Hugo J. Parker, Malcolm E. Cook, et al. 2018. "The Sea Lamprey Germline Genome Provides Insights into Programmed Genome Rearrangement and Vertebrate Evolution." *Nature Genetics*. https://doi.org/10.1038/s41588-017-0036-1.

Sträßer, K., and E. Hurt. 2001. "Splicing Factor Sub2p Is Required for Nuclear MRNA Export through Its Interaction with Yra1p." *Nature*. https://doi.org/10.1038/35098113.

Sung, Young Hoon, In Jeoung Baek, Duk Hyoung Kim, Jisun Jeon, Jaehoon Lee, Kyunghee Lee, Daewon Jeong, Jin Soo Kim, and Han Woong Lee. 2013. "Knockout Mice Created by TALEN-Mediated Gene Targeting." *Nature Biotechnology*. https://doi.org/10.1038/nbt.2477.

Teng, I. Fang, and Stuart A. Wilson. 2013. "Mapping Interactions between MRNA Export Factors in Living Cells." *PLoS ONE*.

https://doi.org/10.1371/journal.pone.0067676.

Tesson, Laurent, Claire Usal, Seq́verine Meq́noret, Elo Leung, Brett J. Niles, Seq́verine Remy, Yolanda Santiago, et al. 2011. "Knockout Rats Generated by Embryo Microinjection of TALENs." *Nature Biotechnology*. https://doi.org/10.1038/nbt.1940.

Teufel, Ashley I., Mackenzie M. Johnson, Jon M. Laurent, Aashiq H. Kachroo, Edward M. Marcotte, and Claus O. Wilke. 2019. "The Many Nuanced Evolutionary Consequences of Duplicated Genes." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msy210.

Tintaru, Aura M., Guillaume M. Hautbergue, Andrea M. Hounslow, Ming Lung Hung, Lu Yun Lian, Jeremy C. Craven, and Stuart A. Wilson. 2007. "Structural and Functional Analysis of RNA and TAP Binding to SF2/ASF." *EMBO Reports*. https://doi.org/10.1038/sj.embor.7401031.

Wang, Haoyi, Hui Yang, Chikdu S. Shivalila, Meelad M. Dawlaty, Albert W. Cheng, Feng Zhang, and Rudolf Jaenisch. 2013. "One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering." *Cell*. https://doi.org/10.1016/j.cell.2013.04.025.

Wapinski, Ilan, Avi Pfeffer, Nir Friedman, and Aviv Regev. 2007. "Natural History and Evolutionary Principles of Gene Duplication in Fungi." *Nature*. https://doi.org/10.1038/nature06107.

Waterston, Robert H., Kerstin Lindblad-Toh, Ewan Birney, Jane Rogers, Josep F. Abril, Pankaj Agarwal, Richa Agarwala, et al. 2002. "Initial Sequencing and Comparative Analysis of the Mouse Genome." *Nature*. https://doi.org/10.1038/nature01262.

Wilkie, G. S., V. Zimyanin, R. Kirby, C. Korey, H. Francis-Lang, D. Van Vactor, and I. Davis. 2001. "Small Bristles, the Drosophila Ortholog of NXF-1, Is Essential for MRNA Export throughout Development." *RNA*.

Wood, Valerie, Antonia Lock, Midori A. Harris, Kim Rutherford, Jürg Bähler, and Stephen G. Oliver. 2019. "Hidden in Plain Sight: What

Remains to Be Discovered in the Eukaryotic Proteome?" *Open Biology*. https://doi.org/10.1098/rsob.180241.

Wu, Yuxuan, Dan Liang, Yinghua Wang, Meizhu Bai, Wei Tang, Shiming Bao, Zhiqiang Yan, Dangsheng Li, and Jinsong Li. 2013. "Correction of a Genetic Disease in Mouse via Use of CRISPR-Cas9." *Cell Stem Cell*. https://doi.org/10.1016/j.stem.2013.10.016.

Yoon, Duk Won, Heuiran Lee, Wongi Seol, Maryann DeMaria, Michael Rosenzweig, and Jae U. Jung. 1997. "Tap: A Novel Cellular Protein That Interacts with Tip of Herpesvirus Saimiri and Induces Lymphocyte Aggregation." *Immunity*. https://doi.org/10.1016/S1074-7613(00)80345-3.

Youn, Ji Young, Wade H. Dunham, Seo Jung Hong, James D.R. Knight, Mikhail Bashkurov, Ginny I. Chen, Halil Bagci, et al. 2018. "High-Density Proximity Mapping Reveals the Subcellular Organization of MRNA-Associated Granules and Bodies." *Molecular Cell*. https://doi.org/10.1016/j.molcel.2017.12.020.

Zhang, Jianzhi. 2003. "Evolution by Gene Duplication: An Update." *Trends in Ecology and Evolution*. https://doi.org/10.1016/S0169-5347(03)00033-8.

Zhang, S., K.D. Ross, G.A. Seidner, M.R. Gorman, T.H. Poon, X. Wang, E.M. Keithley, et al. 2015. "Nmf9 Encodes a Highly Conserved Protein Important to Neurological Function in Mice and Flies." *PLoS Genetics* 11 (7). https://doi.org/10.1371/journal.pgen.1005344.

Zhang, Ying, Irina A. Maksakova, Liane Gagnier, Louie N. Van De Lagemaat, and Dixie L. Mager. 2008. "Genome-Wide Assessments Reveal Extremely High Levels of Polymorphism of Two Active Families of Mouse Endogenous Retroviral Elements." *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1000007.

Zhou, Yuexin, Shiyou Zhu, Changzu Cai, Pengfei Yuan, Chunmei Li, Yanyi Huang, and Wensheng Wei. 2014. "High-Throughput Screening of a CRISPR/Cas9 Library for Functional Genomics in Human Cells." *Nature*. https://doi.org/10.1038/nature13166.

Zhou, Zhaolan, Ming Juan Luo, Katja Straesser, Jun Katahira, Ed Hurt, and Robin Reed. 2000. "The Protein Aly Links Pre-Messenger-RNA Splicing to Nuclear Export in Metazoans." *Nature*. https://doi.org/10.1038/35030160.

Zolotukhin, A. S., D. Michalowski, S. Smulevitch, and B. K. Felber. 2001. "Retroviral Constitutive Transport Element Evolved from Cellular TAP(NXF1)-Binding Sequences." *Journal of Virology*. https://doi.org/10.1128/jvi.75.12.5567-5575.2001.

Zolotukhin, Andrei S., Ralf Schneider, Hiroaki Uranishi, Jenifer Bear, Irina Tretyakova, Daniel Michalowski, Sergey Smulevitch, Sean O'Keeffe, George N. Pavlakis, and Barbara K. Felber. 2008. "The RNA Transport Element RTE Is Essential for IAP LTR-Retrotransposon Mobility." *Virology*. https://doi.org/10.1016/j.virol.2008.04.002.

Zolotukhin, Andrei S., Wei Tan, Jenifer Bear, Sergey Smulevitch, and Barbara K. Felber. 2002. "U2AF Participates in the Binding of TAP (NXF1) to MRNA." *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M107598200.

CHAPTER 2: TARGETED GENOME EDITING TO ACCELERATE REVERSE GENETICS STUDIES

Introduction

Chapter two collects my primary and collaborative efforts in reverse genetics using CRISPR/Cas9 editing. I made the first CRISPR mice at UCSD, using pronuclear injection of Cas9 mRNAs to generate mutant mice quickly and efficiently and made the protocol available to the community through our transgenic core. These first mice precisely defined – to the nucleotide – a genetic modifier of the Intracisternal A particle class of retroviral mutations in mice. Foundational work from our lab showed with conventional transgenes that a wild-derived allele of the bulk mRNA export factor, Nxf1, is able to suppress gene expression and behavioral phenotypes arising from intronic insertions of these retroviruses into certain genes. The animals that I made by genome editing confirmed NXF1 as the genetic suppressor, demonstrated that a single amino acid variant had all of the genetic activity, and that the effect was broadly applicable to strain-specific insertion polymorphisms as well as overt mutations. This experience led to numerous collaborations, and to our application of genome editing to a gene of unknown function in the lab,

Nmf9. This gene has an unusual genealogy among animal lineages. I led

efforts to generate mutations of a conserved motif encoded by homologs

in mice, flies and zebrafish in order to test genetic function across lineages

with different homolog arrangements.

RESEARCH ARTICLE

Nxf1 Natural Variant E610G Is a Semi-dominant Suppressor of IAP-Induced RNA Processing Defects

Dorothy Concepcion[1,2], Kevin D. Ross[3], Kasey R. Hutt[1], Gene W. Yeo[1,3,4], Bruce A. Hamilton[1,2,3]*

1 Department of Cellular and Molecular Medicine, Moores UCSD Cancer Center and Institute for Genomic Medicine, University of California, San Diego School of Medicine, La Jolla, California, United States of America, 2 Department of Medicine, University of California, San Diego School of Medicine, La Jolla, California, United States of America, 3 Biomedical Sciences Graduate Program, University of California, San Diego School of Medicine, La Jolla, California, United States of America, 4 Stem Cell Program, University of California, San Diego School of Medicine, La Jolla, California, United States of America

* bah@ucsd.edu

Abstract

Endogenous retroviruses and retrotransposons contribute functional genetic variation in animal genomes. In mice, Intracisternal A Particles (IAPs) are a frequent source of both new mutations and polymorphism across laboratory strains. Intronic IAPs can induce alternative RNA processing choices, including alternative splicing. We previously showed IAP IΔ1 subfamily insertional mutations are suppressed by a wild-derived allele of the major mRNA export factor, Nxf1. Here we show that a wider diversity of IAP insertions present in the mouse reference sequence induce insertion-dependent alternative processing that is suppressed by $Nxf1^{CAST}$ alleles. These insertions typically show more modest gene expression changes than de novo mutations, suggesting selection or attenuation. Genome-wide splicing-sensitive microarrays and gene-focused assays confirm specificity of Nxf1 genetic modifier activity for IAP insertion alleles. Strikingly, CRISPR/Cas9-mediated genome editing demonstrates that a single amino acid substitution in Nxf1, E610G, is sufficient to recreate a quantitative genetic modifier in a co-isogenic background.

Author Summary

Transposable elements, including endogenous retroviruses, have long been hypothesized as a substrate for creating or modulating gene regulatory networks, particularly through effects on transcription. However, several classes of elements are also known to affect alternative RNA processing events. We previously showed that the major allele of nuclear export factor Nxf1 in Mus musculus castaneus mice acts as a semi-dominant suppressor of de novo mutations caused by intracisternal A particle (IAP) endogenous retroviruses that integrate into introns, disrupting normal RNA processing. Here we show that this suppressor allele of Nxf1 can coordinately modify gene expression phenotypes at several

endogenous loci in the C57BL/6 mouse reference genome that contain IAP sequences in their introns. This quadruples the number of known insertional events modified by *Nxf1* and extends the effect beyond overt mutations. We previously used transgenic mice and viral vector mediated overexpression to demonstrate *Nxf1* as the modifier gene for de novo insertions. Here we use direct genome editing in mouse one cell embryos to create custom germline alleles at the endogenous *Nxf1* locus to show that a specific amino acid substitution, E610G, quantitatively accounts for the *Nxf1* modifier gene activity.

Introduction

Endogenized retroviruses and other molecular parasites frequently influence expression of host genes at sites of insertion. Chromosomal insertions of these mobile elements can alter initiation, splicing, or termination of host gene transcripts, in quality or amount. Remnants of ancient insertion and transposition events that survived selection are thought to have shaped gene expression patterns in modern animals substantially [1–3]. In populations where mobile elements remain highly active, such events can account for a substantial fraction of functional polymorphism and spontaneous mutations. Two families of elements in laboratory mice, Intracisternal A Particle (IAP) and MusD/Early Transposon (ETn) families account for 10–20% of spontaneous mutations [4–6], depending on strain background [6]. While some of these mutations interrupt coding exons [7] or induce novel patterns of transcription [8–11], the majority comprise intronic insertions that introduce alternative splicing or transcript termination (or both), resulting in quantitative loss of normal host gene products.

Alternative processing of nascent transcripts is regulated at several levels [12]. Pre-mRNA splicing typically occurs co-transcriptionally, regulated by a variety of DNA and RNA binding factors that together defines and acts on constitutive exons. Transcriptional initiation complexes may assemble splicing factors on the Pol II complex, resulting in promoter-dependent alternative splicing [13–15]. Elongation rate of the RNA polymerase complex may influence alternative splicing by regulating the appearance of downstream acceptor sites relative to the splicing kinetics for weaker upstream sites [16, 17]. In addition, recent single-molecule imaging data supports post-transcriptional splicing for at least some alternative splice sites [18]. Identification of post-transcriptional alternative splicing events suggests opportunities for regulation by nuclear ribonucleoprotein (RNP)-associated proteins that are not normally found in the nuclear speckles associated with constitutive splicing.

We previously reported a wild-derived variant of the mouse mRNA nuclear export factor gene *Nxf1* within the *Modifier-of-vibrator-1* (*Mvb1*) locus as a genetic modifier for each of six mutations caused by insertion of an IAP element into an intron of a host gene [19, 20]. The suppressing allele, *Nxf1^CAST^*, defines the major haplotype in wild isolates of *Mus musculus castaneus*. For each of the six mutations, genetic suppression altered the balance of alternative RNA processing between virus-dependent isoforms and splicing to the 3' exon. This mechanism appeared highly selective, operating on sense-strand IAPs (six out of seven tested), but not MusD/ETn (0/6), MuLV (0/1), VL30 (0/1), or L1 LINE (0/1) insertions. These studies also identified an exception to the rule: *Atrn^mgL^*, which includes a "full-length" IAP, was not suppressed, in contrast to the more frequent IA1 class [21] in the six suppressed mutations [19]. We proposed that sequences deleted in IA1 elements might in full-length elements mediate additional repressive events that are epistatic to the apparent splicing defects modified by *Nxf1^CAST^*, but acknowledged that local genomic context could also play a role. We also proposed that *Nxf1^CAST^* effects on alternative splicing are highly specific, but genome-wide analysis

of *Nxf1*^CAST^ effects on either general alternative splicing or alternative splicing at genes whose reference allele includes an IAP element have not been reported.

Here we show that *Nxf1*^CAST^ has modifier gene activity toward IAP insertion alleles present in the C57BL/6J (B6) reference genome–this includes several full-length IAPs, an IAP in *Adamts13* with a novel deletion but otherwise high sequence similarity to the non-suppressed element in *Atrn*^mgL^, and two other IAP deletion classes outside of the IΔ1 group. Quantification of well-characterized alternative splicing events, including cassette exon, retained intron, alternative splice donor, and alternative splice acceptor sites detect no *Nxf1* modifier effect at non-IAP introns, further supporting the specificity of *Nxf1*^CAST^ activity toward IAP insertion alleles. We also demonstrate by genome editing that a single nucleotide substitution, encoding a glutamate to glycine amino acid replacement in the carboxyterminal UBA-like domain, is sufficient to explain the semi-dominant modifier phenotype of *Nxf1*^CAST^. These results expand the range of IAP family elements modified by *Nxf1* variant alleles and identify a single residue in Nxf1 protein whose effect on stability, kinetics, or interactions explain the modifier mechanism.

Results

Nxf1^CAST^ modifies *Adamts13*^S^, an atypical deleted-IAP allele in B6 mice

Because congenic *Nxf1*^CAST^ showed modifier activity against several insertions of the IAP IΔ1 subfamily, but not against a single example of a full-length element [19, 20], we tested its activity on a well-studied IAP allele of a different class (Fig. 1). *Adamts13* encodes a large circulating protease responsible for processing multimeric von Willebrand factor (vWF) in vivo [22, 23]; mutations in human *ADAMTS13* cause thrombotic thrombocytopenic purpura [24] and variations in ADAMTS13 activity are associated with other thrombotic abnormalities [25, 26]. In the *Adamts13*^S^ allele present in several inbred strains, an IAP insertion into intron 24 results in an alternatively spliced, stable transcript that terminates in viral sequences, but lacks downstream exons encoding the final two thrombospondin repeats and two CUB (C1r/C1s, Uegf, Bmp1) domains [27]. A very small amount of residual exon 24 to exon 25 splicing can be detected by a quantitative reverse transcription–polymerase chain reaction (qRT-PCR) assays (Ref. 27 and Fig. 1A). The *Adamts13*^S^ IAP sequence is highly similar to the *Atrn*^mgL^ IAP (Fig. 1B), but differs in having a 1131-bp deletion in the *pol* coding region (Fig. 1B), distinct from and smaller than the classical IΔ2 deletion [21]. Of 2329 bp among 107 aligned sites that distinguish the *Atrn*^mgL^ IAP from the IΔ1 elements, only 33 sites (61 bp) are not shared between *Atrn*^mgL^ and *Adamts13*^S^ IAPs.

Congenic *Nxf1*^CAST^ alleles increased the amount of *Adamts13* RNA that is correctly spliced from exon 24 to exon 25. Since *Adamts13* is most robustly expressed by hepatic stellate cells [28, 29], we assayed RNA from replicate sets of liver samples. Measurements in previously collected samples from 29 F2 progeny of a BALB/c X C57BL/6J–*Nxf1*^B6^/*Nxf1*^CAST^ intercross [19], in which the *Adamts13*^S^ allele was fixed, showed a significant increase in full-length *Adamts13* relative to standard reference genes (*Gapdh* or *Ppia*) in animals with the *Nxf1*^CAST^ allele, although variance in the measurement was unusually high (Fig. 1C). To clarify the effect of *Nxf1*^CAST^, we examined qRT-PCR from liver RNA of C57BL/6J–*Nxf1*^CAST^ congenic stock, using *Desmin* as a cell type-specific reference gene to control for proportion of hepatic stellate cells. This confirmed an increase in fully spliced *Adamts13* in *Nxf1*^CAST^ animals, with much lower variance (Fig. 1D). The two experiments together indicate ~2-4-fold elevation in abundance of full-length *Adamts13* in *Nxf1*^CAST^ compared with *Nxf1*^B6^.
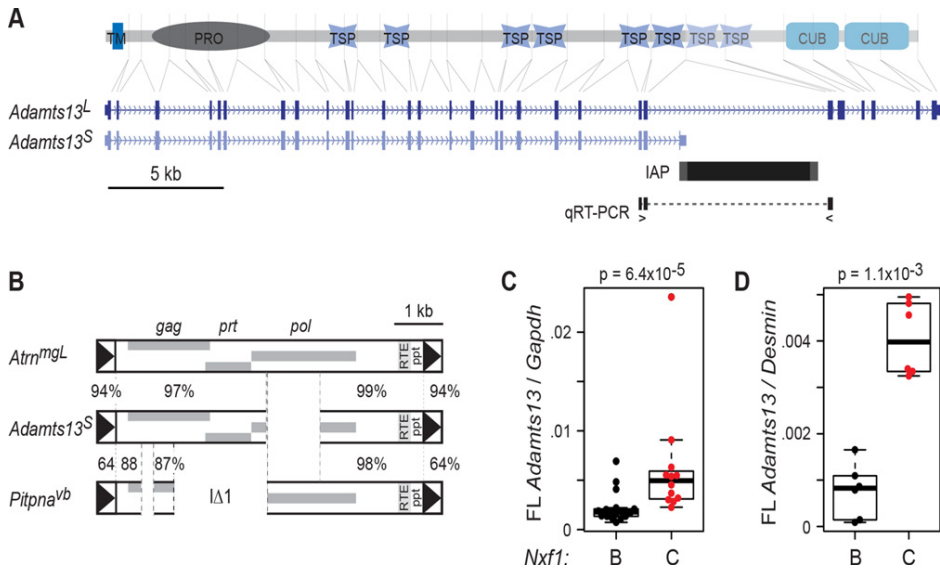
Fig 1. *Adamts13^S* is suppressed by congenic *Nxf1^CAST*. **(A)** Alignment of Adamts13 protein domains to their corresponding exons in *Adamts13^L* and *Adamts13^S* shows loss of two thrombospondin (TSP) and two CUB domains from *Adamts13^S*, which terminates in an intronic IAP sequence. TM, transmembrane segment; PRO, protease domain. Locations of primers to detect exon 24 to exon 25 splicing around the IAP are indicated. **(B)** The *Adamts13^S* IAP shares greater sequence similarity with the *Atrn^mgL* full-length IAP than with IΔ1 elements, an example of which is shown. Percent nucleotide identity is shown for indicated segments. **(C)** Quantitative RT-PCR for exon 24-exon 25 splice junction relative to *Gapdh* in B6 x BALB/c F2 mice homozygous for both *Adamts13^S* and the indicated congenic allele of *Nxf1* shows substantial increase in the *Nxf1^CAST* congenic allele (p = 6.4x10^-5, Wilcoxon rank sum test). **(D)** Similar results were obtained for a smaller number of animals from the B6 congenic line (p = 1.1x10^-3, Wilcoxon rank sum test). Normalization to *Desmin* as a cell type-specific marker increased the separation of values. Each dot in the box plots represents the mean of technical replicates for one biological sample.

Nxf1^CAST suppresses IAPs that correlate with reduced expression

The C57BL/6J reference genome includes more diverse examples of sense-oriented IAPs in host gene introns. To allow a more comprehensive survey of IAP structural classes that can be modified by *Nxf1^CAST*, we systematically identified IAP elements using a combination of sequences that are conserved across IAP subfamilies and RepBase annotations (S1 Fig). Comparison of IAP and host gene strand orientation confirms the well-known orientation bias [20, 30] against sense-strand insertions (323 sense vs. 969 antisense) and shows a significant difference in size distribution, with sense-oriented insertions having a larger proportion of elements ≤ 1 kb (210/323, compared with 525/969 antisense; p = 7.3x10^-4, Fisher's exact test). The set of sense-oriented insertions was manually curated to remove non-intronic events (mostly short IAP sequences in 3' ends that were otherwise homologous with 3' ends in corresponding human or rat genes) and likely annotation errors (mostly tandem gene duplications with the IAP placed between duplicate copies). Of these, 85 were > 3 kb, 58 of which had been previously reported [31] (S2 Fig).

Intronic IAP elements in the reference genome predict decreased expression relative to strains that lack the insertion (Fig 2). We designed unique qRT-PCR assays for exon-exon splicing across 49 IAP-containing introns, of which 41 passed initial quality control measures

A

B

Gene	IAP Length	Inserted Strains	Uninserted Strains	U/I (p)	C/B (p)	Gel (p)
Agtr1a	7107	ABL	S	1.5 (0.01)	1.2 (0.01)	
Cpne8	7154	ABS	L	1.5 (0.003)	1.1 (0.03)	1.1 (0.001)
Fhit	7096	ABL	S	2.3 (0.002)	1.1 (0.03)	
Kcns3	7079	B	ALS	1.0 (0.6)	1.1 (0.14)	
L3mbtl4	7268	B	ALS	1.2 (0.04)	1.7 (0.04)	
Lphn3 (TM)	7103	AB	LS	0.69 (0.13)	0.90 (0.82)	
Me3	7200	B	ALS	$3.2\ (6\times10^{-6})$	1.2 (0.009)	$1.2\ (5\times10^{-4})$
Nsdhl	7154	BS	AL	1.3 (0.02)	1.2 (0.002)	
Nudcd3	7119	B	ALS	$1.6\ (6\times10^{-6})$	0.95 (0.74)	
Pde4d	7367	ABS	L	1.1 (0.31)	0.91 (0.94)	
Pla2g4e	7089	B	ALS	$4.9\ (6\times10^{-6})$	$1.4\ (5\times10^{-4})$	
Scg5	7065	BS	AL	$3.2\ (4\times10^{-5})$	1.1 (0.08)	
Slc20a2	7126	B	ALS	$3.6\ (1\times10^{-5})$	1.1 (0.12)	
Tiam1	7254	ABLS	-	NA	0.94 (0.84)	
Tmem163	7227	B	ALS	0.72 (0.005)	1.1 (0.27)	
Zfp69 (TM)	7105	B	ALS	$176\ (3\times10^{-5})$	7.8 (0.02)	
Adamts13^{S}	6020	ABL	S	~1000	$>2\ (\leq10^{-3})$	
Agbl4	5408	B	ALS	1.9 (0.01)	1.5 (0.04)	
Slc15a2	5298	BS	AL	$5.5\ (2\times10^{-4})$	1.7 (0.008)	
Snx29	5295	ABL	S	$1.6\ (10^{-3})$	1.2 (0.12)	
Trpc6 (TM)	5337	ABL	S	0.49 (0.09)	1.4 (0.002)	1.1 (0.04)
Rgs3	5471	B	ALS	0.90 (0.51)	0.89 (0.98)	
Csmd1	5397	ABL	S	1.2 (0.7)	1.01 (0.35)	
Dph5 (TM)	5313	B	ALS	1.4 (0.009)	1.2 (0.08)	
Cntn5	5371	ABLS	-	NA	1.1 (0.07)	
Gbe1	5349	B	ALS	1.15 (0.22)	1.15 (0.02)	
Rftn1	4357	B	ALS	1.04 (0.2)	1.02 (0.32)	
Adk	5541	ABS	L	0.71 (0.49)	1.1 (0.50)	
Exoc4	5527	ABL	S	0.78 (0.30)	0.97 (0.79)	
Cdh19	4831	B	ALS	1.2 (0.22)	1.1 (0.016)	
Slc9a2	4456	BL	AS	1.3 (0.025)	1.02 (0.51)	
Spon1	4330	ABS	L	1.02 (0.96)	1.03 (0.42)	
Slc25a21	4321	ABL	S	0.93 (0.83)	1.02 (0.29)	
Alk	4321	ABL	S	1.1 (0.27)	1.08 (0.37)	
Col23a1	4321	ABLS	-	NA	1.10 (0.48)	
Lhfpl3	3504	AB	LS	$1.5\ (8\times10^{-4})$	1.04 (0.45)	
Sntg1	4175	ABS	L	$1.7\ (1\times10^{-5})$	1.1 (0.18)	
Phf20	3410	AB	LS	$1.4\ (2\times10^{-5})$	1.02 (0.21)	
Rsrc1	2239	ABS	L	0.87 (0.26)	1.2 (0.80)	
Amph1	349	ABLS	-	NA	1.1 (0.18)	
Bdh (TM)	402	B	ALS	0.99 (0.95)	1.02 (0.39)	
Dtd1	414	ABL	S	0.74 (0.02)	1.01 (0.32)	
Enoph1	499	B	ALS	0.92 (0.68)	1.04 (0.43)	
Tmco3	373	ABLS	-	NA	1.02 (0.45)	

Fig 2. *Nxf1*^{CAST} congenic allele suppresses a diversity of IAP elements in the B6 genome. (A) Schematic diagrams illustrate sequence compositions of IAP elements in this study. Deleted segments in each elements compared with full-length consensus are shown at 10% opacity. **(B)** For each structural class, genes whose expression level was tested are indicated. Genes measured with TaqMan assays are indicated by (TM); all others measured by real-time fluorescence stimulation with SYBR green. Gene symbols are color coded: grey for no significant differences between strains that have (Inserted) and strains that lack (Uninserted) each IAP, black for no evidence of suppression by *Nxf1* of an observed strain difference, brown for a non-significant trend toward suppression, and red for significant evidence of suppression by *Nxf1*^{CAST}. Strains that include (Inserted) or lack (Uninserted) each IAP are indicated: A, A/J; B, C57BL/6J (B6); L, BALB/cJ; S, 129S1/SvImJ. Relative expression ratios for each strain are expressed as Uninserted/Inserted (U/I). Relative expression ratios for *Nxf1* congenic animals homozygous for either *Nxf1*^{CAST} (C) or *Nxf1*^{B6} (B) allele are expressed as C/B, such that insertions suppressed by the *Nxf1*^{CAST} allele have values >1. P-values from a two-tailed Wilcoxon-Mann-Whitney test for strain difference (U/I, unpaired samples) or a one-tailed Wilcoxon signed rank test for *Nxf1*^{CAST}-dependent suppression (C/B, littermate paired samples) are given in parentheses. Three genes were also confirmed by fluorescence intensity of isoform-specific bands on agarose gels using 3-primer assays designed independently of the qRT-PCR assays.

Table 1. Suppression of distinct IAP structural classes by *Nxf1^CAST*.

IAP class*	Group-wise p**	FDR***
>7 kb	2.1×10^{-7}	1.1×10^{-6}
IΔ1	3.4×10^{-4}	8.6×10^{-4}
Δpol[1]	0.43	0.50
Solo LTR	0.50	0.50
other	0.0053	0.0088

* Full list of elements by class is given in S1 Table

** p-values from the Wilcoxon rank sum test for all paired measurements in the class

*** False Discovery Rate q-value from the Benjamini-Hochberg method.

[1] Deletions that remove at least half of the *pol* ORF, but less than half of the adjacent *prt* ORF.

doi:10.1371/journal.pgen.1005123.t001

(4 solo LTRs and 36 larger elements). Commercial hydrolysis probe (TaqMan) assays were purchased for 5 additional sites (1 solo LTR and 4 larger elements). In draft assemblies of A/J, BALB/c, and 129S1 genomes available from the Wellcome Trust Sanger Institute [32, 33], 39 insertions >3 kb and all 5 solo LTR insertions were absent in at least one strain. While other factors may influence strain-dependent expression of any given locus, the presence of sense-strand insertions >3 kb correlated with lower levels of the assayed splice junction. Among 39 polymorphic sites, 20 had significantly lower expression in strains with the insertion, 2 had nominally significant differences in the opposite direction, and 17 were not significantly different. Effect sizes of significant difference ranged from 1.2 to 1000-fold in relative expression. These results confirmed that intronic IAPs fixed in the B6 genome correlate with reduced expression of the correctly processed RNA.

Congenic *Nxf1^CAST* alleles increased relative expression of processed RNA at these same sites. We assayed correctly spliced expression at 44 sites in paired brain samples homozygous for either *Nxf1^B6* or *Nxf1^CAST* alleles (Fig. 2). The effect of *Nxf1* genotype on level of correctly spliced RNA was independently significant at 14 sites, including full-length, IΔ1, and other deletion classes of IAPs. Class-specific statistics and false discovery rates are listed in Table 1 and S1 Table. Notably, *Nxf1^CAST* also suppressed the effect of an IAP in an AT/AC intron in *Gbe1*, suggesting that suppression is not dependent on the major spliceosome. The magnitude of statistically supported effects ranged from 1.1 (*Cdh19*) to 7.8 (*Zfp69*) in these assays. Furthermore, pooled data from full-length and IΔ1 insertion sites that showed a significant difference between inserted and uninserted strains, but not individually between *Nxf1* alleles, nonetheless showed group-wise significance for suppression by *Nxf1^CAST* (p = 0.0012, Wilcoxon signed rank test with continuity correction). These results demonstrate that the *Nxf1* congenic alleles have genetic modifier activity toward a diverse set of IAP elements and genomic contexts.

Nxf1 genotype does not influence alternative splicing choice at non-IAP sites

To test further the specificity of *Nxf1* variant effect on alternative splicing, we examined 12 alternative splicing choices at well-characterized alternative splice sites [34] and alternative splice sites in genes with known functional connections to Nxf1, reasoning that these might be most likely to be influenced by variation in *Nxf1*. None of these introns contain sense-oriented IAP sequences. RT-PCR competition assays that amplify both alternative products were used and the relative ratio of alternative products in each sample quantified by fluorescence imaging
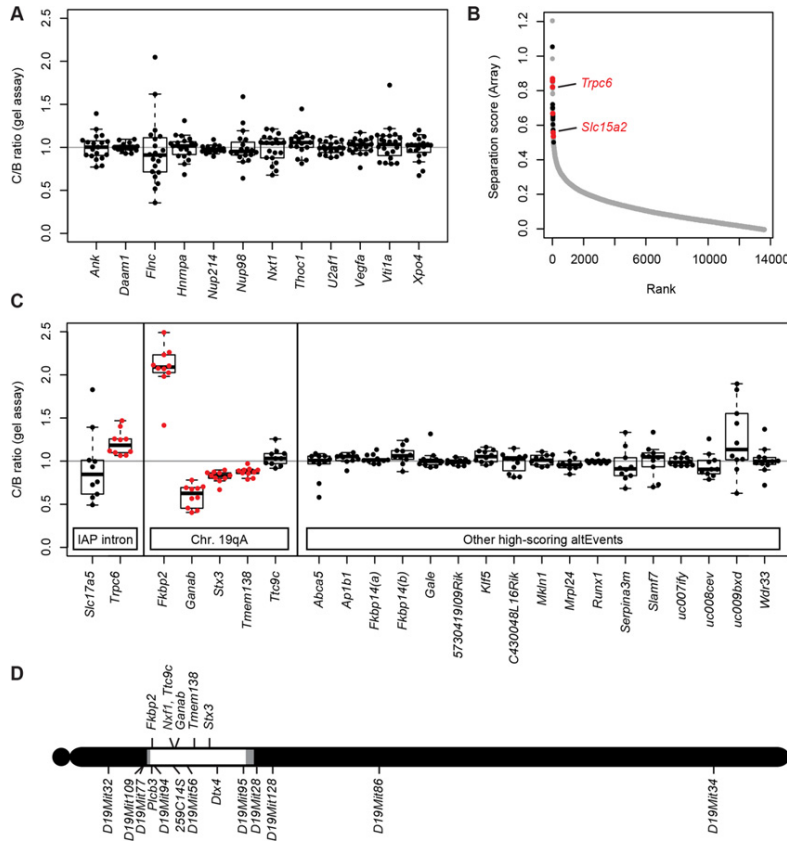
Fig 3. Nxf1 alleles do not affect alternative processing at non-IAP sites. (A) The ratio of alternative splice forms for 12 candidate genes was measured by in-gel fluorescence after PCR from congenic $Nxf1^{CAST}$ or $Nxf1^{B6}$ allele homozygotes. Box-and-whiskers plots show the ratio ratio between paired samples of alternate genotypes (C/B ratio). Ratios of individual sample pairs are plotted as points over each box. **(B)** Three samples of each genotype were hybridized to Affymetrix arrays to interrogate 13,002 alternative RNA events. Absolute value of the separation score (log_2 of the inclusion/exclusion rate for *Nxf1* C/B paired samples) between alleles is plotted by rank across all probe sets. Events that were confirmed in gene-specific assays on additional samples are indicated in red, events that were tested but not confirmed are in black. Positions of *Trpc6* and *Slc15a2* events tested in Fig 2 are indicated. **(C)** Ratios between paired samples, as in panel A, for alternative splicing events predicted by the arrays. Significant differences, plotted in red, are either events at IAP-containing introns or in genes within the Nxf1 congenic region on chromosome 19A. **(D)** Diagram of chromosome 19 shows the positions of genes with alternative outcomes in panel (C) relative to the congenic interval homozygous for CAST/EiJ alleles in all $Nxf1^{CAST}$ samples (white), intervals between informative markers or heterozygous in some samples (gray), and homozygous for B6 alleles (black).

after gel electrophoresis (Figs. 3A and S3). No significant differences were observed among 12 candidate sites tested.

As a more general test of modifier activity on non-IAP alternative splicing choices, we measured brain RNA samples from three same-sex littermate pairs from the *Nxf1* congenic stock using genome-wide splicing-sensitive microarrays (Fig. 3B). The Affymetrix MJAY arrays

interrogated 13,002 known or imputed alternative junctions in mouse genes, including alternative splice site choices and alternative 5' and 3' ends. Approximately 1868 events were called with high confidence (q≤0.05) in this experiment. Using established computational methods [35] to analyze this data, we identified only 43 sites with significant deviations between paired samples (Separation score absolute value >0.5 [~1.4-fold change] and False Discovery Rate q<0.05). Of these, 10 were within the CAST/Ei congenic interval on chromosome 19; due to the high polymorphism rate between B6 and CAST, these are likely due to either mismatch with the probe sequence or to intrinsic allelic differences independent of *Nxf1*. Interestingly, three of the remaining 33 sites included sense-oriented IAP insertions in the corresponding intron: *Trpc6*, *Slc15a2*, and *Slc17a5*, although most IAP-related alternative RNA junctions were not represented in the array.

To test whether the sites implicated by the array analysis were replicable differences rather than probe sequence polymorphisms or statistical fluctuation across a large number of measurements, we designed gel-based RT-PCR assays across each intron for 12 of the most extreme scores on each end of the distribution (Fig. 3C). Of these, four sites tested from chromosome 19 (*Fkbp2*, *Ganab*, *Stx3*, and *Tmem138*), showed significant differences, suggesting an intrinsic processing difference between B6 and CAST/Ei alleles in those genes. Of three IAP-containing introns detected by the arrays *Slc15a2* and *Trpc6* were confirmed by either qRT-PCR (Fig. 2) or both qRT-PCR and the gel assay, while *Slc17a5* could not be confirmed. Thus, 6 out of 8 IAP or chromosome 19 sites identified by the MJAY arrays replicated in an independent assay. By contrast, none of 17 assays for unlinked, non-IAP splice sites showed a nominally significant difference between *Nxf1* alleles. Lack of replication for these events, compared with very high replication rates for IAP-containing events and sites within the *Nxf1*-congenic interval on chromosome 19 (Fig. 3D), further suggest that very few if any non-IAP sites are sensitive to this *Nxf1* allele, consistent with discovery rates seen in previous experiments with this array platform [36–39]. While these data cannot rule out the possibility of a non-IAP alternative processing target for *Nxf1* modifier activity at an unmonitored site or below the limit of detection, together they show that $Nxf1^{CAST}$ modifier activity does not act broadly on alternative splicing and provides further evidence for very strong IAP selectivity.

Construction of *Nxf1* E610G mice by genome editing in mouse one cell embryo

The precise genetic variant or variants responsible for the modifier activity of *Nxf1* have not been previously determined. As *Nxf1* steady-state RNA and protein levels did not appear different between alleles, we reasoned that one or both of two amino acid substitutions that differentiate the alleles was likely causal. Because the E610G variant occurred in a highly-conserved domain and appeared to be the last variant that arose in a 19-marker haplotype before that haplotype rose to high frequency in wild mice [20], we targeted this site for genome editing in mouse one cell embryos, using a synthetic Cas9 mRNA and a single guide RNA (Fig. 4A) co-injected with an oligonucleotide template for homology-dependent repair (Fig. 4B), essentially as described [40]. Two potential founders survived. Each survivor was male and heterozygous for two distinctly edited alleles–(1) a correctly edited allele containing both the E610G polymorphism and a silent polymorphism in the sgRNA homology to deter further cleavage, and (2) either a "pseudo-edited" allele, carrying the induced silent polymorphism, or the E610G edit together with a 3-bp deletion of the adjacent valine codon (Fig. 4C). Both males bred and transmitted each allele. All four transmitted alleles were detected by PCR sequencing tail-clip DNA from the founders and their offspring. Sequencing of 17 predicted [41] off-target cleavage
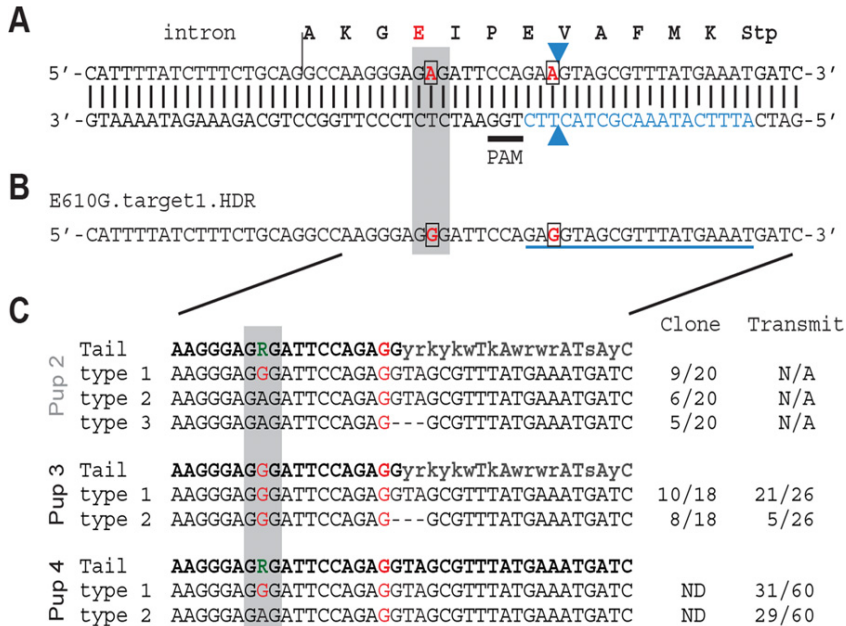
Fig 4. *Nxf1* E610G genome editing in mouse embryos. (A) Translation and genomic sequence surrounding E610G targeted for editing. Codon 610 is shaded grey, targeted bases are in red and boxed, guide RNA sequence is in blue, the presumed Cas9 cut site shown by arrowheads, and the PAM sequence is underlined. **(B)** Partial sequence of the repair template oligonucleotide shows the two targeted base changes, including one in the guide RNA sequence. **(C)** Sequences obtained by PCR from tail DNA of one dead (Pup 2) and two live (Pups 3 and 4) G0 pups. Sequence from cloned PCR products (Pups 2 and 3) confirmed an equivalent 3-bp deletion (ΔVal) on one allele in each animal. Sequence from offspring confirmed transmission of each edited allele from the two viable founders.

sites from the same samples identified no additional mutations, suggesting that off-target editing was not frequent in either founder.

Nxf1 E610G edited allele recapitulates the *Modifier of vibrator*

The hypothesis that the E610G polymorphism is sufficient for the original modifier phenotype makes three explicit predictions: that the edited allele should suppress *vibrator* phenotypes relative to the background *Nxf1^{B6}* (and silent, pseudo-edited) allele, that it should act semi-dominantly (and therefore be measurable in heterozygotes), and that it should be indistinguishable from the congenic *Nxf1^{CAST}* allele in magnitude of effect. We examined 68 sequential *vibrator* (*Pitpn^{vb/vb}*) mutant and 47 control G2 animals in backcrosses of our edited alleles to the B6–*Pitpna^{vb/+}*, *Nxf1^{B6/CAST}* stock (Fig. 5A). The E610G edited allele, but not the pseudo-edited control allele, altered the tremor phenotype of *vibrator* mutants–scored by observers blind to genotype–relative to the *Nxf1^{B6}* allele (S1 Video and S2 Video). This held for heterozygotes in trans to the *Nxf1^{B6}* allele ($p = 7.9 \times 10^{-6}$, Wilcoxon Rank Sum Test) and in trans to the
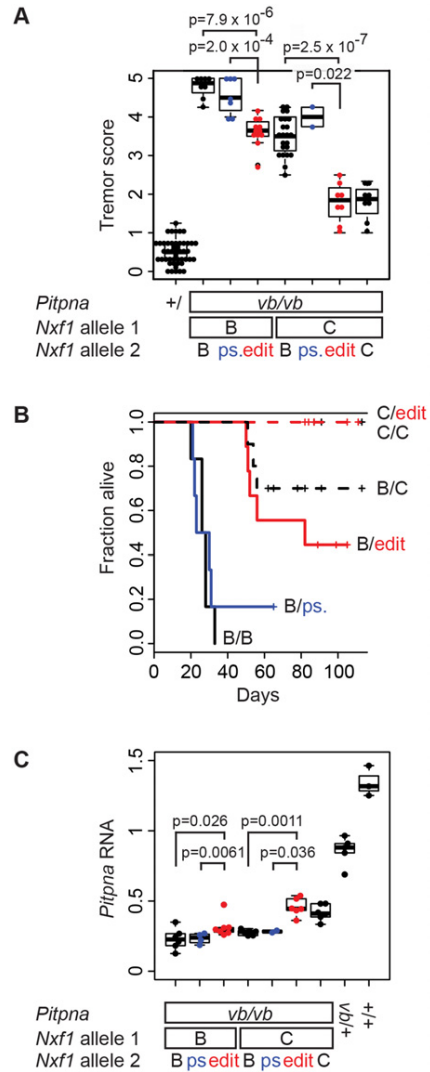
Fig 5. E610G variant accounts for *Modifier-of-vibrator* activity of *Nxf1*. (A) The E610G edited allele decreased tremor severity. Tremor scores for *vibrator* mutant and littermate control animals of the indicated genotypes are plotted. Each point represents the mean observer score for one animal. P-values for the single hypothesis tests that one copy of the edited allele significantly reduced tremor relative to the unedited or pseudo-edited (ps.) B6 allele are shown. **(B)** The E610G edited allele increased genotype-dependent survival of *vibrator* animals. Kaplan-Meier plots show fraction alive over a ~3-month observation period. Heterozygotes for edited alleles are drawn in red, pseudo edited in blue. Censored animals are indicated with vertical lines. **(C)** Brain *Pitpna* gene expression from *vibrator* mutant alleles was increased by E610G alleles.

Measurements were made from animals in (A) not used in (B). Relative quantities compared to three reference genes (*Gapdh*, *Sdha*, and *Ppia*) are plotted. Individual points represent the average of three replicate measurements per biological sample.

$Nxf1^{CAST}$ allele (p = 2.0×10^{-4}). The E610G edited allele was indistinguishable from the congenic $Nxf1^{CAST}$ allele in each context (Fig. 5A).

In a subset of animals followed for survival analysis, the edited allele prolonged the viability of *vibrator* mutants to the same degree as the original congenic allele, while the pseudo-edited allele did not (Fig. 5B). A Cox proportional hazards model showed significantly better survival with the edited allele (p = 0.0013) but not the pseudo-edited allele (p = 0.27) compared with the parental $Nxf1^{B6}$ allele when heterozygous with respect to $Nxf1^{B6}$. Similarly, among animals heterozygous to the congenic $Nxf1^{CAST}$ allele the observed deaths all occurred in unedited animals. Survival in unedited animals was different from non-mutant control and $Nxf1^{CAST}$ homozygotes (p = 0.046) under the proportional hazards model, while edited animals were not (p>0.99).

Most of the backcross animals were processed for gene expression analysis. The edited allele increased the expression of correctly spliced *Pitpna* mRNA from *vibrator* mutant alleles, again comparable by qRT-PCR from brains of the same animals (Fig. 5C). Non-parametric tests supported significantly increased expression in each of the four predicted pair-wise comparisons and in each combination the edited allele was indistinguishable from the congenic allele. These data together show that the single amino acid substitution E610G is sufficient to account quantitatively for the previously observed *Modifier of-vibrator* (*Mvb1*) activities of the $Nxf1^{CAST}$ allele in vivo.

Nxf1 E610G edited allele is a general suppressor of IAP insertions

The discovery of B6-endogenous targets of the modifier activity in our original congenic stock (Figs. 1 and 2) predicts that if E610G amino acid substitution allele explains the modifier activity, it should act on expression level of those genes as well. We tested this prediction with 12 qRT-PCR assays on F1 samples that were heterozygous for the edited allele, but without potentially confounding $Pitpn^{vb}$ mutations (Fig. 6). Both *Adamts13* and *Nsdhl* showed suppression by the edited allele equivalent to the congenic allele in liver, where both are well-expressed (Fig. 6A). In brain cDNA, six of seven genes that were suppressed by the congenic stock (*Agbl4*, *Agtr1a*, *Cdh19*, *L3mbtl4*, *Pla2g4e*, and *Slc15a2*) showed effects of the heterozygous edited allele, while the seventh (*Fhit*) did not approach independent significance in the heterozygote samples (Fig. 6B). In addition, two of three genes for which the congenic experiments showed a trend but did not provide significant support (*Sntg1* and *Dph5*) also showed significant evidence of suppression by the edited allele (Fig. 6B). Each event survived correction for false discovery rate below or near q = 0.05 (S1 Table). The tested sites, including elements closely related to the non-suppressed insertion in $Atrn^{mgL}$, other full-length elements, IΔ1 elements, and other deletion classes, were significantly increased in a semi-dominant manner by the edited allele. Alternative processing events on chromosome 19 that differed between congenic stocks (Fig. 3C,D) were not significantly different between edited and unedited alleles (S4 Fig), confirming that they represented strain variations in each gene rather than effects of the modifier locus. Together, these measurements confirm that the E610G variant is able to suppress IAP-related gene expression changes at the wide array of sites detected for the congenic allele by our genome-wide screen and explains the modifier effect of previous congenic and transgenic strains [19, 20].
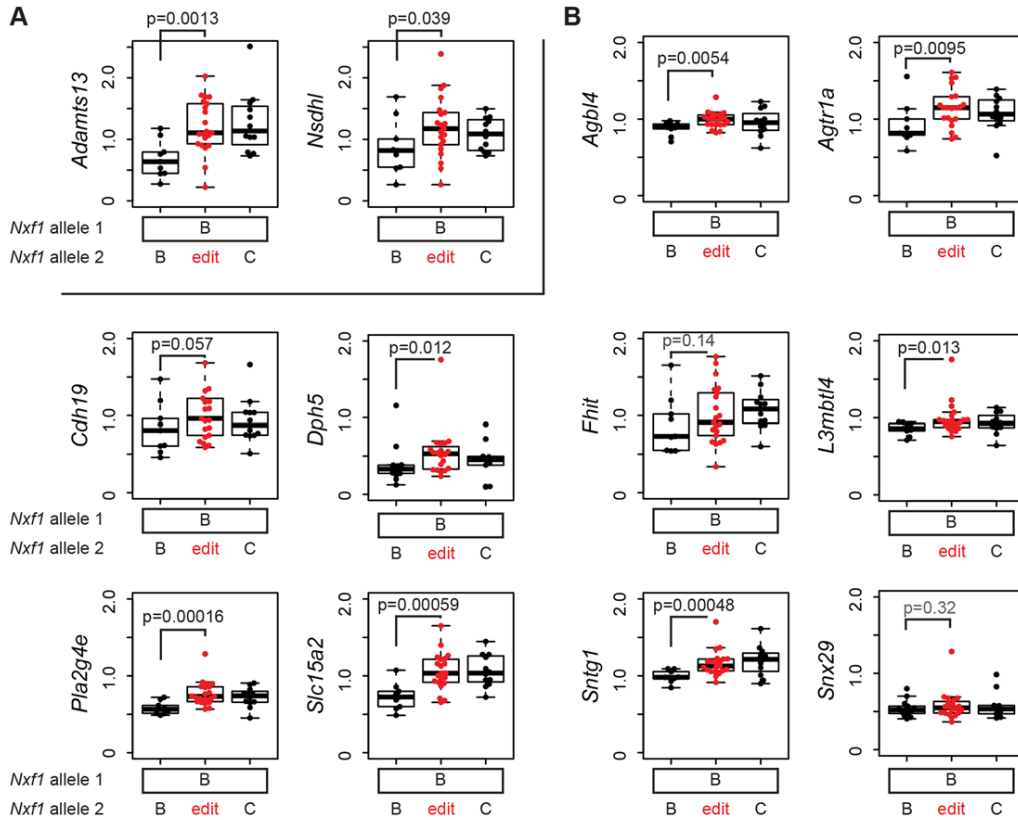
Fig 6. E610G replicates the modifier effect on IAP insertions in the B6 genome. (A) Relative expression levels of *Adamts13* and *Nsdhl* RNAs with the IAP-containing intron correctly spliced were measured by qRT-PCR from livers of animals heterozygous for the indicated *Nxf1* congenic (B, *Nxf1^B6^*; C, *Nxf1^CAST^*) or E610G edited allele. Expression measures are relative to five reference genes (*Desmin*, *Gapdh*, *Sdha*, *Ppia*, *Pitpna*) and divided by average relative value. Individual points represent the average of three replicate measurements per biological sample. **(B)** Similar measurements from brain cDNA of the same animals for the indicated genes. Measures are relative to four reference genes (*Gapdh*, *Sdha*, *Ppia*, *Pitpna*). P values are shown for a one-tailed Wilcoxon-Mann-Whitney test for the predicted increase in E610G edited animals relative to B6 homozygotes.

doi:10.1371/journal.pgen.1005123.g006

Most of the IAP sequence is not required for mutation or suppression by *Nxf1*

We next asked whether identifiable sequence features correlated with expression level in inserted strains or response to *Nxf1* genotype. The ratio of expression in uninserted vs. inserted strains (U/I ratio) was nominally correlated with IAP length (Spearmen's ρ = 0.32, p = 0.05) primarily due to lack of effect among LTR-only events and extreme values for *Adamts13* and *Zfp69* (S5 Fig). Intron size, flanking exon size, distance between the element and each exon, and annotation features of the exons for potential alternative events had no significant correlation.

The expression ratio between $Nxf1^{CAST}$ vs. $Nxf1^{B6}$ alleles (C/B ratio) was positively correlated with U/I ratio (Spearman's $\rho = 0.44$, $p = 0.0058$). Intron length, distance from 5' exon to the IAP, IAP length, length of each flanking exon, and whether 5' or 3' splice site showed evidence for alternative usage in UCSC annotations showed no significant correlation with the expression ratio between *Nxf1* alleles (S6 Fig). However, distance between IAP and 3' exon approached a nominal significance threshold ($\rho = -0.30$, $p = 0.051$) before correction for multiple tests.

A simple alignment among elements that had shown both reduced expression in inserted strains and suppression by either $Nxf1^{CAST}$ or E610G alleles showed that IAP protein coding sequences were not required for either effect, leaving only LTRs, a small sequence 5' to gag, ~60 bp in four small fragments within *pol*, and sequences 3' to pol that include both RNA transport element (RTE) and the polypurine tract (ppt). For classification of elements as suppressed, we used Fisher's combined probability test as a meta-analysis across the congenic homozygous and genome-edited heterozygous samples (S1 Table). We further characterized each element for presence and sequence divergence in eight segments: LTR, segment 2 (between LTR and *gag*), *gag* CDS, *prt* CDS, *pol* CDS, segment 6 (after pol), RTE, and polypurine tract. Phylogenetic trees for either full length or discreet segments of the IAP element did not correlate with either differences between inserted and uninserted strains or effect of *Nxf1* genotype (S7 Fig). Although suppressed insertions appear somewhat clustered with respect to RTE, identical RTE sequences were associated with divergent effects on host gene RNA (S7 Fig, panel J). Predicted RNA folding also did not show any strict requirement for suppression or lack of suppression by *Nxf1* alleles (S8 Fig). These data together show that most of the IAP consensus sequence is not required and suggest that a minimum element for both mutation and *Nxf1*-mediated suppression might require less than 2 kb (837 bp 5' in *Cdh19* and 1108 bp 3' in *Sntg1*), but that efficiency might be influenced by other factors–possibly including elements size, distance to the 3' exon, and factors not readily predicted from sequence alone.

Discussion

As far as we are aware, *Nxf1* remains the most highly-connected node for genetic interactions among modifier genes that act on either natural or chemically-induced mutations in mice [19, 42]. The current results add three times as many allele-specific genetic interactions for *Mvb1* alleles of *Nxf1* (CAST and E610G) as had previously been reported. We previously showed that the most common *Mus musculus castaneus* allele of *Nxf1*, containing many polymorphic sites, suppressed the classical *vibrator* mutation [20] and that a congenic stock including this allele suppressed several other IΔ1 IAP insertional mutations [19]. Our results extend the diversity of IAP insertions susceptible to *Nxf1* substantially beyond the IΔ1 subfamily (Fig. 2) and provide further evidence against activity on alternative processing at non-IAP introns (Fig. 3). Alternative events suppressed by *Nxf1* alleles include known bleeding exons with alternative termination, such as *Slc15a2* [43] and *Trpc6* [30], and IAP-dependent alternative splicing, such as *Adamts13* [27, 44] and *Cpne8*, as well as introns for which public transcriptome data do not indicate a predominant alternative variant. Comparison of IAP elements whose effects can be suppressed by *Nxf1* alleles further suggests minimum sequence requirements for both mutation and suppression lie in <1 kb in the 5' end and ~1 kb in the 3' end of the IAP (Fig. 7), although construction and experimental measurements of a minimal element will be required to test this idea.

In contrast to most de novo mutations caused by IAP insertions, the magnitude of gene expression changes is relatively modest for most of the B6-endogenous IAP sites we tested (Fig. 2). Indeed, recent evaluation of mouse transposable elements concluded that insertion
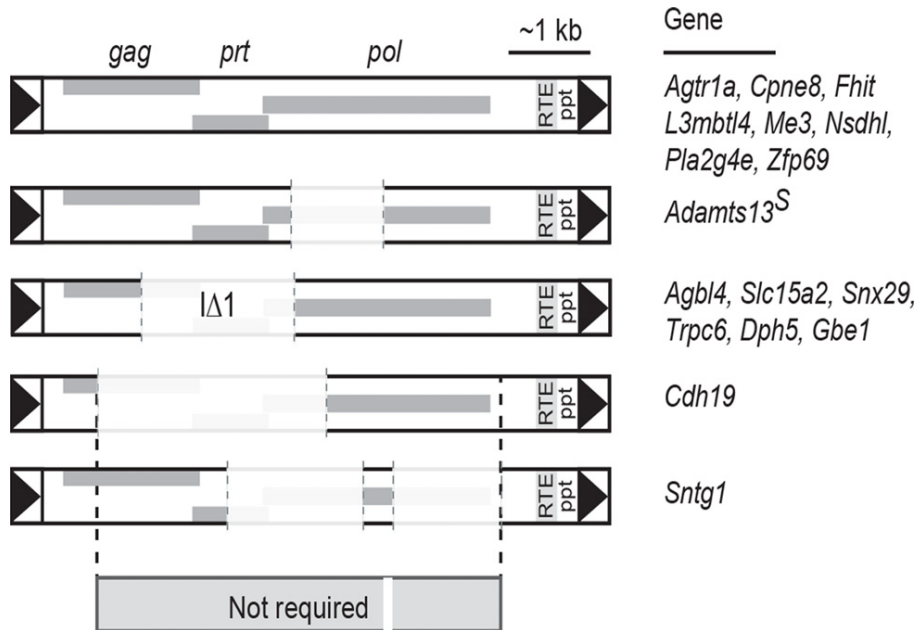
Fig 7. Shared sequence features of IAPs suppressed by *Nxf1* alleles. Alignment of IAPs whose effects on host gene expression were modified by either congenic or genome edited *Nxf1* alleles suggests that most of the IAP sequence is not required for either mutagenic potential or suppression by *Nxf1*. Shared sequence includes just 837 bp of LTR through 5' end of the *gag* gene at *Cdh19*, ~60 bp in four fragments within *pol*, and 1108 bp including the RNA transport element (RTE), polypurine tract (ppt) and 3' LTR.

doi:10.1371/journal.pgen.1005123.g007

events that strongly affected gene expression were rapidly purged from populations, hence remaining elements tend toward more modest effect sizes [45]. *Adamts13* is a notable exception, with ~1000-fold change in completion of the transcript across the inserted intron (Fig. 1). However, as the truncated Adamts13 encoded by the major B6 transcript retains substantial function [44, 46] and the absolute level of full-length transcript recovered by suppression is modest, little physiological consequence is expected from this increase in full-length Adamts13 protein above the level of function provided by the abundant truncated form. Taken all together the data suggest that using IAP insertional mutations in combination with *Nxf1* suppressing alleles should allow relatively clean titrations of intermediate gene doses. This may have application in modeling genetic disorders (and therapeutics) in mice.

Our results using genome-edited mice provide molecular precision to the definition of the modifier allele. Although several polymorphisms distinguish $Nxf1^{B6}$ and $Nxf1^{CAST}$ alleles in both transgenic [20] and congenic [19] experiments, the E610G single amino acid substitution (Fig. 4) was sufficient for the full range of genetic modifier activity previously ascribed to the congenic locus and published 16 kb transgene. The point mutation modified all of the principle features of the *vibrator* mutation on which the modifier activity was first described [47] with

the characteristic semi-dominant effect ([Fig. 5](#)). The ability to assay the edited allele as a heterozygote with respect to congenic $Nxf1^{B6}$ and $Nxf1^{CAST}$ alleles and the specificity of multiple phenotypes assayed (Figs. [5](#), [6](#)) also provides an additional safeguard against potential off-target modifications in the editing process.

Our results demonstrate that a single amino acid substitution in the C-terminal, UBA-like domain of Nxf1 simultaneously enhances expression of full-length mRNA at more than a dozen loci in the mouse reference genome at the level of pre-mRNA processing. This illustrates a potential post-transcriptional mechanism for evolution of gene regulatory networks selected from events generated by molecular parasites [3] and provides a post-transcriptional tool for modulating expression level of any mouse gene into which an IAP element can be introduced.

Methods

Ethics statement

All animal procedures were approved by the University of California, San Diego Institutional Animal Care and Use Committee (IACUC). Tissues were obtained after terminal anesthesia with tribromoethanol (avertin)

Mice

Congenic C57BL/6J (B6)–$Nxf1^{CAST}$ mice were developed and maintained in our laboratory by serial backcross and genotyped by PCR as previously described [19]. Mice in this report were at backcross generation N30 or later. A/J, BALB/cJ and 129S1/SvImJ were purchased from The Jackson Laboratory and bred locally in the same vivarium room with our congenic stocks. *Adamts13* genotypes were determined by PCR using a three primer combination: ACCTCTCAAGTGTTTGGGATGCTA, TCAGCGCCATCTTGTGACGGCGAA, and TGCCAGATGGCCATGATTAACTCT [27]. Tissues were from paired F2 animals from a B6 x BALB/c intercross previously described [19] ([Fig. 1C](#)) or B6 congenic or co-isogenic stocks (all others). Five male and five female pairs were used for measurements from congenic animals. No sex-specific differences were noted. Genome-edited mice were produced in the Moores UCSD Cancer Center Transgenic Mouse Shared Resource by co-injection of mouse C57BL/6 one-cell embryos with sgRNA, Cas9 mRNA, and unpurified oligonucleotide as a homology-dependent repair donor essentially as described [40]. Oligonucleotide sequences for sgRNA, Cas9 amplification, and homology dependent repair are given in [S2 Table](#). Measurements from genome edited mice used unpaired samples collected as a cohort and comprising ~equal numbers of each sex.

Identification of B6 alleles with IAP element insertions

Two independent approaches identified intronic IAP elements. IAP annotations from the Repbase Update definitions were mined from the RepeatMasker track data table in the UCSC Genome Browser ([S1 Fig](#)). A custom table was also developed by iterative BLAT searches with a panel of short (35 bp) sequences conserved in subfamilies of IAP elements for which we had previously observed *Nxf1*-related suppression ([S3 Table](#)). Both sets were filtered by strand orientation in Galaxy [48] and compared to UCSC intron or RefSeq gene boundary annotations. After manual curation to remove likely annotation errors, the oligonucleotide-derived set identified 70 and the Repbase Update annotations identified a superset of 85 IAP-family elements within well-defined introns or 3' ends that were premature relative to human or rat orthologs. Alignments of IAP sequences were performed in MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/) using default parameters.

Quantitative RT-PCR assays

PCR assays for transcripts that are correctly spliced around the IAP insertion and for reference genes were designed relative to exon junctions using Primer3 [49]. Assays were selected for use if they produced a single product by melt profile, single band by gel imaging, and low coefficient of variation among replicate samples. Specific primers used are provided in S4 Table. RNA was prepared from tissues homogenized in Trizol, converted to cDNA by reverse transcription, and assayed by quantitative PCR using SYBR green fluorescence in a Bio-Rad CFX instrument as previously described [19]. Normalization was performed by geometric averaging among the indicated reference genes, as described [50], and implemented in software supplied by the manufacturer. Four reference gene assays, *Gapdh*, *Ppia*, *Sdha*, and *Pitpna* were selected based on high cycle efficiencies, accurate dose titration, and low quantitative variation across tested conditions (only the first three were used for comparisons including *vibrator* samples, which express ~18% normal level of *Pitpna*).

Gel-based assays

Relative quantification of alternative splicing products was carried out by gel electrophoresis of products from non-saturating 3-primer competitive PCR amplifications based on pilot experiments for each assay. Samples for comparison were amplified side-by-side and resolved on the same gel. Relative fluorescence intensities of resolved bands were quantified from non-saturated. tiff image files using ImageJ software.

Splicing-sensitive array profiling

Affymetrix MJAY microarrays (GEO platform GPL13185) were hybridized and processed as described [35, 39], for brain RNA from 3 pairs of littermates (two female pairs, one male pair) with opposite *Nxf1* genotypes. The separation score was defined as the $\log_2$ (ratio of alternative events in $Nxf1^{CAST}$ samples / ratio of alternative events in $Nxf1^{B6}$ samples) [36]. Events with a separation score >0.5 (~1.4-fold change) and false discovery rate q<0.05 were deemed significant.

Statistics

Statistical tests were conducted in R. Non-parametric rank tests were performed in package exactRankTests or coin; both packages gave identical results for repeated analyses. Non-parametric correlation tests with ties were performed in the package Hmisc. Correlation analyses and standard hypothesis tests were conducted in R 2.8.1 or later running on Mac OS 10.5.8 or later.

Supporting Information

S1 Table. False discovery rates and meta-analyses. Non-parametric p-values for the effect of *Nxf1* alleles were corrected for false discovery rate (FDR) using the Benjamini-Hochberg procedure within each IAP class and experimental group (congenic stocks or genome edited allele). For sites measured in both experiments meta-analysis was performed using Fisher's combined probability test. All tests were conducted in R.
(XLSX)

S2 Table. Oligonucleotides used for *Nxf1* E610G genome editing and validation. Sequences are presented for production of single guide RNAs and synthetic Cas9 mRNA, for single-stranded donor templates for homology-dependent repair, and for sequencing target and predicted off-target cleavage sites for the guide RNA used in genome editing.
(XLSX)

S3 Table. IAP seed sequences. Sequences used in genome searches for IAP sequences in introns.
(XLSX)

S4 Table. PCR primers used to assay RNA abundance or isoform variation. Primer sequences are given, by class of assay, for dye intercalation quantitative PCR (Fig. 2B and Fig. 6), cassette exon inclusion and alternative splice site usage (Fig. 3A), and alternative events predicted by microarray data (Fig. 3C).
(XLSX)

S1 Fig. Galaxy workflow to identify intronic IAPs. Mouse reference assembly mm9 was accessed on 1/15/12 and analyzed in Galaxy using the workflow as illustrated. Numbers in parentheses indicate the number of objects contained in each product after each operation. Operations to identify strand specificity are color coded.
(TIF)

S2 Fig. Assay development and quality control workflow. The number of successes and attempts in each step are indicated. PCR primers were selected in Primer3 web interface (http://bioinfo.ut.ee/primer3-0.4.0/primer3/input.htm). Both melt profile after real-time PCR and gel electrophoresis were used to determine that each PCR produced a single significant product. Coefficient of variation after normalization to reference genes was used to eliminate assays with limited power to detect differences by genotype. Gene assays whose measurements differed significantly by *Nxf1* genotype in the congenic strains are shown in red. Inequality symbols in the lower boxes indicate groups with greater or lesser levels of correctly processed RNA based on qRT-PCR assays described in Fig. 2 and accompanying text.
(TIF)

S3 Fig. *Nxf1* alleles do not affect alternative splicing at selected non-IAP splice choices. (A) Gel assays for well-characterized alternative splicing of cassette exons, including several encoding Nxf1-interacting factors, show no difference in general or tissue-specific inclusion rates between $Nxf1^{B6}$ (B) and $Nxf1^{CAST}$ (C) alleles in congenic mice. Splice choice is diagrammed, with PCR product sizes indicated. Bar graphs show proportion of fluorescence intensity in each band among 13–20 independent sample pairs tested for each gene. **(B)** Single examples for a retained intron (*Ank1*), alternative splice acceptor (*Nup214*) and alternative splice donor (*Nup98*) show no difference between *Nxf1* alleles.
(TIF)

S4 Fig. Alternative splicing in the congenic interval is not explained by E610G. The alternative events in the congenic interval on chromosome 19 that were found by the arrays and confirmed in gel assays were re-assayed in co-isogenic E610G genome-edited animals and controls, using the samples from Fig. 6 and the assays from Fig. 3. *Trpc6* gel assay, included as a positive control, showed suppression of the IAP-dependent splice form. None of the four chromosome 19 alternative events that were significant in both array (Fig. 3B) and gel (Fig. 3C) assays showed significant differences between edited and unedited heterozygotes with respect to $Nxf1^{B6}$. Indicated p-vlaues are for the one-tailed Wilcoxon rank sum test for the direction of difference seen in Fig. 3C, with the exception of *Abca5*, included as a negative control, which shows the value for the two-tailed test, as it had no predicted difference. Inclusion of heterozygous congenic samples confirms the effect, detection sensitivity and allele-independents effects predicted from Fig. 3C for *Ganab* (p = 0.00015) and *Tmem138* (p = 0.0036), but not for *Fkbp2* (p = 0.28) or *Stx3* (p = 0.40), albeit comparing new edited with archived congenic samples.
(TIF)

S5 Fig. Lack of strong correlations between genomic context and IAP-associated expression changes. Plots show relationships between indicated attributes of the studied introns (x-axes) and expression ratio between uninserted strain and inserted strains with respect to the IAP (U/I ratio). U/I ratio is expressed as log10 to permit inclusion of exceptionally strong effects at *Adamts13* and *Zfp69*. Dashed line at zero indicates elements with no measureable effect. Spearman correlation coefficients (ρ) and associated p-values are shown. For categorical annotations of alternatively spliced (alt), alternative 3' end formation (altEnd), constitutive (con), and 3' terminal (end) exons, p-values are derived from the Kruskal-Wallis test. The magnitude of U/I ratio is moderately well correlated with its p-value, as expected for events near the threshold for discrimination and in the presence of other potential modifying effects across strains. (TIF)

S6 Fig. Lack of strong correlations between genomic features and *Nxf1*-dependent suppression. Plots show relationships between indicated attributes of the studied introns (x-axes) and expression ratio between $Nxf1^{CAST}$ and $Nxf1^{B6}$ alleles (C/B ratio). For categorical annotations of alternatively spliced (alt), alternative 3' end formation (altEnd), constitutive (con), and 3' terminal (end) exons, p-values are derived from the Kruskal-Wallis test. Nominal evidence for correlation to distance from 3' exon does not survive correction for multiple tests, but may merit further study. As expected, C/B ratio is correlated with U/I ratio, as larger mutational effects should provide a larger phenotypic space for suppression. (TIF)

S7 Fig. Phylogenetic relationships among IAP elements. Neighbor-joining trees were generated from alignments in MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/) for the indicated IAP groups or sequence components. Elements are colored as in Fig 2, red for those with strong evidence for *Nxf1*-mediated suppression, grey for those with no evidence of mutagenic effect across strains. Nucleotide length of the total alignment, as well as lengths of the median, shortest and longest input sequences are indicated for each class or segment of IAP sequence. **(A)** Full alignment of IAP sequences >7kb ("full length"). **(B)** Full alignment of elements containing the IΔ1 deletion. Alignments of specific sequence components are **(C)** LTRs, **(D)** interval between LTR and *gag*, (E) *gag*, **(F)** *prt*, **(G)** *prt* to *pol*, **(H)** *pol*, **(I)** *pol* to the RTE, **(J)** RTE, **(K)** ppt. (PDF)

S8 Fig. RTE secondary structure does not determine *Nxf1*-dependent suppression. Secondary structures of RTE RNAs were predicted using the mfold web server version 3.4 (http://mfold.rna.albany.edu/?q = mfold/download-mfold) under default parameters. **(A)** RTE phylogeny, as in S7 Fig, panel J. **(B)** Optimal (left column) and next best (right column) predicted secondary structures of RTE RNA sequence for each unique phylogeny group, based on calculated lowest free energy, are depicted. **(C)** Free energy of optimal structures, and **(D)** difference in free energy between optimal and next best secondary structures are listed. In addition to identical RTE sequences with divergent outcomes, RTE structures from adjacent positions in the tree also showed dissimilar effects (e.g., *hes*, *mgL*, *Nudcd3*, and *Nsdhl*). (TIF)

S1 Video. *Vibrator* mutant with silent edit only. Moderate tremor of this $Pitpna^{vb/vb}$ $Nxf1^{pseudo/CAST}$ animal is characteristic of $Pitpna^{vb/vb}$ in the presence of heterozygous $Nxf1^{B6/CAST}$, consistent with no effect of the "pseudo-edited" allele that altered only the silent site within the sgRNA target sequence. (MP4)

S2 Video. *Vibrator* **mutant with E610G edited allele.** Mild tremor seen in this *Pitpna*$^{vb/vb}$ *Nxf1*$^{E610G/CAST}$ animal is characteristic of of *Pitpna*$^{vb/vb}$ in the presence of homozygous *Nxf1*$^{CAST/CAST}$, consistent with quantitative effect of the E610G edited allele reproducing the original *Modifier-of-vibrator* effect.
(MP4)

Acknowledgments

Author Contributions

References

1. Cowley M, Oakey RJ. Transposable elements re-wire and fine-tune the transcriptome. PLoS genetics. 2013; 9(1):e1003234. Epub 2013/01/30. doi: 10.1371/journal.pgen.1003234 PMID: 23358118

2. Gifford WD, Pfaff SL, Macfarlan TS. Transposable elements as genetic regulatory substrates in early development. Trends Cell Biol. 2013; 23(5):218–26. Epub 2013/02/16. doi: 10.1016/j.tcb.2013.01.001 PMID: 23411159

3. Peter IS, Davidson EH. Evolution of gene regulatory networks controlling body plan development. Cell. 2011; 144(6):970–85. Epub 2011/03/19. doi: 10.1016/j.cell.2011.02.017 PMID: 21414487

4. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002; 420(6915):520–62. PMID: 12466850

5. Hamilton BA, Frankel WN. Of mice and genome sequence. Cell. 2001; 107(1):13–6. PMID: 11595181

6. Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL. Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS genetics. 2006; 2(1):e2. PMID: 16440055

7. Vasicek TJ, Zeng L, Guan XJ, Zhang T, Costantini F, Tilghman SM. Two dominant mutations in the mouse fused gene are the result of transposon insertions. Genetics. 1997; 147(2):777–86. PMID: 9335612

8. Duhl DM, Vrieling H, Miller KA, Wolff GL, Barsh GS. Neomorphic agouti mutations in obese yellow mice. Nature Genet. 1994; 8(1):59–65. PMID: 7987393

9. Lugani F, Arora R, Papeta N, Patel A, Zheng Z, Sterken R, et al. A retrotransposon insertion in the 5' regulatory domain of Ptf1a results in ectopic gene expression and multiple congenital defects in Danforth's short tail mouse. PLoS genetics. 2013; 9(2):e1003206. Epub 2013/02/26. doi: 10.1371/journal.pgen.1003206 PMID: 23437001

10. Semba K, Araki K, Matsumoto K, Suda H, Ando T, Sei A, et al. Ectopic expression of Ptf1a induces spinal defects, urogenital defects, and anorectal malformations in Danforth's short tail mice. PLoS genetics. 2013; 9(2):e1003204. Epub 2013/02/26. doi: 10.1371/journal.pgen.1003204 PMID: 23436999

11. Vlangos CN, Siuniak AN, Robinson D, Chinnaiyan AM, Lyons RH Jr., Cavalcoli JD, et al. Next-generation sequencing identifies the Danforth's short tail mouse mutation as a retrotransposon insertion affecting Ptf1a expression. PLoS genetics. 2013; 9(2):e1003205. Epub 2013/02/26. doi: 10.1371/journal.pgen.1003205 PMID: 23437000

12. Bentley DL. Coupling mRNA processing with transcription in time and space. Nat Rev Genet. 2014; 15(3):163–75. Epub 2014/02/12. doi: 10.1038/nrg3662 PMID: 24514444

13. Auboeuf D, Honig A, Berget SM, O'Malley BW. Coordinate regulation of transcription and splicing by steroid receptor coregulators. Science. 2002; 298(5592):416–9. PMID: 12376702

14. Cramer P, Pesce CG, Baralle FE, Kornblihtt AR. Functional association between promoter structure and transcript alternative splicing. Proc Natl Acad Sci U S A. 1997; 94(21):11456–60. Epub 1997/10/23. PMID: 9326631

15. Kornblihtt AR. Promoter usage and alternative splicing. Curr Opin Cell Biol. 2005; 17(3):262–8. PMID: 15901495

16. de la Mata M, Kornblihtt AR. RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. Nature structural & molecular biology. 2006; 13(11):973–80.

17. Kornblihtt AR. Chromatin, transcript elongation and alternative splicing. Nature structural & molecular biology. 2006; 13(1):5–7.

18. Vargas DY, Shah K, Batish M, Levandoski M, Sinha S, Marras SA, et al. Single-molecule imaging of transcriptionally coupled and uncoupled splicing. Cell. 2011; 147(5):1054–65. Epub 2011/11/29. doi: 10.1016/j.cell.2011.10.024 PMID: 22118462

19. Concepcion D, Flores-Garcia L, Hamilton BA. Multipotent genetic suppression of retrotransposon-induced mutations by Nxf1 through fine-tuning of alternative splicing. PLoS genetics. 2009; 5(5): e1000484. doi: 10.1371/journal.pgen.1000484 PMID: 19436707

20. Floyd JA, Gold DA, Concepcion D, Poon TH, Wang X, Keithley E, et al. A natural allele of Nxf1 suppresses retrovirus insertional mutations. Nature Genetics. 2003; 35:221–8. PMID: 14517553

21. Kuff EL, Lueders KK. The intracisternal A-particle gene family: structure and functional aspects. Advances in cancer research. 1988; 51:183–276. PMID: 3146900

22. Fujikawa K, Suzuki H, McMullen B, Chung D. Purification of human von Willebrand factor-cleaving protease and its identification as a new member of the metalloproteinase family. Blood. 2001; 98(6):1662–6. PMID: 11535495

23. Gerritsen HE, Robles R, Lammle B, Furlan M. Partial amino acid sequence of purified von Willebrand factor-cleaving protease. Blood. 2001; 98(6):1654–61. PMID: 11535494

24. Levy GG, Nichols WC, Lian EC, Foroud T, McClintick JN, McGee BM, et al. Mutations in a member of the ADAMTS gene family cause thrombotic thrombocytopenic purpura. Nature. 2001; 413(6855):488–94. PMID: 11586351

25. Nolasco LH, Turner NA, Bernardo A, Tao Z, Cleary TG, Dong JF, et al. Hemolytic uremic syndrome-associated Shiga toxins promote endothelial-cell secretion and impair ADAMTS13 cleavage of unusually large von Willebrand factor multimers. Blood. 2005; 106(13):4199–209. PMID: 16131569

26. Fujioka M, Hayakawa K, Mishima K, Kunizawa A, Irie K, Higuchi S, et al. ADAMTS13 gene deletion aggravates ischemic brain damage: a possible neuroprotective role of ADAMTS13 by ameliorating postischemic hypoperfusion. Blood. 2010; 115(8):1650–3. doi: 10.1182/blood-2009-06-230110 PMID: 19965676

27. Banno F, Kaminaka K, Soejima K, Kokame K, Miyata T. Identification of strain-specific variants of mouse Adamts13 gene encoding von Willebrand factor-cleaving protease. J Biol Chem. 2004; 279(29):30896–903. PMID: 15136581

28. Uemura M, Tatsumi K, Matsumoto M, Fujimoto M, Matsuyama T, Ishikawa M, et al. Localization of ADAMTS13 to the stellate cells of human liver. Blood. 2005; 106(3):922–4. PMID: 15855280

29. Zhou W, Inada M, Lee TP, Benten D, Lyubsky S, Bouhassira EE, et al. ADAMTS13 is expressed in hepatic stellate cells. Laboratory investigation; a journal of technical methods and pathology. 2005; 85(6):780–8. PMID: 15806136

30. Zhang Y, Romanish MT, Mager DL. Distributions of transposable elements reveal hazardous zones in mammalian introns. PLoS Comput Biol. 2011; 7(5):e1002046. Epub 2011/05/17. doi: 10.1371/journal.pcbi.1002046 PMID: 21573203

31. Zhang Y, Maksakova IA, Gagnier L, van de Lagemaat LN, Mager DL. Genome-Wide Assessments Reveal Extremely High Levels of Polymorphism of Two Active Families of Mouse Endogenous Retroviral Elements. PLoS genetics. 2008; 4(2):e1000007 doi: 10.1371/journal.pgen.1000007 PMID: 18454193

32. Yalcin B, Wong K, Agam A, Goodson M, Keane TM, Gan X, et al. Sequence-based characterization of structural variation in the mouse genome. Nature. 2011; 477(7364):326–9. Epub 2011/09/17. doi: 10.1038/nature10432 PMID: 21921916

33. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011; 477(7364):289–94. Epub 2011/09/17. doi: 10.1038/nature10413 PMID: 21921910

34. Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, et al. Deciphering the splicing code. Nature. 2010; 465(7294):53–9. doi: 10.1038/nature09000 PMID: 20445623

35. Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, et al. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell reports. 2012; 1(2):167–78. Epub 2012/05/11. doi: 10.1016/j.celrep.2012.02.001 PMID: 22574288

36. Du H, Cline MS, Osborne RJ, Tuttle DL, Clark TA, Donohue JP, et al. Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. Nature structural & molecular biology. 2010; 17(2):187–93. Epub 2010/01/26. doi: 10.1038/nsmb.1720.

37. Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, Shiue L, et al. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. Nat Genet. 2011; 43(7):706–11. Epub 2011/05/31. doi: 10.1038/ng.841 PMID: 21623373

38. Lagier-Tourenne C, Polymenidou M, Hutt KR, Vu AQ, Baughn M, Huelga SC, et al. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. Nature neuroscience. 2012; 15(11):1488–97. Epub 2012/10/02. doi: 10.1038/nn.3230 PMID: 23023293

39. Polymenidou M, Lagier-Tourenne C, Hutt KR, Huelga SC, Moran J, Liang TY, et al. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nature neuroscience. 2011; 14(4):459–68. Epub 2011/03/02. doi: 10.1038/nn.2779 PMID: 21358643

40. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 2013; 153(4):910–8. Epub 2013/05/07. doi: 10.1016/j.cell.2013.04.025 PMID: 23643243

41. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature protocols. 2013; 8(11):2281–308. Epub 2013/10/26. doi: 10.1038/nprot.2013.143 PMID: 24157548

42. Hamilton BA, Yu BD. Modifier genes and the plasticity of genetic networks in mice. PLoS genetics. 2012; 8(4):e1002644. Epub 2012/04/19. doi: 10.1371/journal.pgen.1002644 PMID: 22511884

43. Li J, Akagi K, Hu Y, Trivett AL, Hlynialuk CJ, Swing DA, et al. Mouse endogenous retroviruses can trigger premature transcriptional termination at a distance. Genome Res. 2012; 22(5):870–84. Epub 2012/03/01. doi: 10.1101/gr.130740.111 PMID: 22367191

44. Zhou W, Bouhassira EE, Tsai HM. An IAP retrotransposon in the mouse ADAMTS13 gene creates ADAMTS13 variant proteins that are less effective in cleaving von Willebrand factor multimers. Blood. 2007; 110(3):886–93. PMID: 17426255

45. Nellaker C, Keane TM, Yalcin B, Wong K, Agam A, Belgard TG, et al. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. Genome Biol. 2012; 13(6):R45. Epub 2012/06/19. doi: 10.1186/gb-2012-13-6-r45 PMID: 22703977

46. Banno F, Chauhan AK, Kokame K, Yang J, Miyata S, Wagner DD, et al. The distal carboxyl-terminal domains of ADAMTS13 are required for regulation of in vivo thrombus formation. Blood. 2009; 113(21):5323–9. doi: 10.1182/blood-2008-07-169359 PMID: 19109562

47. Hamilton BA, Smith DJ, Mueller KL, Kerrebrock AW, Bronson RT, van Berkel V, et al. The vibrator mutation causes neurodegeneration via reduced expression of PITP alpha: positional complementation cloning and extragenic suppression. Neuron. 1997; 18(5):711–22. PMID: 9182797

48. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 2005; 15(10):1451–5. Epub 2005/09/20. doi: 10.1101/gr.4086505 PMID: 16169926

49. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods in molecular biology (Clifton, NJ. 2000; 132:365–86. PMID: 10547847

50. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology. 2002; 3(7):RESEARCH0034. Epub 2002/08/20. PMID: 12184808

gene	ratio_CB	class	Congenic: Wilcoxon p-value	Congenic: FDR q-value (within class)	E610G edited: Wilcoxon p-value	E610G edited: FDR q-value (all)	Fisher's combined probability test on both p-values	Fisher's combined probability test: FDR q-value (all)
Agtr1a	1.16116	full	0.012	0.049	0.0095	0.019	0.0011	0.0022
Cpne8	1.08295	full	0.027	0.072				
Fhit	1.05621	full	0.034	0.077	0.14	0.15	0.03	0.033
Kcns3	1.05335	full	0.14	0.19				
L3mbtl4	1.68228	full	0.039	0.078	0.013	0.02	0.0044	0.0066
Lphn3	1.03354	full	0.42	0.52				
Me3	1.20538	full	0.0093	0.049				
Nsdhl	1.18519	full	0.0024	0.02	0.039	0.052	0.00096	0.0022
Nudcd3	0.95069	full	0.74	0.85				
Pde4d	0.90893	full	0.94	0.94				
Pla2g4e	1.39312	full	0.00049	0.0078	0.00016	0.0019	1.36E-06	0.000016
Scg5	1.13561	full	0.087	0.14				
Slc20a2	1.07707	full	0.12	0.17				
Tiam1	0.93726	full	0.84	0.9				
Tmem163	1.1393	full	0.06	0.11				
Zfp69-tm	7.78222	full	0.016	0.052				
Agbl4	1.52591	IΔ1	0.039	0.089	0.0054	0.013	0.002	0.0034
Cntn5_A	1.09424	IΔ1	0.074	0.12				
Csmd1	1.01287	IΔ1	0.35	0.39				
Dph5	1.15929	IΔ1	0.08	0.12	0.012	0.02	0.0076	0.0091
Gbe1	1.14647	IΔ1	0.016	0.048				
Rgs3	0.8845	IΔ1	0.98	0.98				
Slc15a2	1.73589	IΔ1	0.0078	0.035	0.00059	0.0023	6.12E-05	0.00024
Snx29	1.17404	IΔ1	0.13	0.17	0.32	0.32	0.17	0.17
Trpc6-tm	1.39138	IΔ1	0.0024	0.022				
Adk	1.06821	Δpol	0.5	0.6				
Alk	1.08275	Δpol	0.37	0.6				
Col23a1	1.10086	Δpol	0.48	0.6				
Exoc4	0.97296	Δpol	0.79	0.79				
Slc25a21	1.02079	Δpol	0.29	0.6				
Slc9a2	1.02342	Δpol	0.52	0.6				
Spon1	1.03023	Δpol	0.42	0.6				
Amph	1.05644	LTR	0.18	0.46				
Bdh1	1.02238	LTR	0.38	0.46				
Dtd1	1.00936	LTR	0.32	0.46				
Enoph1	1.04249	LTR	0.43	0.46				
Tmco3	1.02042	LTR	0.46	0.46				
Adamts13	2	other	0.001	0.007	0.0013	0.0039	1.89E-05	0.00011
Cdh19	1.11604	other	0.016	0.056	0.057	0.069	0.0073	0.0091
Lhfpl3	1.03573	other	0.45	0.52				
Phf20	1.02206	other	0.21	0.36				
Rftn1	1.0199	other	0.32	0.45				
Rsrc1	1.16345	other	0.8	0.8				
Sntg1	1.08237	other	0.18	0.36	0.00048	0.0023	8.95E-04	0.0022

S1 Table. False discovery rates and meta-analyses.

Non-parametric p-values for the effect of *Nxf1* alleles were corrected for false discovery rate (FDR) using the Benjamini-Hochberg procedure within each IAP class and experimental group (congenic stocks or genome edited allele). For sites measured in both experiments meta-analysis was performed using Fisher's combined probability test. All tests were conducted in R.

Oligos	Notes	Sequence
Cas9/sgRNA Synthesis		
Cas9_F	Addition of T7 Promoter to Cas9 and amplification from pX330	TAATACGACTCACTATAGGGAGAATGGACTATAAGGACCACGAC
Cas9_R	Amplification of Cas9 from pX330	GCGAGCTCTAGGAATTCTTAC
T7_E610G_amp	Amplification of T7sgE610G.R	TAATACGACTCACTATAGTTATCTTTCT
Sp6_E610G_amp	Amplification of Sp6sgE610G.R	ATTTAGGTGACACTATAGATTTCATAAA
sgREV	sgRNA reverse primer	AAAAGCACCGACTCGGTG
T7sgE610G.R	T7E610G sgRNA template	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTTGGCCTGCAGAAAGATAACTATAGTGAGTCGTATTA
SP6sgE610G.R	Sp6E610G sgRNA template	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAAGTAGCGTTTATGAAATCTATAGTGTCACCTAAAT
Targeted Repair		
T7E610G-HDR	T7sgE610G HDR ssODN	GGGAAGGTGGGAGAGAGGTGCCTGGTTCTACTTACATTTTATCTTTCTGCAGGCTAAGGGAGGGATTCCAGAAGTAGCGTTTATGAAATGATCACAAGAAAAGTCCCCTTGTAAATAGCCCCTGGATAA
Sp6E610G-HDR	Sp6sgE610G HDR ssODN	GGGAAGGTGGGAGAGAGGTGCCTGGTTCTACTTACATTTTATCTTTCTGCAGGCCAAGGGAGGGATTCCAGAGGTAGCGTTTATGAAATGATCACAAGAAAAGTCCCCTTGTAAATAGCCCCTGGATAA
Sequencing		
E610G PCR F	Genomic E610G Sequencing	GGTGACTGACTGCATTTTGG
E610G PCR R	Genomic E610G Sequencing	GTCAGATGAAGGCTGGGTGT
E610G_Clone_F	Add BamHI site	ACGTGGATCCAAGGCTGGGTGTGGTGATAC
E610G_Clone_R	Add EcoRV site	ATTCGATATCGCACCATGAGGCTCAATCTT
T7-Promoter	E610G Clone Sequencing	AATACGACTCACTATAGG
T3-Promoter	E610G Clone Sequencing	ATTAACCCTCACTAAAGG
Sp6_OT1_L	Off-Target Sequencing	TCACTCCTAGCTTTTGGTCCT
Sp6_OT2_F	Off-Target Sequencing	AGGACACTGGATGCTTTCAA
Sp6_OT3_F	Off-Target Sequencing	ATGGCTGTGTTGGGAGATGA
Sp6_OT4_F	Off-Target Sequencing	GCTAGACACTCTGCCACTCA
Sp6_OT5_F	Off-Target Sequencing	AGGGATGCGCTTACTGTTTC
Sp6_OT6_F	Off-Target Sequencing	CCATCGTGCACCAGAATGAC
Sp6_OT7_F	Off-Target Sequencing	TTCTTAGGGCCCTCTTTGCT
Sp6_OT8_F	Off-Target Sequencing	CTGGGATAAACATCATGCGCT
Sp6_OT9_F	Off-Target Sequencing	CATTTGCCATCCACCCTTCC
Sp6_OT10_F	Off-Target Sequencing	AGTGCTGAAATTGGCTTGCA
Sp6_OT11_F	Off-Target Sequencing	TCTTTGTTACTTTCCAAGCCTGT
Sp6_OT12_F	Off-Target Sequencing	GTCAGGTGTCTCTTAACATGTGT
Sp6_OT13_F	Off-Target Sequencing	CAGAGAACATTGGTGGCTGA
Sp6_OT14_F	Off-Target Sequencing	TATTGGGCAGATGGTGAGGG
Sp6_OT15_F	Off-Target Sequencing	CCTCCTGCAAGAAACTGCTG
Sp6_OT16_F	Off-Target Sequencing	AGGGAATGGTGAAATGGCAG
Sp6_OT17_F	Off-Target Sequencing	TGCCTCGATGCTTCGTTTCT
Sp6_OT1_R	Off-Target Sequencing	CCCACTGCTTAGTTTCTCTCC
Sp6_OT2_R	Off-Target Sequencing	TACCCAGAAGTGTGTTGCCA
Sp6_OT3_R	Off-Target Sequencing	GCAGAGCTTTAAGACAGAAACAT
Sp6_OT4_R	Off-Target Sequencing	GACCCAGAGCCAGACTAGAC
Sp6_OT5_R	Off-Target Sequencing	TCAACCAATGTGAGTCCTTCC
Sp6_OT6_R	Off-Target Sequencing	TTCCAGCCCACTCAGAACAT
Sp6_OT7_R	Off-Target Sequencing	GCCAATGCCATCACAAGTCA
Sp6_OT8_R	Off-Target Sequencing	GCCTAGTCCTGAGCTAAGACC
Sp6_OT9_R	Off-Target Sequencing	TGGGAATAGTGTCATTGACCTCA
Sp6_OT10_R	Off-Target Sequencing	GAGATCAGAGAGCCCCTGAC
Sp6_OT11_R	Off-Target Sequencing	TCTTCTGTTCACACTCTGCTTT
Sp6_OT12_R	Off-Target Sequencing	CACCAGCGTTAAAATCATTCTGT
Sp6_OT13_R	Off-Target Sequencing	GGTGCGTTCTCTGATGCAG
Sp6_OT14_R	Off-Target Sequencing	TCCCTTTACCGGCCAAATCT
Sp6_OT15_R	Off-Target Sequencing	TCATTCAGCAACAGGACCCT
Sp6_OT16_R	Off-Target Sequencing	TGCTGCTAACGTGTAACCCA
Sp6_OT17_R	Off-Target Sequencing	TTCTCTGGTCGTATGTGGCT

S2 Table. Oligonucleotides used for *Nxf1* E610G genome editing and validation.
Sequences are presented for production of single guide RNAs and synthetic Cas9 mRNA, for single-stranded donor templates for homology-dependent repair, and for sequencing target and predicted off-target cleavage sites for the guide RNA used in genome editing.

IAP_Conserved_Queries

Name	Sequence
>IAP.1	GACGAGAAAAAACTCCGGACTGGCGCAGGAGGGAT
>IAP.2	CCGTAACACAGACTGTTGAGAAGGATTCAACTGCC
>IAP.3	CAGAACTCATCAGCTGGGGAACGACGGTGATAAAG
>IAP.4	CTTTCCAAGGGTCTTGAACCCGAGGAAAAGTTAAG
>IAP.5	GGCAGTGCGCAGACTCATTCATTCCCCGAGAGGAA
>IAP.6	GGAGCCGAGGGTGGGCGTGTCCACGCTCCGGTAGA
>IAP.7	AGAGAGATTGGACTTTTGACTTGTTAACGGGTCAG
>IΔ1.1	TCTCTCGAGCAGGTGAAACCACTGGTCAGTTAACA
>IΔ1.2	GATTTTGTGGCCAGAATGACAGAGGCAGCAGAGCG
>IΔ1.3	CGAGCAAGCCACAAAGGAGTGCCGAGCGGCCATAG
>IΔ1.4	CCTCTCAGCAATGCAGGTTTAGCGGCTGCCATCCT
>IΔ1.5	AGAGCTCCAGATAAACAGGGAGGGACTCTCACTCT
>IΔ1.6	AAAACGGGTCACCGGGCCCTCGGTCCCAGGGCCCT
>IΔ1.7	CAGGACCCAGGAAGCAGTCAGAGAGACGACCCAGG
>IΔ1.8	TTCAGCCGGTGCCGGTGGAGCCTATACCATCCTTG
>IΔ1.9	AACATTCCCCTGAAATACAGGTCCTGTGCTCAAGC
>IΔ1.10	GACCTGAGATAAAGAAAATGGGCTCCACAGGAAAT
>IΔ1.11	CAGAGTCATCTCATTCATTACAGGGCCTAGGATAT
>IΔ1.12	ATGAAAACGCCCCATCAGGAGGGTATTCAGCTAAA
>IΔ1.13	ACCCATACCATGGAAAACAGGGGACCCAGTGTGGG
>IΔ1.14	CTCTACCTCACCCTGGAATACTCCAATTTTTGTAA
>IΔ1.15	GGGTCTCCCTGTACTTTCCGCCTTACCACGTGGCT
>IAP.8	GGAACAATTCCCCTCTTTAATTTTGCTCCTTTACA
>IAP.9	CCATAAAGACCTTACCATGCTACAAAAGGCATATC
>IAP.10	CAAATTTCTGATACAGGACAATTCTTGGGCTCTGT
>IAP.11	CTTAAATGATTTTCAAAAGCTGTTGGGAGATATTA
>IAP.12	CCCTAGGACTCTTACTCTAGCTGCTAACCAGGCCT
>IAP.13	GAATGCACAATTACAACGTATTGAGGATTCGCAGC
>IAP.14	CGTATCCCCAGCTAAAATAATAGATTGGTATCCTG
>IAP.15	TACCGCTGCACAGGTTCAAACCTTGGCAGCCGCAT
>IAP.16	AAATCAATCTGTTGTGTTTCCACAAATAACAGTAA
>IAP.17	TAGGTGCCTATGTGGCTAATGGTAAAGTGGTATCC
>IAP.18	ATCTTTTAGAAGTGGCTGGAGTGATTAAGCCTTCC
>IAP.19	GAGCCCATTCAGGCCTACCTGGCCCCATGGCTCTG
>IAP.20	CTAAAGTGGTGGCTGCTGCCCTATCATCCCCGGTA
>IAP.21	GTGACATTGTTACTCAATGTCAAAGCTGCTGTGAG
>IAP.22	CCTTTGGAAAACTTCAATATCTCCATGTGTCCATT
>IAP.23	GTGCTTGGGGGAAACCCAGACTCCTTAAGACTGAT
>IAP.24	CTCAAGGACAGGGTATTGTTGAGCGTGCGCATCGC
>IAP.25	CGTCATTGTTCAGAGCCAGATAGGCCCAATGAGAT
>IAP.26	AATAAATGGTATGGCCCGGATCCTATCTTGATAAG
>IAP.27	TGACCAAGGGAATACTAATGTCCCTCGTCTTGGTG
>IAP.28	TGCTCAAGTATTCTCCTGCTTTTTTACCACTAACT
>IAP.29	AACAGGCCTCAGCTAATGTCAAGATACAGAGAGGT
>IAP.30	ACTCAGCAGGGATGTGAACAAAAGTTTCCGGGATT
>IAP.31	ATTTGTCAAAAAGTCTTTTTCAGTATATGTTACAG
>IAP.32	ACCCAATTGGATCTCCTCAGCATTTTCTTTCTTTA

S3 Table. IAP seed sequences.

Sequences used in genome searches for IAP sequences in introns.

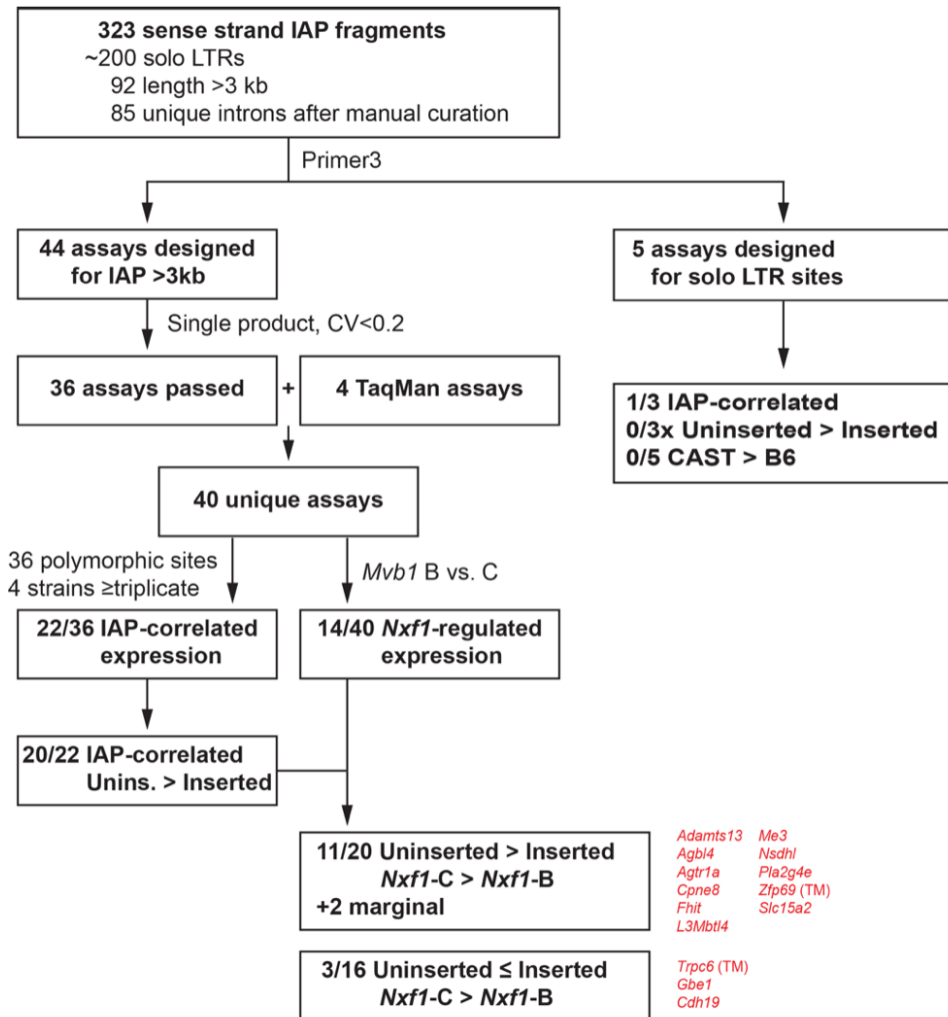| Target	RefSeq	Primer 1	Primer 2	Primer 3	Product size	Alternative product size	Majority ESTs	Comment	EST Ratio	Rationale	Note
IAP-intron alternative processing, qRT-PCR, Figure 2B											
Agbl4 (4931433A07Rik)		TGATGGGCGTTTGACGTGAAT	GAGTTGAGCGTGCATGTCAAT		145	NA	NA			sense IAP in intron	
Agpat4		TGCTCTCCGGACTTAACAT	AAGGGCCATTTTGGTTTTCT		156	NA	NA			sense IAP in intron	
Aik		GCCTGATGCCTGAGAGAAG	AGGAGGAGAAGGGCGTTTCTG		149	NA	NA			sense IAP in intron	
Amph		GGGCATTCTCGCCAAAAAC	CGGAGTTCTCGCTGAAGTCT		164	NA	NA			sense IAP in intron	
Bdh1		TCCCAGGAAAAGCCCTAAGT	GAACCCAAATCCAGAGTCACA		170	NA	NA			sense IAP in intron	
Cdh19		GGAAGGCGTGTGGAAACTGAA	CATTGGCGTCACCTTGATG		147	NA	NA			sense IAP in intron	
Cnih3		AGGAGAGCCGTGTGAGCTAAG	CACACAAAGGAGAGGAGATCA		121	NA	NA			sense IAP in intron	
Col23a1		GCCTGTTACGGAGAAGTTG	GGTAGCCATCTCGTCGTGAT		159	NA	NA			sense IAP in intron	
Cpne8		CACAAGGAGTTGGGAAATAAGA	CAACATCATATAAGTCAAAACGAAGA		147	NA	NA			sense IAP in intron	
Csmd1		CTCTGGTTCACCACGGACTT	CCAAGATGTAGCCGGAGAGA		175	NA	NA			sense IAP in intron	
Dph5		GGGGCTAACAACACAGTGA	AGTGTCCGTCCAAAACACAA		159	NA	NA			sense IAP in intron	
Dtd1		CAGCTTCGCAAGAGCTACAG	CTCAGCTGCTTTGGGTCAG		149	NA	NA			sense IAP in intron	
Enoph1		CAAAGTTCACGTAATTCCTGTT	TGAACGTCCTTGCTGACAATC		147	NA	NA			sense IAP in intron	
Fhit		TTTGGCCAACATCTCATCAA	AAACAAATCGGCCACTTCAT		165	NA	NA			sense IAP in intron	
Gbe1		CACAGGCCTAGGTGAGTGAA	CCGCTGCGTCAGAATCTAGT		168	NA	NA			sense IAP in intron	
Kcns3		AGGAGGAGCTTCTTGGATGA	TGGATTGAGAGGCGAGAGTG		135	NA	NA			sense IAP in intron	
L3mbtl4		GCGGGACACTGGGATATG	TCTCCAGCGCGTTCTTATCT		167	NA	NA			sense IAP in intron	
Lfp3		TTAGACGCCGTGATCTCTC	TTGGCTTAACAGAAATGTGTGAG		153	NA	NA			sense IAP in intron	
Lpin3		CCTCCAGAGACATTCCTTCA	GCTCAATGGGTTAGCTCCTCA		138	NA	NA			sense IAP in intron	
Me3		CTGTACATCGGCCTGAAACA	CATTGGGATTGGCAAAGTCT		136	NA	NA			sense IAP in intron	
Nsdhl		TGCTGGAGCGAGAGCTATACT	GCAGTGGAAAACTGTGCTTACA		143	NA	NA			sense IAP in intron	
Nudcd3		TGGCATCCGTGAGAACTACA	GAGGACACGTCTCCATTCT		160	NA	NA			sense IAP in intron	
Pde4d		TTGTTGGAGGCCATCTCACT	CAACCAACCATCCATCAACA		147	NA	NA			sense IAP in intron	
Pih2D		TCCTTCTGTCGGTTGGTCTG	GCTCCCTTCAAAGCTGAT		131	NA	NA			sense IAP in intron	
Pla2g4e		TGTTGCCGTCCTCACCTTTAAC	CAGAGGGTTGACAAAGCAGTCT		142	NA	NA			sense IAP in intron	
Rftn1		AGCTTCACGAACGAAGGCTAC	GGGCACAGCGTACGAACTT		156	NA	NA			sense IAP in intron	
Rgs3		CATGGCCTTTTCCCTTTGTA	TCCAGCCAGAGAAAATGGAC		168	NA	NA			sense IAP in intron	
Rsrc1		ATGGCTTCCAGAAGGAAAAGA	TGCTACGGGGTTCTTCTGTGA		149	NA	NA			sense IAP in intron	
Scg5		ACATCCAGAGGCTGCTTCAT	AAAAGGACCCAGATGCTGAA		135	NA	NA			sense IAP in intron	
Slc15a2		CATGCTAAGAGGGGATGTGAA	TCAGGAGGCGGGTTTCTGTA		147	NA	NA			sense IAP in intron	
Slc20a2		AGGAGGACCCTGTTCCAAAC	CGAAGGATATGAGCGCTATTG		150	NA	NA			sense IAP in intron	
Slc25a21		CTGGGTTAAGGATGCGGGACTA	ACCCCAAGCCTTCCTTCTTA		153	NA	NA			sense IAP in intron	
Slc9a2		CCGTGTTCACCCTGGATTAC	GAAGCCCGACCATTATGAGA		150	NA	NA			sense IAP in intron	
Snig1		AACATCCCGGTTGTCATTTC	AGCATTCCGAAGAACCTGAA		118	NA	NA			sense IAP in intron	
Snon2b (4933437K13Rik)		CCAATGCGTTCCATGTGTAT	TGCGGGAAAGGACTCTGTGAA		159	NA	NA			sense IAP in intron	
Spon1		ACTTGCTAAGACTGGGCTGAC	CGTGTCCTGTCCACTGAAGTA		154	NA	NA			sense IAP in intron	
Spnk2		GCTCCTGATGTCAGTTAACTCTG	CCCAGAATCTCCTCCTCTG		148	NA	NA			sense IAP in intron	
Tmco3		CATGGAGCCTATCCGTGACT	CAGCACTGCCAACACAAACT		146	NA	NA			sense IAP in intron	
Tmem163		AGTGGCCGCCTTCACTGT	ACTCCCTGTTTGCAGAGTGC		130	NA	NA			sense IAP in intron	
Trpc6		TGAATGGGGATGCAGAGACT	GACAGGAGCTGTTGCTGACA		144	NA	NA			sense IAP in intron	
Zbp89		AGGATCCCAGGAATTTGGTTGA	TATCCTGCCACTGACACGAG		153	NA	NA			sense IAP in intron	
Tssn1		GGTCACTGGTTGATGAGCTAC	CTAGAATTGTGTCCGCGGACTG			NA	NA			sense IAP in intron	
NonIAP alternative splice sites, Figure 3A											
Hmpaq1	NM_001039129	GTGGGGATGGCTAATGGA	CACCACCTCCAAAATTGCTT	CGTCAGGAGACTGGGAAGA	194	35	even	even cassette exon	even	Nxf1 interactor, known alternative splice	acrylamide gel
Tbcc1	NM_153552	CTGATGAGCAATCACTCTGGA	TCCATGAGTTCCAATTTTCCTC		138	335	138	retained intron		Nxf1 interactor, known alternative splice	
Nup98	NM_022979_1	GGATTTTGGTGCAAGTTGGTTC	AGCTGCTCGGTTTACTTCCAA		213	164	164	alternative donor		Nxf1 interactor, known alternative splice	
Nxf1	NM_001110159	GGGAGGAAGGAGGAGACAAGG	CCTCCTCCTTAGGGATCCAG		262	127	262	cassette exon		Nxf1 interactor, known alternative splice	
U2af1	multiple	TTACCACTTTGGGGATCTGC	GCACTGTTTTGGGGATTCCA		75	89,156	75,89	cassette exon		Nxf1 interactor, known alternative splice	
Nup214	NM_172268.2	ACTGGTGGGTCAGTCTTTGG	CCCTCTTGTAGCATTTCCA	CAAGTGATGTCTCGGTGGTG	79	82	82 NAGNAG acceptor		11:0	Nxf1 interactor, known alternative splice	3-primer assay
Dsam1	NM_026102.2	TCTGCCTACCAGAAGGACAGCA	AATTCTGAGCGTCTCCAAACA		136	109	even	even cassette exon	3:4	Barash et al. 2010 Nature 465: 53-59	
Xpo4	NM_020506.1	TGCAGTTTCACAGAGCATATTT	ATGCAGAGCACTGGCAAATT		101	256	101	cassette exon	20:6	Barash et al. 2010 Nature 465: 53-59	
V6ta	multiple	AAGGTCACGGATCGCCTAC	TCCAGCCTCTCCGTGTTATC	GCATGATGGGCTTCATTCAC	111	132	111	cassette exon	18:7	Barash et al. 2010 Nature 465: 53-59	3-primer assay
Flnc	NM_001081185.1	GAAGGGGATGACCAAGTTGA	GAGAACCATGATGGCACCTT		288	189	288	cassette exon	10:4	Barash et al. 2010 Nature 465: 53-59	
Vegfa	NM_009505.4	GAAAGACACGAACAAAGCCAGA	GCGGATCTTGGACAAACAA	TTTGGAAGTTTTTCTTAAGGCTTTTCT	77	149	77	cassette exon	13:5	Known alternative splice	3-primer assay
Ank1	NM_001110783.1	GAAGGACACACAGTGGATT	CCCTGTTTATCCAGAGTCG		240	76	240	retained intron	6:4	Known alternative splice	
Array-predicted Nxf1-responding altEvents, Figure 3C											
uc009oxc+1_(67)		CTCCCAGTGAGAGCCTCAAA	CACACAGCCAATCCATTCTG	CGGACATTTTGCTGCTGAAG	222	288		alt_cassette		Array hit	3-primer assay
Kff5_1_s6e9		CTGGAGCTGGCTATAATGGA	AGGATGAATGGCGCTTAGAC		215	183		alt_end		Array hit	3-primer assay
Gale_1_s9e4		GGAGAGCTTTTTGGTTTTCT	GGAAGGCGATTATGGAGAGT		158	226		alt_start		Array hit	3-primer assay
Mrp2d_1_s8e6		GGACTCGCGGACGTAGAGAAC	GTAGAAGGGGAAGGAAGTGG		275	152		alt_5'splice		Array hit	
Slamf7_(104)		TTACCACTTTGGGGATCTGC	TCCTGTGCTTCTCAGCAGTG		147	252		alt_cassette		Array hit	
uc007fiy+1_1_s1e8		TCACCAGTGCATGGTTTCTC	ATCCGAAAACTGCCACACAT	CAAGTGATGTCTCGGTGGTG	160	287		alt_end		Array hit	3-primer assay
5730419l9Rik_(125)		TGTCTGCCTTGTGACTTCAT	CTTTCCTGGGAGGATCTGC		204	178		alt_cassette		Array hit	
Abca8_(1008)		TGCTCACCTTTGGTGAGTCAT	CGAAGTTCCATAGGACATGGTG	GCATGATGACCAAATAAGA	156	271		retained_int		Array hit	3-primer assay
Tic9c_(334)		AAAGGAGCGAAGGCTTTTTG	GTAGGGGCTGGTTTCCTTCCT		128	440		retained_int		Array hit	
Mkln1_(2565)		AGATGTTCCCTTGTCGAATTCA	GGGCACAAAGCAGTCAAA		183	248		complex		Array hit	
Fkbp14_(90) a		GGGAAGAAAGGAAAAGTATTTTAGTA	CTCCTGGAATTCGAAATGG		300	153		alt_cassette		Array hit	
Fkbp14_(90) b		TGGGATCACTGGAGGTTTCTC	CAAATTGGAAGTTGCAGAG		265	175		alt_cassette		Array hit	
Fkbp2_alt3_1		GGGCAGAAGTGAGAGCAC	CTGTGAAGATCCGCATCGTC		285	249		alt_3'splice		Array hit	
uc007fiy+1_1_s8e8		CTTGGACAACAGCACCACCAG	GAGTTGGAGATGGAGGGTTG		309	133		alt_end		Array hit	
Slc17a5_2_s10e25		ATTTCCTATGGGAGTGAGTG	ACCTTCTGGAAGAAAGGTG	TTTGGAGTTTTTCTTAAGGCTTTTCT	236	148		alt_end		Array hit	3-primer assay
C430048L16Rik_1_s9e8		AGGAGAGTGTGGTCTCCAAT	CTGTCAAGATCGCAGGAAGC		284	264		complex		Array hit	
Tmem138_(72)		TCGGAATGGCTCCTGTTATC	AAGCTGGTGGAGTTTTTCCA		165	239		alt_3		Array hit	
Ap1b1_(63)		GGGTCAGCGAGTAGTGGACTA	TCTCCCCTTTCCTTGGTTGTG	CGGACATTTTGCTGCTGAAG	168	129		alt_cassette		Array hit	3-primer assay
Ganab_(66)		GGACTTATGGCCTTTGAAGCA	CCTTTTCCTGGGTCCTCTC		199	133		alt_cassette		Array hit	
Rwox_1_s9e5		TTTTGGAAAGCGAAGGTATGG	ATGACGGTGACCAGAGTGC		375	334		alt_start		Array hit	
Stx3_2_s17e30		CCAGCCATCTTCAACTCTGG	CTGTGTGATGACATTCAACTCT		199	180		alt_end		Array hit	
Trpc6_1_s6e27		TTAGCAAGTGAGCTGGCAGTG	GCTCTTTCCAGCTTGGCATA	TGCTGACAGTTTGGATGAGC	299	256		alt_end		Array hit	3-primer assay
Wdr53_(227)		CCTCATCAAGGTTGGGAAGGA	GGGCAGAGAGGTTTGGAGTCGTA	GATGGACCCAGGATCTCG	242	155		retained_int		Array hit	3-primer assay
Serpina3m_1_s0e15		CAGGAGTCAAGCTATCACAGAGG	CCCTGGTGGATGTGTGCT	ACACCAAGACAATGGGAGAC	389	246		alt_start		Array hit	3-primer assay

S4 Table. PCR primers used to assay RNA abundance or isoform variation.
Primer sequences are given, by class of assay, for dye intercalation quantitative
PCR (Fig. 2B and Fig. 6), cassette exon inclusion and alternative splice site usage
(Fig. 3A), and alternative events predicted by microarray data (Fig. 3C).
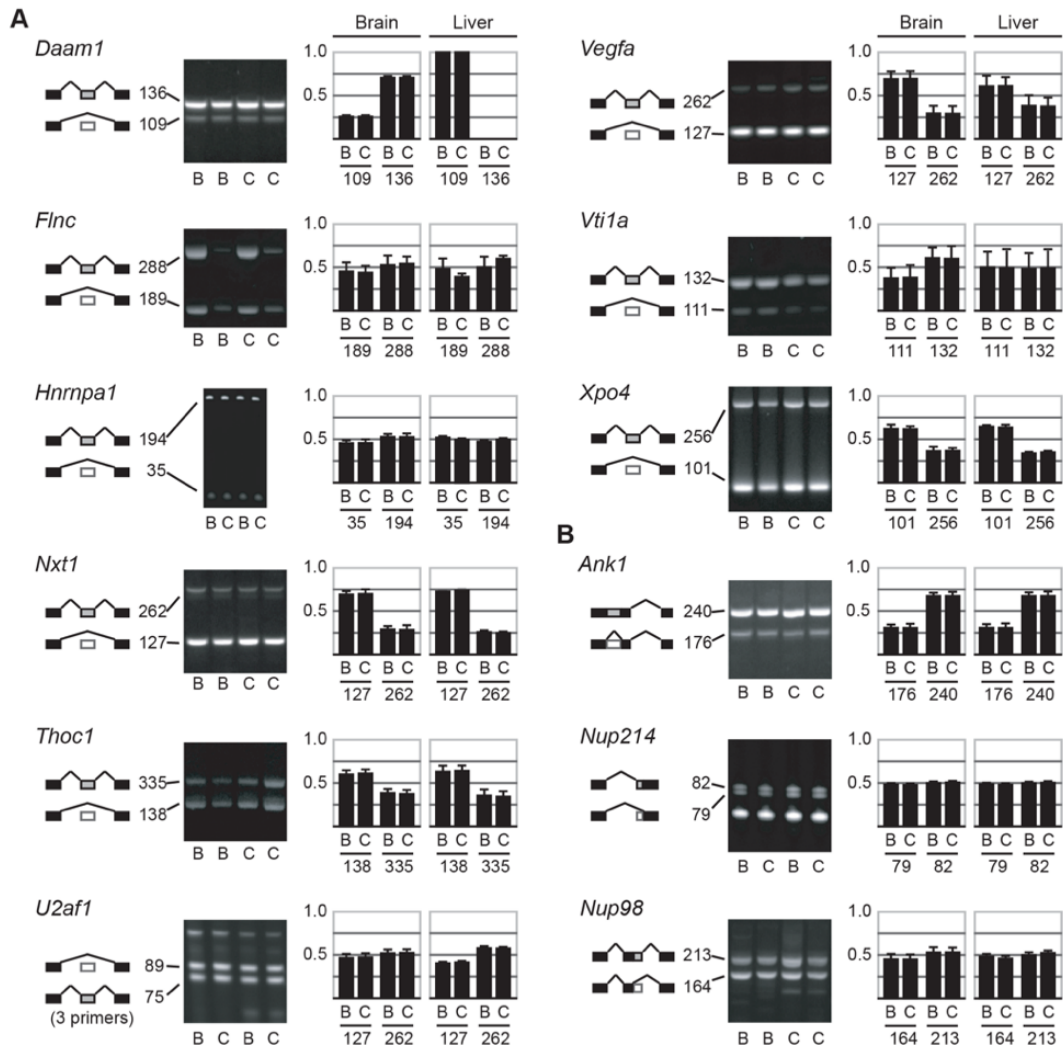
S1 Fig. Galaxy workflow to identify intronic IAPs.
Mouse reference assembly mm9 was accessed on 1/15/12 and analyzed in Galaxy using the workflow as illustrated. Numbers in parentheses indicate the number of objects contained in each product after each operation. Operations to identify strand specificity are color coded.
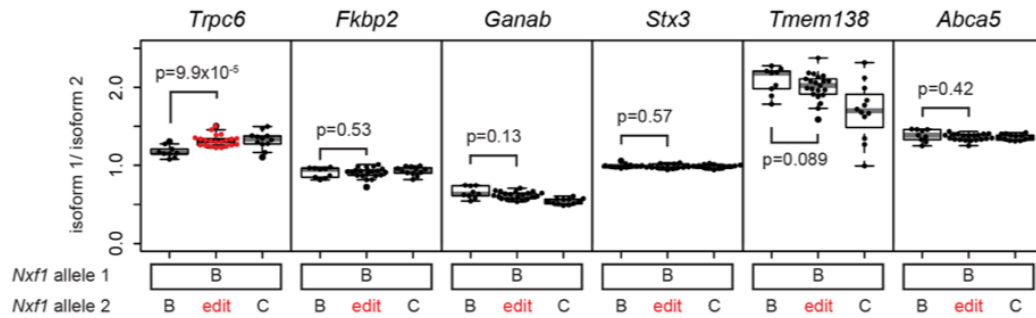
323 sense strand IAP fragments
~200 solo LTRs
  92 length >3 kb
  85 unique introns after manual curation

Primer3

44 assays designed for IAP >3kb

Single product, CV<0.2

36 assays passed + **4 TaqMan assays**

5 assays designed for solo LTR sites

1/3 IAP-correlated
0/3x Uninserted > Inserted
0/5 CAST > B6

40 unique assays

36 polymorphic sites
4 strains ≥triplicate

Mvb1 B vs. C

22/36 IAP-correlated expression

14/40 *Nxf1*-regulated expression

20/22 IAP-correlated Unins. > Inserted

11/20 Uninserted > Inserted
***Nxf1*-C > *Nxf1*-B**
+2 marginal

Adamts13   *Me3*
Agbl4   *Nsdhl*
Agtr1a   *Pla2g4e*
Cpne8   *Zfp69* (TM)
Fhit   *Slc15a2*
L3Mbtl4

3/16 Uninserted ≤ Inserted
***Nxf1*-C > *Nxf1*-B**

Trpc6 (TM)
Gbe1
Cdh19

S2 Fig. Assay development and quality control workflow.
The number of successes and attempts in each step are indicated. PCR primers
were selected in Primer3 web interface (http://bioinfo.ut.ee/primer3-
0.4.0/primer3/input.htm). Both melt profile after real-time PCR and gel
electrophoresis were used to determine that each PCR produced a single
significant product. Coefficient of variation after normalization to reference genes
was used to eliminate assays with limited power to detect differences by
genotype. Gene assays whose measurements differed significantly
by *Nxf1* genotype in the congenic strains are shown in red. Inequality symbols in
the lower boxes indicate groups with greater or lesser levels of correctly
processed RNA based on qRT-PCR assays described in Fig. 2 and accompanying
text.

S3 Fig. *Nxf1* alleles do not affect alternative splicing at selected non-IAP splice choices.

(A) Gel assays for well-characterized alternative splicing of cassette exons, including several encoding Nxf1-interacting factors, show no difference in general or tissue-specific inclusion rates between $Nxf1^{B6}$ (B) and $Nxf1^{CAST}$ (C) alleles in congenic mice. Splice choice is diagrammed, with PCR product sizes indicated. Bar graphs show proportion of fluorescence intensity in each band among 13–20 independent sample pairs tested for each gene. **(B)** Single examples for a retained intron (*Ank1*), alternative splice acceptor (*Nup214*) and alternative splice donor (*Nup98*) show no difference between *Nxf1* alleles.

S4 Fig. Alternative splicing in the congenic interval is not explained by E610G.

The alternative events in the congenic interval on chromosome 19 that were found by the arrays and confirmed in gel assays were re-assayed in co-isogenic E610G genome-edited animals and controls, using the samples from Fig. 6 and the assays from Fig. 3. *Trpc6* gel assay, included as a positive control, showed suppression of the IAP-dependent splice form. None of the four chromosome 19 alternative events that were significant in both array (Fig. 3B) and gel (Fig. 3C) assays showed significant differences between edited and unedited heterozygotes with respect to *Nxf1^{B6}*. Indicated p-values are for the one-tailed Wilcoxon rank sum test for the direction of difference seen in Fig. 3C, with the exception of *Abca5*, included as a negative control, which shows the value for the two-tailed test, as it had no predicted difference. Inclusion of heterozygous congenic samples confirms the effect, detection sensitivity and allele-independents effects predicted from Fig. 3C for *Ganab* (p = 0.00015) and *Tmem138* (p = 0.0036), but not for *Fkbp2* (p = 0.28) or *Stx3* (p = 0.40), albeit comparing new edited with archived congenic samples.

S5 Fig. Lack of strong correlations between genomic context and IAP-associated expression changes.

Plots show relationships between indicated attributes of the studied introns (x-axes) and expression ratio between uninserted strain and inserted strains with respect to the IAP (U/I ratio). U/I ratio is expressed as log10 to permit inclusion of exceptionally strong effects at *Adamts13* and *Zfp69*. Dashed line at zero indicates elements with no measurable effect. Spearman correlation coefficients (ρ) and associated p-values are shown. For categorical annotations of alternatively spliced (alt), alternative 3' end formation (altEnd), constitutive (con), and 3' terminal (end) exons, p-values are derived from the Kruskal-Wallis test. The magnitude of U/I ratio is moderately well correlated with its p-value, as expected for events near the threshold for discrimination and in the presence of other potential modifying effects across strains.

S6 Fig. Lack of strong correlations between genomic features and *Nxf1*-dependent suppression.
Plots show relationships between indicated attributes of the studied introns (x-axes) and expression ratio between $Nxf1^{CAST}$ and $Nxf1^{B6}$ alleles (C/B ratio). For categorical annotations of alternatively spliced (alt), alternative 3' end formation (altEnd), constitutive (con), and 3' terminal (end) exons, p-values are derived from the Kruskal-Wallis test. Nominal evidence for correlation to distance from 3' exon does not survive correction for multiple tests, but may merit further study. As expected, C/B ratio is correlated with U/I ratio, as larger mutational effects should provide a larger phenotypic space for suppression.

68

A: Full length IAPs

Alignment length: 7521
Input median: 7119
Input range: 7065-7268

B: IΔ1 IAPs

Alignment length: 5604
Input median: 5343
Input range: 5295-5471

C: LTRs

Alignment length: 609
Input median: 367
Input range: 339-483

S7 Fig. Phylogenetic relationships among IAP elements.
Neighbor-joining trees were generated from alignments in MUSCLE
(http://www.ebi.ac.uk/Tools/msa/muscle/) for the indicated IAP groups or
sequence components. Elements are colored as in Fig. 2, red for those with strong
evidence for *Nxf1*-mediated suppression, grey for those with no evidence of
mutagenic effect across strains. Nucleotide length of the total alignment, as well
as lengths of the median, shortest and longest input sequences are indicated for
each class or segment of IAP sequence. (**A**) Full alignment of IAP sequences
>7kb ("full length"). (**B**) Full alignment of elements containing the IΔ1 deletion.
Alignments of specific sequence components are (**C**) LTRs, (**D**) interval between
LTR and *gag*, (E) *gag*, (**F**) *prt*, (**G**) *prt* to *pol*, (**H**) *pol*, (**I**) *pol* to the RTE, (**J**)
RTE, (**K**) ppt.

69

S8 Fig. RTE secondary structure does not determine *Nxf1*-dependent suppression.
Secondary structures of RTE RNAs were predicted using the mfold web server version 3.4 (http://mfold.rna.albany.edu/?q=mfold/download-mfold) under default parameters. **(A)** RTE phylogeny, as in S7 Fig, panel J. **(B)** Optimal (left column) and next best (right column) predicted secondary structures of RTE RNA sequence for each unique phylogeny group, based on calculated lowest free energy, are depicted. **(C)** Free energy of optimal structures, and **(D)** difference in free energy between optimal and next best secondary structures are listed. In addition to identical RTE sequences with divergent outcomes, RTE structures from adjacent positions in the tree also showed dissimilar effects (e.g., *hes*, *mgL*, *Nudcd3*, and *Nsdhl*).

RESEARCH ARTICLE

Nmf9 Encodes a Highly Conserved Protein Important to Neurological Function in Mice and Flies

Shuxiao Zhang[1], Kevin D. Ross[1], Glen A. Seidner[2], Michael R. Gorman[3,4], Tiffany H. Poon[1], Xiaobo Wang[5¤a], Elizabeth M. Keithley[5], Patricia N. Lee[6¤b], Mark Q. Martindale[6¤c], William J. Joiner[1,2,4], Bruce A. Hamilton[1,7]*

1 Biomedical Sciences Graduate Program, University of California, San Diego School of Medicine, La Jolla, California, United States of America, 2 Department of Pharmacology, University of California, San Diego School of Medicine, La Jolla, California, United States of America, 3 Department of Psychology, University of California, San Diego School of Medicine, La Jolla, California, United States of America, 4 Center for Circadian Biology, University of California, San Diego School of Medicine, La Jolla, California, United States of America, 5 Division of Otolaryngology-Head and Neck Surgery, University of California, San Diego School of Medicine, La Jolla, California, United States of America, 6 Kewalo Marine Lab, University of Hawaii, Honolulu, Hawaii, United States of America, 7 Department of Cellular and Molecular Medicine, Department of Medicine, University of California, San Diego Moores Cancer Center, and Institute for Genomic Medicine, University of California, San Diego School of Medicine, La Jolla, California, United States of America

¤a Current Address: Otonomy, Inc., San Diego, California, United States of America
¤b Current Address: University of Hawai'i at Mānoa, Honolulu, Hawaii, United States of America
¤c Current Address: The Whitney Laboratory for Marine Science, University of Florida, Florida, United States of America
* bah@ucsd.edu

Abstract

Many protein-coding genes identified by genome sequencing remain without functional annotation or biological context. Here we define a novel protein-coding gene, Nmf9, based on a forward genetic screen for neurological function. ENU-induced and genome-edited null mutations in mice produce deficits in vestibular function, fear learning and circadian behavior, which correlated with Nmf9 expression in inner ear, amygdala, and suprachiasmatic nuclei. Homologous genes from unicellular organisms and invertebrate animals predict interactions with small GTPases, but the corresponding domains are absent in mammalian Nmf9. Intriguingly, homozygotes for null mutations in the Drosophila homolog, CG45058, show profound locomotor defects and premature death, while heterozygotes show striking effects on sleep and activity phenotypes. These results link a novel gene orthology group to discrete neurological functions, and show conserved requirement across wide phylogenetic distance and domain level structural changes.

Author Summary

Genome sequencing projects have identified large numbers of genes that encode proteins of unknown function. Many of these genes show strong evolutionary conservation, predicting important and well-conserved functions. A fraction of these show strong

conservation of core domains but dynamic changes in other domains, predicting both conserved and lineage-dependent functions. Here we identify neurological functions associated with one such gene identified by a forward genetic screen in mice. We use recently developed genome editing tools both to confirm the mouse studies and to test comparative functions in a model insect, the fruit fly *Drosophila melanogaster*. Each of these species has a single homolog of this gene family, but differ by inclusion of a ras-association (RA) domain present in most invertebrate species but missing in mammals. Null mutations in both mice and flies produce neurological phenotypes, but while the mouse phenotypes are comparatively mild (vestibular deficits, mild tremor, hyperactivity, mild circadian phenotypes and abnormal fear learning–but normal viability and breeding), null flies rarely survive to adulthood and surviving flies have severe locomotor deficits. Interestingly, heterozygous flies have significant sleep-related phenotypes. Together, our results provide a detailed first look at comparative function for a gene lineage with an unusual evolutionary history.

Introduction

The biological functions of many protein-coding genes remain unknown, often despite conservation across considerable evolutionary distance. Such "orphan" molecules may include genes whose functions affect fitness but not overt phenotypes in experimental settings, genes that affect overt but under-studied phenotypes, and/or genes that have been difficult to study due to unusual molecular properties, such as low level expression or poor physical annotation in genomes of well-studied organisms. Forward genetics offers one entry point into functional studies by highlighting those genes whose alterations produce measurable effects [1, 2]. A key challenge in de-orphanizing novel genes is to integrate their functions with observable sites of action, organism-level phenotypes and known cellular activities.

Nmf9 and its homologs were essentially unknown prior to our studies. The *nmf9* mutation was recovered in a N-ethyl-N-nitrosourea (ENU) screen at the Neuroscience Mutagenesis Facility (NMF) of the Jackson laboratory [3, 4] based on tremor and vestibular phenotypes. Several exons of the gene we identify as *Nmf9* had been systematically annotated as *Ankfn1* (for ANK and FN3 domain containing), while other exons annotated as separate genes were identified only by clone names, based on non-overlapping partial cDNA clones from high-throughput screens [5, 6]. We selected *nmf9* for study as part of a long-standing interest in mice with unusual tremors and ataxias [7–11]. The unusual conservation pattern we identified among *Nmf9* homologs, along with recent innovations in genome editing [12–18], motivated additional studies to test conservation of function in mice and flies, including the generation of equivalent mutations at the same amino acid position in both species.

While our work was in progress, two groups reported identification of the *Drosophila* homolog in very different contexts. A transposon-based screen for low-sleeping mutants by Mark Wu and colleagues found insertions in a poorly conserved, variably included 5' region of the gene, but concluded based on antibody studies that the resulting *wide-awake* (*wake*) alleles were null [19]. The same group also found that the fly gene interacts with and promotes cell surface localization of the fly GABA-A receptor encoded by *Resistance-to-dieldrin* (*Rdl*), through physical and genetic interactions. Independently, an RNAi screen for genes required for proper segregation of Numb during asymmetric cell division of sensory organ precursors carried out by Juergen Knoblich and colleagues identified the same gene, which they named *Banderuola* (*Bnd*). Intragenic deletion of most of the *Bnd* coding sequence resulted in pupal and early adult lethality [20]. This group also showed physical, genetic and functional

interactions with *Discs-large 1* (*Dlg1*), a prototypical membrane-associated guanylate kinase (MAGUK). We will refer to the *Drosophila* gene by its current annotation symbol, *CG45058*, and to 5' P element alleles as *wake*.

Here we use positional cloning, expression-guided behavioral studies, deep evolutionary constraint, and genome editing to define novel activities of *Nmf9* and its *Drosophila* homolog. Positional cloning of mouse *nmf9* identified a single point mutation at a splice donor sequence, resulting in exon skipping of a frame-shifting exon in *Nmf9*. The expression pattern of *Nmf9* in mouse brain suggested neural circuits that might be compromised in mutant animals. Behavioral tests confirmed deficits in vestibular function, fear learning, and circadian behavior. Female mice were more severely affected than males for several phenotypes. Sliding window analysis of relative constraint across the protein coding sequence showed that the skipped exon encodes the most evolutionarily constrained peptide sequence among Nmf9 homologs and genome editing of a single glycine to alanine at this site was sufficient to generate a non-complementing allele. Genome editing of the *Drosophila* homolog at three different sites produced null alleles that resulted in premature lethality and severe locomotor retardation in homozygotes. Remarkably, heterozygous flies showed mild activity and sleep-related phenotypes. Our studies thus show the first genetic, behavioral and molecular information for mammalian *Nmf9*, highlight functional importance of the most conserved sequence among its homologs, and resolve competing views regarding *CG45058* null phenotypes. These genetic, behavioral, and comparative studies provide a foundation for understanding activities of *Nmf9* homologues in broad context.

Results

Vestibular dysfunction in *nmf9* mice is sexually dimorphic and progressive

The *nmf9* mutation was recognized in a chemical mutagenesis screen based on tremor and vestibular phenotypes. In our evaluation, vestibular signs included circling, nodding, and head tilt (S1 Video) and abnormal landing and forced swim tests. Not every mutant was abnormal in every test, but all mutant animals were abnormal in at least one test and by visual inspection relative to co-isogenic non-mutant littermates. Visible head nodding, hyperactivity, and tremor were enhanced by light vertical acceleration or extended handling. In the landing test, animals were suspended by their tails and scored for trunk curling and attempts to rotate, rather than reaching for ground, by an investigator blinded to genotype. Mutant animals showed a significant increase in frequency of both trunk curling and rotation compared to littermate controls (Fig 1A). In the forced swim test, most control littermates swam with their snout above water for $\geq$1 min., while *nmf9* mutants typically were unable to remain righted above the surface and had to be rescued before 30 sec. to prevent drowning (Fig 1B). Differences between genotypes were more pronounced in females than males for vestibular phenotypes. The frequency and severity of circling and head nodding in mutant animals increased progressively from 21 days to 6 months (Fig 1C and 1D), though hyperactivity and tremor did not (Fig 1E and 1F); these phenotypes were essentially absent from control littermates. Histologically, however, mutant animals had grossly normal inner ear structures and did not show hearing impairment (S1 Fig), suggesting a mature functional, rather than a gross morphological, basis for vestibular defects.

Positional cloning of *nmf9* reveals a novel gene expressed at low abundance

We mapped the *nmf9* mutation by recombination in ~1000 F2 progeny from crosses between B6-*nmf9* and four mapping strains (Fig 2A). Intercross progeny included 332 from AKR/J, 245
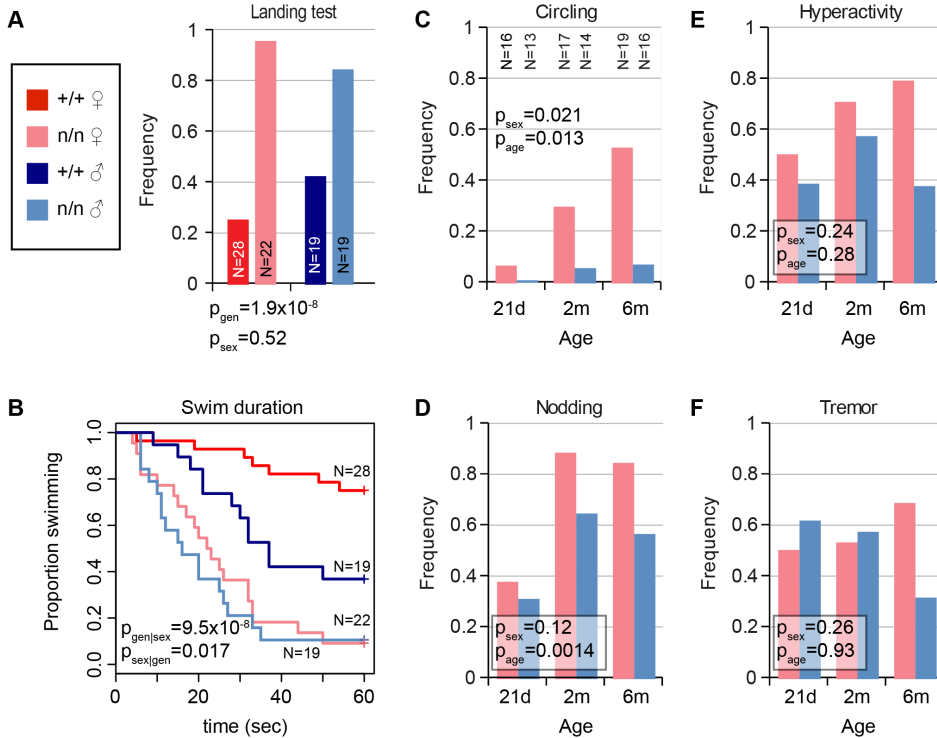
73

Fig 1. Mutant *nmf9* animals show age and sex-dependent vestibular dysfunction. (A) Landing test showed increased frequency of rotation or curling by *nmf9* homozygotes relative to control (+/+) littermates (Fisher's exact test, p = 1.9x10^{-8} between genotypes, 0.52 between sexes). (B) Forced swim test further showed vestibular dysfunction in *nmf9* relative to littermates (Asymptotic logrank test, p = 9.5x10^{-8} for geneotype stratified by sex) with a greater performance difference in females (p = 0.017 for sex stratified by genotype) due to lower performance among control males. Red, control; pink, mutant females. Navy, control; blue, mutant males. (C) Circling by mutant animals also showed significant sex bias (Fisher's exact test, p = 0.021) and age-dependent progression (p = 0.013) in frequency. Sample sizes apply to panels C-F. (D) Head nodding also supported progression (p = 0.0014), but not significant sex bias (p = 0.12). Although (E) hyperactivity and (F) tremor were pronounced in mutant animals, effects of sex and age on did not reach statistical significance.

from BALB/cJ, 371 from C3H/HeJ, and 119 from DBA/2J that were typed by PCR ([21] and S1 Table). The data confirmed a fully penetrant, recessive phenotype with no evidence of segregating modifier genes. Exclusion mapping placed *nmf9* within a 1.3 Mb interval (Chr11: 88,240,426–89,538,743 in GRCm38/mm10 assembly) that included several well-annotated genes. We previously estimated the frequency of induced nucleotide mutations in the NMF screen as ~1/Mb [4], which predicts a very low probability of confounding functional mutations in an interval of this size [22]. Sanger sequencing of all canonical and EST exons (Fig 2B) in mutant and littermate control identified a single mutation: a G-to-A transition in the splice donor U1 binding site of a frame-shifting exon of predicted gene *Ankfn1* (Fig 2C). RT-PCR, homologous cDNAs from other species, and detailed in situ hybridization patterns indicated that the major transcript of this gene comprised both *Ankfn1* and *4932411E22Rik* annotations
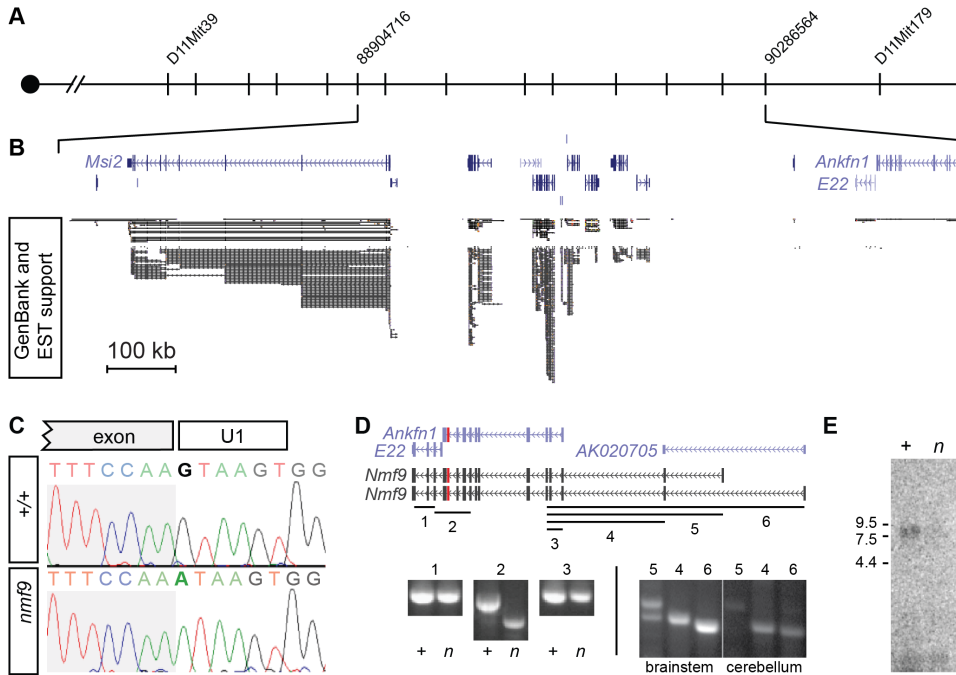
Fig 2. Positional cloning identifies *Nmf9*. (A) Recombination mapping places the *nmf9* mutation within a 1.3 Mb interval on Chr. 11. (B) Aligned UCSC Genome Browser window shows known and predicted genes in the non-recombining *nmf9* interval. Extent of EST support indicates poor sampling of putative genes *4932411E22Rik* and *Ankfn1*. (C) Sanger sequencing across the exon-intron junction shows a G-to-A transition in the essential GU splice donor adjacent to a frame-shifting exon in *Ankfn1*. (D) Schematic shows alignment of gene annotations from the UCSC genome browser (purple) with empirically determined transcripts (black) based on RT-PCR assays (numbered lines). Amplification and sequence analysis of assay 2 showed both inclusion of *Ankfn1* and *E22* in a single transcript in littermate control samples (+) and selective skipping of the mutated exon (red) in mutant samples (*n*). Additional assays indicated based on conservation, predicted exons and 5' RACE showed variable inclusion of exons 5' to the *Ankfn1* annotation. (E) Northern blot shows a single major band for *Nmf9* poly(A)⁺ RNA from brain ~9 kb, with reduced level in *nmf9* mutant (*n*) compared to control (+) littermate brain RNA.

as well as additional 5' exons (Figs 2D and S2). We refer to this transcript and locus as *Nmf9* to avoid confusion with the narrower definition of *Ankfn1*. RT-PCR across EST and predicted exons showed low overall abundance (requiring polyA⁺ RNA for detection from whole brain), skipping of the frame-shifting exon adjacent to the mutated U1 site, and variable utilization of alternative 5' ends (Fig 2D). RNA gel blot hybridization showed a single major size form and confirmed that while the transcript had very low abundance in control brain, the mutant transcript had still lower levels in *nmf9* homozygote brains, consistent with predicted nonsense-mediated decay for the frame-shifted splice product (Fig 2E). While it remains possible that other transcripts exist at low levels, our data support a single major open reading frame, with some variation at the 5' end.
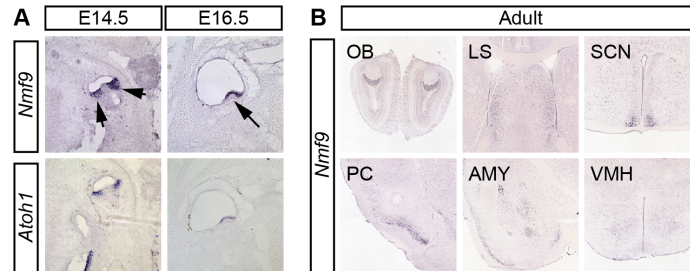
Fig 3. *Nmf9* expression pattern predicts new sites of function. (A) *In situ* hybridization to adjacent sections in E14.5 and E16.5 embryos show that *Nmf9* and *Atoh1*, a marker of hair cell progenitor were expressed in the same population of cells (indicated with arrows). Scale bar = 500 μm. (B) Panels from the Allen Brain Atlas [25] show *Nmf9* RNA enriched in distinct structures of adult brain, including accessory olfactory bulb (OB), piriform cortex (PC), lateral septum (LS), amygdala (AMY), suprachiasmatic nucleus (SCN) and Ventromedial hypothalamus (VMH).

Nmf9 is expressed in discrete sensory and CNS structures

To define potential sites of *Nmf9* action, we examined its pattern of expression. In situ hybridization to both male and female embryos and RT-PCR from dissected tissue showed *Nmf9* expression largely restricted to the nervous system. In vestibular inner ear, *Nmf9* expression fully overlapped *Atoh1* (Fig 3A), a well-studied marker for hair-cell progenitors [23, 24]. In situ hybridization showed that *Nmf9* was expressed as early as E14.5 in inner ear, nasal epithelium, ventricular zone, and the spinal cord (S2 Fig). In adult brain, both our data and data from the Allen Brain Atlas [25] showed expression enriched in a few centers (Fig 3B), including the accessory olfactory bulb (OB), piriform cortex (PC), lateral septum (LS), amygdala (AMY), suprachiasmatic nucleus (SCN), and modest enrichment in ventral medial hypothalamus (VMH). Probes corresponding to *Ankfn1* and *4932411E22Rik* annotated exons showed the same detailed pattern in our data (S2 Fig) and in comparable data from the Allen Brain Atlas, further supporting a single transcription unit. Identification of *Nmf9* pattern in mouse brain allows behavioral tests of neurological function for each of the major sites of expression, providing a structured approach to defining additional phenotypes in *nmf9* mice.

Fear learning and circadian behaviors are impaired in *nmf9* mice

To probe the functional significance of *Nmf9* expression in neuroanatomical structures outside of the vestibular system, we tested *nmf9* and littermate (+/+) control mice on selected behavioral tasks. Each task included enough male and female animals to assess potential sex differences, as female mice had shown earlier and stronger vestibular defects than males. A summary of statistical analyses is presented in S2 Table.

AMY. The amygdala is essential for fear learning. In both contextual and cued fear conditioning paradigms, *nmf9* mutants showed diminished freezing behavior after training compared to littermates, indicating decreased fear learning (Fig 4A). Similar to vestibular defects, fear learning was more strongly affected in females than in males.

SCN. The suprachiasmatic nuclei are essential for classic circadian behaviors, in which *nmf9* mutant animals showed significant defects. Animals were monitored for activity on a wheel running apparatus in 12 h light: 12 h dark (LD) and in constant dark (DD) conditions. The most severely affected *nmf9* mice showed less robust consolidation of activity into
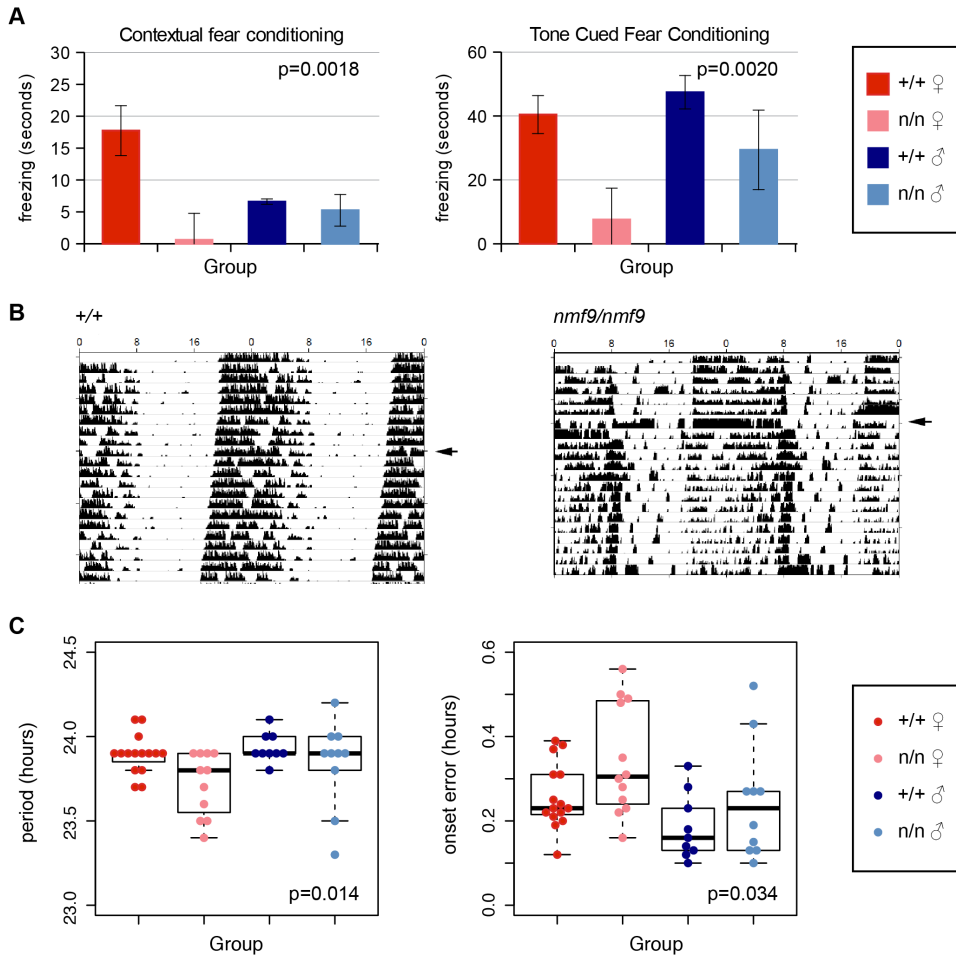
Fig 4. *Nmf9* is important to fear learning and circadian rhythm. (A) Mutant animals showed decreased freezing behavior in both tone cued and contextual fear conditioning. (2-factor ANOVA p = 0.0018 for genotypes in contextual conditioning, 0.002 in tone-cued conditioning. Sex effects were non-significant, p = 0.29 in contextual, p = 0.09 in tone-cued, N = 14 +/+ female, 8 mutant female, 11 +/+ male, 8 mutant male, although the interaction between genotype and sex was supported for contextual fear conditioning, p = 0.026.) (B) Double-plotted actograms of running-wheel counts of mutant and wildtype from the same litter housed in 12 hour light: 12 hour dark (LD) cycle and continuous dark (DD) illustrates less robust circadian rhythm in a more severely affected animal. Arrowheads indicate transfer to DD. (C) Mutants showed a small but significant decrease in period length (p = 0.014) and an increase in period onset error (p = 0.034) during DD (2-factor ANOVA; effects of sex p = 0.14 for period length, p = 0.012 for onset error, with larger deviation in females. N = 14 wildtype females, 12 mutant females, 9 wildtype males, 10 mutant males).

nighttime hours when compared to +/+ littermates (Fig 4B). Mutants in aggregate showed both a slight shortening of period and an increase in onset error in DD (Fig 4C). Hyperactivity

did not significantly affect activity counts in either LD or DD housing conditions (Fig 4B and S3 Table).

LS. Measures of innate anxiety, which are dependent on the lateral septum, appeared intact in *nmf9* mutants. Behavior on an elevated plus maze detected no significant difference between genotypes (Fig 5A). In an open field test, *nmf9* mutants spent more time in the center of the open field, which could indicate a functional change in LS, but the number of line crossings was also significantly higher, suggesting that the hyperactivity and circling behaviors in *nmf9* confound a simple interpretation of the open field results (Fig 5B). Marble burying, a third test of anxiety-related behavior that is less affected by activity, showed no significant difference between genotypes (Fig 5C).

PC, OB and olfactory epithelium. Piriform cortex plays an important role in olfactory sensing and behavior. Although PC, OB and olfactory epithelium showed significant *Nmf9* expression, mutants showed no significant difference from littermates in behavioral tests of olfactory function, which included a buried food finding task (Fig 5D) and odor habituation / dishabituation tests with either social or non-social odors (Fig 5E).

VMH. Ventromedial nuclei of the hypothalamus are important to satiety and feeding behavior, among other functions. Mutant animals show a slight but significant decrease in weight compared to littermates at 2–6 months (Fig 5F) but no significant difference in the amount of food consumed (Fig 5G). While these results were also potentially confounded by hyperactivity and energy expenditure, the small effect sizes suggested limited, if any, impact on feeding behavior.

Taken together, results of these behavioral data support selective functional impairments in anatomical circuits that express *Nmf9*.

Pattern of conservation among *Nmf9* homologs indicates modular architecture and novel domains

Homologs of *Nmf9* across wide taxonomic boundaries showed strong patterns of conservation that highlight select domains, as well as lineage-restricted modulation of domain architecture (Figs 6A and S4). Homologs were identified by reciprocal BLAST/BLAT searches in sequenced genomes of nearly all metazoan lineages, including placozoa and porifera, and in at least some choanoflagellates and filasterea, sister groups to animals that diverged from the lineage leading to animals after the split between animals and fungi. Choanoflagellate (*M. brevicollis*, *S. rosetta*) and filasterea (*C. owczarzaki*) homologs included an N-terminal CRIB domain (associated with binding to Cdc42/Rac subfamily small GTPases), a C-terminal Ras-association (RA) domain, or neither. A single instance in fungi–comprising only a choanoflagellate-like copy of the conserved, non-motif domains–was found in *Mortierella verticillata*, but not in basal or sister lineages, and might therefore represent a horizontal gene transfer event rather than an earlier origin of the gene genealogy. Sequenced invertebrate animal genomes had a single *Nmf9* homolog, except for an apparent loss in the urochordate lineage (0/4 species). Invertebrate homologs included an RA domain, except for the single placozoa sequence and a few genomes with incomplete assembly, but none included a CRIB domain. Jawed vertebrates basal to mammals contained two homologs: an ancestral copy with the RA domain and a derived copy without it. Mammalian genomes had only the derived copy. Most animal homologs occur in poorly annotated regions of their respective genome assemblies, limiting the number of complete sequences available among homologs we examined.

Analysis of evolutionary constraint among 14 animal species spaced by known evolutionary distance [26] confirmed conservation of the annotated ankyrin and fibronectin type 3 motifs, but also predicts three additional regions of unknown biological function under equal or
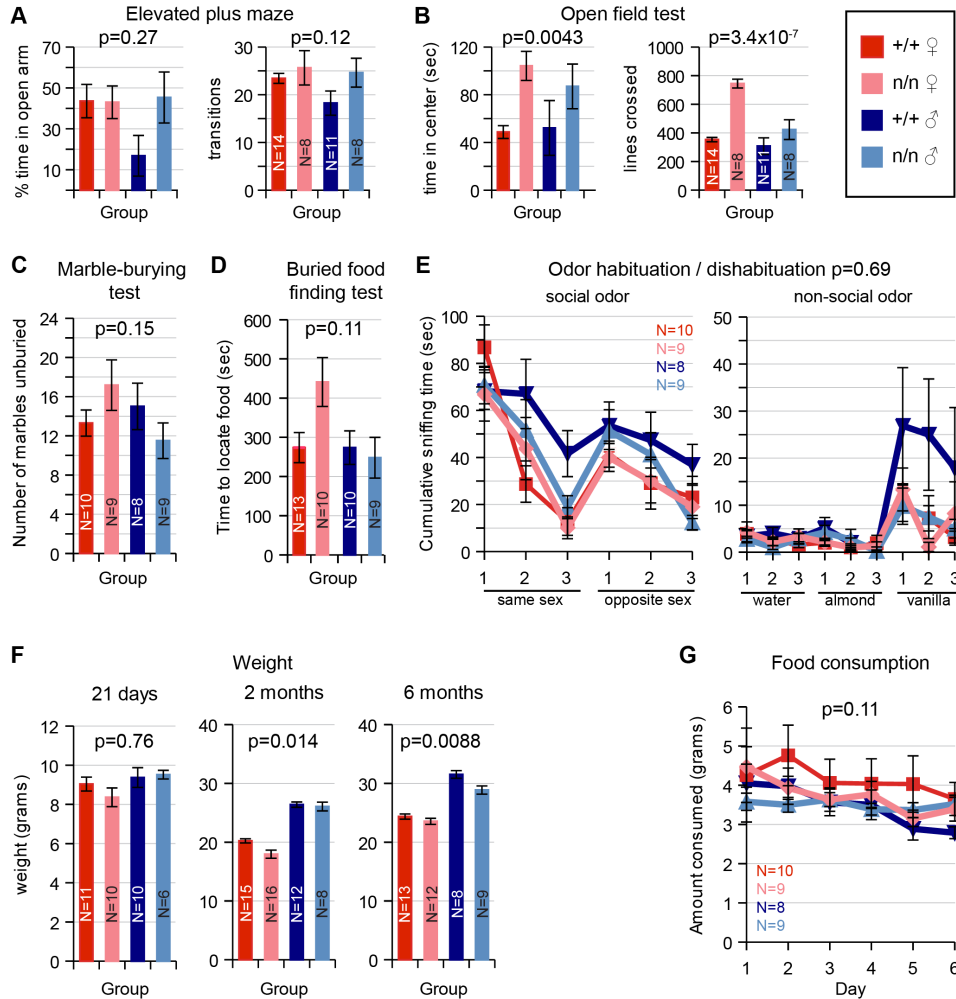
Fig 5. Tests of LS, PC/olfactory pathway, and VMH. Behaviors that depend on *Nmf9*-expressing circuits were assayed and assessed by 2-factor ANOVA or MANOVA for sex and genotype. Number tested for each group (N) is indicated in each bar or a legend to line graphs. (A) Elevated Plus Maze behaviors were not significantly different between control (+/+) and mutant (n/n) littermates (Percent time in open arm p = 0.27 for genotype, p = 0.22 for sex; Number of transitions p = 0.12 for genotype, p = 0.23 for sex). (B) In the Open Field Test, mutant animals showed increased time spent in center, but potentially confounded by increased number of line crosses due to hyperactivity (Time in center p = 0.0043 for genotype, p = 0.82 for sex; number of transitions, 3.4x10^{-7} for genotype, p = 0.0004 for sex). (C) No significant difference between genotypes was detected for the Marble Burying Test (p = 0.15 and 0.68 for genotype and sex, respectively). (D) While mutant females took somewhat longer in olfactory-dependent Buried Food Finding Test, this was did not reach conventional significance (p = 0.11 for genotype, 0.09 for sex). (E) Similarly, Odor Habituation / Dishabituation Tests did not support a significant difference (MANOVA, p = 0.69 for genotype, p = 0.30 for sex). (F) Weight was significantly different by genotype at 2 months (ANOVA, p = 0.014) and remained significant at 6 months (p = 0.0088), including well-known sex differences, while food consumption (G) was not significantly different (p = 0.11 for genotype, 0.75 for sex).

doi:10.1371/journal.pgen.1005344.g005

79

A

Ancestral copy lost

Gene duplicated; RA domain lost in derived copy

Gene lost in Urochordates

RA domain lost

Origin of Nmf9 orthology group

?

Variable inclusion of RA & CRIB domains

Mammals

Birds, Reptiles, Frogs, Fish

Lamprey

Lancelet

Tunicates (0/4)

Echinoderms

Arthropods, Mollusks, Roundworms, Flatworms, Annelids, Cnidaria

Placozoa

Porifera, Ctenophora

Choanoflagellates Filasterea

Fungi (Not found among >25 diverse fungal genomes) **HGT to Mortierella ?**

B

codons

C

choanoflagellate	IEPQLE GLYLSVLV PRSTVGGLQVTVLKHQAAVLPLVR
tricoplax	CVTFSR GLYLGYVK LRCTINDIAVMTDY---------
sea anemone	RKPLER GMYIGYLK LRTFVDAIHVVVQESNPNVLPHVK
nematode	QIQLQR GLYVGYLK LHSSLNSIRVVVPDVLPSMLPFTR
fruitfly	LQRLEP GLYLGYLK MQCSIDQIQVVVPVKTPNVLPHCK
sea urchin	SLPLTR GLYVGYLK LRSSLGNIRIQVPERHPSMLPHIK
lamprey	RQTLPP GLYLGYLK LCSSVDQIKVLVPQRIPNSLVHLK
zebrafish	RKCLPP GLYLGYLK LYTSVDQIRVLVSHKHPNVLCHVK
chicken	SHALPP GLYLGYLK LCSAVDQIRVLVPEQLPNILCHVK
mouse	HQRLSP GLYLGYLK LCSSVDQIKVLVTQKLPNILCHVK
human	HQRLFP GLYLGYLK LCSSVDQIKVLVTQKLPNILCHVK

Fig 6. Nmf9 protein is highly conserved in Metazoans. (A) Nmf9 homologs exist for all branches of Metazoa (except Urochordates) and several non-animal Holozoa. A single example in fungi (*Mortierella*

verticillata) may represent horizontal gene transfer. (B) Sliding window analysis of 14 approximately full-length metazoan homologs highlights known motifs and predicts three novel domains where the most highly conserved individual residues tend to cluster. Threshold line (red) indicates 25% percentile of values. Aligned schematic shows motif annotations (rectangles) and constrained intervals (ovals) in relation to the sliding window plot. (C) The most highly conserved amino acid sequence (red box) was in domain 2 and encoded by the exon skipped in *nmf9* mRNA.

stronger constraint ([Fig 6B](#)). Novel domain 2 was the most conserved region in the entire protein, including a GLYLGYLK sequence that is nearly invariant among animals and whose first glycine was the most highly conserved residue ([Fig 6C](#)). This region contained no motif annotation in current databases nor predicted post-translational modification sites. In a detailed analysis of 113 homologous sequences, Domain 2 was the most highly conserved sequence both among ancestral homologs and among derived homologs, with the GLYLGYLK signature providing the strongest sliding window support in both ([S5 Fig](#)).

In situ hybridization studies in diverse animal species suggested the potential for a conserved neuronal function. In contrast to mouse, the homolog in *D. melanogaster* appeared more broadly expressed during development, including both CNS and tissues outside of the nervous system ([S6 Fig](#)). Expression in the sea anemone *N. vectensis* also showed neural expression: in the planula, highest expression was in the apical tuft, a larval chemosensory organ, and in the pharyngeal nerve ring; in the polyp stage highest expression was in the endoderm, the predominant region of the net-like adult nervous system [27]. Ancestral and derived homologs may also differ in expression pattern. In the model fish *D. rerio*, the ancestral homolog appeared broadly expressed, as in *D. melanogaster*, while the derived homolog appeared more strictly localized, including distinct expression in the developing inner ear.

Genome-edited alleles confirm functional importance of conserved domain 2

To confirm the gene identification of *Nmf9* and to test the functional importance of conserved domain 2, we used CRISPR/Cas9-mediated genome editing in mouse one-cell embryos [17, 28] to target mutations to the conserved GLYLGYLK region. We recovered 26 G0 animals, of which 24 survived to adulthood. Sequencing of a 550-bp PCR product encompassing the target site showed that 21 G0 animals were edited on both alleles, 4 were edited on only one allele, and one was not determined. Among 21 edited on both alleles, 17 appeared to be homozygous (or possibly heterozygous to an allele that precluded amplification, such as a large deletion). G0 mice with both alleles edited to frameshift or other clearly deleterious lesions phenocopied the overt vestibular phenotypes of *nmf9* ([Table 1](#)). Four predicted pathogenic mutations were used for complementation tests with the original *nmf9* allele ([Fig 7A](#) and [Table 2](#)). In aggregate, G1 and later progeny carrying one edited allele heterozygous to the original *nmf9* mutation failed to complement in both the landing test and forced swim test ([Fig 7B and 7C](#)), confirming correct gene identification of *nmf9*. The first glycine of the GLYLGYLK sequence was conserved through metazoa and holozoan sister groups. Strikingly, an alanine substitution at this residue failed to complement *nmf9*. Although homozygotes for this allele did not show a strong phenotype, G-to-A/*nmf9* heterozygotes had the tremor, hyperactivity, and vestibular dysfunction characteristic of *nmf9* homozygotes ([Table 2](#) and [S2 Video](#)). These results support the functional importance of the highly conserved domain at a neurological level, even under the relaxed constraints of inbred laboratory mice.

Table 1. Mouse *nmf9* alleles generated by CRISPR/Cas9-mediated genome editing.

Mouse	Sex	Apparent Zygosity	Genotype	Mutations	Hyper	Nod	Tremor	Landing	Fig 7 label
1 (Dead)	-	Homozygous	-6	CTCTCAC>A	-	-	-	-	
2	F	Homozygous	+13	-CTCACCTGGACTGTATCT + 31	2	2	0	1	-13,+26
3	M	Homozygous	substitution	C>A	0	0	0	0	
4	M	Homozygous	substitution	TGG>CGC	0	0	0	0	G592A
5	F	Heterozygous	not determined	not determined	0	0	0	0	
6	F	Heterozygous	-33	-TCTCACCTGGACTGTATCTAGGTTACCTAAAGC	2	2	0	2	
7	F	Homozygous	-7	GGCTCTCA>T	1	1	1	0	
8	M	Homozygous	+24	+TGACTGCTTTAGGTAACCTAGATA	2	1	1	2	
9	M	Homozygous	-10	-GCTCTCACCT	0	1	1	2	
10	M	Trans-Heterozygous	+1/-4	+C/-CTCT	2	2	2	2	
11	F	Homozygous	>300bp insertion	[not depicted]	1	2	0	2	
12	F	not determined	not determined	not determined	0	1	2	2	
13	F	Homozygous	-12	-CTCACCTGGACT	2	2	1	2	-ΔLSPG
14	F	Homozygous	-4 & substitution	-ACCT and 2 substition	2	2	1	2	
15	F	Homozygous	+98 & substitution	+98 & TGG>CGC	2	2	1	2	
16	M	Heterozygous	-3	CTCTCACCTGG>ATACCCGC	1	0	0	0	
17	M	Homozygous	+1	+T	1	2	0	2	
18 (Dead)	F	Heterozygous	-12	-TCAGCGGCTCTC	-	-	-	-	
19	F	Homozygous	-220 & substitution	TGG>CGC & -220	1	0	0	0	
20	F	Homozygous	-169	[not depicted]	2	1	0	2	
21	F	Heterozygous	-6	CTCTCACCT>ACA	2	2	1	2	
22	F	Homozygous	-4	-TCTC	2	2	2	1	
23	F	Homozygous	inversion	[not depicted]	0	2	1	2	
24	F	Trans-Heterozygous	-11/complex	CCATCAGCGGCTCTCACCT>CTCTCCCT	2	1	1	1	
25	F	Trans-Heterozygous	+1/-4	+C/-CTCA	0	1	0	1	
26	F	Homozygous	-3	GGCTCTC>TCAT	0	0	0	2	RLS>LI

Phenotypes are noted as 2 for strong, 1 for mild, or 0 for not evident.

Null alleles of the *Drosophila* homolog have viability and locomotor deficits and heterozygous effects on sleep

To test conservation of function at an organismal level, we similarly edited the *Drosophila* homolog of *Nmf9*, *CG45058*, at conserved sites in three different exons: the first ANK repeat, the FN3 domain, and conserved domain 2 (Fig 8A). Apparent null and other predicted deleterious mutations induced at each site showed a consistent, severe adult locomotor phenotype and reduced viability. Newly eclosed homozygotes were predominantly stationary and most could be placed unanesthetized on a cardboard surface without the animal attempting to fly or

Fig 7. Genome edited alleles show importance of conserved domain 2 and confirm identity of *nmf9*. (A) CRISPR/Cas9 target sequence is underlined, PAM site boxed, and predicted cut site shown as an arrow. Sequences of four induced alleles selected for further analysis are shown; deleted bases are indicated by dashes and insertions/substitutions by red lower-case letters. The first glycine residue, targeted for replacement, was at position 592 relative to RefSeq protein NP_001074402. (B) Results of the landing test with induced mutations heterozygous either to a coisogenic nonmutant allele (+/n) or to the original *nmf9* Jackson allele (n^J/n^C) showed non-complementation (p = 5.3x10^{-6}, Fisher's exact test). Similarly, head nodding (p = 1.3x10^{-5}), tremor (3.6x10^{-4}), and hyperactivity (p = 0.0053) were more frequently observed in trans-heterozygotes for *nmf9* and CRISPR-induced alleles than in controls. (C) Time-to-event in the forced swim tests also showed non-complementation between edited and the original *nmf9* alleles (Asymptotic logrank test, p = 0.0011).

doi:10.1371/journal.pgen.1005344.g007

walk (Fig 8B, 8C and 8D). Most held their wings at unusual postures, either up or slightly down from the horizontal plane. When these animals did move, they typically fell over onto their backs and kicked their legs without apparent coordination. Mutant adults had very poor survival, even in uncrowded, horizontal vials (Fig 8E). In addition to three independent sets of alleles with consistent phenotype, deficiency mapping with two distinct alleles induced at conserved domain 2 confirmed that the severe locomotor defects map to the *CG45058* locus as a loss of function (S7 Fig). These data conflict with an interpretation that *wake* alleles of this locus, which do not have strong locomotor phenotypes, are null [19]. In contrast, these data are fully consistent with the *bnd* intragenic deletion allele as null [20].

Since *wake* alleles were previously reported to have reduced sleep and *nmf9* mice had abnormal circadian behaviors, we made related measurements for heterozygotes of our engineered null mutations. Consistent with previous reports, *wake* alleles showed decreased sleep in both sexes (Fig 9). While male null heterozygotes also showed reduced daily sleep, surprisingly, females showed slightly increased daily sleep. Null heterozygotes also had increased latencies to first sleep bout after lights on (daytime sleep latency), and heterozygote females had increased latency to first sleep bout after lights off (nighttime sleep latency) as well. Increased

Table 2. Allele-level data and p-values for non-complementation of *nmf9* by CRISPR-induced mutations.

Genotype	Landing	Nodding	Tremor	Hyper.	Swim ≥60s	Avg. # Tests Failed
G592A/+	0/5	0/5	1/5	2/5	0/5	0.66
G592A/*nmf9*	4/5	4/5	4/5	4/5	3/5	3.80
RLS->LI/+	1/7	1/7	1/7	1/7	2/7	0.86
RLS->LI/*nmf9*	1/1	0/2	0/2	0/2	1/2	1.00
ΔLSPG/+	0/3	0/3	0/3	1/3	0/3	0.33
ΔLSPG/*nmf9*	3/4	4/4	3/4	4/4	3/4	4.25
-13,+26/+	0/2	0/2	0/2	0/2	0/2	0.00
-13,+26/*nmf9*	5/5	5/5	5/5	4/5	5/5	4.60
SUM mut/+	1/17	1/17	2/17	4/17	2/17	0.83
SUM mut/*nmf9*	13/15	13/16	12/16	12/16	11/16	3.74
Fisher exact p-value	5.3×10^{-6}	1.3×10^{-5}	3.6×10^{-4}	0.0053	0.0013	

Frequency with which animals of the indicated genotype were called abnormal by a trained observer blind to genotype is indicated for each test and allele combination, as is the average number of the five tests for which a given genotype was considered abnormal. P-values are for the one-sided Fisher's Exact Test of the null hypothesis that mutant/*nmf9* heterozygotes are not more often affected than mutant/+ heterozygotes, based on sums across all five induced alleles.

doi:10.1371/journal.pgen.1005344.t002

latencies were consistent with, but less severe than, *wake* alleles (Fig 9). Reduced sleep and increased sleep latencies could not be accounted for by an increase in rate of movement since heterozygotes for engineered null alleles showed moderately reduced waking activity levels, as did P-element *wake* alleles (Fig 8F). Statistical summaries for these tests are given in S5 Table. Together, these results show a strong requirement for *CG45058* function for viability, activity level, and sleep-related measures; show sexual dimorphism for impact on daily sleep; and resolve a conflict between previous reports of *CG45058* null phenotypes.

Discussion

Functional annotation of a novel gene genealogy

Through a combination of genetic and molecular approaches, we showed that *Nmf9* encodes a novel yet highly conserved protein that is functionally important to a distinct set of neurological phenotypes in both mice and flies. Starting with the mutation, we identified *Nmf9* transcripts using positional cloning and experimental validation of predicted exons. RT-PCR and in situ hybridization experiments defined timing and location of gene expression. Sites of expression predicted, in addition to readily observed vestibular abnormalities, substantial phenotypes in fear learning and circadian behavior that were not obvious from the initial description or gross observation. Sequence information from >100 animal homologs identified distinct regions of strong evolutionary constraint, including ANK and FN3 motifs and three novel segments that lack motif annotation. Induced mutations at most conserved peptide sequence among the novel conserved domains produced strong phenotypes in both mice and flies, including a single glycine to alanine substitution in mice that failed to complement the original *nmf9* allele. Among animal and non-animal Holozoan homologues some strongly predicted functional motifs appeared to be modular across phyla, including two small GTPase-binding motifs, CRIB and RA. Among animals, the Nmf9 homology group appears to have been lost only in urochordates, based on four genomes available in that taxon. The N-terminal CRIB domain was found only in choanoflagellate and filasterean sequences and was either

A


```
                                          Drosophila CG6954

ANK       T  I  L  E  S  T  D  V  D  V  N  S intron--------------->    coding      locomotor    reduced
          ACTATTCTAGAGTCCACCGACGTAGATGTAAACAGGTGAGCTTAGATCCTTATAACTT    effect      phenotype    viability

16-1  +7  ACTATTCTAGAGTCCACCGtaaacagat--GTAGATGTAAACAGGTGAGCTTAGATCC    fs +1       nd           yes
36-1  +8  ACTATTCTAGAGTtcaccggtggactct-CGTAGATGTAAACAGGTGAGCTTAGATCC    fs +2       nd           yes
32-1  +12 ACTATTCTAGAGTCCACCGtaggagtctgtcACGTAGATGTAAACAGGTGAGCTTAGA     D>VGVCH     yes          yes
23-4  +15 ACTATTCTAGAGTCCACCGtagatgtagatgtagACGTAGATGTAAACAGGTGAGCTT     D>VDVDVD    no           yes
29-3  -3  ACTATTCTAGAGTCCACC---GTAGATGTAAACAGGTGAGCTTAGATCCTTATAACTT     ΔD          no           no
27-1  -4  ACTATTCTAGAGTCCAC----GTAGATGTAAACAGGTGAGCTTAGATCCTTATAACTT     fs +2       yes          yes
42-2  -13 ACTATTCTAGAG------------ATGTAAACAGGTGAGCTTAGATCCTTATAACTT      fs +2       yes          yes


FN3       K  G  W  G  T  Y  R  T  S  V  P  A  S  V  V  P intron--->
          AAGGGATGGGGCACCTATCGAACCTCTGTGCCGGCTAGTGTAGTGCCCTCAAGTGAGT

3-10  -4  AAGGGATGGGGCACCTATCGAACCT----GCCGGCTAGTGTAGTGCCCTCAAGTGAGT     fs +2       yes          yes


Domain 2  Q  L  H  Y  Q  Q  A  A  L  Q  R  L  E  P  G  L  Y  L  G  Y
          CAATTGCACTATCAACAGGCGGCGCTGCAGCGTTTGGAGCCCGGTCTTTACCTGGGCT

66-7  +2  CAATTGCACTATCAACAGGCGGCGCTGCAGCGTTTGGAGCCtaCGGTCTTTACCTGGG     fs +2       yes          yes
48-17 -49 --------------------------------------CCGGTCTTTACCTGGGCT       fs +2       yes          yes
```

B **C** **D**

66-7/TM6   66-7/66-7

F

E

Domain	Line	Indel	Genotype	Survivors	
				day 1	day 7
ANK	16-1	+7	n/+	212	212
			n/n	0	0
	27-1	-4	n/+	661	658
			n/n	17	0
	36-1	+8	n/+	263	261
			n/n	1	0
	42-2	-13	n/+	409	404
			n/n	14	1
	23-4	+15	n/+	126	126
			n/n	16	8
	29-3	-3	n/+	149	148
			n/n	51	51
	32-1	+12	n/+	155	154
			n/n	11	2
FN3	3-10	-4	n/+	485	479
			n/n	33	0
Domain 2	66-7	+2	n/+	302	298
			n/n	1	0
	48-17	-49	n/+	277	270
			n/n	1	0

Fig 8. Genome editing of *Drosophila CG45058* exons results in severe locomotor deficits and early death. (A) Target sites and recovered alleles are shown for separately targeted exons encoding the first ANK repeat, FN3 domain, and the GLYLGYLK peptide. (B) Posture of a heterozygous control fly. (C) Posture of a mutant fly homozygous for a frame-shifting mutation in conserved domain 2. Note wings held up at an angle with respect to the body. (D) The same mutant fly subsequently fell over and moved legs in an uncoordinated fashion. (E) Survival frequencies at eclosion and 7 days post-eclosion. (F) Total waking activity (beam breaks per minutes awake) for males (blue) and females (pink) of the indicated genotypes. Heterozygotes for framshifting mutations in *CG45058* ANK, FN3 or conserved domain 2 consistently showed reduced activity relative to +/+ control, similar in magnitude to *wake* (EY02219) homozygotes. *** $p < 2.2 \times 10^{-16}$.

doi:10.1371/journal.pgen.1005344.g008

gained in these lineages or lost in the lineage leading to animals. The RA domain appears to have been lost in a derived paralog during or soon after a gene duplication event at the basal vertebrate lineage and the ancestral paralog was lost in the basal mammalian lineage.
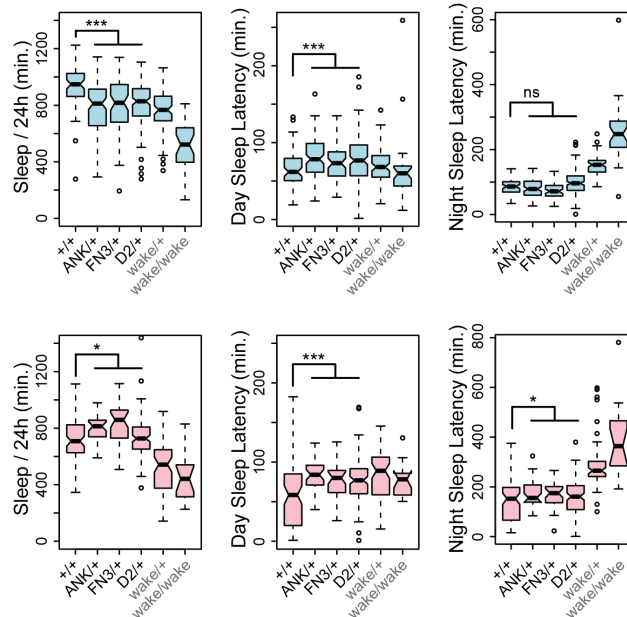
Fig 9. *CG45058* null heterozygotes have abnormal sleep patterns. Male heterozygotes (blue graphs) recorded less total sleep per 24 hr. period and slightly longer latency to first sleep bout after lights on (daytime sleep latency) than +/+ controls, but no significant difference in latency after lights off (nighttime sleep latency). By contrast female heterozygotes (pink graphs) recorded slightly more overall sleep than +/+ controls and slightly longer daytime and nighttime sleep latencies. Heterozygous *wake* P-element alleles (EY02219, KG02188, KG08407 and MI02905 insertion lines) and a homozygous *wake* allele (EY02219) were recorded as a comparison. * p<0.025, *** p<5x10^{-4}, ns p = 0.78, Wilcoxon rank sum test.

doi:10.1371/journal.pgen.1005344.g009

Expression and function of *Nmf9* in mammalian brain

Nmf9 showed a complex pattern of expression in the vestibular system, olfactory system, and regions of the brain implicated in satiety and metabolism, innate anxiety, fear learning, and circadian rhythm. We tested the functional integrity of each of these systems with a battery of standardized behavioral assays. Deficits in vestibular, circadian and fear learning measures were clear despite significant sex differences. Abnormalities in measures of anxiety and appetitive behavior were nominally significant, but confounded by hyperactivity in the mutant mice– particularly after sustained handling, which may suggest either an anxiety-related aspect or perhaps feedback from disturbed sensory input from the vestibular system. Although mutants and non-mutant littermates were easily distinguished by overall behavior, mutant phenotypes for any single formalized test were not fully penetrant and tended toward higher variance than control littermate values, which suggests that *Nmf9* may be important to the robustness of these pathways, but not essential to their basic function. On this interpretation, loss of *Nmf9* activity in AMY, LS, and the olfactory system may not be disruptive enough to these circuits for behavioral phenotypes to be detected on a standard inbred background, but might have consequences that would be subject to selection under the more competitive fitness constraints

of wild populations. Although we observed striking *Nmf9* expression in cortical ventricular zones, the adult cortex appeared normal on gross inspection, with no obvious loss of cells in any of the *Nmf9*-enriched sites.

Females were more severely affected than males by the *nmf9* mutation in several tests (Figs 1 and 4). While both sexes express *Nmf9* RNA, we have not explicitly tested for quantitative differences in specific pattern elements. Sex-dependence is known for several mouse behaviors both at baseline and in response to perturbations [29]. Interestingly, mutations of *Ntrk2* (TrkB) also induce vestibular defects that are substantially more pronounced in female mice [30], although any shared mechanism remains to be explored.

Functions of *CG45058* in *Drosophila*

Our analysis of *CG45058* in *Drosophila* resolves a conflict in the current literature and extends both sets of previous findings. While our work was nearing completion, two other groups reported putative null mutations in *CG45058*, but with very different primary outcomes. In a P-element screen for sleep phenotypes, Mark Wu and colleagues reported viable P-element and imprecise excision alleles, which they termed *wide-awake* (*wake*) based on increased latency to sleep and decreased total sleep by mutant animals [19]. In contrast, through an RNAi screen for regulators of asymmetric cell division, Juergen Knoblich and colleagues reported pupal-lethal gene deletion alleles of *CG45058*, which they termed *banderuola* (*bnd*) based on the cytological appearance of dividing sensory organ precursor cells [20]. A major difference between these studies is the nature of the alleles examined. *CG45058* has several annotated transcripts, which primarily differ in transcriptional start site and utilization of 5' exons that are not well-conserved outside of Diptera. The *wake* alleles occur in the variant 5' region of the gene and appear likely to affect only those transcripts, but not several others that include all of the conserved protein coding sequences. Alternatively, the *bnd* gene deletion allele might have removed regulatory sequences or barriers that influence expression of neighboring genes resulting in a synthetic phenotype. Our results resolve this conflict by creating discrete loss-of-function alleles induced at three distinct, functionally important sites in the predicted protein, encoded by separate exons that are each included in all known *CG45058* transcripts. These alleles strikingly reduced adult viability and locomotor function of surviving adults, demonstrating that this is the null phenotype, in support of Mauri et al. Heterozygotes for these null alleles had reduced waking activity, increased latencies to first sleep bout, and effects on total sleep, confirming and extending the behavioral results of Liu et al. for *wake* alleles. Our data show sexually dimorphic effects on total sleep in null heterozygotes, in contrast to decreased sleep found in both sexes of *wake* mutants. This allelic difference suggests either distinct cellular functions of *CG45058* isoforms or, perhaps more likely, isoform-specific expression in cell types that impinge on sleep regulation.

Conservation, divergence, and pleiotropy

While the *Nmf9* orthology group shows extraordinary conservation–including highly constrained sequence domains that had no prior functional or motif annotation–the reorganization of putative GTPase-binding domains across major phylogenetic boundaries, loss of the homology group in urochordates, and duplication and rescission events in the vertebrate lineage all suggest a degree of adaptive plasticity that may be reflected in both shared and lineage-specific requirements. Indeed, mutations in the *Nmf9* homologues of both mouse and fly resulted in phenotypes related to daily cycles of activity, as well as locomotor abnormalities, but with some clear differences. While mouse *nmf9* null alleles showed hyperactivity and tremor that increased over the course of six months, fly mutations at distinct conserved

domains (ANK, FN3, and conserved domain 2) produced severe locomotor retardation and early death, within one week of eclosion. These studies lay a foundation for understanding both common functions of *Nmf9* homologs and lineage-restricted activities that might relate to derived loss of the RA domain or other lineage-specific features. How differences in domain architecture and expression patterns reshape the functional networks in which Nmf9/CG45058 proteins act will surely be of interest in the evolutionary development of the nervous system.

Materials and Methods

Nmf9 genetic mapping

Coisogenic C57BL/6J–*nmf9* mice were obtained from the Neuroscience Mutagenesis Facility (NMF) and AKR/J, BALB/cJ, C3H/HeJ and DBA/2J mapping partners from production colonies at the Jackson Laboratory. Conventional exclusion mapping was performed as described [11, 31]. PCR primers for new genetic markers are given in S1 Table. New markers were also typed on a smaller C57BL/6J–*nur12* x BALB/cJ cross obtained from NMF. C57BL/6J–*nmf9* mutant line was maintained on C57BL/6J and genotyped by sequencing.

Gene expression detection and measurements

Northern blots were performed by standard methods [32], essentially as described [33]. RNA from whole brains was extracted with Trizol reagent (Life Technologies). Poly(A)$^+$ RNA was isolated on Oligo(dT) cellulose type 7 (Amersham Biosciences). Concentrations and integrity were verified by spectrophotometry and gel electrophoresis. 8.5 μg poly(A)$^+$ RNA per sample was electrophoresed through a denaturing formaldehyde/agarose gel and transferred to Hybond-N+ membrane (Thermo Fisher). Size standard was removed prior to transfer and imaged by ethidium bromide fluorescence. Probes were synthesized from two cloned fragments of *Nmf9* by random priming in the presence of ^{32}P-dCTP and hybridized overnight. Blots were exposed to a phosphor screen for 5 days before quantitative imaging with a Storm 860 instrument (Molecular Dynamics). Quantitative RT-PCR was performed using SybrGreen fluorescence quantification on a BioRad CFX96 instrument. Expression was relative to *Gapdh* and *RP49* in *Drosophila*, and to *GAPDH* and *SDHA* in zebrafish.

In situ hybridization

Mouse tissues were processed by a standard method [34] as previously modified [8]. Briefly, embryos were fixed in formalin, adult mouse brain in 4% PFA. Samples were cryoprotected in 30% sucrose and sectioned at 20 μm. Slides were treated 15' with boiling sodium citrate, followed by acetic anhydride in triethanolamine prior to hybridization with dioxigenin-labeled RNA probes at 65°C overnight. After hybridization and washing, sections were blocked with 5% normal donkey serum (Jackson ImmunoResearch) in PBSTX (0.2% Triton-X100 in PBS) for 1 hour and incubated with anti-digoxigenin-AP Fab fragments (Roche) at 1:2000 overnight at 4°C. Whole mount *in situ* in zebrafish was performed as described [35]. Samples were fixed in 4% PFA. Proteinase K was used for antigen retrieval at 10 μg/mL for 1 hour at 37°C. Hybridization was at 65°C for 48 hours. Samples were blocked with 2% normal donkey serum, 2mg/mL BSA (NEB) in PBTx and incubated with 1:5000 anti-dig-AP Fab fragments overnight at 4°C. Hybridization to whole mount fly embryos was performed as described [36]. Briefly, embryos were fixed in methanol and treated with Proteinase K at 37C for 7 minutes. Embryos were hybridized at 55°C overnight. Samples were blocked with 1:10 Western Blocking Reagent (Roche) in PBTwx (0.1% Tween 20 and 0.1% Triton-X in PBS) and incubated with 1:1000 anti-dig-AP Fab fragments for 2 hours at 4°C. All probes were prepared by in vitro transcription

from linearized plasmid templates and diluted in hybridization buffer prior to use. Sea anemone in situ hybridization was performed as described [37]. Probe templates for all species were generated by RT-PCR and cloned into appropriate vectors for in vitro transcription.

Mouse behavioral assays

All behavioral tests were performed on mice between 2 to 6 months of age, with wild-type littermates as control to the mutant animals, by investigators blinded to the animals' genotypes.

Vestibular function. For the forced swim test, animals were individually placed in a 4L beaker filled with 3L 25°C water and the swim time measured as the length of time an animal was able to keep its snout above water [38]. For the landing test, animals were individually lifted by their tails. Normal animals extended their head and forelimbs toward the ground, exhibiting landing behavior, whereas mutant animals tend to curl their trunk sideways or ventrally and occasionally trying to rotate [39]. The presence of the curling / rotation behavior was scored. For nodding, tremor, circling, and hyperactivity, each animal was placed in a fresh cage and observed for one minute to score each phenotype.

Auditory function. The startle response was performed in startle chambers (SR-Lab, San Diego Instruments) in the La Jolla Neuroscience Blueprint Behavioral Core Facility at The Scripps Research Institute (TSRI Behavioral Core). The chambers consisted of nonrestrictive Plexiglas cylinders 5 cm in diameter resting on a Plexiglas platform in a ventilated chamber. High-frequency speakers mounted 33 cm above the cylinders produced all acoustic stimuli, which were controlled by SR-LAB software. Piezoelectric accelerometers mounted under the cylinders transduce movements of the animal, which were digitized and stored by an interface and computer assembly. Startle pulses started at 70 dB, 40ms in duration, and increased to 120 dB in 5 dB increments and then ramped back down, interspersed by no stimulus trials. The background noise level was at 70 dB.

Fear learning. Conditioning was performed by the TSRI Behavioral Core and took place in Freeze Monitor chambers (Med Associates) stationed in sound-proof boxes. The Plexiglas conditioning chambers (26 x 26 x 17 cm) had speakers and lights mounted on two opposite walls and were installed with an electric grid floor. On day 1, mice were habituated in the conditioning chamber for five minutes. On day 2, mice were exposed to the context and conditioned stimulus (30 seconds, 3000 Hz, 80 dB sound) in association with foot shock (0.70 mA, 2 second, scrambled current). Each mouse received 2 shock exposures, each in the last 2 seconds of a 30 second tone exposure during a 5.5 minute session. On day 3, contextual conditioning in animals was measured in a 5 minute test in the training chamber. On day 4, cued conditioning was measured in a novel context, where the chamber previously used was disguised (new opaque plastic rounded compartment replacing old clear square compartment, new opaque plastic floor replacing metal grid, novel odor with a drop of orange extract under floor). The animals were allowed to habituate for 3 minutes before exposure to the conditioned stimulus (tone) for 3 minutes. Freezing behavior was measured in all sessions by a validated computer-controlled recording of photocell beam interruptions [40].

Circadian function. Mutant and wildtype littermates of either sex, between three-five months of age, were individually housed with food and water *ad libitum*. Activity records were plotted as actograms and analyzed using ClockLab (Actimetrics, Evanston, IL). Unless otherwise noted, all calculations were performed across an interval of approximately 14 days. The period estimates were analyzed with the activity onset predicted by ClockLab, after manually checking and adjusting when necessary. Onset tau and onset error of the period of daily activity in DD were calculated using the adjusted activity onset data. Power and tau in LD 12:12 and DD were additionally determined with Chi-square periodogram analysis.

Innate anxiety. The elevated plus-maze and open field tests were performed at the TSRI Behavioral Core. The plus maze apparatus had four arms (5 x 30 cm) at right angles and was elevated 30 cm from the floor. The two closed arms had 16 cm high walls. The two open arms had a 0.5 cm lip and no walls. Animals were placed in the center of the maze and allowed free access to all arms for 5 minutes. Behaviors were recorded with mounted cameras. The open field apparatus was a 50 x 50 cm plexiglass square surrounded by 22 cm high walls. The field was divided into 16 squares (12 outer and 4 inner) of equal areas. Animals were placed in the center of the field and recorded for 10 to 60 minute with mounted cameras. The marble burying test was performed by placing individual animals in a standard mouse cage containing 5 cm deep bedding with 20 small marbles arranged in an evenly spaced grid of 4 rows of 5 marbles on top of the bedding material. After 30 min mice were removed and the number of marbles that were at least 2/3 covered by bedding was counted.

Olfaction. The buried food finding test was performed on individually housed mice. Beginning 3 days before testing, animals received one piece of Froot Loops cereal (Kellog's, the same color/flavor was used for all animals) per day with *ad libitum* access to water and normal chow. Fourteen hours (overnight) before testing, animals were deprived of food while water remained ad libitum. Each animal was tested in a clean cage containing 2 cm of bedding, in the corner opposite to the corner where a piece of cereal was hidden. The time taken to locate and hold the cereal in both of its paws was recorded with a stopwatch. The odor habituation / dishabituation test was performed by placing individual animals in a clean cage and allowing 30 minutes for habituation. Non-social odors were presented before social odors. For non-social odors, a cotton-tipped applicator dipped in either de-ionized water or 1:100 dilution of almond or vanilla extract (McCormick &Co) was presented. For social odors, each applicator was used to swab the inside of a cage that had housed a single non-littermate animal for three days and then presented to the animal being tested. The length of time each animal spent investigating the tip applicator was recorded, with two minutes given per applicator trial. The same odor was presented in three consecutive trials, with the same order of odor presentation for all animals.

Satiety and metabolism. Animals were weighed at 21 days, 2 months, and 6 months of age. For food consumption test, a known mass of food was provided to the animals *ad libitum* at the start of experiment. The mass of the food was weighed each subsequent evening at the same time and monitored for spillage (Huszar et al. 1997, Crawley 1999).

Conservation analysis

Protein sequences homologous to mouse Nmf9 were retrieved through NCBI tblastn, UCSC Genome Browser, Ensembl, and JGI web sites and made use of GNOMON and ENSEMBL transcript predictions. Identified fragments were used to search the full length sequences in surrounding genomic sequences where available (see S4 Table). For homologs inferred purely through genomic sequences, a combination of BLAT [41], BLAST [42] and GENSCAN [43] programs were used to predict open reading frames. All database searches were performed before February 2013. For the full-length protein analysis, the best-annotated sequence of each evolutionary branch was used. For the high resolution analysis sequences with more than 50% gaps in that domain were excluded. Predicted cDNA sequences were translated using ExPASy Translate [44] and aligned using MUSCLE [45, 46] and the result of the amino acid alignment used to manually correct the nucleic acid alignment. Conservation rates were calculated with Datamonkey [47] using codon data type and universal genetic code with neighbor joining tree. Motif search was performed with Motif Scan [48], Scansite 2.0 Scan Motif [49], Scan Prosite (ExPASy), InterProScan 4 [50], Pfam [51], and SMART 6 [52]. Protein modification scan was

performed using The Sulfinator (ExPASy) for sulfation and YinOYang [53] for glycosylation and phosphorylation.

Genome edited animals

Genome-edited mice were generated essentially as described [17, 28]. Briefly, in-vitro synthesized Cas9 mRNA, sgRNA, and ssDNA homology-directed repair oligos were co-injected as a cocktail into C57BL/6 one-cell embryos at the Moores UCSD Cancer Center Transgenic Mouse Shared Resource. Oligonucleotide sequences are listed in S6 Table. sgRNA templates and ssDNA repair oligonucleotides were synthesized as Ultramers by IDT. All procedures were approved by the UCSD IACUC. RNA and DNA reagents for fly injection were prepared as above. Conserved Domain 2 mutants were generated by co-injection of in-vitro synthesized Cas9, sgRNA, and ssDNA repair oligo into w^{1118} embryos. ANK and FN3 mutants were generated by co-injection of sgRNA and repair oligonucleotides into Cas9-expressing embryos (PBac<y[+mDint2] = vas-Cas9>VK00027, Bloomington Stock Number 51324) as described by [54]. All fly embryo injections were performed by Rainbow Transgenic Flies, Inc. (Camarillo, CA). Oligonucleotide sequences are listed in S6 Table.

Drosophila stocks

Mutations were verified by Sanger sequencing of PCR products encompassing targeted genomic loci. Transposon mutant and deficiency lines for CG45058 were obtained from the Bloomington Stock Center according to the following stock numbers: $wake^{EY02219}$ [15858], $wake^{KG02188}$ [14082], $wake^{KG08407}$ [15129], $wake^{MI02905}$ [37162], Df(3R)Exel6273 [7740], Df(3R)ED6085 [150049], Df(3R)ED6091 [9092], Df(3R)BSC527 [25055], Df(3R)Exel6192 [7671], Df(3R)BSC618 [25693], and Df(3R)ED6090 [150614].

Drosophila behavior assays

Flies were raised at room temperature (~21°C) on standard cornmeal/molasses. Within 1–2 days of eclosion, flies were assayed for wing orientation and the ability to maintain a standing position. For sleep assays, 1–5 day old flies were loaded into 65x5 mm glass tubes containing 5% sucrose and 2% agarose and entrained to a 12hr:12hr light:dark (LD) cycle for 2 days before recording sleep/wake patterns using the *Drosophila* Activity Monitoring System (Trikinetics, Waltham, MA). Sleep was defined as 5 minutes of inactivity and, along with waking activity, was measured at 25°C using custom software as previously described [55].

Ethics statement

Mice were euthanized by CO2 inhalation or by perfusion or organ removal under deep anesthesia with tribromoethanol (avertin). Fish embryos were euthanized by fixation or snap-frozen in E3 media. All vertebrate animal procedures were approved by the University of California San Diego Institutional Animal Care and Use Committee (IACUC). The University of California San Diego is AAALAC accredited, AAALAC institutional number 000503.

Supporting Information

S1 Table. Genetic markers used for genetic fine mapping. Dinucleotide repeat sequences were identified from an early draft public mouse reference sequence (MGSCv2) and tested for length polymorphism in mapping strains.
(XLSX)

S2 Table. Statistical analyses of mouse behavioral tests.
(XLSX)

S3 Table. Circadian phenotype of *Nmf9*. Animals were housed in 12:12 LD cycles for a minimum of 7 days prior to transition to DD. Tau, power, and onset were calculated as average over ~14 days. N = 14 +/+ females, 12 mutant females, 9 +/+ males, 10 mutant males. Data was analyzed using two-way ANOVA.
(XLSX)

S4 Table. Species used in evolution conservation analysis. Homologs of *Nmf9* were identified in various databases and hand curated. Due to quality of genomes available, sequences that contain more than 50% gaps were removed and not included in this list. Due to poor assembly, all homolog sequences, with the exception of *H. sapiens*, was manually assembled utilizing the BLAST, BLAST, TBLASTN tools available in many of the databases, the ID of the seed sequence was listed when available. Species highlighted in yellow indicate the sequences used on whole-protein conservation analysis. 52 species containing ancestral versions of the homolog and 61 species containing derived version of the homolog were used for the region specific analysis of conservation.
(XLSX)

S5 Table. Statistical analyses of *Drosophila* behavioral tests.
(XLSX)

S6 Table. Oligonucleotides used in genome editing of flies and mice.
(XLSX)

S1 Fig. Mutant animals have normal inner ear histology and hearing. (A) Ampulla, saccule, Scarpa's ganglion, utricle, and semi-circular canals were grossly normal in *nmf9* animals. (B) Acoustic brainstem responses were within normal range (waveform threshold ≤25 dB) for 5 of 6 mutant mice between 2–4 months age. (C) Acoustic startle responses were also normal in a separate cohort animals between 2–6 months age. Two-factor MANOVA p = 0.30 for genotype, p = 0.28 for sex. N = 12 +/+ females, 12 +/+ males, 13 mutant females, 7 mutant males. Error bars, s.e.m.
(TIF)

S2 Fig. *Nmf9* mRNA includes both *Ankfn1* and *4932411E22Rik*. (A) UCSC Genome Browser shows minus-strand transcripts for the 3' exons of *Ankfn1* and 5' exons of *E22* compared to BLAT alignment of empirically determined *Nmf9* mRNA sequence. Positions of primers used for RT-PCR are indicated. (B) Relative quantities for real-time fluorescence measurements (SYBR) of RT-PCR products from the indicated assays from either brain cDNA or cloned plasmid. (C) Publicly deposited cDNA sequences for insect (*D. melanogaster*, A. mellifera), echinoderm (*S. purpuratus*), nematode (*C. elegans*) and cnidarian (*N. vectensis*) also support a single transcript across segments homologous to *Ankfn1* and *E22*. (D) In situ hybridization patterns were indistinguishable for *Ankfn1* and *E22* probes in ventromedial hypothalamus (VMH), amygdala (AMY), lateral septum (LS), piriform cortex (PC), suprachiasmatic nuclei (SCN) and spinal cord (SpC). This mirrors data for both annotations in the Allen Brain Atlas ([Fig 3B](#)).
(TIF)

S3 Fig. *Nmf9* expression in developing brain and sensory structures. *Nmf9* RNA was detected by *in situ* hybridization as early as E14.5 in ventricular zones, inner ear, and nasal epithelium.
(TIF)

S4 Fig. Modular organization of *Nmf9* homologs. Known motifs are indicated in green. Highly conserved domains novel to the *Nmf9* homology group are shown in blue. An amino-terminal CRIB domain was found only in the two choanoflagellae species, while inclusion or loss of a ras-association domain (RA) appears not to be monophyletic. Vertebrate species basal to mammals had paralogous copies, one with and one without the RA domain in each species. All mammals examined had only one copy, which never had an identifiable RA domain. Conservation of the first ANK repeat, and consensus match to the motif definition were poor or absent in some invertebrate and choanozoa species.
(TIF)

S5 Fig. Core *Nmf9* domains have lineage-dependent patterns of conservation. High-resolution sliding window analysis of domains, with inclusion of species that lacked full-length sequence. Separate analyses were performed for ancestral homologs and vertebrate-derived homologs. Y-axis shows dN/dS values calculated using HyPhy [56]. The 10 residues with smallest dN in each analysis are plotted in orange. Light blue points indicate gaps in sequence. Amino acid residues are given for regions of highest conservation. (A) Annotated ankyrin and fibronection type III motifs are plotted. For the ankyrin repeats, N = 53 species for derived homologs, 47 for ancestral homologs. For fibronectin type III repeats, N = 50 derived, 50 ancestral homologs. (B) The three highly conserved domains that do not align with known motif definitions are plotted. For domain 1, N = 60 species for derived homologs, 52 for ancestral. For domain 2, N = 54 derived, 49 ancestral. For domain 3, 49 derived, 48 ancestral.
(TIF)

S6 Fig. Expression of *Nmf9* homologs suggests both conserved and adapted functions. *Nmf9* homologues were expressed during development in (A) fruit fly and (B) zebrafish. Samples included 3 biological replicates for fruit fly, but only technical replicates of single samples for each zebrafish stage and paralogous gene. (C) *In situ* hybridization to *Drosophila* embryos showed broad *CG45058* expression, including CNS, agreeing with stage-specific expression in the BDGP in situ database (http://insitu.fruitfly.org/). (D) Expression in sea anemone *N. vectensis* planulae (left) includes endoderm (arrow) and some apical tuft cells. In polyps, expression appears uniform throughout the endoderm (arrow) and mesenteries. (E) In zebrafish embryos, ancestral paralog retained a broad but weak expression pattern that includes expression in the developing inner ear (box and inset). The derived paralog showed a cleaner, more restricted pattern.
(TIF)

S7 Fig. *CG45058* allele locations and deficiency mapping for severe phenotypes. (A) Schematic of the *CG45058* locus shows exon position of encoded protein domains (blue) across transcripts (navy) from the UCSC browser. Chromatin accesibility to DNase I at stage 5 (green), 9 (orange), 10 (red), 11(blue) and 14 (purple) embryos suggests dynamic regulation of alternative transcripts. (B) UCSC transcripts (navy) in the interval around *CG45058* (red) on *Drosophila* chromosome 3R are shown relative to deficiencies that complement (green bars) or fail to complement (red bars) frameshift mutation at conserved domain 2 (GLYLGYLK). The inferred interval (yellow box) defined by the shared breakpoint of Df(3R)Exel6273 and Df(3R)Exel6192 and the right breakpoint of Df(3R)ED6090 contains parts of 6 transcription units.
(TIF)

S1 Video. Comparison of homozygous *nmf9* female to littermate control on Coplin jars at 10 months of age shows tremor, tilted head posture and head nodding in the mutant. Behavior in the home cage shows circling by a strongly affected mutant female compared to

non-circling littermate control.
(MOV)

S2 Video. Trans-heterzyogotes for the G-to-A edited and *nmf9* chemically induced alleles show agitation, tremor and mild vestibular defects compared to same-sex littermates that are heterozygous for either mutant allele over wild type. Birth date and identifier is indicated for each animal.
(MP4)

Acknowledgments

Author Contributions

References

1. Beutler B, Du X, Xia Y. Precis on forward genetics in mice. Nat Immunol. 2007; 8(7):659–64. Epub 2007/06/21. PMID: 17579639

2. Justice MJ, Noveroske JK, Weber JS, Zheng B, Bradley A. Mouse ENU mutagenesis. Hum Mol Genet. 1999; 8(10):1955–63. Epub 1999/09/02. PMID: 10469849

3. Clark AT, Goldowitz D, Takahashi JS, Vitaterna MH, Siepka SM, Peters LL, et al. Implementing large-scale ENU mutagenesis screens in North America. Genetica. 2004; 122(1):51–64. Epub 2004/12/29. PMID: 15619961

4. Concepcion D, Seburn K, Wen G, Frankel WN, Hamilton BA. Mutation rate and predicted phenotypic target sizes in ENU-treated mice. Genetics. 2004; 168(2):953–9. PMID: 15514066

5. Kawai J, Shinagawa A, Shibata K, Yoshino M, Itoh M, Ishii Y, et al. Functional annotation of a full-length mouse cDNA collection. Nature. 2001; 409(6821):685–90. Epub 2001/02/24. PMID: 11217851

6. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature. 2002; 420(6915):563–73. PMID: 12466851

7. Alcaraz WA, Chen E, Valdes P, Kim E, Lo YH, Vo J, et al. Modifier genes and non-genetic factors reshape anatomical deficits in Zfp423-deficient mice. Hum Mol Genet. 2011; 20(19):3822–30. Epub 2011/07/07. doi: 10.1093/hmg/ddr300 PMID: 21729880

8. Alcaraz WA, Gold DA, Raponi E, Gent PM, Concepcion D, Hamilton BA. Zfp423 controls proliferation and differentiation of neural precursors in cerebellar vermis formation. Proc Natl Acad Sci U S A. 2006; 103(51):19424–9. PMID: 17151198

9. Gold DA, Baek SH, Schork NJ, Rose DW, Larsen DD, Sachs BD, et al. RORa coordinates reciprocal signaling in cerebellar development through sonic hedgehog and calcium-dependent pathways. Neuron. 2003; 40(6):1119–31. PMID: 14687547

10. Hamilton BA, Frankel WN, Kerrebrock AW, Hawkins TL, FitzHugh W, Kusumi K, et al. Disruption of the nuclear hormone receptor RORa in *staggerer* mice. Nature. 1996; 379:736–9. PMID: 8602221

11. Hamilton BA, Smith DJ, Mueller KL, K A.W., B R.T., van Berkel V, et al. The *vibrator* mutation causes neurodegeneration via reduced expression of PITPa: positional complementation cloning and extragenic suppression. Neuron. 1997; 18:711–22. PMID: 9182797

12. Bassett AR, Tibbit C, Ponting CP, Liu JL. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. Cell Rep. 2013; 4(1):220–8. Epub 2013/07/06. doi: 10.1016/j.celrep.2013.06.020 PMID: 23827738

13. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013; 339(6121):819–23. Epub 2013/01/05. doi: 10.1126/science.1231143 PMID: 23287718

14. Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, et al. Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. Genetics. 2013; 194(4):1029–35. Epub 2013/05/28. doi: 10.1534/genetics.113.152710 PMID: 23709638

15. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. Elife. 2013; 2:e00471. Epub 2013/02/07. doi: 10.7554/eLife.00471 PMID: 23386978

16. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013; 339(6121):823–6. Epub 2013/01/05. doi: 10.1126/science.1232033 PMID: 23287722

17. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 2013; 153 (4):910–8. Epub 2013/05/07. doi: 10.1016/j.cell.2013.04.025 PMID: 23643243

18. Yu Z, Ren M, Wang Z, Zhang B, Rong YS, Jiao R, et al. Highly efficient genome modifications mediated by CRISPR/Cas9 in *Drosophila*. Genetics. 2013; 195(1):289–91. Epub 2013/07/09. doi: 10.1534/genetics.113.153825 PMID: 23833182

19. Liu S, Lamaze A, Liu Q, Tabuchi M, Yang Y, Fowler M, et al. WIDE AWAKE mediates the circadian timing of sleep onset. Neuron. 2014; 82(1):151–66. Epub 2014/03/19. doi: 10.1016/j.neuron.2014.01.040 PMID: 24631345

20. Mauri F, Reichardt I, Mummery-Widmer JL, Yamazaki M, Knoblich JA. The conserved discs-large binding partner Banderuola regulates asymmetric cell division in *Drosophila*. Curr Biol. 2014; 24(16):1811–25. Epub 2014/08/05. doi: 10.1016/j.cub.2014.06.059 PMID: 25088559

21. Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, et al. A comprehensive genetic map of the mouse genome. Nature. 1996; 380(6570):149–52. PMID: 8600386

22. Keays DA, Clark TG, Flint J. Estimating the number of coding mutations in genotypic- and phenotypic-driven N-ethyl-N-nitrosourea (ENU) screens. Mamm Genome. 2006; 17(3):230–8. Epub 2006/03/07. PMID: 16518690

23. Bermingham NA, Hassan BA, Price SD, Vollrath MA, Ben-Arie N, Eatock RA, et al. Math1: an essential gene for the generation of inner ear hair cells. Science. 1999; 284(5421):1837–41. Epub 1999/06/12. PMID: 10364557

24. Kelley MW. Hair cell development: commitment through differentiation. Brain Res. 2006; 1091(1):172–85. Epub 2006/04/22. PMID: 16626654

25. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature. 2007; 445(7124):168–76. Epub 2006/12/08. PMID: 17151600

26. Pond SL, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics. 2005; 21(10):2531–3. Epub 2005/02/17. PMID: 15713735

27. Marlow HQ, Srivastava M, Matus DQ, Rokhsar D, Martindale MQ. Anatomy and development of the nervous system of *Nematostella vectensis*, an anthozoan cnidarian. Dev Neurobiol. 2009; 69(4):235–54. Epub 2009/01/27. doi: 10.1002/dneu.20698 PMID: 19170043

28. Concepcion D, Ross KD, Hutt KR, Yeo GW, Hamilton BA. Nxf1 natural variant E610G is a semi-dominant suppressor of IAP-induced RNA processing defects. PLoS Genet. 2015; 11(4):e1005123. Epub 2015/04/04. doi: 10.1371/journal.pgen.1005123 PMID: 25835743

29. Crawley JN. What's wrong with my mouse?: behavioral phenotyping of transgenic and knockout mice. 2nd ed. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2007.

30. Lucas EK, Jegarl A, Clem RL. Mice lacking TrkB in parvalbumin-positive cells exhibit sexually dimorphic behavioral phenotypes. Behav Brain Res. 2014; 274:219–25. Epub 2014/08/17. doi: 10.1016/j.bbr.2014.08.011 PMID: 25127683.

31. Floyd JA, Gold DA, Concepcion D, Poon TH, Wang X, Keithley E, et al. A natural allele of Nxf1 suppresses retrovirus insertional mutations. Nat Genet. 2003; 35(3):221–8. PMID: 14517553

32. Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.

33. Concepcion D, Flores-Garcia L, Hamilton BA. Multipotent genetic suppression of retrotransposon-induced mutations by Nxf1 through fine-tuning of alternative splicing. PLoS Genet. 2009; 5(5): e1000484. Epub 2009/05/14. doi: 10.1371/journal.pgen.1000484 PMID: 19436707

34. Wilkinson DG, Nieto MA. Detection of messenger RNA by in situ hybridization to tissue sections and whole mounts. Methods Enzymol. 1993; 225:361–73. Epub 1993/01/01. PMID: 8231863

35. Clements WK, Kimelman D. LZIC regulates neuronal survival during zebrafish development. Dev Biol. 2005; 283(2):322–34. Epub 2005/06/04. PMID: 15932753

36. Bergson C, McGinnis W. An autoregulatory enhancer element of the *Drosophila* homeotic gene Deformed. EMBO J. 1990; 9(13):4287–97. Epub 1990/12/01. PMID: 1979945

37. Wolenski FS, Layden MJ, Martindale MQ, Gilmore TD, Finnerty JR. Characterizing the spatiotemporal expression of RNAs and proteins in the starlet sea anemone, *Nematostella vectensis*. Nat Protoc. 2013; 8(5):900–15. Epub 2013/04/13. doi: 10.1038/nprot.2013.014 PMID: 23579779

38. Paffenholz R, Bergstrom RA, Pasutto F, Wabnitz P, Munroe RJ, Jagla W, et al. Vestibular defects in head-tilt mice result from mutations in Nox3, encoding an NADPH oxidase. Genes Dev. 2004; 18 (5):486–91. Epub 2004/03/12. PMID: 15014044

39. Lessenich A, Lindemann S, Richter A, Hedrich HJ, Wedekind D, Kaiser A, et al. A novel black-hooded mutant rat (ci3) with spontaneous circling behavior but normal auditory and vestibular functions. Neuroscience. 2001; 107(4):615–28. Epub 2001/11/27. PMID: 11720785

40. Contarino A, Baca L, Kennelly A, Gold LH. Automated assessment of conditioning parameters for context and cued fear in mice. Learn Mem. 2002; 9(2):89–96. Epub 2002/05/07. PMID: 11992019

41. Kent WJ. BLAT—the BLAST-like alignment tool. Genome Res. 2002; 12(4):656–64. Epub 2002/04/05. Article published online before March 2002. PMID: 11932250

42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215(3):403–10. Epub 1990/10/05. PMID: 2231712

43. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Mol Biol. 1997; 268(1):78–94. Epub 1997/04/25. PMID: 9149143

44. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, et al. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 2012; 40(Web Server issue):W597–603. Epub 2012/06/05. doi: 10.1093/nar/gks400 PMID: 22661580

45. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004; 5:113. Epub 2004/08/21. PMID: 15318951

46. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32(5):1792–7. Epub 2004/03/23. PMID: 15034147

47. Pond SL, Scheffler K, Gravenor MB, Poon AF, Frost SD. Evolutionary fingerprinting of genes. Mol Biol Evol. 2010; 27(3):520–36. Epub 2009/10/30. doi: 10.1093/molbev/msp260 PMID: 19864470

48. Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, Hau J, et al. MyHits: improvements to an interactive resource for analyzing protein sequences. Nucleic Acids Res. 2007; 35(Web Server issue): W433–7. Epub 2007/06/05. PMID: 17545200

49. Obenauer JC, Cantley LC, Yaffe MB. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res. 2003; 31(13):3635–41. Epub 2003/06/26. PMID: 12824383

50. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. Nucleic Acids Res. 2005; 33(Web Server issue):W116–20. Epub 2005/06/28. PMID: 15980438

51. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, et al. The Pfam protein families database. Nucleic Acids Res. 2010; 38(Database issue):D211–22. Epub 2009/11/19. doi: 10.1093/nar/gkp985 PMID: 19920124

52. Letunic I, Doerks T, Bork P. SMART 6: recent updates and new developments. Nucleic Acids Res. 2009; 37(Database issue):D229–32. Epub 2008/11/04. doi: 10.1093/nar/gkn808 PMID: 18978020

53. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput. 2002:310–22. Epub 2002/04/04.

54. Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, et al. Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in *Drosophila*. Genetics. 2014. Epub 2014/01/31.

55. Wu M, Robinson JE, Joiner WJ. SLEEPLESS is a bifunctional regulator of excitability and cholinergic synaptic transmission. Curr Biol. 2014; 24(6):621–9. Epub 2014/03/13. doi: 10.1016/j.cub.2014.02.026 PMID: 24613312

56. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005; 21 (5):676–9. Epub 2004/10/29. PMID: 15509596

S1 Fig. Mutant animals have normal inner ear histology and hearing.
(A) Ampulla, saccule, Scarpa's ganglion, utricle, and semi-circular canals were grossly normal in *nmf9* animals. (B) Acoustic brainstem responses were within normal range (waveform threshold ≤25 dB) for 5 of 6 mutant mice between 2–4 months age. (C) Acoustic startle responses were also normal in a separate cohort animals between 2–6 months age. Two-factor MANOVA p = 0.30 for genotype, p = 0.28 for sex. N = 12 +/+ females, 12 +/+ males, 13 mutant females, 7 mutant males. Error bars, s.e.m.

A

RefSeq Genes   4932411E22Rik   Ankfn1
UCSC Genes   4932411E22Rik   Ankfn1
YourSequence   Nmf9
RT-PCR assays   E22   Ankfn1
                Both

10 kb

B

cDNA ■ plasmid ■

Relative level

Ankfn1   E22   Both

C

Species	Annotation	Evidence
Human	split	sparse ESTs
Mouse	split	singleton cDNAs
Human	single	RT-PCR: DR731397
Mouse	single	RT-PCR: this work
D. melanogaster	single	multiple cDNAs
A. mellifera	single	multiple cDNAs
S. purpuratus	single	single cDNA
C. elegans	single	multiple cDNAs
N. vectensis	single	RT-PCR: this work

D

VMH | AMY | LS | PC | SCN | SpC

Ankfn1

E22

S2 Fig. *Nmf9* mRNA includes both *Ankfn1* and *4932411E22Rik*.
(A) UCSC Genome Browser shows minus-strand transcripts for the 3' exons of *Ankfn1* and 5' exons of *E22* compared to BLAT alignment of empirically determined *Nmf9* mRNA sequence. Positions of primers used for RT-PCR are indicated. (B) Relative quantities for real-time fluorescence measurements (SYBR) of RT-PCR products from the indicated assays from either brain cDNA or cloned plasmid. (C) Publicly deposited cDNA sequences for insect (*D. melanogaster*, A. mellifera), echinoderm (*S. purpuratus*), nematode (*C. elegans*) and cnidarian (*N. vectensis*) also support a single transcript across segments homologous to *Ankfn1* and *E22*. (D) In situ hybridization patterns were indistinguishable for *Ankfn1* and *E22* probes in ventromedial hypothalamus (VMH), amygdala (AMY), lateral septum (LS), piriform cortex (PC), suprachiasmatic nuclei (SCN) and spinal cord (SpC). This mirrors data for both annotations in the Allen Brain Atlas (Fig 3B).

S3 Fig. *Nmf9* expression in developing brain and sensory structures.
Nmf9 RNA was detected by *in situ* hybridization as early as E14.5 in ventricular zones, inner ear, and nasal epithelium.

S4 Fig. Modular organization of *Nmf9* homologs.
Known motifs are indicated in green. Highly consereved domains novel to
the *Nmf9* homology group are shown in blue. An aminoterminal CRIB domain
was found only in the two choanoflagellae species, while inclusion or loss of a
ras-association domain (RA) appears not to be monophyletic. Vertebrate species
basal to mammals had paralogous copies, one with and one without the RA
domain in each species. All mammals examined had only one copy, which never
had an identifiable RA domain. Conservation of the first ANK repeat, and
consensus match to the motif definition were poor or absent in some invertebrate
and choanozoa species.

S5 Fig. Core *Nmf9* domains have lineage-dependent patterns of conservation.
High-resolution sliding window analysis of domains, with inclusion of species
that lacked full-length sequence. Separate analyses were performed for ancestral
homologs and vertebrate-derived homologs. Y-axis shows dN/dS values
calculated using HyPhy [56]. The 10 residues with smallest dN in each analysis
are plotted in orange. Light blue points indicate gaps in sequence. Amino acid
residues are given for regions of highest conservation. (A) Annotated ankyrin and
fibronection type III motifs are plotted. For the ankyrin repeats, N = 53 species for
derived homologs, 47 for ancestral homologs. For fibronectin type III repeats, N =
50 derived, 50 ancestral homologs. (B) The three highly conserved domains that
do not align with known motif definitions are plotted. For domain 1, N = 60
species for derived homologs, 52 for ancestral. For domain 2, N = 54 derived, 49
ancestral. For domain 3, 49 derived, 48 ancestral.

S6 Fig. Expression of *Nmf9* homologs suggests both conserved and adapted functions.

Nmf9 homologues were expressed during development in (A) fruit fly and (B) zebrafish. Samples included 3 biological replicates for fruit fly, but only technical replicates of single samples for each zebrafish stage and paralogous gene. (C) *In situ* hybridization to *Drosophila* embryos showed broad *CG45058* expression, including CNS, agreeing with stage-specific expression in the BDGP in situ database (http://insitu.fruitfly.org/). (D) Expression in sea anemone *N. vectensis* planulae (left) includes endoderm (arrow) and some apical tuft cells. In polyps, expression appears uniform throughout the endoderm (arrow) and mesenteries. (E) In zebrafish embryos, ancestral paralog retained a broad but weak expression pattern that includes expression in the developing inner ear (box and inset). The derived paralog showed a cleaner, more restricted pattern.

S7 Fig. *CG45058* allele locations and deficiency mapping for severe phenotypes.

(A) Schematic of the *CG45058* locus shows exon position of encoded protein domains (blue) across transcripts (navy) from the UCSC browser. Chromatin accesibility to DNase I at stage 5 (green), 9 (orange), 10 (red), 11(blue) and 14 (purple) embryos suggests dynamic regulation of alternative transcripts. (B) UCSC transcripts (navy) in the interval around *CG45058* (red) on *Drosophila* chromosome 3R are shown relative to deficiencies that complement (green bars) or fail to complement (red bars) frameshift mutation at conserved domain 2 (GLYLGYLK). The inferred interval (yellow box) defined by the shared breakpoint of Df(3R)Exel6273 and Df(3R)Exel6192 and the right breakpoint of Df(3R)ED6090 contains parts of 6 transcription units.

ORIGINAL ARTICLE

Aging Cell WILEY

The lipid elongation enzyme ELOVL2 is a molecular regulator of aging in the retina

Daniel Chen[1] | Daniel L. Chao[1] | Lorena Rocha[1] | Matthew Kolar[2] | Viet Anh Nguyen Huu[1] | Michal Krawczyk[1] | Manish Dasyani[1] | Tina Wang[3] | Maryam Jafari[1] | Mary Jabari[1] | Kevin D. Ross[4] | Alan Saghatelian[2] | Bruce A. Hamilton[4,5] | Kang Zhang[1] | Dorota Skowronska-Krawczyk[1,6] (ID)

[1]Shiley Eye Institute, Viterbi Family Department of Ophthalmology, University of California San Diego, La Jolla, CA, USA

[2]The Salk Institute for Biological Studies, Clayton Foundation Laboratories for Peptide Biology, La Jolla, CA, USA

[3]Department of Medicine, University of California San Diego, La Jolla, CA, USA

[4]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA

[5]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA

[6]Atkinson Laboratory for Regenerative Medicine, University of California San Diego, La Jolla, CA, USA

Correspondence
Dorota Skowronska-Krawczyk, Shiley Eye Institute, Viterbi Family Department of Ophthalmology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA.
Email: DorotaSK@health.ucsd.edu

Funding information
NIH, Grant/Award Number: R01 EY02701, R01 GM086912, K12 EY024225, T32 GM008; UCSD Vision Research Center Core

Abstract

Methylation of the regulatory region of the elongation of very-long-chain fatty acids-like 2 (*ELOVL2*) gene, an enzyme involved in elongation of long-chain polyunsaturated fatty acids, is one of the most robust biomarkers of human age, but the critical question of whether *ELOVL2* plays a functional role in molecular aging has not been resolved. Here, we report that *Elovl2* regulates age-associated functional and anatomical aging in vivo, focusing on mouse retina, with direct relevance to age-related eye diseases. We show that an age-related decrease in *Elovl2* expression is associated with increased DNA methylation of its promoter. Reversal of *Elovl2* promoter hypermethylation in vivo through intravitreal injection of 5-Aza-2'-deoxycytidine (5-Aza-dc) leads to increased *Elovl2* expression and rescue of age-related decline in visual function. Mice carrying a point mutation C234W that disrupts *Elovl2*-specific enzymatic activity show electrophysiological characteristics of premature visual decline, as well as early appearance of autofluorescent deposits, well-established markers of aging in the mouse retina. Finally, we find deposits underneath the retinal pigment epithelium in *Elovl2* mutant mice, containing components found in human drusen, a pathologic hallmark of age related macular degeneration. These findings indicate that ELOVL2 activity regulates aging in mouse retina, provide a molecular link between polyunsaturated fatty acids elongation and visual function, and suggest novel therapeutic strategies for the treatment of age-related eye diseases.

KEYWORDS
age-related macular degeneration, aging, DNA methylation, ELOVL2, retina, PUFA

Chen and Chao authors contributed equally

1 | INTRODUCTION

Chronological age predicts relative levels of mental and physical performance, disease risks across common disorders, and mortality (Glei, 2016). The use of chronological age is limited, however, in explaining the considerable biological variation among individuals of a similar age. Biological age is a concept that attempts to quantify different aging states influenced by genetics and a variety of environmental factors. While epidemiological studies have succeeded in providing quantitative assessments of the impact of discrete factors on human longevity, advances in molecular biology now offer the ability to look beyond population-level effects and to hone in on the effects of specific factors on aging within single organisms.

A quantitative model for aging based on genome-wide DNA methylation patterns by using measurements at 470,000 CpG markers from whole-blood samples of a large cohort of human individuals spanning a wide age range has recently been developed (Hannum, 2013; Horvath, 2013; Levine, 2018). This method is highly accurate at predicting age and can also discriminate relevant factors in aging, including gender, genetic variants, and disease (Gross, 2016; Hannum, 2013). Several models work in multiple tissues (Horvath, 2013; Levine, 2018), suggesting the possibility of a common molecular clock, regulated in part by changes in the methylome. In addition, these methylation patterns are strongly correlated with cellular senescence and aging (Xie, Baylin, & Easwaran, 2019). The regulatory regions of several genes become progressively methylated with increasing chronological age, suggesting a functional link between age, DNA methylation, and gene expression. The promoter region of *ELOVL2*, in particular, was the first to be shown to reliably show increased methylation as humans age (Garagnani, 2012), and confirmed in one of the molecular clock models (Hannum, 2013).

ELOVL2 (elongation of very-long-chain fatty acids-like 2) encodes a transmembrane protein involved in the elongation of long-chain (C22 and C24) omega-3 and omega-6 polyunsaturated fatty acids (LC-PUFAs; Leonard, 2002). Specifically, ELOVL2 is capable of converting docosapentaenoic acid (DPA) (22:5n-3) to 24:5n-3, which can lead to the formation of very-long-chain PUFAs (VLC-PUFAs) as well as 22:6n-3, docosahexaenoic acid (DHA; Gregory, Cleland, & James, 2013). DHA is the main polyunsaturated fatty acid in the retina and brain. Its presence in photoreceptors promotes healthy retinal function and protects against damage from bright light and oxidative stress. *ELOVL2* has been shown to regulate levels of DHA (Pauter, 2014), which in turn has been associated with age-related macular degeneration (AMD), among a host of other retinal degenerative diseases (Bazan, Molina, & Gordon, 2011). In general, LC-PUFAs are involved in crucial biological functions including energy production, modulation of inflammation, and maintenance of cell membrane integrity. It is, therefore, possible that *ELOVL2* methylation plays a role in the aging process through the regulation of these diverse biological pathways.

In this study, we investigated the role of ELOVL2 in molecular aging in the retina. We find that the *Elovl2* promoter region is increasingly methylated with age in the retina, resulting in age-related decreases in *Elovl2* expression. These changes are associated with decreasing visual structure and function in aged mice. We then demonstrate that loss of ELOVL2-specific function results in the early-onset appearance of sub-RPE deposits that contain molecular markers found in drusen in AMD. This phenotype is also associated with visual dysfunction as measured by electroretinography, and it suggests that ELOVL2 may serve as a critical regulator of a molecular aging clock in the retina, which may have important therapeutic implications for diseases such as age-related macular degeneration.

2 | RESULTS

2.1 | Elovl2 expression is downregulated with age through methylation and is correlated with functional and anatomical biomarkers in aged wild-type mice

Previous studies showed that methylation of the promoter region of *ELOVL2* is highly correlated with human age (Hannum, 2013). Methylation of regulatory regions is thought to prevent the transcription of neighboring genes and serves as a method to regulate gene expression. We first wished to characterize whether the age-associated methylation of the *ELOVL2* promoter previously found in human serum also occurs in the mouse. First, we analyzed ELOVL2 promoter methylation data obtained using bisulfite sequencing in mouse blood and compared it to the available human data for the same region (Wang, 2017) and observed similar age-related increase in methylation level in the compared regions (Figure S1a). To assay methylation of the *Elovl2* promoter in retina, we used methylated DNA immunoprecipitation (MeDIP) method (Weber, 2005) and tested the methylation levels in the CpG island in the *Elovl2* regulatory region by quantitative PCR with *Elovl2*-specific primers (Table S1). MeDIP analysis of the CpG island in the *Elovl2* regulatory region showed increasing methylation with age in the mouse retina (Figure 1a). This was well-correlated with age-related decreases in expression of *Elovl2* as assessed by Western blot and qPCR (Figure 1b and Figure S1b,c) indicating the potential role of age-related changes in DNA methylation in *Elovl2* expression.

To understand the cell-type and age-specific expression of *Elovl2*, we performed in situ hybridization with an *Elovl2* RNAscope probe on mouse retina sections (Stempel, Morgans, Stout, & Appukuttan, 2014). In three-month-old and in 22-month-old mice, we noticed *Elovl2* expression in the photoreceptor layer, particularly in the cone layer as well as the RPE (Figure 1c and Figure S1e). We observed that the expression of Elovl2 on mRNA level in RPE was lower than in the retina (Figure S1d). Importantly, at older stages (22-month-old animals), we noticed *Elovl2* mRNA in the same locations but dramatically reduced in expression (Figure 1c). As *Elovl2* is also highly expressed in the liver, we performed a time course of *Elovl2* expression in this tissue. We observed similar age-related decreases in *Elovl2* expression correlated with increases in methylation of the *Elovl2* promoter in mouse liver, indicating that age-associated methylation of *Elovl2* occurs in multiple tissues in mice (Figure S1f).

Visual function is highly correlated with age, including age-related decreases in rod function in both humans and mice (Birch & Anderson,

FIGURE 1 ELOVL2 expression is downregulated with age through methylation of its promoter and is correlated with age-related increases in autofluorescence aggregates and decreased scotopic response. (a) Methylation of ELOVL2 promoter region measured using immunoprecipitation of methylated (MeDIP) followed by qPCR. ELOVL2 promoter is increasingly methylated with age. (b) Time course of retinal ELOVL2 protein expression by Western blot. ELOVL2 protein is expression decreases with age. ns, nonspecific signal produced by ELOVL2 antibodies (c) Images of mouse retina sections from young—3mo (top panels) and old—22mo (bottom panels) animals stained with RNAscope probes designed for *Elovl2* and *Arrestin 3*, counterstained with DAPI. ONL, outer nuclear layer, INL, inner nuclear layer, RGC, retinal ganglion cells. Bar—100um. (d) Time course of representative fundus autofluorescence pictures of C57BL/6J mice. Arrows denote autofluorescent deposits. (e) Quantification of autofluorescent deposits in fundus images. $N = 4$. (f) Scotopic responses by ERG over mouse lifespan. For panels A, E, and F, $N = 4$, *$p < .5$, ** $p < .01$, 1-way ANOVA. Error bars denote *SD*
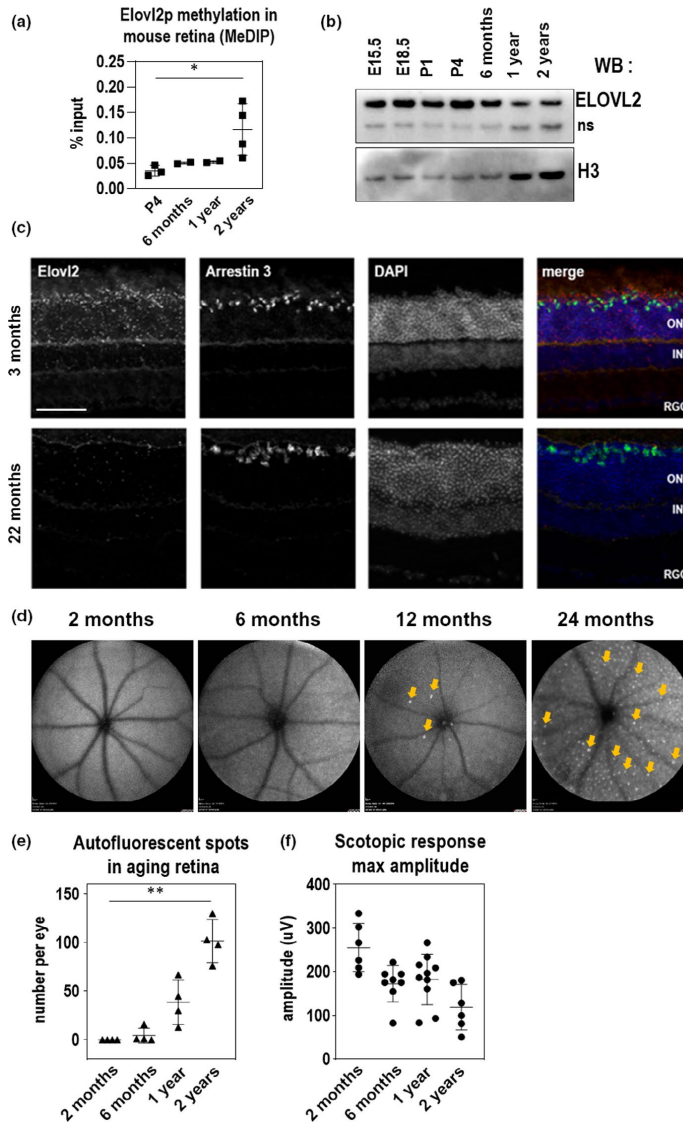
1992; Kolesnikov, Fan, Crouch, & Kefalov, 2010). In addition, autofluorescent aggregates have been observed in the fundus of aged mice, suggesting that these aggregates may also be an anatomical surrogate of aging in the mouse retina (Chavali, 2011; Xu, Chen, Manivannan, Lois, & Forrester, 2008). To measure and correlate these structural and visual function changes with age in mice, we performed an analysis of wild-type C57BL/6J mice at various timepoints through development, using fundus autofluorescence and electroretinography (ERG) as structural and functional readouts for vision. We observed increasing amounts of autofluorescent aggregates on fundus autofluorescence imaging with increasing mouse age, most prominently at two years (Figure 1d,e and Figure S1g). We also detected an age-associated decrease in visual function, as measured by maximum scotopic amplitude by ERG (Figure 1f and Figure S1h), as shown in previous studies (Kolesnikov et al., 2010; Williams & Jacobs, 2007). These data show that an age-associated accumulation of autofluorescent spots and decrease in visual function as detected by ERG correlate with *Elovl2* downregulation in the mouse retina.

2.2 | Manipulating ELOVL2 expression causes age-related changes in cells

The WI38 and IMR90 cell lines are well-established cell models of aging (Hayflick, 1965). We used these cell lines to further explore the effect of *ELOVL2* promoter methylation on cell health. First, using MeDIP, we found that promoter methylation increased with cell population doubling (Figure 2a) further confirming strong correlation between increased *ELOVL2* methylation and aging. Since the methylation of the promoter region was shown to be inhibitory for transcription (Jones, 1998), we investigated whether the expression level of *ELOVL2* inversely correlated with *ELOVL2* promoter methylation. Using qRT–PCR, we found that the expression level of the gene decreased with increasing population doubling (PD) number (Figure 2b)). We conclude that *ELOVL2* expression is downregulated in aging cells, with a correlated increase in *ELOVL2* promoter methylation.
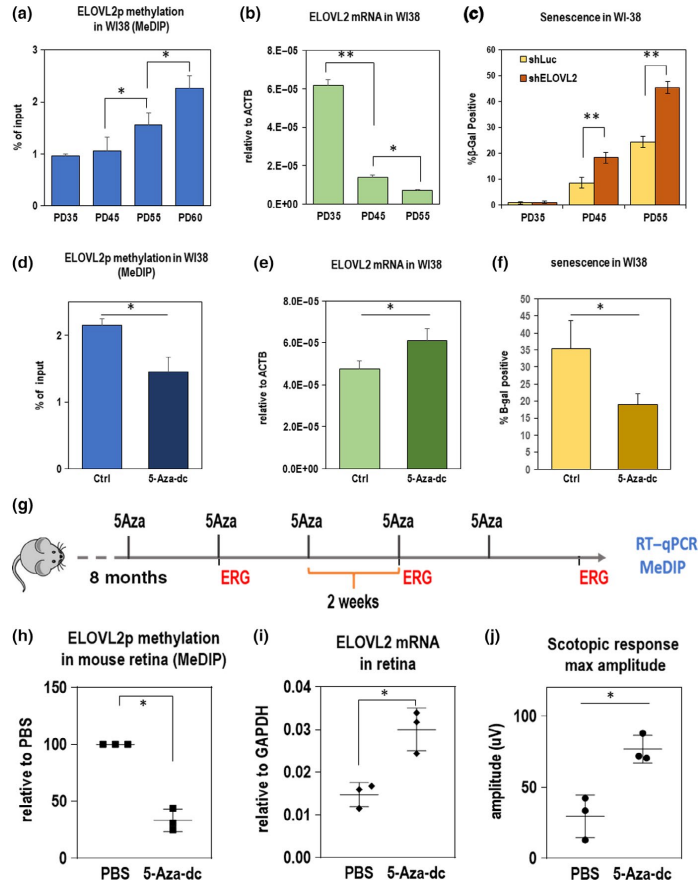
FIGURE 2 (a-c) *ELOVL2* expression, methylation, and senescence in WI38 csells. (a) Methylation level in *ELOVL2* promoter region in human normal lung cell line WI38 by MeDIP/qPCR. Amplicons contain CpG markers cg16867657, cg24724428, and cg21572722. *N* > 3. (b) *ELOVL2* expression by qPCR in WI38 cells at PD35, PD45, and PD55. (c) Fraction of senescent cells measured by beta-galactosidase staining in WI38 cells at given population doubling upon shRNA-mediated knockdown of ELOVL2 gene or control Luc. (d-f) Manipulating DNA methylation in PD52 WI38 cells. (d) *ELOVL2* promoter methylation as measured by MeDIP followed by qPCR in untreated control and 5-Aza-dc-treated WI38 cells. (e) *ELOVL2* expression by qPCR in untreated control and 5-Aza-dc-treated WI38 cells. (f) Percent senescence by beta-galactosidase staining in WI38 cells treated with 2μM 5-Aza-dc. (g-j) Manipulating DNA methylation in mice. (g) Experimental setup. Eight-month-old mice were injected intravitreally with of 5-Aza-dc five times every two weeks. ERG measurements were taken at indicated time points. At 11 months, expression and methylation levels were measured in 5-Aza-dc treated and control (PBS-treated) mice. (h) Methylation of ELOVL2 promoter by MeDIP at 11 months after 5-Aza injection. (i) *ELOVL2* expression by qPCR after 5-Aza injection. (j) Maximum amplitude scotopic response by ERG after 5-Aza injection. For panels A-F, *N*>=3, *p < .05, **p < .01, *t* test. Error bars denote *SD*; for panels H-J, *N* = 3, *p < .05, **p < .01, *t* test. Error bars denote *SD*

We then asked whether modulating the expression of *ELOVL2* could influence cellular aging. First, using shRNA delivered by lentivirus, we knocked down *ELOVL2* expression in WI38 and another model cell line, IMR-90, and observed a significant decrease in proliferation rate (Figure S2a,b), an increased number of senescent cells in culture as detected by SA-β-gal staining (Figure 2c and Figure S2e), and morphological changes consistent with morphology of high PD cells (Figure S2f). Altogether, these data suggest that decreasing ELOVL2 expression results in increased aging and senescence in vitro.

Next, we tested whether we could manipulate *Elovl2* expression by manipulating the *Elovl2* promoter methylation. We treated WI38 fibroblasts with 5-Aza-2'-deoxycytidine (5-Aza-dc), a cytidine analog that inhibits DNA methyltransferase (Momparler, 2005). Cells were treated for two days with 2 μM 5-Aza-dc followed by a five-day washout period. Interestingly, we found that upon treatment with 5-Aza-dc, *Elovl2* promoter methylation was reduced (Figure 2d), and *Elovl2* expression was upregulated (Figure 2e). Moreover, upon 5-Aza-dc treatment, a lower percentage of senescent cells were observed in culture (Figure 2f). To assess whether the decrease in senescence is caused at least in part by the ELOVL2 function, we knocked down the *ELOVL2* expression in aged WI38 cells and treated them with 5-Aza-dc as previously described. Again, significantly lower proportion of senescent cells was detected upon the drug treatment, but the effect of drug treatment was significantly reduced by shRNA-mediated knockdown of *ELOVL2*, using either of two *ELOVL2* shRNAs compared with a control shRNA (Figure S2b). This indicates an important role of ELOVL2 in the process. Altogether, these data suggest that the reversing *ELOVL2* promoter methylation increases its expression and decreases senescence in vitro.

2.3 | DNA demethylation in the retina by intravitreal injection of 5-Aza-dc increases Elovl2 expression and rescues age-related changes in scotopic function in aged mice

We next explored whether demethylation of the *Elovl2* promoter could have similar effects on *Elovl2* expression in vivo. To accomplish this, we performed intravitreal injection of 5-Aza-dc, known to affect DNA methylation in nondividing neurons (Choi, Lee, Kim, Choi, & Lee, 2018; Christman, 2002; Miller & Sweatt, 2007; Wang, 2018), into aged wild-type mice. Eight-month-old C57BL/6J mice were injected with 1 μl of 2 μM 5-Aza-dc in one eye and 1 μl of PBS in the other eye as a control, every other week over a period of 3 months (total of 5 injections) (Figure 2g). After the treatment, tissues were collected, and RNA and DNA were extracted. We found, using the MeDIP method, that methylation of the *Elovl2* promoter decreased after treatment (Figure 2h), with a corresponding upregulation of *Elovl2* expression (Figure 2i). Notably, we observed that the scotopic response was significantly improved in the 5-Aza-dc-injected eyes compared with vehicle controls (Figure 2j). These data show that DNA demethylation, which included demethylation

of the *Elovl2* promoter region, influence and potentially delay age-related changes in visual function in the mouse retina.

2.4 | Elovl2[C234W] mice demonstrate a loss of ELOVL2-specific enzymatic activity

We next sought to investigate the in vivo function of *Elovl2* in the retina. Since C57BL/6 *Elovl2* knockout heterozygous mice display defects in spermatogenesis and are infertile (Zadravec, 2011), we developed an alternative strategy to eliminate ELOVL2 enzymatic activity in vivo. Using CRISPR-Cas9 technology, we generated *Elovl2*-mutant mice encoding a cysteine-to-tryptophan substitution (C234W). This mutation selectively inactivates enzymatic activity of *ELOVL2* required to process C22 PUFAs, to convert docosapentaenoic acid (DPA) (22:5n-3) to 24:5n-3, while retaining elongase activity for other substrates common for ELOVL2 and the paralogous enzyme ELOVL5 (Figure 3a, Figure S3a; Gregory et al., 2013; Gregory et al., 2013; Zadravec, 2011). A single-guide RNA against the *Elovl2* target region, a repair oligonucleotide with a base pair mutation to generate the mutant C234W, and Cas9 mRNA were injected into C57BL/6N mouse zygotes (Figure 3b). One correctly targeted heterozygous founder with the C234W mutation was identified. No off-target mutations were found based on DNA sequencing of multiple related DNA sequences in the genome (Figure S3b). The C234W heterozygous mice were fertile, and C234W homozygous mice developed normally and showed no noticeable phenotypes. We analyzed the long-chain fatty levels in the retinas of homozygous *Elovl2*[C234W] mice to determine whether there was a loss of enzymatic activity specific to ELOVL2. We observed that *Elovl2*[C234W] mice had higher concentrations of C22:5 fatty acid (a selective substrate of ELOVL2 elongation) and lower levels of C24:5 (primary product of ELOVL2 enzymatic activity) and C22:6 (DHA—the secondary product of ELOVL2) (Figure 3c). We also observed similar changes in fatty acid levels in livers of *Elovl2*[C234W] mice as well as lower levels of longer fatty acids that require primary product of Elovl2 as a substrate (Figure S4) This suggests that the *Elovl2*[C234W] mice have altered ELOVL2 substrate specificity and inhibited ELOVL2-specific C22 elongase activity.

2.5 | Loss of ELOVL2-specific activity results in early vision loss and accumulation of sub-RPE deposits

We next investigated whether the *Elovl2*[C234W] mutation affected the retinal structure and/or function in vivo. First, we observed a significant number of autofluorescent spots on fundus photography in animals at six months of age, which were not found in wild-type littermates (Figure 4a,b). This phenotype was consistently observed in 6-, 8-, and 12-month-old mutant animals and in both animal sexes, but the phenotype was consistently more pronounced in male mice (Figure S5). Importantly, ERG analysis revealed that 6-month-old
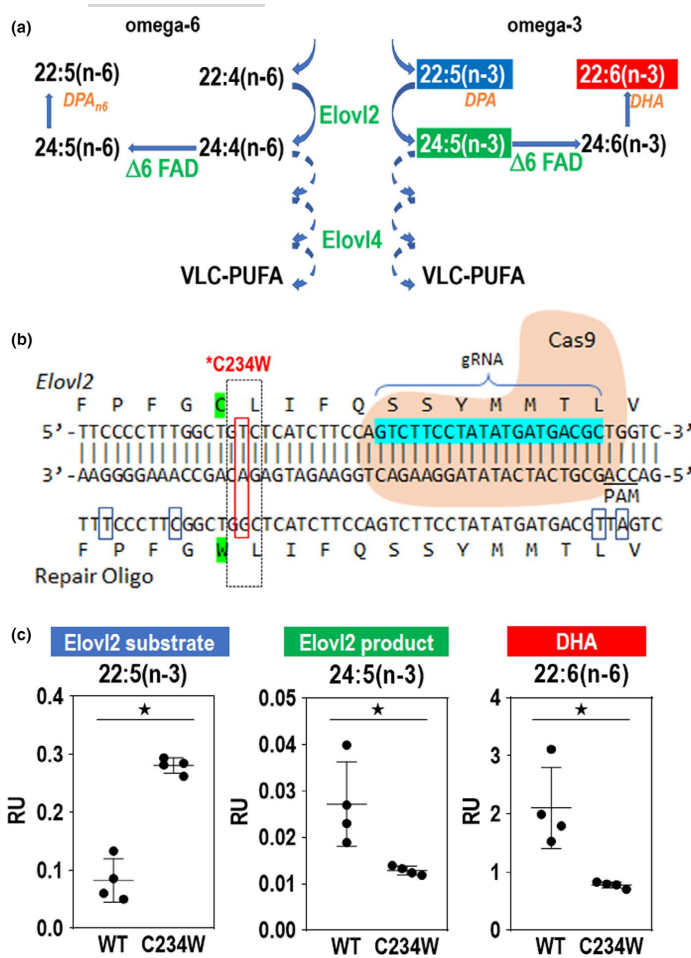
FIGURE 3 $Elovl2^{C234W}$ mice show a loss of ELOVL2 enzymatic activity. (a) Schematic of ELOVL2 elongation of omega-3 and omega-6 fatty acids. ELOVL2 substrates 22:5 (n-3) and 22:4(n-6) are elongated by ELOVL2 to 24:5(n-3) and 24:4(n-6). This leads to other products such as DHA, $DPA_{n6,}$ and VLC-PUFAs, which are elongated by ELOVL4. (b) CRISPR-Cas9 strategy to create $Elovl2^{C234W}$ mice. $Elovl2$ gRNA, Cas9, and repair oligo are used to create the $Elovl2^{C234W}$ mutant. (c) Lipid levels of ELOVL2 substrate DPA (22:5(n-3)), ELOVL2 product (24:5(n-3)), and DHA (22:6(n-3)) in retinas of $Elovl2^{C234W}$ mice and wild-type littermates. $N = 4$, $*p < .05$ by Mann–Whitney U test. Error bars represent SD

$Elovl2^{C234W}$ mice displayed a decrease in visual function as compared to wild-type littermates (Figure 4c, Figure S5).

To determine the impact of the mutation on the morphology of the retina on the microscopic level, we performed an immunohistological analysis of tissue isolated from wild-type and $Elovl2^{C234W}$ littermates. Although we did not observe gross changes in morphology of the retinas in mutant animals, we have observed the presence of small aggregates underneath the RPE and found that these sub-RPE aggregates contained the complement component C3 as well as the C5b-9 membrane attack complex, proteins found in human drusenoid aggregates (Figure 4d). In addition, in the mutant sub-RPE aggregates, we also identified other components found in human deposits such as HTRA1 (Cameron, 2007), oxidized lipids/T15 (Shaw, 2012), and ApoE, an apolipoprotein component of drusen (Li, Clark, Chimento, & Curcio, 2006; Figure 4e). This suggests that the sub-RPE deposits found in the $Elovl2^{C234W}$ mouse contain some drusen-specific components found in early nonexudative AMD. Taken together, these data implicate ELOVL2-specific activity as a potential functional target in age-related eye diseases.

3 | DISCUSSION

3.1 | ELOVL2 as a critical regulator of molecular aging in the retina

This work is the first demonstration, to our knowledge, of a functional role for $Elovl2$ in regulating age-associated phenotypes in the retina. Methylation of the promoter region of $ELOVL2$ is well-established as a robust prognostic biomarker of human aging (Garagnani, 2012; Gopalan, 2017), but whether ELOVL2 activity contributes to aging phenotypes had not yet been documented.
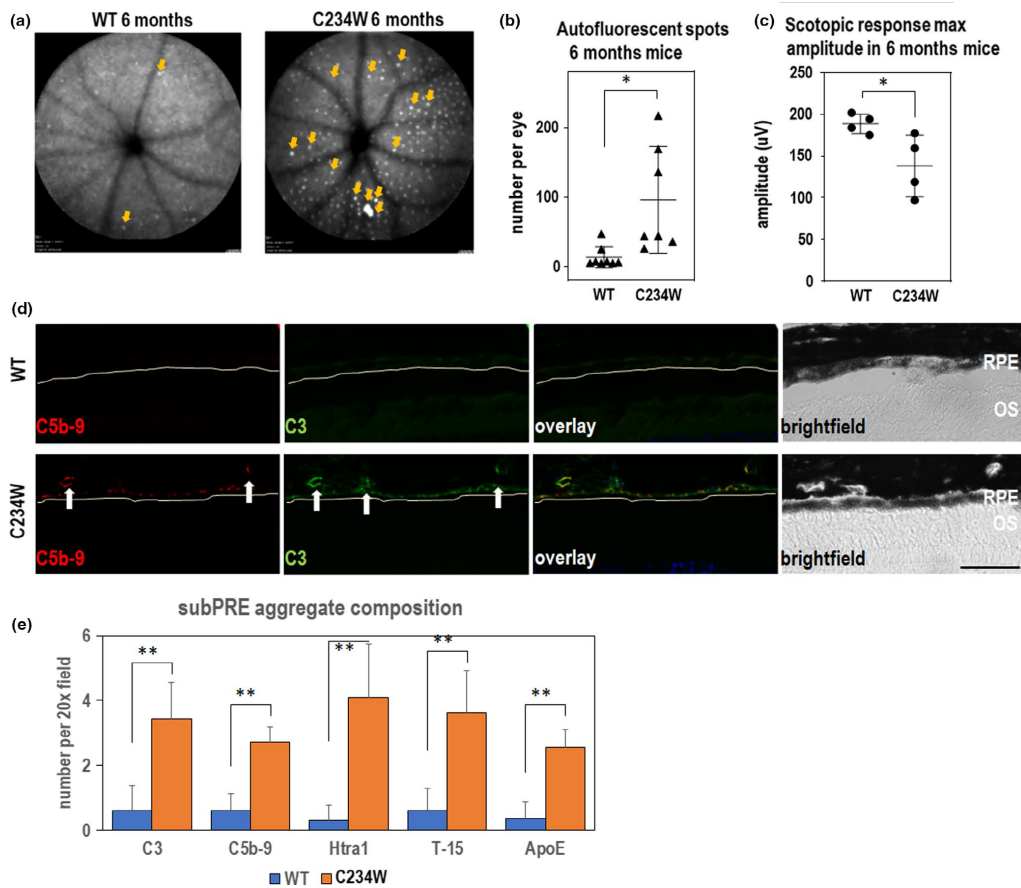
FIGURE 4 *Elovl2*[C234W] mice show autofluorescent deposits and vision loss. (a) Representative fundus autofluorescence images of WT and *Elovl2*[C234W] mice at 6 months with representative scotopic ERG waveforms. Note multiple autofluorescent deposits (arrows) in *Elovl2*[C234W] mice which are almost absent in wild-type littermates. (b) Quantification of the autofluorescent spots in 6mo wild-type and C234W mutant mice. *N* = 8. *p < .05, *t* test. Error bars denote *SD*. (c) Maximum scotopic amplitude by ERG at 6 months between WT and *Elovl2*[C234W] mice. *N* = 4, *p < .05, *t* test. Error bars represent *SD*. (d) Immunohistochemistry of sub-RPE deposits found in *Elovl2*[C234W] mice. Deposits are found underneath the RPE (yellow line), which colocalize with C3 and C5b-9, which is not present in WT controls. Bar—50um. (e) Quantification of sub-RPE aggregates stained with C3, C5b-9, Htra1, T-15, and ApoE, all components found in drusen in AMD. *N* = 4, ** p < .01, *t* test. Error bars represent *SD*

In this work, we demonstrated that the age-related methylation of regulatory regions of *Elovl2* occurs in the rodent retina and results in age-related decreases in the expression of *Elovl2*. We show that inhibition of ELOVL2 expression by transfection of ELOVL2 shRNA in two widely used cell models results in increased senescence and decreased proliferation, endpoints associated with aging. Conversely, we show that the administration of 5-Aza-dc leads to demethylation of ELOVL2 promoter and prevents cell proliferation and senescence compared with controls.

Next, we explored whether *Elovl2* expression affected age-related phenotypes in vivo. Intravitreal injection of 5-Aza-dc in rodents increased *Elovl2* expression and reversed age-related changes in visual function by ERG. Next, we showed a decrease in visual function as assessed by ERG as well as increased accumulation of autofluorescent white spots in *Elovl2*[C234W] mice, with ELOVL2-specific activity eliminated, compared with littermates controls. These physiologic and anatomical phenotypes are well-established markers of aging in the mouse retina, suggesting that loss of *Elovl2* may be accelerating aging on a molecular level in the retina. Finally, in *Elovl2*[C234W] mice, we observed the appearance of sub-RPE deposits, which colocalize with markers found in human drusen in macular degeneration, a pathologic hallmark

of a prevalent age-related disease in the eye. Taken together, we propose that *Elovl2* plays a critical role in regulating a molecular aging in the retina, which may have therapeutic implications for age-related eye diseases.

3.2 | Methylation of the regulatory region as a mechanism of age-dependent gene expression

DNA methylation at the 5-position of cytosine (5-methylcytosine, 5mC) is catalyzed and maintained by a family of DNA methyltransferases (DNMTs) in eukaryotes (Law & Jacobsen, 2010) and constitutes ~2%–6% of the total cytosines in human genomic DNA (28). Alterations of 5mC patterns within CpG dinucleotides within regulatory regions are associated with changes in gene expression (Jones, 1998; Telese, Gamliel, Skowronska-Krawczyk, Garcia-Bassets, & Rosenfeld, 2013). Recently, it has been shown that one can predict human aging using DNA methylation patterns. In particular, increased DNA methylation within the CpG island overlapping with the promoter of *ELOVL2* was tightly correlated with the age of the individual (Gopalan, 2017). We attempted to demethylate this region using 5-Aza-dc, known to inhibit the function of DNMTs also in nondividing neurons (Choi et al., 2018; Miller & Sweatt, 2007; Wang, 2018). We reported that upon intravitreal injection of the compound, the DNA methylation is reduced, gene expression is upregulated, and visual function is maintained in the treated eye compared with the contralateral control. These data suggest that *Elovl2* is actively methylated by enzymes inhibited by 5-Aza-dc and that age-related methylation either directly or indirectly regulates *Elovl2* expression. Further studies are needed to fully address the directness and specificity of methylation effects on *Elovl2* expression and visual function.

3.3 | A molecular link between long-chain PUFAs in age-related eye diseases

Our data show that *Elovl2*[C234W] animals display accelerated loss of vision and the appearance of macroscopic autofluorescent spots in fundus images. The exact identity of such spots in mouse models of human diseases is unclear, as they have been suggested to be either protein-rich, lipofuscin deposits or accumulating microglia (Chavali, 2011; Combadiere, 2007). Rather than deciphering the identity of these macroscopic spots, we used the phenotype as a potential sign of age-related changes in the retina, as suggested by others (Chavali, 2011; Kim, 2014).

The composition of aggregates visible on the microscopic level in sub-RPE layers in the retina is potentially informative with regard to human parallels. Using immunofluorescence, we observed the accumulation of several proteins described previously as characteristic for drusen in human AMD samples. Although our analysis did not exhaust the documented components of drusen in human disease (Crabb, 2014), nevertheless, our data show the appearance of these sub-RPE deposits, even in the absence of known confounding mutations or variants correlating with the risk of the disease.

What may be the mechanism by which *Elovl2* activity results in drusen-like deposits and loss of visual function? ELOVL2 plays an essential role in the elongation of long-chain (C22 and C24) omega-3 and omega-6 polyunsaturated acids (LC-PUFAs) (Figure 3a). LC-PUFAs are found primarily in the rod outer segments and play essential roles in retinal function. These PUFAs include both long-chain omega-3 (n-3) and omega-6 (n-6) fatty acids such as docosahexaenoic acid (DHA) and arachidonic acid (AA). DHA is the major polyunsaturated fatty acid found in the retina and has been shown to play diverse roles in photoreceptor function, protection in oxidative stress, and retinal development (Leeuwen, 2018). While DHA has been well-studied in the human retina, the function of other LC-PUFAs in the ELOVL2 elongation pathway is unknown. Further experiments to dissect the roles of specific LC-PUFAs in this pathway and which of these lipid species are implicated in this phenotype are still required.

Multiple lines of evidence have linked PUFAs to age-related macular degeneration (AMD). AMD is the leading cause of blindness in developed countries (Ambati & Fowler, 2012) among the elderly. There are two advanced subtypes of AMD, an exudative form due to neovascularization of the choroidal blood vessels, and a nonexudative form which results in gradual retinal pigment epithelium (RPE) atrophy and photoreceptor death. While there are currently effective therapies for exudative AMD, there are no treatments which prevent photoreceptor death from nonexudative AMD. A pathologic hallmark of nonexudative AMD is the presence of drusen, lipid deposits found below the RPE, which leads to RPE atrophy and photoreceptor death, termed geographic atrophy. The pathogenesis of macular degeneration is complex and with multiple pathways implicated including complement activation, lipid dysregulation, oxidative stress, and inflammation (Ambati & Fowler, 2012). Despite intense research, the age-related molecular mechanisms underlying drusen formation and geographic atrophy are still poorly understood.

Analysis of AMD donor eyes showed decreased levels of multiple LC-PUFAs and VLC-PUFAs in the retina and RPE/choroid compared with age-matched controls (Liu, Chang, Lin, Shen, & Bernstein, 2010). Epidemiologic studies suggest that low dietary intake of LC-PUFAs such as omega-3 fatty acids was associated with a higher risk of AMD (Sangiovanni, 2009; Seddon, George, & Rosner, 2006). Furthermore, mutations in ELOVL4, a key enzyme in the synthesis of VLC-PUFAs, have been identified in Stargardt-like macular dystrophy (STGD3), a juvenile retinal dystrophy with macular deposits reminiscent of AMD (Bernstein, 2001; Edwards, Donoso, & Ritter, 2001; Zhang, 2001). Despite the biochemical, epidemiologic, and genetic evidence implicating PUFAs in AMD, the molecular mechanisms by which LC and VLC-PUFAs are involved in drusen formation and AMD pathogenesis are still poorly understood. The finding that loss of ELOVL2 activity results in early accumulation of sub-RPE deposits strengthens the relationship between PUFAs and macular degeneration. Since *Elovl2* is expressed in both photoreceptors and RPE, whether these phenotypes of visual loss and sub-RPE deposits are due to cell-autonomous function in the photoreceptors and RPE, respectively, or require interplay between photoreceptors and RPE still needs to be established.

3.4 | Role of *Elovl2* in aging

DNA methylation of the regulatory region of *Elovl2* gene is well-established to be a cell-type-independent molecular aging clock (Garagnani, 2012; Hannum, 2013; Slieker, Relton, Gaunt, Slagboom, & Heijmans, 2018) with *Elovl2* expression detectable in many tissues and highest levels observed in liver, testis, and central nervous system including retina (https://www.proteinatlas.org). The high metabolic activity and critical role of PUFAs, reflecting a high metabolic demand for the products of the ELOVL2 enzyme in the photoreceptors, is the most probably the reason why the ocular phenotype is first to be observed in the *Elovl2*[C234W] animals. Further studies are required to establish the role of the gene in other tissues than the retina and impact of the lack of the ELOVL2 products in the lipid bilayers in aged organisms.

4 | CONCLUSIONS

In summary, we have identified the lipid elongation enzyme ELOVL2 as a critical component in regulating molecular aging in the retina. Further studies may lead to a better understanding of molecular mechanisms of aging in the eye, as well as lead to therapeutic strategies to treat a multitude of age-related eye diseases.

5 | METHODS

5.1 | Cell culture and treatment

WI38 (ATCC Cat# CCL-75, RRID:CVCL_0579) and IMR-90 (ATCC Cat# CCL-186, RRID:CVCL_0347) human fibroblasts were cultured in EMEM (ATCC) supplemented with 10% fetal bovine serum (Omega) and 1% penicillin/streptomycin (Gibco), and kept in a humidified incubator at 5% CO_2 and 37°C. Confluence was calculated via ImageJ imaging software, including three fields of view per sample (10×). Upon confluence, cells were split and seeded at a 1:3 ratio. Population doublings (PDs) were calculated by cell count. Knockdown lentivirus was generated using MISSION shRNA (Sigma) according to the manufacturer's instructions. 5-Aza-2'-deoxycytidine was purchased from TSZ Chem (CAS#2353-33–5) and dissolved in cell culture medium at a concentration of 2μM. Cells were treated every day for a period of 48 hr. The medium was then replaced with regular cell culture medium, and the cells were cultured for 5 more days.

5.2 | Senescence-associated β-galactosidase (SA-β-gal) activity

The SA-β-gal activity in cultured cells was determined using the Senescence β-Galactosidase Staining Kit (Cell Signaling Technology), according to the manufacturer's instructions. Cells were stained with DAPI afterward, and percentages of cells that stained positive were calculated with imaging software (Keyence), including three fields of view (10×).

5.3 | Nucleic acid analysis

DNA and RNA were isolated from human fibroblasts and mouse tissues with TRIzol (Ambion) according to the manufacturer's instructions. RNA was converted to cDNA with iScript cDNA Synthesis Kit (Bio-Rad). qPCR was performed using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad).

Methylated DNA immunoprecipitation (MeDIP) was performed by shearing 1μg DNA by Bioruptor (Diagenode) for 8 cycles on the high setting, each cycle consisting of 30 s on and 30 s off. Sheared DNA was denatured, incubated with 1 μg 5mC antibody MABE146 (Millipore) for 2 hr, and then with SureBeads protein G beads (Bio-Rad) for 1 hr. After washing, DNA was purified with QIAquick PCR Purification Kit (Qiagen). qPCR was then performed as above. List of primers can be found in Table S1.

5.4 | Western blotting

10μg of total protein isolated with TRIzol (Invitrogen) from retinas of WT mice of varying stages of development was subject to SDS-PAGE followed by Western blotting (see Table S2 for antibodies used in the study). H3 served as loading control.

5.5 | Quantification of western blots

WB ECL signals were imaged using Bio-Rad ChemiDoc system. Background-subtracted signal intensities were calculated using ImageJ separately for ELOVL2 bands and H3 loading-control bands. ELOVL2 levels were calculated by dividing ELOVL2 signals by corresponding H3 signals, and then normalized to E15.5.

5.6 | RNAscope® In situ hybridization

In situ hybridization was performed using the RNAscope® Multiplex Fluorescent Assay v2 (ACD Diagnostics). Mouse *Elovl2 Rpe65* and *Arr3* probes (p/n 542711, p/n 410151, and p/n 486551, respectively) were designed by the manufacturer. Briefly, fresh frozen histologic sections of mouse eyes were pretreated per manual using hydrogen peroxide and target retrieval reagents such as protease IV. Probes were then hybridized according to the protocol and then detected with TSA Plus® Fluorophores fluorescein, cyanine 3, and cyanine 5 (Perkin Elmer). Sections were mounted with DAPI and Prolong Gold Antifade (Thermo Fisher) with coverslip for imaging and imaged (Keyence BZ-X700).

5.7 | CRISPR-Cas9 design

CRISPR-Cas9 reagents were generated essentially as described (Wang, 2013) and validated in our facility (Concepcion, Ross, Hutt, Yeo, & Hamilton, 2015). T7 promoter was added to cloned Cas9 coding sequence by PCR amplification. The T7-Cas9 product was then gel-purified and used as the template for in vitro transcription (IVT) using mMESSAGE mMACHINE T7 ULTRA Kit (Life Technologies). T7 promoter and sgRNA sequence were synthesized as a long oligonucleotide (Ultramer, IDT) and amplified by PCR. The T7-sgRNA PCR product was gel-purified and used as the template for IVT using the MEGAshortscript T7 Kit (Life Technologies). A repair template encoding the C234W variant was synthesized as a single-stranded oligonucleotide (Ultramer, IDT) and used without purification. Potential off-targets were identified using Cas-OFFinder (Bae, Park, & Kim, 2014), selecting sites with fewest mismatches (http://www. rgenome.net/cas-offinder/). The founder mouse and all F1 mice were sequenced for off-targets. List of primers is in Table S1.

5.8 | Animal injection and analysis

All animal procedures were conducted with the approval of the Institutional Animal Care Committee at the University of California, San Diego (protocol number: S17114). All studies were performed on equal number of females and males. Number of animals nedeed for each experiment was estimated using power analysis.

5.9 | CRISP/Cas9 injection

C57BL/6N mouse zygotes were injected with CRISPR-Cas9 constructs. Oligos were injected into the cytoplasm of the zygotes at the pronuclei stage. Mice were housed on static racks in a conventional animal facility and were fed ad libitum with Teklad Global 2020X diet.

5.10 | Genotyping, mice substrains

To test for the potentially confounding Rd8 mutation, a mutation in the *Crb1* gene which can produce ocular disease phenotypes when homozygous, we sequenced all mice in our study for Rd8. C57BL/6J mice in the aging part of the study were purchased from the Jax Laboratory and confirmed to be negative for mutation in *Crb1* gene. All C234W mutant animals and their littermates were heterozygous for Rd8 mutation. To test RPE65 gene, all animals were tested for the presence of the variants. All animals in the study harbor homozygous RPE65 variant Leu/Leu.

5.11 | Intravitreal injections

For the 5-Aza-dc injection study, mice were anesthetized by intraperitoneal injection of ketamine/xylazine (100 mg/kg and 10 mg/

kg, respectively), and given an analgesic eye drop of proparacaine (0.5%, Bausch & Lomb). Animals were intraocularly injected with 1μl of PBS in one eye, and 1 μl of 2 μM 5-Aza-dc dissolved in PBS in the contralateral eye, every other week over a period of 3 months. Drug dosage was estimated based on our cell line experiments and on previously published data (Gore, 2018).

Autofluorescence imaging was performed using the Spectralis ® HRA + OCT scanning laser ophthalmoscope (Heidelberg Engineering) as previously described (16) using blue light fluorescence feature (laser at 488 nm, barrier filter at 500 nm). Using a 55-degree lens, projection images of 10 frames per fundus were taken after centering around the optic nerve. The image that was most in focus was on the outer retina was then quantified blindly by two independent individuals.

Electroretinograms (ERGs) were performed following a previously reported protocol (Luo, 2014). Briefly, mice were dark-adapted for 12 hr, anesthetized with a weight-based intraperitoneal injection of ketamine/xylazine, and given a dilating drop of tropicamide (1.5%, Alcon) as well as a drop of proparacaine (0.5%, Bausch & Lomb) as analgesic. Mice were examined with a full-field Ganzfeld bowl setup (Diagnosys LLC), with electrodes placed on each cornea, with a subcutaneous ground needle electrode placed in the tail, and a reference electrode in the mouth (Grass Telefactor, F-E2). Lubricant (Goniovisc 2.5%, HUB Pharmaceuticals) was used to provide contact of the electrodes with the eyes. Amplification (at 1–1,000 Hz bandpass, without notch filtering), stimuli presentation, and data acquisition are programmed and performed using the UTAS-E 3000 system (LKC Technologies). For scotopic ERG, the retina was stimulated with a xenon lamp at −2 and −0.5 log cd·s/m^2. For photopic ERG, mice were adapted to a background light of 1 log cd·s/m^2, and light stimulation was set at 1.5 log cd·s/m^2. Recordings were collected and averaged in manufacturer's software (Veris, EDI) and processed in Excel.

5.12 | Immunostaining

Eyeballs were collected immediately after sacrificing mice, fixed in 4% paraformaldehyde for 2 hr, and stored in PBS at 4°C. For immunostainings, eyeballs were sectioned, mounted on slides, and then incubated with 5% BSA 0.1% Triton-X PBS blocking solution for 1 hr. Primary antibodies (see Table S2 for antibodies used in the study) were added 1:50 in 5% BSA PBS and incubated at 4°C for 16 hr. Following 3× PBS wash, secondary antibodies were added 1:1,000 in 5% BSA PBS for 30 min at room temperature. Samples were then washed 3x with PBS, stained with DAPI for 5 min at room temperature, mounted, and imaged (Keyence BZ-X700).

5.13 | Lipid analysis

Lipid extraction was performed by homogenization of tissues in a mixture of 1 ml PBS, 1 ml MeOH, and 2 ml CHCl3. Mixtures were vortexed and then centrifuged at 2,200 g for 5 min to separate the aqueous and organic layer. The organic phase containing

the extracted lipids was collected and dried under N2 and stored at −80°C before LC-MS analysis. Extracted samples were dissolved in 100 μl CHCl3; 15 μl was injected for analysis. LC separation was achieved using a Bio-Bond 5U C4 column (Dikma). The LC solvents were as follows: buffer A, 95:5 water:methanol + 0.03% NH4OH; buffer B, 60:35:5 isopropanol:methanol: water + 0.03% NH4OH. A typical LC run consisted of the following for 70 min after injection: 0.1 ml/min 100% buffer A for 5 min, 0.4 ml/min linear gradient from 20% buffer B to 100% buffer B over 50 min, 0.5 ml/min 100% buffer B for 8 min and equilibration with 0.5 ml/min 100% buffer A for 7 min. FFA analysis was performed using a Thermo Scientific Q Exactive Plus fitted with a heated electrospray ionization source. The MS source parameters were 4kV spray voltage, with a probe temperature of 437.5°C and capillary temperature of 268.75°C. Full-scan MS data were collected with a resolution of 70k, AGC target 1x106, max injection time of 100 ms, and scan range 150–2000 m/z. Data-dependent MS (top 5 mode) was acquired with a resolution of 35 k, AGC target 1×105, max injection time of 50 ms, isolation window 1 m/z, scan range 200 to 2,000 m/z, and stepped normalized collision energy (NCE) of 20, 30, and 40. Extracted ion chromatograms for each FFA were generated using a m/z ± 0.01 mass window around the calculated exact mass (i.e., palmitic acid, calculated exact mass for M-H is 255.2330 and the extracted ion chromatogram was 255.22–255.24). Quantification of the FFAs was performed by measuring the area under the peak and is reported as relative units (R.U.).

5.14 | Analysis of ELOVL2 promoter DNA methylation in mice and humans

Reduced representation bisulfite sequencing (RRBS) in mouse blood was downloaded from Gene Expression Omnibus (GEO) using accession number GSE80672 (Petkovich, 2017). For each sample, reads obtained from sequencing were verified using FastQC (Andrews 2010), then trimmed 4bp using TrimGalore (Bioinformatics B 2016) (4bp), and aligned to a bisulfite-converted mouse genome (mm10, Ensembl) using Bismark (v0.14.3) (Krueger & Andrews, 2011), which produced alignments with Bowtie2 (v2.1.0) (Langmead, Trapnell, Pop, & Salzberg, 2009) with parameters "-score_min L,0,-0.2." Methylation values for CpG sites were determined using MethylDackel (v0.2.1).

To explore methylation of the promoter region of ELOLV2, we first designated the promoter as −1000bp to + 300bp with respect to the strand and transcription start site (TSS) and then identified profiled methylation CpGs using BEDtools (v2.25.0) (Quinlan & Hall, 2010). We then binned each profiled CpG in the promoter region according to 30-bp nonoverlapping windows considering CpGs with at least 5 reads. We then grouped the 136 C57BL/6 control mice according to five quantile age bins and took the average methylation for each age bin and each window. All analysis was performed using custom python (version 3.6) scripts, and plots were generated using matplotlib and seaborn.

To explore the homologous region in humans, we accessed human blood methylome data generated using the Human Illumina methylome array downloaded from GEO, using accessions GSE36054 (Alisch, 2012) and GSE40279 (Hannum, 2013) for a total of 736 samples. Methylation data were quantile-normalized using Minfi, (Aryee, 2014) and missing values were imputed using the Impute package in R. These values were adjusted for cell counts as previously described (Gross, 2016). To enable comparisons across different methylation array studies, we implemented beta-mixture quantile dilation (BMIQ; Gross, 2016; Teschendorff, 2013) and used the median of the Hannum *et al.* dataset as the gold standard (Hannum, 2013).

We then identified probes within the promoter region of ELOLV2 in the human reference (hg19, UCSC), identifying 6 total probes in the commonly profiled region. We then grouped the 787 individuals according to 5 quantile age bins and grouped probes into 10bp non-overlapping windows. These data were then analyzed and plotted identically as for mice.

CONFLICT OF INTEREST
DSK and DC are scientific cofounders of Visgenx.

AUTHOR CONTRIBUTIONS
D.S.K. designed the study with the contribution of D.C., D.L.C., and K.Z. D.C. performed most of the experiments. L.R. and V.A.N.H. performed animal experiments. D.L.C. and M.D. performed RNAscope analysis. D.L.C., M.Kolar, and A.S. performed lipidomic studies. M. Krawczyk, M. Jafari, and M. Jabari performed additional experiments in the study. M. Krawczyk performed a statistical analysis of the data. T.W. contributed a bioinformatic analysis of the methylation data. K.D.R. and B.A.H. proposed and constructed the Elovl2[C234W] mouse mutation. D.C., D.L.C., and D.S.K. wrote the manuscript with helpful edits from B.A.H.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in Dryad at https://doi.org/10.6075/J0TX3CQ9

ORCID
Dorota Skowronska-Krawczyk https://orcid.org/0000-0002-5758-4225

REFERENCES

Alisch, R. S., Barwick, B. G., Chopra, P., Myrick, L. K., Satten, G. A., Conneely, K. N., & Warren, S. T. (2012). Age-associated DNA methylation in pediatric populations. *Genome Research*, *22*, 623–632. https://doi.org/10.1101/gr.125187.111

Ambati, J., & Fowler, B. J. (2012). Mechanisms of age-related macular degeneration. *Neuron*, *75*, 26–39. https://doi.org/10.1016/j.neuron.2012.06.018

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, *30*, 1363–1369. https://doi.org/10.1093/bioinformatics/btu049

Bae, S., Park, J., & Kim, J. S. (2014). Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, *30*, 1473–1475. https://doi.org/10.1093/bioinformatics/btu048

Bazan, N. G., Molina, M. F., & Gordon, W. C. (2011). Docosahexaenoic acid signalolipidomics in nutrition: Significance in aging, neuroinflammation, macular degeneration, Alzheimer's, and other neurodegenerative diseases. *Annual Review of Nutrition*, *31*, 321–351. https://doi.org/10.1146/annurev.nutr.012809.104635

Bernstein, P. S., Tammur, J., Singh, N., Hutchinson, A., Dixon, M., Pappas, C. M., ... Allikmets, R. (2001). Diverse macular dystrophy phenotype caused by a novel complex mutation in the ELOVL4 gene. *Investigative Ophthalmology & Visual Science*, *42*, 3331–3336.

Bioinformatics B (2016). Trim Galore! Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

Birch, D. G., & Anderson, J. L. (1992). Standardized full-field electroretinography. Normal values and their variation with age. *Archives of Ophthalmology*, *110*, 1571–1576. https://doi.org/10.1001/archopht.1992.01080230071024

Cameron, D. J., Yang, Z., Gibbs, D., Chen, H., Kaminoh, Y., Jorgensen, A., ... Zhang, K. (2007). HTRA1 variant confers similar risks to geographic atrophy and neovascular age-related macular degeneration. *Cell Cycle*, *6*, 1122–1125. https://doi.org/10.4161/cc.6.9.4157

Chavali, V. R. M., Khan, N. W., Cukras, C. A., Bartsch, D.-U., Jablonski, M. M., & Ayyagari, R. (2011). A CTRP5 gene S163R mutation knock-in mouse model for late-onset retinal degeneration. *Human Molecular Genetics*, *20*, 2000–2014. https://doi.org/10.1093/hmg/ddr080

Choi, I. A., Lee, C. S., Kim, H. Y., Choi, D. H., & Lee, J. (2018). Effect of inhibition of DNA methylation combined with task-specific training on chronic stroke recovery. *International Journal of Molecular Sciences*, *19*(7), 2019. https://doi.org/10.3390/ijms19072019

Christman, J. K. (2002). 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: Mechanistic studies and their implications for cancer therapy. *Oncogene*, *21*, 5483–5495. https://doi.org/10.1038/sj.onc.1205699

Combadière, C., Feumi, C., Raoul, W., Keller, N., Rodéro, M., Pézard, A., ... Sennlaub, F. (2007). CX3CR1-dependent subretinal microglia cell accumulation is associated with cardinal features of age-related macular degeneration. *J Clin Invest*, *117*, 2920–2928. https://doi.org/10.1172/JCI31692

Concepcion, D., Ross, K. D., Hutt, K. R., Yeo, G. W., & Hamilton, B. A. (2015). Nxf1 natural variant E610G is a semi-dominant suppressor of IAP-induced RNA processing defects. *PLoS Genetics*, *11*, e1005123. https://doi.org/10.1371/journal.pgen.1005123

Crabb, J. W. (2014). The proteomics of drusen. *Cold Spring Harbor Perspectives in Medicine*, *4*, a017194. https://doi.org/10.1101/cshperspect.a017194

Edwards, A. O., Donoso, L. A., & Ritter, R. 3rd (2001). A novel gene for autosomal dominant Stargardt-like macular dystrophy with homology to the SUR4 protein family. *Investigative Ophthalmology & Visual Science*, *42*, 2652–2663.

Garagnani, P., Bacalini, M. G., Pirazzini, C., Gori, D., Giuliani, C., Mari, D., ... Franceschi, C. (2012). Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*, *11*, 1132–1134. https://doi.org/10.1111/acel.12005

Glei, D. A., Goldman, N., Risques, R. A., Rehkopf, D. H., Dow, W. H., Rosero-Bixby, L., & Weinstein, M. (2016). Predicting survival from telomere length versus conventional predictors: A multinational population-based cohort study. *PLoS ONE*, *11*, e0152486. https://doi.org/10.1371/journal.pone.0152486

Gopalan, S., Carja, O., Fagny, M., Patin, E., Myrick, J. W., McEwen, L. M., ... Henn, B. M. (2017). Trends in DNA methylation with age replicate across diverse human populations. *Genetics*, *206*, 1659–1674. https://doi.org/10.1534/genetics.116.195594

Gore, A. V., Tomins, K. A., Iben, J., Ma, L. I., Castranova, D., Davis, A. E., ... Weinstein, B. M. (2018). An epigenetic mechanism for cavefish eye degeneration. *Nature Ecology & Evolution*, *2*, 1155–1160. https://doi.org/10.1038/s41559-018-0569-4

Gregory, M. K., Cleland, L. G., & James, M. J. (2013). Molecular basis for differential elongation of omega-3 docosapentaenoic acid by the rat Elovl5 and Elovl2. *Journal of Lipid Research*, *54*, 2851–2857. https://doi.org/10.1194/jlr.M041368

Gross, A. M., Jaeger, P. A., Kreisberg, J. F., Licon, K., Jepsen, K. L., Khosroheidari, M., ... Ideker, T. (2016). Methylome-wide analysis of chronic HIV infection reveals five-year increase in biological age and epigenetic targeting of HLA. *Molecular Cell*, *62*, 157–168. https://doi.org/10.1016/j.molcel.2016.03.019

Hannum, G., Guinney, J., Zhao, L., Zhang, L. I., Hughes, G., Sadda, S. V., ... Zhang, K. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, *49*, 359–367. https://doi.org/10.1016/j.molcel.2012.10.016

Hayflick, L. (1965). The limited in vitro lifetime of human diploid cell strains. *Experimental Cell Research*, *37*, 614–636. https://doi.org/10.1016/0014-4827(65)90211-9

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*, R115. https://doi.org/10.1186/gb-2013-14-10-r115

Jones, P. L., Jan Veenstra, G. C., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., ... Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics*, *19*, 187–191. https://doi.org/10.1038/561

Kim, S.-Y., Yang, H.-J., Chang, Y.-S., Kim, J.-W., Brooks, M., Chew, E. Y., ... Swaroop, A. (2014). Deletion of aryl hydrocarbon receptor AHR in mice leads to subretinal accumulation of microglia and RPE atrophy. *Investigative Ophthalmology & Visual Science*, *55*, 6031–6040. https://doi.org/10.1167/iovs.14-15091

Kolesnikov, A. V., Fan, J., Crouch, R. K., & Kefalov, V. J. (2010). Age-related deterioration of rod vision in mice. *Journal of Neuroscience*, *30*, 11222–11231. https://doi.org/10.1523/JNEUROSCI.4239-09.2010

Krueger, F., & Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, *27*, 1571–1572. https://doi.org/10.1093/bioinformatics/btr167

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, *11*, 204–220. https://doi.org/10.1038/nrg2719

Leonard, A. E., Kelder, B., Bobik, E. G., Chuang, L.-T., Lewis, C. J., Kopchick, J. J., ... Huang, Y.-S. (2002). Identification and expression

of mammalian long-chain PUFA elongation enzymes. *Lipids*, *37*, 733–740. https://doi.org/10.1007/s11745-002-0955-6

Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Bandinelli, S., … Horvath, S. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*, *10*, 573–591. https://doi.org/10.18632/aging.101414

Li, C.-M., Clark, M. E., Chimento, M. F., & Curcio, C. A. (2006). Apolipoprotein localization in isolated drusen and retinal apolipoprotein gene expression. *Investigative Opthalmology & Visual Science*, *47*, 3119–3128. https://doi.org/10.1167/iovs.05-1446

Liu, A., Chang, J., Lin, Y., Shen, Z., & Bernstein, P. S. (2010). Long-chain and very long-chain polyunsaturated fatty acids in ocular aging and age-related macular degeneration. *Journal of Lipid Research*, *51*, 3217–3229. https://doi.org/10.1194/jlr.M007518

Luo, J., Baranov, P., Patel, S., Ouyang, H., Quach, J., Wu, F., … Zhang, K. (2014). Human retinal progenitor cell transplantation preserves vision. *Journal of Biological Chemistry*, *289*, 6362–6371. https://doi.org/10.1074/jbc.M113.513713

Miller, C. A., & Sweatt, J. D. (2007). Covalent modification of DNA regulates memory formation. *Neuron*, *53*, 857–869. https://doi.org/10.1016/j.neuron.2007.02.022

Momparler, R. L. (2005). Pharmacology of 5-Aza-2'-deoxycytidine (decitabine). *Seminars in Hematology*, *42*, S9–16. https://doi.org/10.1053/j.seminhematol.2005.05.002

Pauter, A. M., Olsson, P., Asadi, A., Herslöf, B., Csikasz, R. I., Zadravec, D., & Jacobsson, A. (2014). Elovl2 ablation demonstrates that systemic DHA is endogenously produced and is essential for lipid homeostasis in mice. *Journal of Lipid Research*, *55*, 718–728. https://doi.org/10.1194/jlr.M046151

Petkovich, D. A., Podolskiy, D. I., Lobanov, A. V., Lee, S.-G., Miller, R. A., & Gladyshev, V. N. (2017). Using DNA methylation profiling to evaluate biological age and longevity interventions. *Cell Metabolism*, *25*(4), 954–960.e6. https://doi.org/10.1016/j.cmet.2017.03.016

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Sangiovanni, J. P., Agrón, E., Meleth, A. D., Reed, G. F., Sperduto, R. D., Clemons, T. E., … Age-Related Eye Disease Study Research Group (2009). {omega}-3 Long-chain polyunsaturated fatty acid intake and 12-y incidence of neovascular age-related macular degeneration and central geographic atrophy: AREDS report 30, a prospective cohort study from the Age-Related Eye Disease Study. *American Journal of Clinical Nutrition*, *90*, 1601–1607. https://doi.org/10.3945/ajcn.2009.27594

Seddon, J. M., George, S., & Rosner, B. (2006). Cigarette smoking, fish consumption, omega-3 fatty acid intake, and associations with age-related macular degeneration: The US Twin Study of age-related macular degeneration. *Archives of Ophthalmology*, *124*, 995–1001. https://doi.org/10.1001/archopht.124.7.995

Shaw, P. X., Zhang, L., Zhang, M., Du, H., Zhao, L., Lee, C., … Zhang, K. (2012). Complement factor H genotypes impact risk of age-related macular degeneration by interaction with oxidized phospholipids. *Proceedings of the National Academy of Sciences USA*, *109*, 13757–13762. https://doi.org/10.1073/pnas.1121309109

Slieker, R. C., Relton, C. L., Gaunt, T. R., Slagboom, P. E., & Heijmans, B. T. (2018). Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. *Epigenetics Chromatin*, *11*, 25. https://doi.org/10.1186/s13072-018-0191-3

Stempel, A. J., Morgans, C. W., Stout, J. T., & Appukuttan, B. (2014). Simultaneous visualization and cell-specific confirmation of RNA and protein in the mouse retina. *Molecular Vision*, *20*, 1366–1373.

Telese, F., Gamliel, A., Skowronska-Krawczyk, D., Garcia-Bassets, I., & Rosenfeld, M. G. (2013). "Seq-ing" insights into the epigenetics of neuronal gene regulation. *Neuron*, *77*, 606–623. https://doi.org/10.1016/j.neuron.2013.01.034

Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., & Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*, *29*, 189–196. https://doi.org/10.1093/bioinformatics/bts680

van Leeuwen, E. M., Emri, E., Merle, B. M. J., Colijn, J. M., Kersten, E., Cougnard-Gregoire, A., … Lengyel, I. (2018). A new perspective on lipid research in age-related macular degeneration. *Progress in Retinal and Eye Research*, *67*, 56–86. https://doi.org/10.1016/j.preteyeres.2018.04.006

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, *153*, 910–918. https://doi.org/10.1016/j.cell.2013.04.025

Wang, Q., Xu, L., Chen, P., Xu, Z., Qiu, J., Ge, J., … Zhuang, J. (2018). Brca1 Is upregulated by 5-Aza-CdR and promotes DNA repair and cell survival, and inhibits neurite outgrowth in rat retinal neurons. *International Journal of Molecular Sciences*, *19*, https://doi.org/10.3390/ijms19041214

Wang, T., Tsui, B., Kreisberg, J. F., Robertson, N. A., Gross, A. M., Yu, M. K., … Ideker, T. (2017). Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biology*, *18*, 57. https://doi.org/10.1186/s13059-017-1186-2

Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L., & Schübeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genetics*, *37*, 853–862. https://doi.org/10.1038/ng1598

Williams, G. A., & Jacobs, G. H. (2007). Cone-based vision in the aging mouse. *Vision Research*, *47*, 2037–2046. https://doi.org/10.1016/j.visres.2007.03.023

Xie, W., Baylin, S. B., & Easwaran, H. (2019). DNA methylation in senescence, aging and cancer. *Oncoscience*, *6*, 291–293. https://doi.org/10.18632/oncoscience.476

Xu, H., Chen, M., Manivannan, A., Louis, N., & Forrester, J. V. (2008). Age-dependent accumulation of lipofuscin in perivascular and subretinal microglia in experimental mice. *Aging Cell*, *7*(1), 58–68.

Zadravec, D., Tvrdik, P., Guillou, H., Haslam, R., Kobayashi, T., Napier, J. A., … Jacobsson, A. (2011). ELOVL2 controls the level of n-6 28:5 and 30:5 fatty acids in testis, a prerequisite for male fertility and sperm maturation in mice. *Journal of Lipid Research*, *52*, 245–255. https://doi.org/10.1194/jlr.M011346

Zhang, K., Kniazeva, M., Han, M., Li, W., Yu, Z., Yang, Z., … Petrukhin, K. (2001). A 5-bp deletion in ELOVL4 is associated with two related forms of autosomal dominant macular dystrophy. *Nature Genetics*, *27*, 89–93. https://doi.org/10.1038/83817

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Off-targets	Sequence (5' -> 3')	Sequence (5' -> 3')
E2off-1	ATTGCCTTATTAGGAGGAAAC	GAGAGTGATGAGCTTAATTTG
E2off-2	GCTGGAAATTCACTGAAGAC	TTGGCAACACCAAAGAAC
E2off-3	TCTGGAGGTACTGGTTAG	GAACACCTGCAGTCATAG
E2off-4	GCCTCATTTACTACAGTAGTC	TCCATAGAGGAGTGAGAGTAAG
E2off-5	CTCCCAAGTGTCGGGATTA	CTACTTCCCCAGCCCTTATAG
E2off-6	CCAGCTATTGAGCGTGAAG	ACATTCCCTGAGTGCCTAC
E2off-7	TCTATGAGGGTGCTGAGTC	CACCCAGGATCTTCATATAGG
E2off-8	GACATTCTATTGGAGGGTTTAC	CTGCCTTGCTATATCTTTCTAC
E2off-9	CCAAAGAGCATCACTAAGG	TCGGTTATGTCTTCGACTG
E2off-10	GGAGGTCAGAAAGTCATTG	GAAGTCGATCACTGGAAAG
MeDIP primers		
hELOVL2 prom.	CGATTTGCAGGTCCAGCCG	CAGCGGGTGGGTATTCCTG
mELOVL2 prom.	AGCTCCTCCGCTACTC	CCAGCCCTTGGTCATC
qPCR primers		
mRhodopsin	ACCTGGATCATGGCGTTG	TCGTTGTTGACCTCAGGCTT
mOpsin	TGTACATGGTCAACAATCGGA	ACACCATCTCCAGAATGCAAG
mNrl	AGTCTCCAGGGAAGCTGTGC	TGGGACTGAGCAGAGAGAGG
mGAPDH	TCAACAGCAACTCCCACTCTTCCA	ACCCTGTTGCTGTAGCCGTATTCA
hELOVL2	GCGGATCATGGAACATCTAA	CCAGCCATATTGAGAGCAGA
hACTB	CACCATTGGCAATGAGCGGTTC	AGGTCTTTGCGGATGTCCACGT
mRpe65	ACCAGAAATTTGGAGGGAAAC	CCCTTCCATTCAGAGCTTCA
mRd8	GGTGACCAATCTGTTGACAATCC	GCCCCATTTGCACACTGATGAC
Megamer	CAAAGCCCTTTCAAGCCAAGTCACGATTCTAATTTTAATTTGGCTAAGGAACACAGAAATAGGCCCAAACTCAAGTTGAACAAGGGCTTCTGACACTTTGATAACTTCGTATAGCATACATTATACGAAGTTATGATAGGCAGAGCATCAGTGCGAGAACAACGTTCTCAGGTTCTGGCAGGGTAGAGAAGCAGTAACTGCAACCTTGCACAGCTCTGGAAGGTCCACCTAACCAGTTCTCAAACTCAGATGCGTGCTGAGTGAAGCAAGCCAGAGTGCAGGCCACAGCCACTTACCTCCACCAGCATATACGCAGAAAGAAGTGTGATTGCGAGGTTATACAAGGTGAGGATGCCCCTGAGAGACAGAGCAGGCCTGTTCTTCATGTACTTGTTACCCAGCCATATCGAGAGCAGGTACGTGATGGTGAGGATGAAGGTGGGAAGGTAAGAGTCCAGCAGGAACCACCCGCGAACTCGAGAATCTGTAAAGAAATGCTTACGGTGAGGAGCCCAAGGAGGGATGTCCCTGTAATTAGCAACTCTAACACACATACTCATCTGTGCAACGGGCGTCCCCATCGATGGTGCACGCCCCACGATAACTTCGTATAGCATACATTATACGAAGTTATGTCTCCGTGAACACAAGCCCCGTCCTCACCGAGTACTCGGTGTTTGAGATGAACTCACATAAACAGGTCAAGTTGCTTTTCCCTGAACAAATTCTCAGAC

Table S1

Immunostaining	Company, Cat#	RRID
TEPC 15	Sigma M1421	AB_1163630
HtrA	Santa Cruz sc-377050	AB_2813838
C3	Santa Cruz sc-58926	AB_1119819
C5-b9	Santa Cruz sc-66190	AB_1119840
ApoE	Santa Cruz sc-13521	AB_626691
MeDIP		
5-methylcytosine	Millipore MABE146	AB_10863148
Western blot		
ELOVL2	Santa Cruz sc-54874	AB_2262364
Histone H3	Cell Signaling 9715	AB_331563

Table S2

Figure S1.

Figure S1 continued.

F

ELOVL2 mRNA in mouse liver

ELOVL2p methylation in mouse liver

G

H

scotopic responses

120

A WI-38 growth

B Senescence in aged WI-38 cells

C *ELOVL2* mRNA in IMR90

D IMR90 growth

E Senescence in IMR90

F

G *Rhodopsin* | *Arrestin-3* | *Nrl*

Figure S2.

121
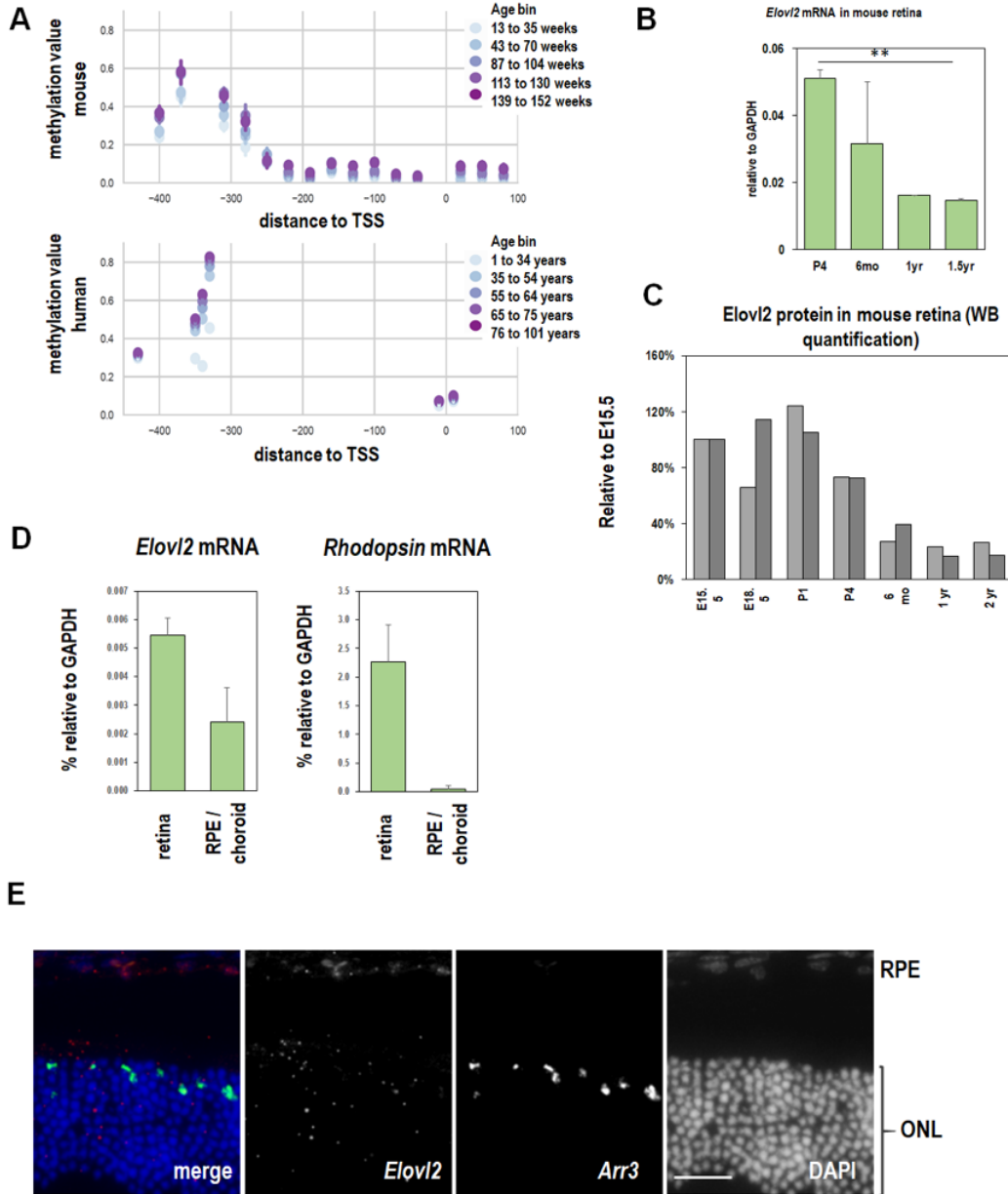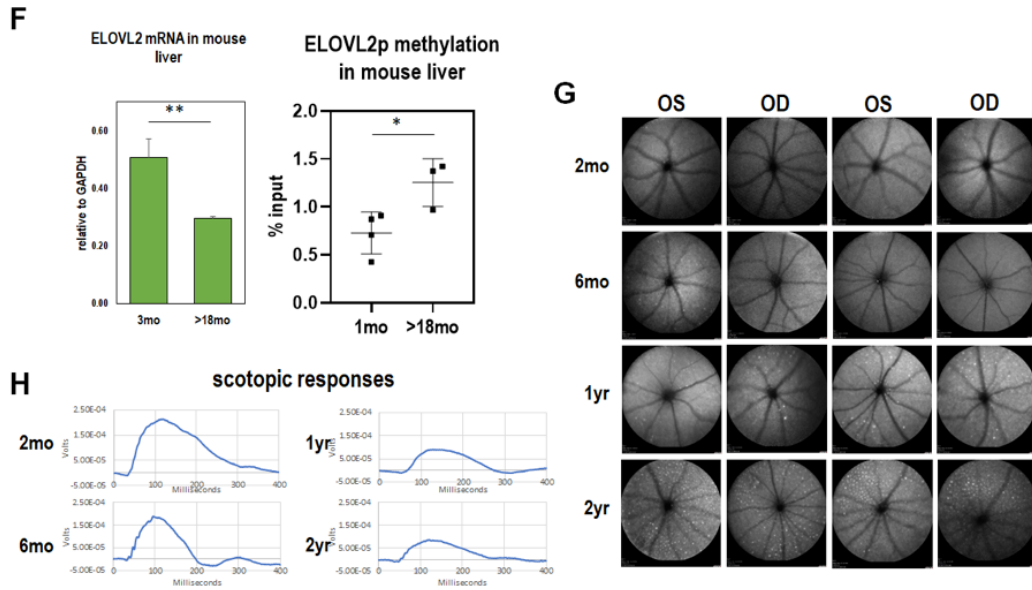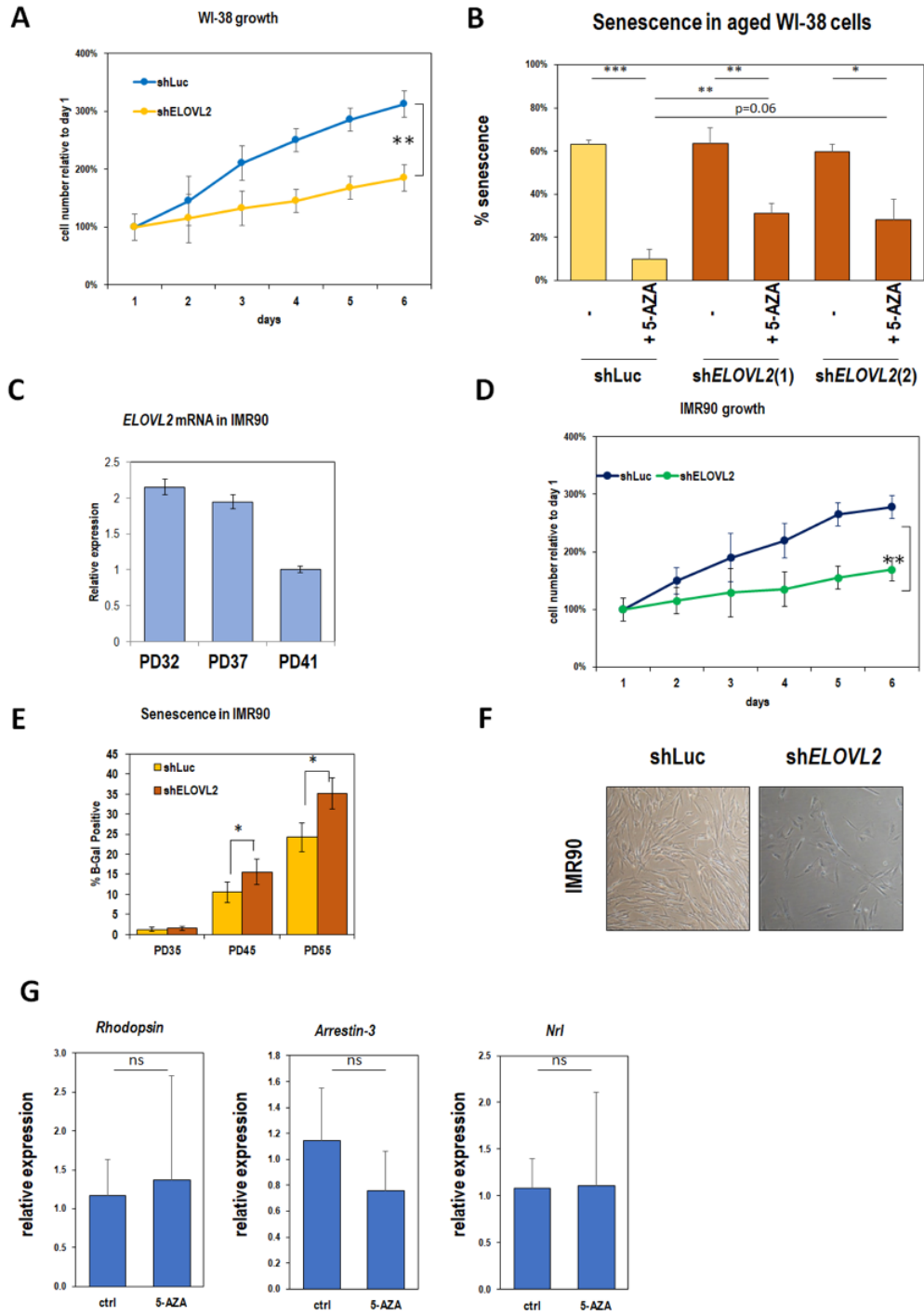
Figure S3

A


```
ELOVL5[Human]  vLMYSYYGLSsvPSMrpYLWWKKYITQgQLlQFVLTIIqTsCgVIwPCtFPlGwLyFQiG
Elovl5[Mouse]  vLMYSYYGLSsiPSMrpYLWWKKYITQgQLVQFVLTIIqTtCgVfwPCsFPlGwLfFQiG
ELOVL2[Human]  ILMYSYYGLSvFPSMHKYLWWKKYlTQAQLVQFVLTItHTmsAVVKPCGFPFGCLIFQss
Elovl2[Mouse]  ILMYSYYGLSvFPSMHKYLWWKKYlTQAQLVQFVLTItHTlsAVVKPCGFPFGCLIFQss
```

B

Type of target	Coordinates	strand	MM	target_seq (MM)	PAM	distance		gene name
specific target	chr13:41186882-41186904	-	0	GTCTTCCT [ATATGATGACGC]	TGG	0	E	Elovl2
off-target #1	chr5:99464327-99464349	+	4	GTAGTCAG[ATATGATGACGC]	TGG	44040	I	Gm43253
off-target #2	chr2:123737444-123737466	-	4	ATCCTTCT[ATTTGATGACGC]	AGG	NA	-	NA
off-target #3	chr13:96284276-96284298	+	4	GTTTTTCT[GTCTGATGACGC]	AGG	72813	-	Gm25213
off-target #4	chr12:51659250-51659272	+	4	GGGTTTCT[ATATGATAACGC]	TGG	1879	I	Stm3
off-target #5	chr2:80525660-80525682	+	4	GATCTCCT[ATATGATGGCGC]	TGG	0	E	Nckap1
off-target #6	chr1:179559339-179559361	+	4	GTCTTGCA[CTATGATGTCGC]	AGG	12723	I	Cnst
off-target #7	chr17:68269790-68269812	+	4	GAGTTCCT[CTATGATGATGC]	TGG	3985	-	L3mbtl4
off-target #8	chr8:17617658-17617680	+	3	GTCTTCAT[CTATGATGAAGC]	TGG	82072	-	Csmd1
off-target #9	chr6:137543559-137543581	+	4	GTCTTGTT[ACATGATGAAGC]	TGG	4158	I	Eps8
off-target #10	chr11:101141681-101141703	-	4	ATCTACCT[ATATCATGAGGC]	AGG	1172	-	Gm27626

★ ★ introduced by HDR oligo

122
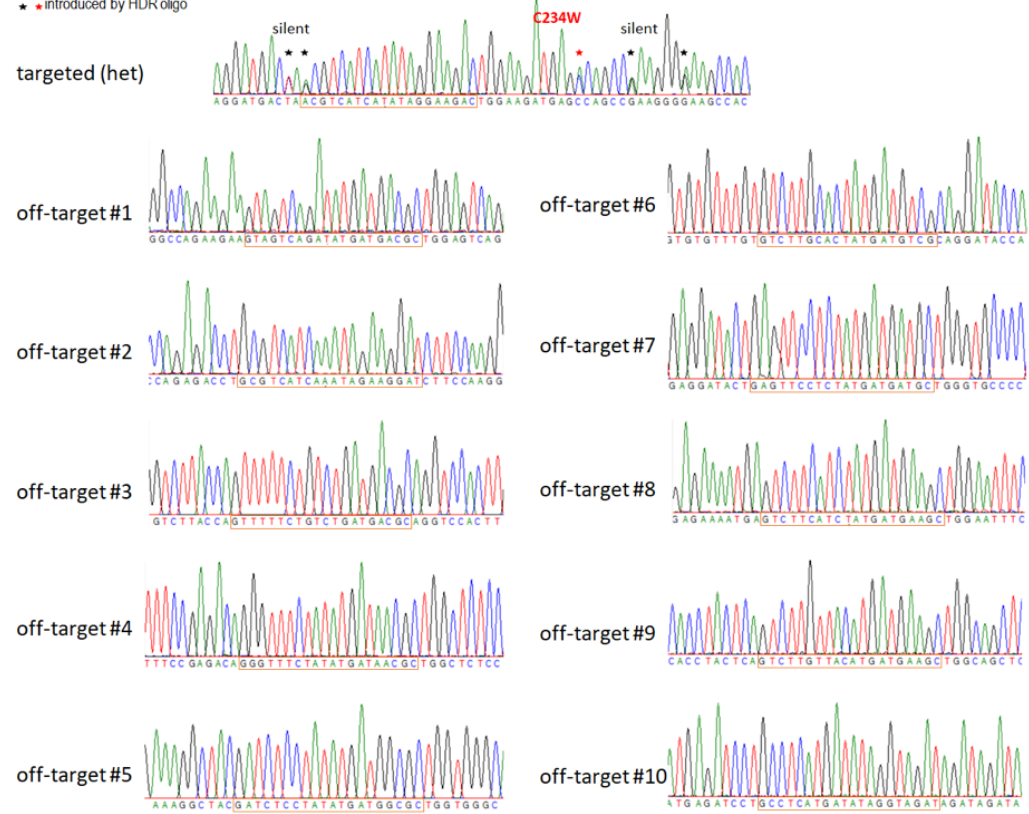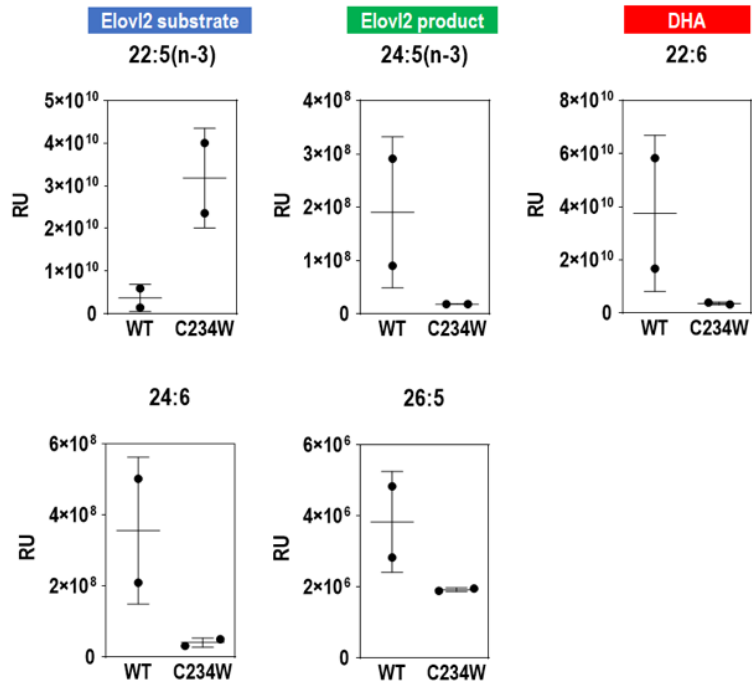
Figure S4

Figure S5
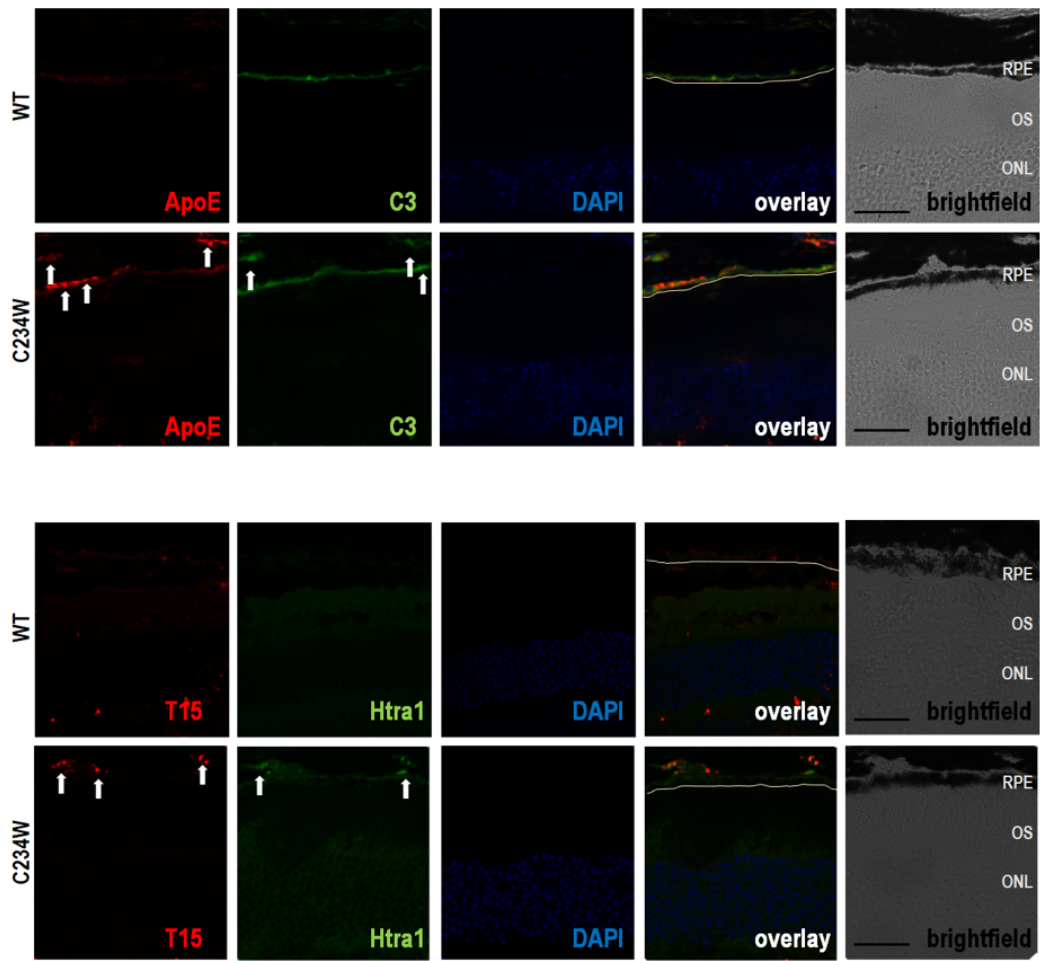
124

Figure S6

Article

Deleting an *Nr4a1* Super-Enhancer Subdomain Ablates Ly6C^{low} Monocytes while Preserving Macrophage Gene Function

Graphical Abstract
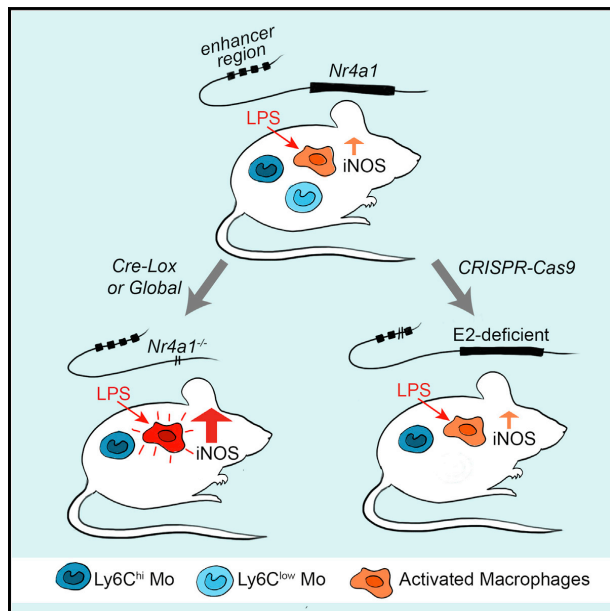
Highlights

- The *Nr4a1* enhancer domain regulates Ly6C^{low} monocyte development

- E2 domain-deficient resident tissue macrophages show wild-type *Nr4a1* expression

- E2 domain-deficient macrophages respond normally to LPS stimulation

- KLF2 acts via E2 to regulate *Nr4a1* in Ly6C^{low} monocyte development

Authors

Graham D. Thomas, Richard N. Hanna, Neelakatan T. Vasudevan, ..., Mukesh K. Jain, Christopher K. Glass, Catherine C. Hedrick

Correspondence

hedrick@lji.org

In Brief

Identifying tissue-specific functions for myeloid cells has been hindered by a lack of approaches to disrupt gene expression within individual subsets. Thomas et al. (2016) show that deleting an individual enhancer ablates *Nr4a1*-dependent Ly6C^{low} monocytes while retaining both *Nr4a1* expression in tissue macrophages and normal macrophage responses to inflammatory stimuli.

Accession Numbers

GSE80031
GSE80036
GSE80040

CellPress

Article

Deleting an *Nr4a1* Super-Enhancer Subdomain Ablates Ly6C[low] Monocytes while Preserving Macrophage Gene Function

Graham D. Thomas,[1] Richard N. Hanna,[1] Neelakatan T. Vasudevan,[2] Anouk A. Hamers,[1] Casey E. Romanoski,[3] Sara McArdle,[1] Kevin D. Ross,[4] Amy Blatchley,[1] Deborah Yoakum,[1] Bruce A. Hamilton,[4] Zbigniew Mikulski,[1] Mukesh K. Jain,[2] Christopher K. Glass,[4] and Catherine C. Hedrick[1],[5],*

[1]Division of Inflammation Biology, La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037, USA
[2]Case Cardiovascular Research Institute, Department of Medicine, Harrington Heart and Vascular Institute, University Hospitals Case Medical Center, Cleveland, OH 44106, USA
[3]Keating Biosearch Building, 1657 E. Helen St, Tucson, AZ 85721, USA
[4]Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[5]Lead Contact
*Correspondence: hedrick@lji.org
http://dx.doi.org/10.1016/j.immuni.2016.10.011

SUMMARY

Mononuclear phagocytes are a heterogeneous family that occupy all tissues and assume numerous roles to support tissue function and systemic homeostasis. Our ability to dissect the roles of individual subsets is limited by a lack of technologies that ablate gene function within specific mononuclear phagocyte sub-populations. Using *Nr4a1*-dependent Ly6C[low] monocytes, we present a proof-of-principle approach that addresses these limitations. Combining ChIP-seq and molecular approaches we identified a single, conserved, sub-domain within the *Nr4a1* enhancer that was essential for Ly6C[low] monocyte development. Mice lacking this enhancer lacked Ly6C[low] monocytes but retained *Nr4a1* gene expression in macrophages during steady state and in response to LPS. Because *Nr4a1* regulates inflammatory gene expression and differentiation of Ly6C[low] monocytes, decoupling these processes allows Ly6C[low] monocytes to be studied independently.

INTRODUCTION

Mononuclear phagocytes (MP) are an ontologically diverse family of cells composed of macrophages and monocytes (Perdiguero and Geissmann, 2016). From a traditional standpoint, the primary functions of MP involve recognition and clearance of invading pathogens, maintaining tissue integrity, and resolving inflammation (Sica and Mantovani, 2012). To perform these diverse functions, MP occupy all bodily tissues and display a tremendous degree of phenotypic plasticity. In recent years, transcriptomic profiling efforts have revealed the full extent of this heterogeneity, and functional studies have unveiled diverse, specialized, and often unexpected roles for individual MP subsets (Amit et al., 2016). In this context, tissue macrophages and blood monocytes may be considered accessory cells that allow optimal perfor-

mance of the host tissue (Okabe and Medzhitov, 2016). In complement to their roles in immunity and host defense, tissue MP are now also recognized as major regulators of higher-order physiological processes including systemic energy balance (Odegaard and Chawla, 2011), intestinal peristalsis (Muller et al., 2014), and cognitive function (Parkhurst et al., 2013). Delineating the full extent of MP functions in health and disease represents a largely unexplored frontier in both immunology and physiology.

Monocytes constitute the blood-borne phase of the MP system and are composed of at least two subsets in both mouse and human that are believed to be conserved (Cros et al., 2010). Mouse monocyte subsets can be discriminated using the Ly6C surface antigen (Geissmann et al., 2010), Ly6C[hi] monocytes are progenitors of inflammatory, and some tissue-resident macrophages. Ly6C[low] monocytes exhibit a patrolling behavior on the vascular endothelium and contribute to vascular homeostasis by maintaining the endothelial layer (Carlin et al., 2013). In addition, Ly6C[low] monocytes have recently been shown to play a crucial role protecting against the seeding of tumor metastases in the lung (Hanna et al., 2015).

An indispensable approach to understand cellular behavior is through loss of function. The identification of lineage defining transcription factors (LDTFs) for MP subsets has led to new insights into the functions of these cells. For example, the transcription factor *Nr4a1* is the master regulator of the Ly6C[low] monocyte subset (Hanna et al., 2011). *Nr4a1* is uniquely highly expressed in Ly6C[low] monocytes and is required in a cell-intrinsic fashion. Accordingly, *Nr4a1*[−/−] mice were instrumental in revealing a role for Ly6C[low] monocytes in tumor metastasis (Hanna et al., 2015). In another example, loss of the transcription factor *Spic* selectively ablates splenic red pulp macrophages revealing a role for this population in iron homeostasis (Kohyama et al., 2009). Similarly macrophage-specific deletion of *Gata6* impairs the maturation of peritoneal macrophages leading to the discovery that these cells modulate B-1 cell IgA production (Okabe and Medzhitov, 2014).

Pleiotropy, exemplified by the differential action of a single gene in multiple cell types, is a major obstacle to understanding cell-specific gene function. To overcome this problem, the Cre recombinase-*loxP* (Cre-Lox) system is routinely used to ablate

genes containing *loxP*-flanked exons by controlling Cre enzyme expression with cell-specific promoters. The ability of the Cre-Lox system to excise gene expression in a cell-specific manner is limited by the cell specificity of Cre transgenes and the time taken for recombination to occur (Yona et al., 2012). These considerations present problems when using the Cre-Lox system to study gene function within populations of closely related cells. A prime example of this problem is *Nr4a1*. *Nr4a1*$^{-/-}$ mice lack Ly6C^{low} monocytes, macrophage *Nr4a1* is also induced by LPS and represses inflammatory gene expression (Hanna et al., 2012). These confounding influences limit the utility of *Nr4a1*$^{-/-}$ to study Ly6C^{low} monocytes. Furthermore, the current MP Cre transgenes (*Lyz2-cre*, *Csf1r-cre*, *Cx3cr1-cre*) cannot delete Ly6C^{low} monocyte *Nr4a1* without also disrupting *Nr4a1* across MP subsets. These pan-myeloid effects of the current myeloid Cre transgenes limit the ability of conditional deletion approaches to determine gene- and cell-type function within the diverse MP compartment.

Enhancers are genomic sequences that positively influence promoter activity and are critical determinants of signal-dependent and cell-specific gene expression (Andersson et al., 2014). While such elements are rigorously defined by deletion or mutation, they are associated with a distinct chromatin signature that includes enrichment of histone H3 lysine 4 monomethylation (H3K4me1) and (H3K4me2) (Heintzman et al., 2007; Heinz et al., 2010). MP enhancers are enriched for the LDTFs PU.1 and C/EBPβ, which instruct enhancer selection (Heinz et al., 2010). Enhancers are subject to additional regulation leading to their further classification as "poised" or "active." Enhancers are activated upon binding of signal dependent transcription factors (SDTFs), leading to acetylation at H3 lysine 27 (H3K27ac) and the increased expression of associated genes (Creyghton et al., 2010; Heinz et al., 2013). These chromatin features can be used to map enhancer-like regions on a genome-wide scale through the use of ChIP-sequencing. The enhancer landscapes between MP subsets show considerable diversity (Gosselin et al., 2014; Lavin et al., 2014) and result from the myriad environmental niches these populations are exposed to. Because *Nr4a1* shows unique expression characteristics in Ly6C^{low} monocytes, we hypothesized a SDTF acting at a cell-specific enhancer regulates transcription of the Ly6C^{low} monocyte LDTF *Nr4a1*.

Here we explore this hypothesis by mapping the *Nr4a1* enhancer locus in Ly6C^{low} monocytes. We identified a single sub-domain 4 kb upstream of the *Nr4a1* transcription start site that was essential for Ly6C^{low} monocyte development. Dissection of this element provided insight into the transcriptional processes driving Ly6C^{low} monocyte development. Furthermore, by using mice that lack this enhancer, we have shown that macrophage *Nr4a1* gene expression was unaffected both in response to inflammatory signaling and during steady state.

RESULTS

Ly6C^{low} Monocytes Are Ontological Neighbors to Ly6C^{hi} Monocytes

Lineage tracing and BrdU pulse-chase studies have established a consensus that Ly6C^{hi} monocytes give rise to Ly6C^{low} monocytes (Yona et al., 2012). Yet it remains possible that Ly6C^{low} monocytes arise independently of the Ly6C^{hi} population. The

monocyte dendritic cell precursor (MDP) that expresses *Flt3* (CD135), *Cx3cr1*, and *Kit* (CD117) gives rise to all mouse monocytes (Hettinger et al., 2013). MDPs undergo further restriction toward the monocyte lineage upon their differentiation into the common monocyte progenitor (cMoP), at which point the expression of FLT3 is lost and Ly6C is gained (Hettinger et al., 2013).

We reasoned that if Ly6C^{low} monocytes arise independently of the Ly6C^{hi} population that we would observe a signature that was common to the Ly6C^{low} monocyte and their progenitor, but not the Ly6C^{hi} monocyte. To test this hypothesis, we performed ChIP-seq on MDP, cMoP, Ly6C^{hi}, and Ly6C^{low} monocytes (Figure 1A) using H3K4me2 to define enhancers, and activity with H3K27ac. To assess data quality, we visually inspected loci associated with prototypical marker genes for the profiled cell types (Figure 1B). As expected, MDPs showed enrichment for H3K27ac at the *Flt3* locus, whereas both MDPs and cMoPs showed H3K27ac at the *Kit* locus, both used to sort these populations. In contrast to Ly6C^{low} monocytes, Ly6C^{hi} monocytes and cMoPs show enhancer activity at the *Ly6c2* locus encoding for the Ly6C surface antigen. MDPs also showed activity at the *Ly6c2* locus in spite of low mRNA and cell-surface protein expression (Figures 1A and 1B), perhaps indicating a priming of this locus prior to transcription. Finally, in Ly6C^{low} monocytes *Cx3cr1* shows marked H3K27ac reflecting the robust Cx3cr1 expression in this population (Carlin et al., 2013).

To gain a global overview of enhancer dynamics during monocyte development, we quantitated H3K4me2 and H3K27ac at all 99,462 enhancers defined as H3K4me2-enriched regions greater than 2.5 kb from the nearest transcription start site. Differential enrichment (DE) of histone marks was then computed between all pairwise comparisons identifying a total of 620 DE H3K4me2 regions (0.62% of total, Figure S1A) and 9,879 H3K27ac regions (9.93% of total, Figure S1B). Unbiased hierarchical clustering of DE enhancers (Figure 1C, S1C) showed that these changes in H3K27ac are mirrored by more subtle shifts in H3K4me2. De novo motif analysis within these defined clusters identified transcriptional processes driving monocyte differentiation (Figure S1D). All enhancer clusters were enriched for ETS (PU.1) motifs. Progenitor-associated enhancers showed restricted over-representation of Myb motifs, which were lost upon Ly6C^{hi} monocyte differentiation, concomitant with the acquisition of CEBP and KLF motifs. At the global acetylation level, Ly6C^{low} monocytes enhancers were uniquely characterized by over-represented NR4A and Mef2a motifs. Of note, we did not observe a signature of enhancer or motif usage that was shared between Ly6C^{low} monocytes and either progenitor but not Ly6C^{hi} monocytes, nor did such a pattern emerge at the transcriptomic level (Figure S2). Instead, the patterns of differential enhancer usage were consistent with a continuous transition between the MDP and Ly6C^{low} monocytes with both the cMoP and Ly6C^{hi} monocytes falling as intermediaries in this cascade. Collectively, these data strongly support the consensus model of monocyte development in which Ly6C^{low} monocytes derive directly from the Ly6C^{hi} monocyte population.

Ly6C^{low} Monocytes Possess a Cell-Specific Super-Enhancer at the *Nr4a1* Locus

Super-enhancers (SE) are a recently defined class of highly cell-type-specific enhancer (Whyte et al., 2013) that selectively mark
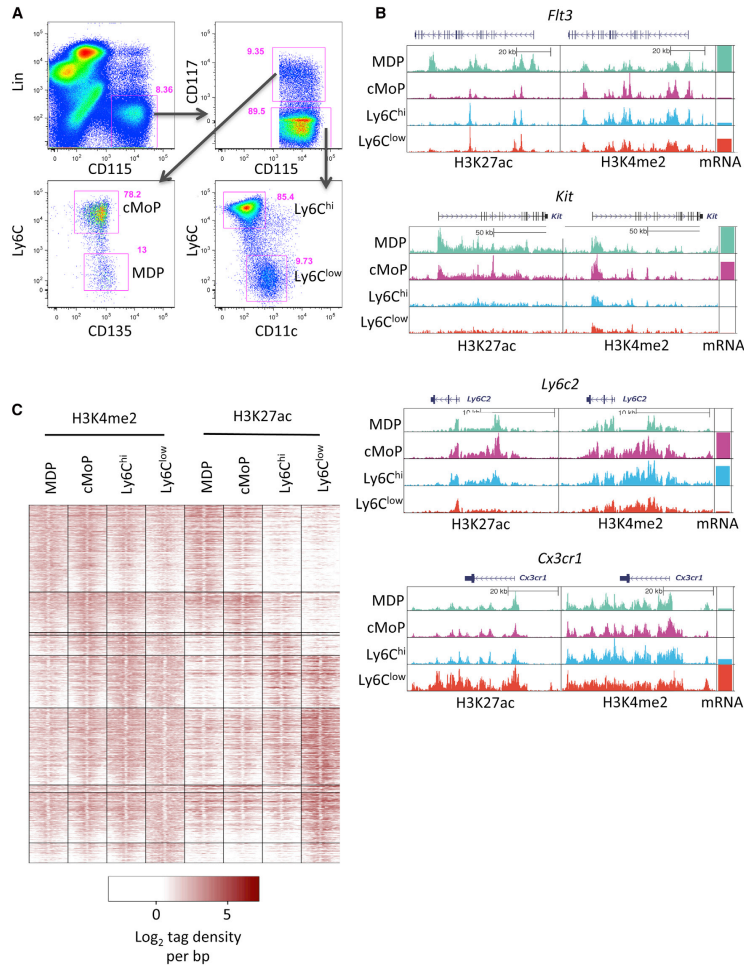
128

Figure 1. Epigenomic Profiling of Mo Subsets and Progenitors Supports the Model of Ly6C[hi] to Ly6C[low] Mo Conversion
(A) Gating strategy used to sort monocytes and upstream progenitors. Cells were previously gated on live singlets (using a FSC-W versus FSC-A gate).
(B) UCSC genome browser screenshots showing H3K27ac and H3K4me2 tag distributions at key lineage genes.
(C) Distribution of H3K4me2 and H3K27ac tags ± 1kb from PU.1 peak centers in DE enhancers. Please see Figures S1 and S2.

genes involved in lineage specification and function. The defining features of SEs include their extended size, adornment with LDTFs and genomic correlates of high gene expression (Hnisz et al., 2013). Because Ly6C[low] monocytes highly express *Nr4a1* and require it for their development (Hanna et al., 2011), we postulated that it is a SE-regulated gene. We used

A

Ly6C^low Mo
- PU.1
- Cebpb

H3K27ac
- Ly6C^low
- Ly6C^hi

H3K4me2
- Ly6C^low
- Ly6C^hi

B H3K4me2

C H3K27ac

E H3K27ac

D Lavin / Thomas

fixed enhancer signature at *Nr4a1se*, or if Ly6C^low monocytes possessed a unique H3K27ac profile. We visualized H3K27ac at *Nr4a1se* in MP populations corresponding to those in Figure 2D (Figure 2E). We observed comparable H3K27ac tag distributions between Ly6C^hi monocyte populations, facilitating a qualitative comparison between datasets. Each population was found to differ qualitatively in both the enrichment of histone acetylation and, to varying degrees, also in the histone acetylation "fingerprint" at *Nr4a1se* (Figure 2E). Notably, the most intense H3K27ac signature at *Nr4a1se* was observed in Ly6C^low monocytes (Figures 2D and 2E). Taken together, these findings imply that distinct mechanisms might regulate steady-state *Nr4a1* expression between MP populations.

H3K27ac to define SEs in Ly6C^low monocytes, confirming that *Nr4a1* does overlap a SE (*Nr4a1se;* Figure 2A and S2). Validating the SE characteristics of *Nr4a1se*, we also observed a high density of the MP LDTFs PU.1 and CEBPβ (Figure S3). Enumeration of H3K4me2 and H3K27ac tag counts at *Nr4a1se* in MDP, cMoP, Ly6C^hi, and Ly6C^low monocytes revealed a selective and robust (5-fold) acquisition of H3K27ac in Ly6C^low monocytes, demonstrating that *Nr4a1se* is a Ly6C^low monocyte-specific SE region (Figures 2B and 2C).

The formation of SEs represents a confluence of developmental and environmental cues (Hnisz et al., 2015). We leveraged the phenotypic heterogeneity of the diverse MP family to address whether steady-state *Nr4a1* expression might be regulated by distinct mechanisms. First we assessed *Nr4a1* expression in MP populations taken from Lavin et al. (2014) and our own monocyte subset data. We observed steady-state expression of *Nr4a1* spanning three orders of magnitude (Figure 2D). Importantly *Nr4a1* expression in Ly6C^hi monocytes from both datasets (labeled Ly6C^low and mononocytes) were consistent, facilitating the comparison between datasets. Consistent with previous observations, we also observed by far the highest *Nr4a1* expression in Ly6C^low monocytes (Hanna et al., 2011).

Next we questioned whether the range of *Nr4a1* mRNA expression was associated with the continuous acquisition of a

Nr4a1se Domain E2 Is a Conserved SE Sub-domain Essential for Ly6C^low Monocyte Development

We sought to identify regions of *Nr4a1se* that regulate *Nr4a1* gene expression in Ly6C^low monocytes. On the basis of PU.1 and C/EBPβ binding in Ly6C^low monocytes (Figure S4A), we cloned 12 candidate sequences into luciferase reporters containing a minimal *Nr4a1* promoter (Figure S4B). *Nr4a1se* sub-domains E2, E6, E8, and E9 induced modest but consistent luciferase activity in the myeloid RAW264.7 cell line, indicating that these regions might regulate *Nr4a1* in Ly6C^low monocytes (Figure 3A). It has been suggested that human CD14^dimCD16^hi monocytes are homologous to mouse Ly6C^low monocytes. This notion is based on global gene-expression profiling and observation of the characteristic patrolling behavior in CD14^dimCD16^hi monocytes (Cros et al., 2010). We reasoned that if CD14^dimCD16^hi and Ly6C^low monocytes are truly orthologous, then *Nr4a1* should be regulated by a conserved mechanism. To test this hypothesis, we assessed the genetic regulatory state of human DNA sequences orthologous to mouse E2, E6, E8, and E9 in human monocytes using publicly available datasets (Boyle et al., 2008; Schmidl et al.,

each containing a deletion of enhancer sequence to give E2 domain-deficient, E6 domain-deficient, and E9 domain-deficient mice, respectively (Figure 3C, S4C). Flow cytometry analysis of peripheral monocytes revealed a reduction in monocyte frequencies in E2 domain-deficient, no difference in E9 domain-deficient, and a modest increase in E6 domain-deficient mice (Figure 3D). WT-like monocyte subset ratios were preserved in both E6 domain-deficient and E9 domain-deficient mice, however a striking deficit in Ly6C^{low} monocytes phenocopying the *Nr4a1$^{-/-}$* strain was present in E2 domain-deficient mice (Figures 3E and 3F) (Hanna et al., 2011). Therefore the conserved super-enhancer sub-domain E2 is indispensable for Ly6C^{low} monocyte development.

Deletion of Enhancer Domain E2 Decouples Inflammation-Associated *Nr4a1* Gene Expression from Ly6C^{low} Monocyte-Associated *Nr4a1* Expression

Current genetic models to study MP *Nr4a1* are based on global *Nr4a1$^{-/-}$* mice or Cre recombinase-mediated deletion of *Nr4a1$^{flox/flox}$* alleles using myeloid specific Cre transgenes, which include, but are not restricted to, *Csf1r, Lyz2* (LysM), and *Cx3cr1*. Although these tools ablate MP *Nr4a1*, they are non-specific with respect to individual MP subsets (Chow et al., 2011; Yona et al., 2012). Consequently, in models that lack Ly6C^{low} monocytes, *Nr4a1* is also disrupted in macrophages (Figure 4A). Because

2014). Human monocyte DNase-Seq clearly shows open chromatin at orthologous regions to mouse E2, E6, and E9 (Figure 3B); however, no ortholog for E8 was identified by nucleotide sequence alignment using BLAT. Human E2, E6, and E9 also possessed H3K27ac, and H3K27ac enrichment was greater at E2 and E6 in CD14^{dim}CD16^{hi} monocytes than CD14^{hi}CD16^{neg} monocytes, consistent with the pattern observed between mouse monocyte subsets (Figure 3B). This provides evidence that E2, E6, and E9 are functionally conserved enhancer elements between species, supporting the notion that these regions are important regulators of MP *Nr4a1* gene expression.

To test the in vivo functions of E2, E6, and E9, we used the CRISPR-*Cas9* system to generate three mouse strains,

A

B

C

D

E

F

G

Ly6C^low^ monocytes are absent in E2 domain-deficient mice and *Nr4a1* is a key suppressor of macrophage inflammatory gene expression, we questioned whether E2 domain-deficient macrophages possess wild-type-like responses to inflammatory stimuli.

During endotoxic shock, macrophage-mediated production of inflammatory cytokines ultimately precipitates organ failure and mortality (Jacob et al., 2007). Because *Nr4a1^−/−^* mice are more sensitive to LPS-induced organ failure (Li et al., 2015), we decided to test the role of E2 in this setting. We administered a single high dose of LPS (2.5 mg/kg) to E2 domain-deficient, *Nr4a1^−/−^*, and WT mice by IP injection. All three groups displayed disheveled fur and shivering within the first 72 hr; however, *Nr4a1^−/−^* mice showed a significantly higher mortality than both WT and E2 domain-deficient groups, whose sympt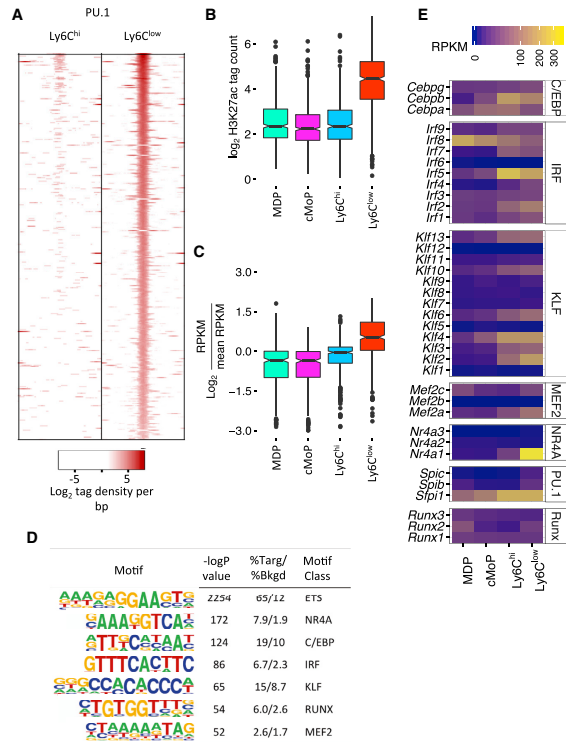oms resolved in the 72–120 hr time window (Figure 4B). To understand the role of E2 in the regulation of *Nr4a1* gene expression, we measured *Nr4a1* responses to LPS stimulation in thioglycollate-elicited macrophages. WT, E2 domain-heterozygous, and E2 domain-deficient macrophages all showed the well-characterized peak of *Nr4a1* mRNA at 1 hr followed by a return to baseline by 3 hr (Figure 4C, Pei et al., 2005). Immunoblot analysis also confirmed protein induction in WT and E2 domain-deficient macrophages at 1 hr following stimulation (Figure 4D). Furthermore, in a dose-response setting (Figure 4E) we observed no differences in *Nr4a1* mRNA expression between WT, E2 domain-heterozygous, and E2 domain-deficient macrophages. Finally, following LPS challenge we detected higher expression of *Il12*, *Il1b*, and *Nos2* mRNA and iNOS activity in *Nr4a1^−/−^* mice relative to E2 domain-deficient mice, which were comparable to wild-type controls (Figures 4F and 4G). These studies confirm that the E2 region does not regulate *Nr4a1* expression in response to TLR4 stimulation.

Differential PU.1 Binding Reveals Candidate Regulators of Ly6C^low^ Monocyte Gene Expression

We sought to determine the molecular mechanisms regulating *Nr4a1* expression in Ly6C^low^ monocytes. Cooperative interactions between PU.1 and secondary co-factors establish and maintain macrophage enhancer repertoires (Heinz et al., 2010, 2013). Furthermore, differences in SDTF activity elicited within tissue microenvironments are responsible for the diverse range

Figure 5. Identification of Motifs Associated with Ly6C^{low} Mo Development

(A) Normalized PU.1 ChIP-seq signal in Ly6C^{hi} and Ly6C^{low} Mo of Ly6C^{low} Mo-specific PU.1 sites ± 400 bases from PU.1 peak center.

(B) H3K27ac ± 500 bases of Ly6C^{low} Mo-specific PU.1 peaks. Box and whisker plots in (B) and (C) show interquartile range.

(C) Relative expression of genes proximal to Ly6C^{low} Mo-specific PU.1 peaks.

(D) Overrepresented motifs in Ly6C^{low} Mo specific PU.1 peaks identified by de novo enrichment analysis.

(E) RNA-seq expression amounts of TFs predicted to bind motifs in (D). Please refer to Figure S5A.

of MP enhancers observed in vivo (Gosselin et al., 2014; Lavin et al., 2014). One mechanism for enhancer acquisition involves SDTF driving the formation of latent enhancers associated with de novo H3K4me2 and PU.1 binding (Kaikkonen et al., 2013; Ostuni et al., 2013). Thus the subset of de novo PU.1 binding events provides valuable information concerning secondary SDTF identity by virtue of motif enrichment (Gosselin et al., 2014).

To define motifs associated with Ly6C^{low} monocytes, we determined PU.1 binding profiles in Ly6C^{hi} and Ly6C^{low} monocytes, identifying a total of 65,070 PU.1 peaks. Strict criteria for differential binding led to the identification of 345 Ly6C^{low} monocyte-specific PU.1 peaks (Figure 5A) that possess the hallmarks of a latent enhancer repertoire. These include increased H3K4me2 and nucleosome phasing at regions immediately surrounding the PU.1 peaks (Figure S5A), elevated H3K27ac (Figure 5B), and increased expression of proximal mRNA transcripts (Figure 5C) in Ly6C^{low} monocytes relative to upstream progenitors. Thus, the Ly6C^{low} monocyte PU.1-specific peak profile
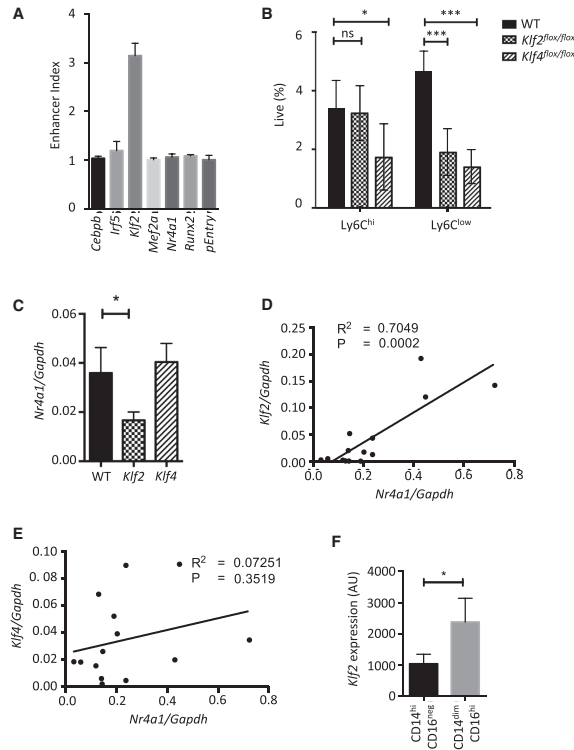
was associated with enhancers that regulate the Ly6C^{low} monocyte gene-expression program. Motifs implicated in Ly6C^{low} monocyte gene expression were identified using de novo motif enrichment analysis on this PU.1 peak set. In strong agreement with our earlier analysis (Figure S1C) we recovered ETS, C/EBP, and NR4A motifs (Figure 5D) as expected, alongside IRF, KLF, RUNX, and MEF2 motifs. Using our RNA-seq dataset, we identified transcription factors that were expressed in Ly6C^{low} monocytes and could bind to these motifs (Figure 5E), considering the most abundant as our primary candidate. This approach defined *Cebpb*, *Irf5*, *Klf2*, *Mef2a*, *Nr4a1*, *Sfpi1* (PU.1), and *Runx2* as principal regulators of Ly6C^{low} monocyte gene expression.

Klf2 Regulates Ly6C^{low} Monocyte Conversion via E2

To identify transcription factors that regulate *Nr4a1* expression via E2, we overexpressed each candidate with the E2 reporter. *Cebpb*, *Irf5*, *Klf2*, *Mef2a*, *Nr4a1*, and *Runx2* were co-transfected into RAW264.7 macrophages alongside either the *Nr4a1-TSS* or E2-*Nr4a1-TSS* luciferase reporter plasmids. To account for promoter-dependent effects of the cDNA and the intrinsic activity of the E2 sequence, we calculated an enhancer index that denotes the difference in the ratio of induced luciferase activity between E2-*Nr4a1-TSS* and *Nr4a1-TSS*. We found that only *Klf2* drove E2-dependent luciferase expression (Figure 6a). Supporting this finding, a motif search for the enriched IRF, KLF, RUNX, and MEF2 motifs within the E2 region identified a cluster of 3 KLF motifs but failed to find any instances of the other motif classes (Figure S5B).

We chose to investigate the role of KLF factors in Ly6C^{low} monocyte development. KLFs are broadly expressed in Ly6C^{low} monocytes with *Klf2* and *Klf4* being the most abundant (Figure 5E). Owing to the perinatal and embryonic lethality of *Klf2*$^{-/-}$ and *Klf4*$^{-/-}$ mice, we assessed monocyte frequencies in myeloid-specific *Lyz2-cre Klf2*$^{flox/flox}$ and *Lyz2-cre Klf4*$^{flox/flox}$

from *Lyz2-cre Klf2^flox/flox* and *Lyz2-cre Klf4^flox/flox* mice using primers that span the *loxP*-flanked exon to determine recombination efficiency. In *Lyz2-cre Klf2^flox/flox* mice, 72% and 91% loss of *Klf2* mRNA was observed for Ly6C[hi] and Ly6C[low] monocytes, respectively (Figure S6C). Thus the selective loss of Ly6C[low] monocytes in *Lyz2-cre Klf2^flox/flox* mice is not consistent with the incomplete deletion of *Klf2* in Ly6C[hi] monocytes. A role for *Klf2* in Ly6C[hi] to Ly6C[low] monocyte conversion is also consistent with the induction of *Klf2* gene expression in the transition between subsets (Figure 5E). In *Lyz2-cre Klf4^flox/flox* mice, *Klf4* mRNA was reduced by 63% and 69% in Ly6C[hi] and Ly6C[low] monocytes respectively (Figure S6D). Although *Lyz2-cre Klf4^flox/flox* mice possessed fewer Ly6C[low] monocytes it is not clear whether this resulted from fewer precursor Ly6C[hi] monocytes,

mice (Liao et al., 2011; Mahabeleshwar et al., 2011). Flow cytometric analysis of *Lyz2-cre Klf4^flox/flox* blood monocytes showed a decrease in both populations (Figure 6B), consistent with previous reports (Alder et al., 2008). In *Lyz2-cre Klf2^flox/flox* mice Ly6C[hi] monocytes were unaffected however Ly6C[low] monocytes were partially reduced (Figure 6B). Similar results were also observed in the bone marrow of *Lyz2-cre Klf2^flox/flox* and *Lyz2-cre Klf4^flox/flox* mice and in bone marrow chimeric mice retrovirally transduced with shRNA targeting *Klf2* and *Klf4* (Figures S6A and S6B).

We have previously observed incomplete deletion of *loxP*-flanked genes in blood monocytes using the *Lyz2-cre* system (Hanna et al., 2015). Therefore, the differences in monocyte frequencies between *Lyz2-cre Klf2^flox/flox* and *Lyz2-cre Klf4^flox/flox* mice may either reflect intrinsic differences in the ability of each gene to regulate monocyte development, or simply the abundance of each TF in each subset. To address this, we measured *Klf2* and *Klf4* mRNA in Ly6C[hi] and Ly6C[low] monocytes sorted

a decrease in Ly6C[hi] to Ly6C[low] monocyte conversion, or both. We reasoned that if *Klf2* or *Klf4* regulated Ly6C[hi] to Ly6C[low] monocyte conversion via *Nr4a1*, then *Klf2* or *Klf4* gene expression would predict *Nr4a1* expression. Indeed, *Nr4a1* mRNA expression was lower in Ly6C[hi] monocytes derived from *Lyz2-cre Klf2^flox/flox* mice, but not *Lyz2-cre Klf4^flox/flox* mice (Figure 6C). Subsequent investigation revealed a significant positive correlation between *Klf2* and *Nr4a1* transcript expression in both monocyte subsets, however no such relationship was found between *Klf4* and *Nr4a1* gene expression (Figures 6D and 6E and S6E and S6F).

While of a preliminary nature, these correlative findings suggest that *Klf2* regulates Ly6C[low] monocyte development via E2. Given our evidence for a conserved mechanism of *Nr4a1* gene expression between species (Figure 3B), we predict that *KLF2* follows a similar pattern in human monocyte subsets. Interrogation of microarray data for *KLF2* in human monocytes confirmed this hypothesis, showing significantly

A

PEC

Lin / CD11b

F4/80 / MHCII

LPM 89.3

SPM 6.32

41.8

Lung

CD11b / CD11c

SSC-A / CD11c

7.86

Spleen

Lin / FSC-A

F4/80 / CD11b

40.4

4.44

B

ns

WT

E2-def

$Nr4a1/Gapdh$

Lung   Spleen   LPM   SPM

C

**

*

WT

E2-def

$Nr4a1/Gapdh$

Ly6C^{hi} Ly6C^{low}

D

WT $Nr4a1^{-/-}$

E2-def

E

*

*
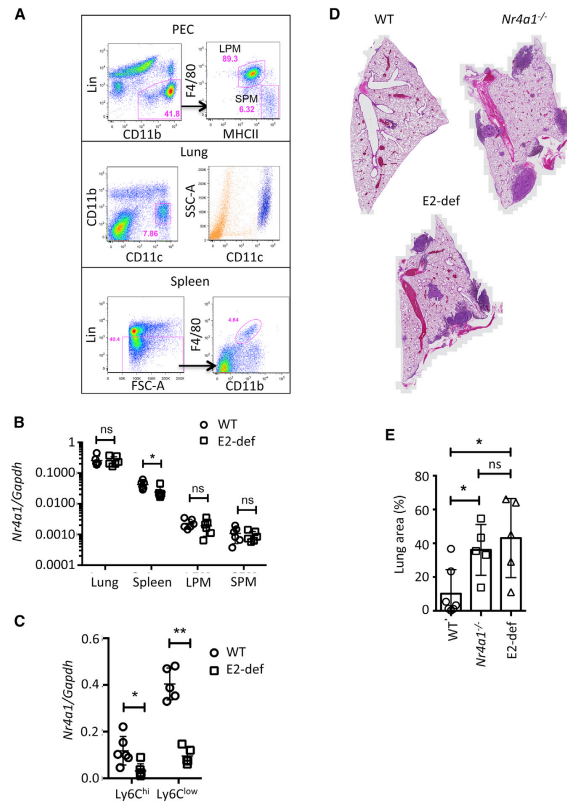
ns

Lung area (%)

WT $Nr4a1^{-/-}$ E2-def

Figure 7. E2 Is a Monocyte-Specific Enhancer

(A) Gating scheme for tissue macrophage sorting. All cells were previously gated on live singlets. The side scatter high profile of lung CD11c[+] macrophages is shown in the right panel (blue).

(B and C) Relative *Nr4a1* mRNA expression in tissue macrophages and blood Ly6C^{hi} and Ly6-C^{low} Mo. Statistics for (B) and (C) measured by Student's t test *p < 0.05, **p < 0.01.

(D) Representative sections of B16F10 tumors in WT, $Nr4a1^{-/-}$, and E2 domain-deficient mice.

(E and F) Quantification of cancer metastasis area; error bars represent SD statistics and were analyzed by ANOVA. *p < 0.05. Please refer to Figure S7.

except for moderately fewer F4/80^{hi} LPM (Figure S7A). RT-PCR analysis captured the expected range of *Nr4a1* expression between macrophage populations (Figure 7B). Strikingly, however, no difference in *Nr4a1* mRNA was detected between tissue macrophages, except for moderately lower expression in E2 domain-deficient splenic macrophages. In contrast to this a substantial reduction in steady-state *Nr4a1* mRNA expression was observed in both Ly6C^{hi} and Ly6C^{low} monocytes (Figure 7C), showing that E2 is a monocyte-specific enhancer.

We recently reported that Ly6C^{low} monocytes prevented cancer metastasis to the lung (Hanna et al., 2015). To demonstrate the utility of the E2 domain-deficient model to study Ly6C^{low} monocytes, we assessed tumor burden in mice using a well-established model of metastasis. B16F10 melanoma cells were intravenously injected into WT, $Nr4a1^{-/-}$, and E2 domain-deficient mice. 18 days after challenge, mice were sacrificed and tumor burden was

higher *KLF2* expression in human CD14^{dim}CD16^{hi} monocytes (Figure 6F).

E2 Is a Monocyte-Specific Enhancer

E2 regulates *Nr4a1* expression in Ly6C^{low} monocytes but not in response to LPS stimulation. Furthermore, at steady state, macrophage *Nr4a1* expression spanned three orders of magnitude. To determine the extent to which E2 was Ly6C^{low} monocyte specific, we measured *Nr4a1* expression in various MP populations from E2 domain-deficient mice. Blood monocytes, F4/80^{hi} large, F4/80^{int} small peritoneal macrophages (LPM and SPM respectively), CD11c^{+} lung alveolar macrophages, and F4/80^{+} splenic red pulp macrophages were selected (Figure 7A). In line with previous observations (Tacke et al., 2015) macrophage frequencies were largely unchanged in $Nr4a1^{-/-}$ mice,

measured by histology. Both $Nr4a1^{-/-}$ and E2 domain-deficient mice showed a loss of Ly6C^{low} monocytes (Figure S7B) and significantly higher tumor burden than WT (Figure 7D and 7E); however, no difference was observed between $Nr4a1^{-/-}$ and E2 domain-deficient mice. Consistent with a critical role for Ly6C^{low} monocytes in controlling metastasis, the differences are explained by fewer tumor regions rather than tumor size (Figures S7C and S7D). Therefore the E2 domain-deficient model effectively decouples *Nr4a1*-dependent inflammatory phenotypes from Ly6C^{low} monocyte function, providing an improved tool to study the function of these cells. In contrast to conditional deletion strategies using the Cre-Lox system, enhancer targeting confers an unprecedented degree of specificity, enabling the causal inference of individual MP subsets in disease.

DISCUSSION

During our investigation into the transcriptional regulatory pathways of Ly6C[low] monocyte development, we identified E2 as a single domain that is absolutely required for Ly6C[low] monocyte development. We also observed distinct, cell-subset specific, patterns of H3K27ac at the *Nr4a1se* locus leading us to investigate whether *Nr4a1* gene expression is controlled by distinct mechanisms between MP subsets. E2 domain-deficient mice lack Ly6C[low] monocytes, yet macrophage *Nr4a1* expression was largely unaffected during steady state and during inflammation. This implies a modular structure for the *Nr4a1* locus control region *Nr4a1se* and is consistent with a recent analysis of several SEs showing that these regions consist of multiple sub-domains (Hnisz et al., 2015). Disrupting these sub-domains differentially affected target gene expression depending upon the locus. For example at *Sik1*, additive effects between sub-domains contribute toward the stable induction of gene expression, whereas at *Prdm14*, mRNA expression was regulated almost entirely by one of five sub-domains (Hnisz et al., 2015). Our study has identified three active and conserved MP enhancers upstream of *Nr4a1*; E2, E6, and E9, yet only E2 regulates Ly6C[low] monocyte development. This highlights the importance of detailed molecular analyses to establish the relevance of individual enhancer elements, and TF binding motifs within these elements. Whether E6 and E9 regulate MP *Nr4a1* gene expression in response to other stimuli is a line of future enquiry.

Our studies suggest that KLF transcription factors regulate *Nr4a1* gene expression via E2. Ly6C[low] monocyte enhancers were enriched for KLF motifs, and these motifs were present in E2. Multiple KLF family members are expressed in Ly6C[low] monocytes, which may act redundantly at E2. In Ly6C[low] monocytes, *Klf2* and K*lf4* were the most abundant family members, thus we investigated the roles of these genes. Both *Klf2* and *Klf4* regulate macrophage inflammatory gene expression by inhibiting p300 and PCAF recruitment to NF-κB (Das et al., 2006; Liao et al., 2011). The different roles ascribed to each factor might arise from the different contexts in which they are expressed. *Klf4* was induced during Ly6C[hi] monocyte differentiation and was further upregulated in macrophages by IL-4. In this context, *Klf4* regulates Ly6C[hi] monocyte development and facilitates the M2 program of macrophage activation (Feinberg et al., 2007; Liao et al., 2011). *Klf2*, however, is expressed in circulating monocytes and strongly reduced by hypoxia leading to increased NF-κB and Hif-1 activity (Mahabeleshwar et al., 2011).

Our findings suggest non-redundant roles for *Klf2* and *Klf4* in monocyte development. *Klf4* expression is high in both monocyte subsets, and *Lyz2-cre Klf4*[flox/flox] mice have fewer monocytes. However, the ratio of Ly6C[hi] to Ly6C[low] monocytes and Ly6C[low] monocyte *Nr4a1* expression were not affected in *Lyz2-cre Klf4*[flox/flox] mice. However, depletion of *Klf2* facilitated a selective loss of Ly6C[low] monocytes and *Klf2* mRNA expression was predictive of *Nr4a1* expression, implying a causal relationship between *Klf2* and *Nr4a1*. Despite repeated attempts, we were not able to obtain successful immunoprecipitation of KLF2 at the *Nr4a1se* locus due to an absence of high-quality commercial antibodies; neither was overexpression of *Klf2*

sufficient to upregulate *Nr4a1* mRNA in vitro (data not shown). These preliminary findings suggest that *Klf2* acts via E2 locus and is necessary, but not sufficient, for Ly6C[low] monocyte development. The identification of additional critical co-factors acting via E2, the *Nr4a1* promoter, or elsewhere will pave the way for a more detailed mechanistic understanding of how E2 regulates *Nr4a1* gene expression in Ly6C[low] monocyte development.

MPs occupy all tissues of the body where these cells perform specialized functions to facilitate homeostasis. For example, Ly6C[low] monocytes maintain vascular integrity by facilitating removal of damaged endothelium (Carlin et al., 2013), alveolar macrophages maintain lung function by clearing surfactants (Nakamura et al., 2013) and red pulp macrophages contribute to iron homeostasis by recycling erythrocytes (Kohyama et al., 2009), whereas the microglia's dedicated functions include supporting brain function by synaptic pruning (Paolicelli et al., 2011). These specialized functions are imparted by tissue-specific LTDFs, Ly6C[low] monocytes require *Nr4a1* (Hanna et al., 2011), alveolar macrophages rely upon PPARγ (Schneider et al., 2014), splenic red pulp-macrophages macrophages need *Spic* (Kohyama et al., 2009), while resident peritoneal macrophages depend upon the transcription factor GATA6 (Okabe and Medzhitov, 2014).

In some cases, such as *for Gata6* and *Spic*, LDTF expression within the MP system is cell-specific such that macrophage-specific deletion of the factor is sufficient to study the function of the associated subset. However, for the most part both the expression and requirement of key MP TFs are not cell-specific. Epigenetic analyses of MP subsets reveals shared motif usage and co-expression of cognate TFs over multiple subsets (Lavin et al., 2014). For example PPAR motifs are enriched in alveolar and renal macrophage enhancers and PPARγ regulates both alveolar macrophage and osteoclast differentiation (Schneider et al., 2014; Wan et al., 2007). Disease processes complicate matters further by invoking dynamic transcription factor expression and altering requirements; in this context, PPARγ is upregulated by IL-4 and controls alternative macrophage activation (Odegaard et al., 2007). Similarly, *Nr4a1* is required for Ly6C[low] monocyte development and is also induced by LPS (Hanna et al., 2011; Pei et al., 2005). These pleiotropic effects limit our ability to study individual MP subsets and the lack of specificity provided by current myeloid Cre transgenes presents a problem when attempting to target individual subsets.

We have shown that enhancer targeting can overcome these limitations. Our approach leverages the unique enhancer repertoires resulting from the exclusive environment each MP subset occupies. We have shown that targeting cell-specific SE sub-domains within key LDTFs functionally decouples cell-specific aspects of gene expression while retaining physiological characteristics of gene function relevant to neighboring cell subsets. In our case, targeting the E2 sub-domain provides a loss-of-function tool to study Ly6C[low] monocytes that preserves both *Nr4a1* expression in other MP populations and the rapid kinetics of *Nr4a1* expression in response to LPS signaling. Conceptually, this strategy can be applied to any scenario in which gene expression is regulated by distinct enhancer sequences as an approach to study gene function among closely related cell types.

136

EXPERIMENTAL PROCEDURES

Mice

C57BL/6J mice were purchased from The Jackson Laboratory. Congenic $Nr4a1^{-/-}$ mice previously described (Hanna et al., 2011) and enhancer deficient mice were maintained in-house. All experiments followed guidelines of the La Jolla Institute for Allergy and Immunology (LIAI) Animal Care and Use Committee, and approval for use of rodents was obtained from LIAI according to criteria outlined in the Guide for the Care and Use of Laboratory Animals from the National Institutes of Health. Mice were euthanized by CO_2 inhalation.

Cell Sorting

Mice were sacrificed by CO_2 inhalation and bone marrow or blood extracted into DPBS + 2 mM ETDA. Following RBC lysis cells were blocked and stained for surface antigens and purified by flow cytometry as described in the figures as described in the Supplemental Experimental Procedures.

Thioglycollate Elicited Macrophages

Thioglycollate elicited macrophages were elicited by intraperitoneal injection of 1 ml 4% Brewer's Thioglycollate Medium. 5 days after injection macrophages were harvested and cultured as described in the Supplemental Experimental Procedures.

RNA Isolation and RNA-Seq

Sorted cells were immediately spun down and stored in Trizol. RNA was extracted using DirectZol columns (Zymo Research). RNA was prepared for sequencing using the TruSeq v2 kit (Illumina).

ChIP-Seq

Histone ChIP was performed using the protocol described in (Gilfillan et al., 2012) and transcription factor ChIP performed essentially as described in (Gosselin et al., 2014), with modifications described in the extended methods. ChIP-seq libraries were prepared using the ThruPlex-FD kit (Rubicon Genomics).

Molecular Cloning and Overexpression Studies

Enhancers were amplified by PCR and cloned into the Sall and Bamh1 sites of a pGL4.10 series vector described in (Heinz et al., 2013) modified to contain an $Nr4a1$ minimal promoter (300 bp) and 5′UTR (please see Table S1). Transfection assays were performed in murine RAW264.7 macrophages using Lipofectamine LTX. cDNA expression vectors were obtained from Origene. Full details are in the Supplemental Experimental Procedures.

CRISPR Mouse Generation

Mouse genome editing was performed essentially as described with minor modifications (Concepcion et al., 2015; Wang et al., 2013). Four sgRNA (two pairs targeting each enhancer flanking region) were injected into embryos of supervoulated C57BL/6 mice along with $Cas9$ mRNA (Life Technologies) at the UCSD transgenic core facility (see Table S2). Founder mice containing the desired deletions identified by PCR (Table S3) were mated with C57BL/6 and offspring thereafter maintained via sibling mating.

Bone-Marrow Chimeras

Bone marrow was shipped on ice as described in the Supplemental Experimental Procedures. Recipient mice were lethally irradiated and bone-marrow transplanted by retroorbital injection. Mice were allowed to recover for at least 6 weeks before performing experiments.

Cancer Study

300,000 B16F10 melanoma cells were injected via tail vein into recipient mice. Lungs were harvested in zinc buffered formalin and mounted into paraffin blocks. Sections were stained with H&E and slides were scanned with an AxioScan Z1 (Zeiss); see Supplemental Experimental Procedures.

Endotoxin Sensitivity Assay

Ultrapure LPS-EB (Invivogen) was made up in PBS and injected i.p. Mice were monitored three times daily for the first 72 hr and twice daily thereafter.

Bioinformatics Data Analysis

FASTQ files were mapped to the mouse mm10 reference genome using RNA-STAR for RNA-seq experiments, or Bowtie for ChIP-seq studies. RNA-seq expression was quantitated using featureCounts and differential expression analyzed using edgeR. ChIP-seq analysis was performed using HOMER. For full details, see Supplemental Experimental Procedures.

ACCESSION NUMBERS

ChIP-seq and RNA-seq data for this project have been deposited at NCBI's Gene Expression Omnibus under the following accession numbers GEO: GSE80031; GEO: GSE80036 (SuperSeries GEO: GSE80040).

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, three tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.immuni.2016.10.011.

REFERENCES

Alder, J.K., Georgantas, R.W., 3rd, Hildreth, R.L., Kaplan, I.M., Morisot, S., Yu, X., McDevitt, M., and Civin, C.I. (2008). Kruppel-like factor 4 is essential for inflammatory monocyte differentiation in vivo. J. Immunol. *180*, 5645–5652.

Amit, I., Winter, D.R., and Jung, S. (2016). The role of the local environment and epigenetics in shaping macrophage identity and their effect on tissue homeostasis. Nat. Immunol. *17*, 18–25.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311–322.

Carlin, L.M., Stamatiades, E.G., Auffray, C., Hanna, R.N., Glover, L., Vizcay-Barrena, G., Hedrick, C.C., Cook, H.T., Diebold, S., and Geissmann, F. (2013). Nr4a1-dependent Ly6C(low) monocytes monitor endothelial cells and orchestrate their disposal. Cell *153*, 362–375.

137

Chow, A., Brown, B.D., and Merad, M. (2011). Studying the mononuclear phagocyte system in the molecular age. Nat. Rev. Immunol. *11*, 788–798.

Concepcion, D., Ross, K.D., Hutt, K.R., Yeo, G.W., and Hamilton, B.A. (2015). Nxf1 natural variant E610G is a semi-dominant suppressor of IAP-induced RNA processing defects. PLoS Genet. *11*, e1005123.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Cros, J., Cagnard, N., Woollard, K., Patey, N., Zhang, S.-Y., Senechal, B., Puel, A., Biswas, S.K., Moshous, D., Picard, C., et al. (2010). Human CD14dim monocytes patrol and sense nucleic acids and viruses via TLR7 and TLR8 receptors. Immunity *33*, 375–386.

Das, H., Kumar, A., Lin, Z., Patino, W.D., Hwang, P.M., Feinberg, M.W., Majumder, P.K., and Jain, M.K. (2006). Kruppel-like factor 2 (KLF2) regulates proinflammatory activation of monocytes. Proc. Natl. Acad. Sci. USA *103*, 6653–6658.

Feinberg, M.W., Wara, A.K., Cao, Z., Lebedeva, M.A., Rosenbauer, F., Iwasaki, H., Hirai, H., Katz, J.P., Haspel, R.L., Gray, S., et al. (2007). The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation. EMBO J. *26*, 4138–4148.

Geissmann, F., Manz, M.G., Jung, S., Sieweke, M.H., Merad, M., and Ley, K. (2010). Development of monocytes, macrophages, and dendritic cells. Science *327*, 656–661.

Gilfillan, G.D., Hughes, T., Sheng, Y., Hjorthaug, H.S., Straub, T., Gervin, K., Harris, J.R., Undlien, D.E., and Lyle, R. (2012). Limitations and possibilities of low cell number ChIP-seq. BMC Genomics *13*, 645.

Gosselin, D., Link, V.M., Romanoski, C.E., Fonseca, G.J., Eichenfield, D.Z., Spann, N.J., Stender, J.D., Chun, H.B., Garner, H., Geissmann, F., and Glass, C.K. (2014). Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. Cell *159*, 1327–1340.

Hanna, R.N., Carlin, L.M., Hubbeling, H.G., Nackiewicz, D., Green, A.M., Punt, J.A., Geissmann, F., and Hedrick, C.C. (2011). The transcription factor NR4A1 (Nur77) controls bone marrow differentiation and the survival of Ly6C- monocytes. Nat. Immunol. *12*, 778–785.

Hanna, R.N., Shaked, I., Hubbeling, H.G., Punt, J.A., Wu, R., Herrley, E., Zaugg, C., Pei, H., Geissmann, F., Ley, K., and Hedrick, C.C. (2012). NR4A1 (Nur77) deletion polarizes macrophages toward an inflammatory phenotype and increases atherosclerosis. Circ. Res. *110*, 416–427.

Hanna, R.N., Cekic, C., Sag, D., Tacke, R., Thomas, G.D., Nowyhed, H., Herrley, E., Rasquinha, N., McArdle, S., Wu, R., et al. (2015). Patrolling monocytes control tumor metastasis to the lung. Science *350*, 985–990.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Heinz, S., Romanoski, C.E., Benner, C., Allison, K.A., Kaikkonen, M.U., Orozco, L.D., and Glass, C.K. (2013). Effect of natural genetic variation on enhancer selection and function. Nature *503*, 487–492.

Hettinger, J., Richards, D.M., Hansson, J., Barra, M.M., Joschko, A.-C., Krijgsveld, J., and Feuerer, M. (2013). Origin of monocytes and macrophages in a committed progenitor. Nat. Immunol. *14*, 821–830.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hnisz, D., Schuijers, J., Lin, C.Y., Weintraub, A.S., Abraham, B.J., Lee, T.I., Bradner, J.E., and Young, R.A. (2015). Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. Mol. Cell *58*, 362–370.

Jacob, A., Zhou, M., Wu, R., Halpern, V.J., Ravikumar, T.S., and Wang, P. (2007). Pro-inflammatory cytokines from Kupffer cells downregulate hepatocyte expression of adrenomedullin binding protein-1. Biochim. Biophys. Acta. *1772*, 766–772.

Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C., and Glass, C.K. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol. Cell *51*, 310–325.

Kohyama, M., Ise, W., Edelson, B.T., Wilker, P.R., Hildner, K., Mejia, C., Frazier, W.A., Murphy, T.L., and Murphy, K.M. (2009). Role for Spi-C in the development of red pulp macrophages and splenic iron homeostasis. Nature *457*, 318–321.

Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S., and Amit, I. (2014). Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell *159*, 1312–1326.

Li, L., Liu, Y., Chen, H.Z., Li, F.W., Wu, J.F., Zhang, H.K., He, J.P., Xing, Y.Z., Chen, Y., Wang, W.J., et al. (2015). Impeding the interaction between Nur77 and p38 reduces LPS-induced inflammation. Nat. Chem. Biol. *11*, 339–346.

Liao, X., Sharma, N., Kapadia, F., Zhou, G., Lu, Y., Hong, H., Paruchuri, K., Mahabeleshwar, G.H., Dalmas, E., Venteclef, N., et al. (2011). Krüppel-like factor 4 regulates macrophage polarization. J. Clin. Invest. *121*, 2736–2749.

Mahabeleshwar, G.H., Kawanami, D., Sharma, N., Takami, Y., Zhou, G., Shi, H., Nayak, L., Jeyaraj, D., Grealy, R., White, M., et al. (2011). The myeloid transcription factor KLF2 regulates the host response to polymicrobial infection and endotoxic shock. Immunity *34*, 715–728.

Muller, P.A., Koscsó, B., Rajani, G.M., Stevanovic, K., Berres, M.-L., Hashimoto, D., Mortha, A., Leboeuf, M., Li, X.-M., Mucida, D., et al. (2014). Crosstalk between muscularis macrophages and enteric neurons regulates gastrointestinal motility. Cell *158*, 300–313.

Nakamura, A., Ebina-Shibuya, R., Itoh-Nakadai, A., Muto, A., Shima, H., Saigusa, D., Aoki, J., Ebina, M., Nukiwa, T., and Igarashi, K. (2013). Transcription repressor Bach2 is required for pulmonary surfactant homeostasis and alveolar macrophage function. J. Exp. Med. *210*, 2191–2204.

Odegaard, J.I., and Chawla, A. (2011). Alternative macrophage activation and metabolism. Annu. Rev. Pathol. *6*, 275–297.

Odegaard, J.I., Ricardo-Gonzalez, R.R., Goforth, M.H., Morel, C.R., Subramanian, V., Mukundan, L., Red Eagle, A., Vats, D., Brombacher, F., Ferrante, A.W., and Chawla, A. (2007). Macrophage-specific PPARgamma controls alternative activation and improves insulin resistance. Nature *447*, 1116–1120.

Okabe, Y., and Medzhitov, R. (2014). Tissue-specific signals control reversible program of localization and functional polarization of macrophages. Cell *157*, 832–844.

Okabe, Y., and Medzhitov, R. (2016). Tissue biology perspective on macrophages. Nat. Immunol. *17*, 9–17.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. Cell *152*, 157–171.

Paolicelli, R.C., Bolasco, G., Pagani, F., Maggi, L., Scianni, M., Panzanelli, P., Giustetto, M., Ferreira, T.A., Guiducci, E., Dumas, L., et al. (2011). Synaptic pruning by microglia is necessary for normal brain development. Science *333*, 1456–1458.

Parkhurst, C.N., Yang, G., Ninan, I., Savas, J.N., Yates, J.R., 3rd, Lafaille, J.J., Hempstead, B.L., Littman, D.R., and Gan, W.-B. (2013). Microglia promote learning-dependent synapse formation through brain-derived neurotrophic factor. Cell *155*, 1596–1609.

Pei, L., Castrillo, A., Chen, M., Hoffmann, A., and Tontonoz, P. (2005). Induction of NR4A orphan nuclear receptor expression in macrophages in response to inflammatory stimuli. J. Biol. Chem. *280*, 29256–29262.

Perdiguero, E.G., and Geissmann, F. (2016). The development and maintenance of resident macrophages. Nat. Immunol. *17*, 2–8.

138

Schmidl, C., Renner, K., Peter, K., Eder, R., Lassmann, T., Balwierz, P.J., Itoh, M., Nagao-Sato, S., Kawaji, H., Carninci, P., et al. (2014). Transcription and enhancer profiling in human monocyte subsets. Blood *123*, e90–9.

Schneider, C., Nobs, S.P., Kurrer, M., Rehrauer, H., Thiele, C., and Kopf, M. (2014). Induction of the nuclear receptor PPAR-γ by the cytokine GM-CSF is critical for the differentiation of fetal monocytes into alveolar macrophages. Nat. Immunol. *15*, 1026–1037.

Sica, A., and Mantovani, A. (2012). Macrophage plasticity and polarization: in vivo veritas. J. Clin. Invest. *122*, 787–795.

Tacke, R., Hilgendorf, I., Garner, H., Waterborg, C., Park, K., Nowyhed, H., Hanna, R.N., Wu, R., Swirski, F.K., Geissmann, F., and Hedrick, C.C. (2015). The transcription factor NR4A1 is essential for the development of a novel macrophage subset in the thymus. Sci. Rep. *5*, 10055.

Wan, Y., Chong, L.-W., and Evans, R.M. (2007). PPAR-gamma regulates osteoclastogenesis in mice. Nat. Med. *13*, 1496–1503.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910–918.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Yona, S., Kim, K.-W., Wolf, Y., Mildner, A., Varol, D., Breker, M., Strauss-Ayali, D., Viukov, S., Guilliams, M., Misharin, A., et al. (2012). Fate Mapping Reveals Origins and Dynamics of Monocytes and Tissue Macrophages under Homeostasis. Immunity *38*, 79–91.

Supplemental Information

Deleting an *Nr4a1* Super-Enhancer Subdomain

Ablates Ly6C^{low} Monocytes while Preserving

Macrophage Gene Function

Graham D. Thomas, Richard N. Hanna, Neelakatan T. Vasudevan, Anouk A. Hamers, Casey E. Romanoski, Sara McArdle, Kevin D. Ross, Amy Blatchley, Deborah Yoakum, Bruce A. Hamilton, Zbigniew Mikulski, Mukesh K. Jain, Christopher K. Glass, and Catherine C. Hedrick
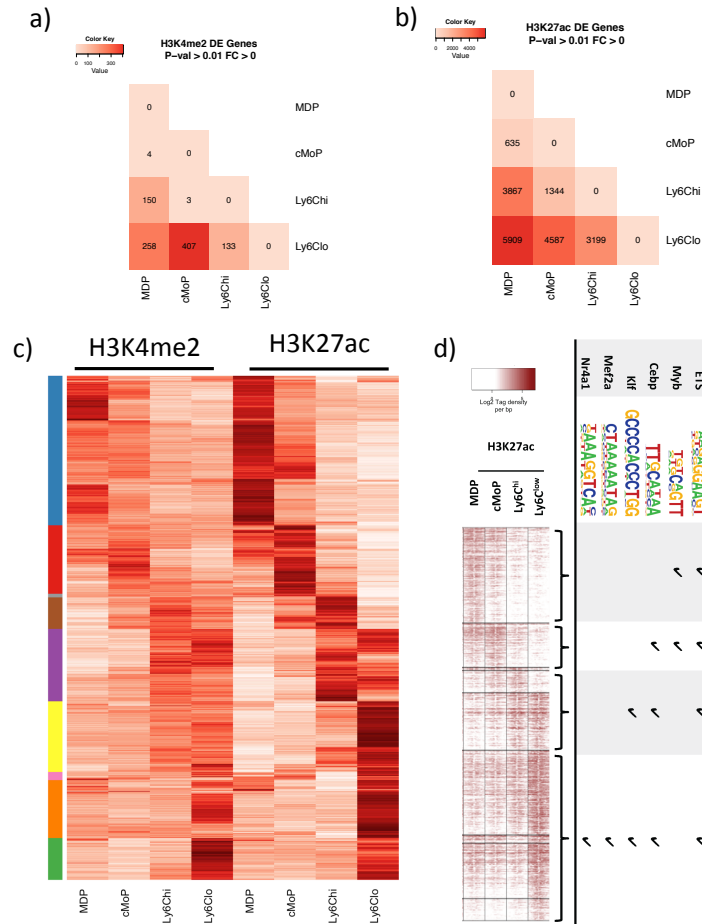
Figure S1, related to figure 1. a) Differentially bound H3K4me2 peaks (P,1*10^-5). In total 640 enhancer regions were found to be differentially regulated between all conditions. B) Differentially bound H3K27ac peaks (P,1*10^-5) . In total 9,879 enhancer regions were found to be differentially regulated between all conditions. c) Hierarchical clustering of DE enhancer regions. Bands on the left show clusters used to define groups in Figure 1c. d) Presence/ absence of *de novo* motif instances in clusters associated with MDP, MDP/cMoP, Ly6C[hi]/ Ly6C[low] Mo, or Ly6C[low] Mo only, as defined in Figures 1c and S1c. Only motif instances with a significance p–value 1e-40 are shown, alongside the representative motif identified by HOMER..
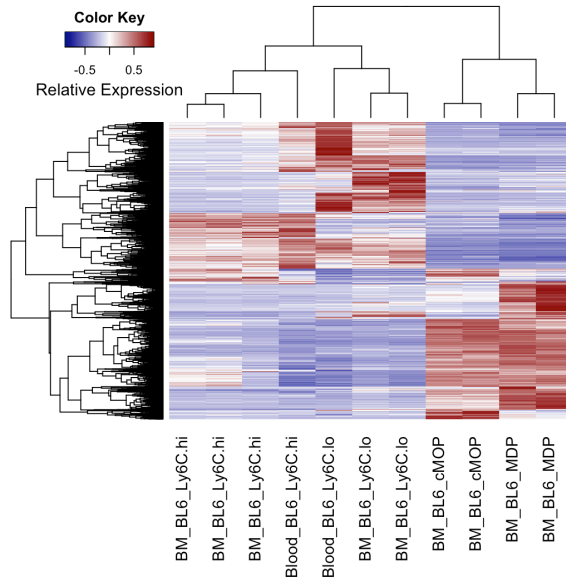
Figure S2, related to figure 1. Gene expression profiles Differential gene expression pairwise comparisons in mRNA-Seq data.
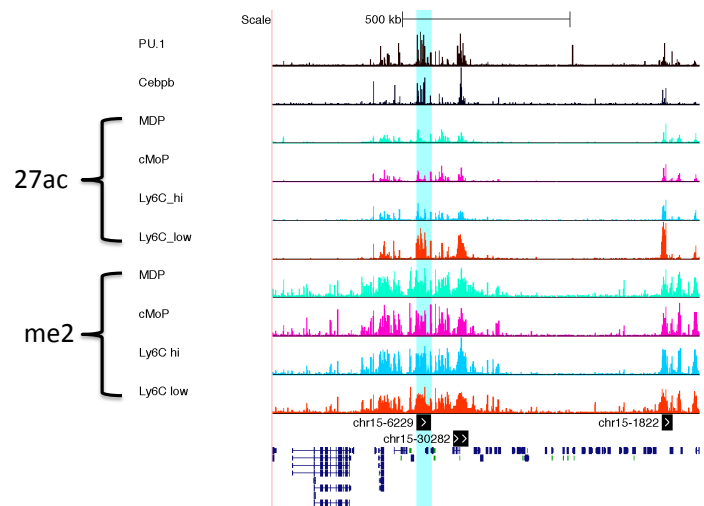
Figure S3, related to figure 2. UCSC genome browser screenshot of *Nr4a1* and surrounding region. The *Nr4a1*-associated super-enhancer (*Nr4a1se*) is highlighted in blue. Super-enhancer predictions for Ly6C[low] monocytes are shown in black directly above the gene prediction track.
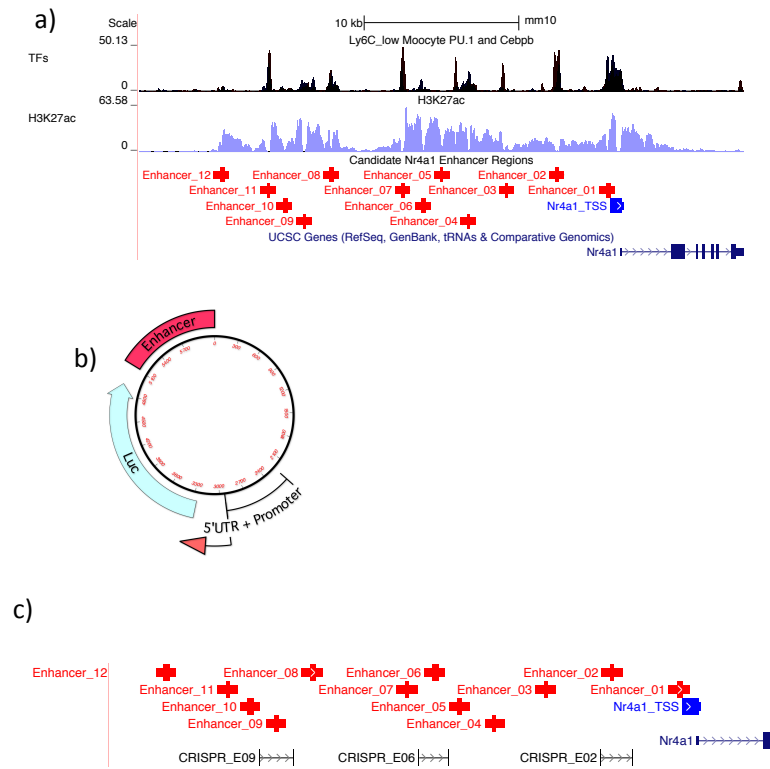
Figure S4, related to figure 3. a) UCSC genome browser screenshot showing positions of CRISPR sgRNA sites for E2, E6 and E9 domain-deficient mice. The TFs track shows superimposed PU.1 and CEBPb transcription factor binding profiles in Ly6C[low] Mo. b) Schematic of pGL4.10 luciferase reporter vector containing *Nr4a1* 300bp of *Nr4a1* promoter upstream of the TSS and the 5' UTR sequence. c) sgRNA design.

a)
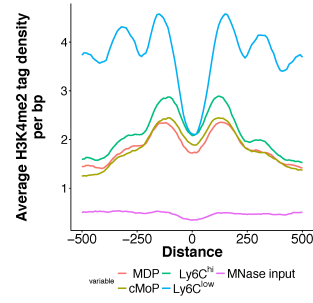
b)
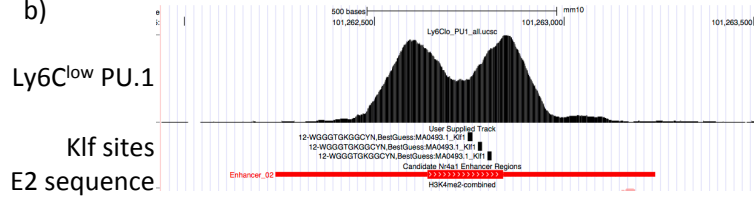
Figure S5, related to figures 5 and 6. a) mRNA expression levels of transcription factors belonging to family of over-represented motifs present in Ly6C^{low} monocyte enhancer regions based on RNA-Seq data. b) Klf motifs (identified using HOMER) present in the *Nr4a1* E2 enhancer sub-sequence.
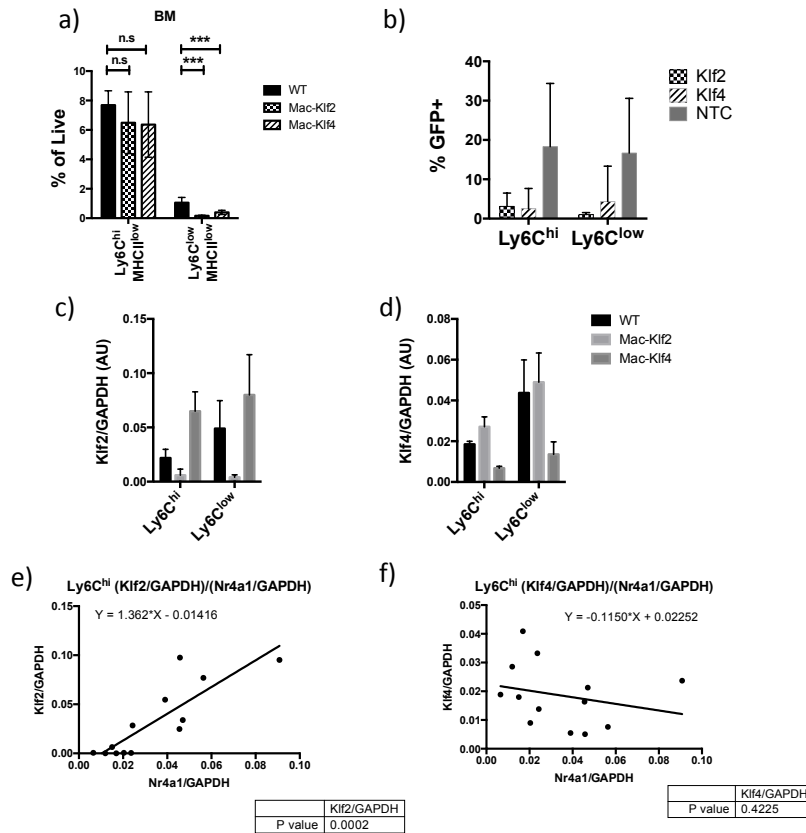
Figure S6, related to figure 6. a) Flow cytometric analysis of Mo subset frequencies in the bone marrow of Mac-Klf2 and Mac-Klf4 mice. b) Frequencies of GFP+ cells transduced with GFP and shRNA expressing retrovirus targeting *Klf2*, *Klf4*, or non-targeting control (NTC) sequence. c, d) *Klf2* (c) and *Klf4* (d) mRNA expression in blood monocyte subsets sorted from Mac-Klf2 and Mac-Klf4 mice. e,f) *Klf2* (e) and *Klf4* (f) mRNA expression levels correlated against *Nr4a1* mRNA expression in Ly6C[hi] Mo.
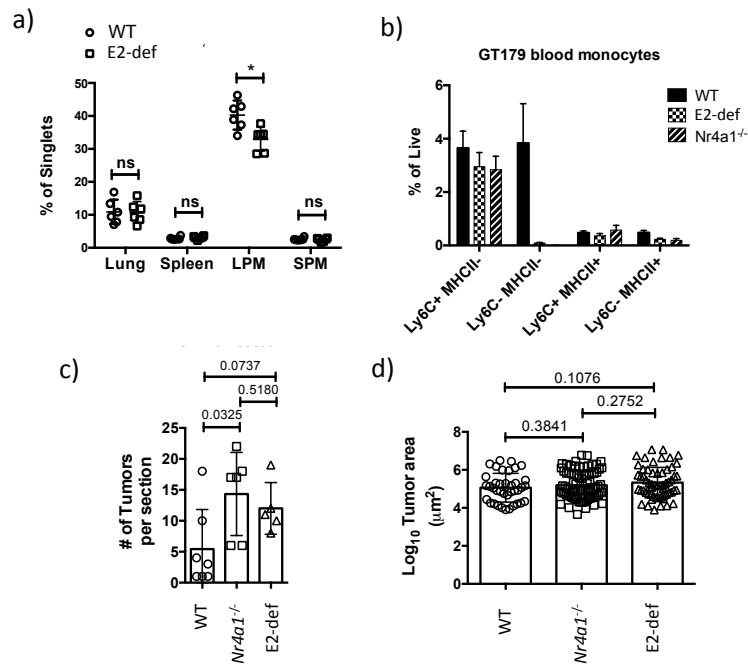
Figure S7, related for figure 7. a) Tissue macrophage subset frequencies in WT and E2 domain-deficient mice as measured by FACS. B) Blood monocyte subset frequencies in WT, $Nr4a1^{-/-}$ and E2 domain-deficient mice 18 days after injection with 300,000 B16F10 melanoma cells. c) Quantification of histological sections for B16F10 tumors d) Size distribution of individual tumors within lungs of mice relating to figures 7d. Statistics for c) and d) performed using students t-test, p values are shown.

Materials and Methods

Cell preparation and isolation

Bone marrow monocytes were harvested from femurs, tibias and hip bones of 8-12 week old male C57BL/6 mice (strain 006664) obtained from the Jackson laboratory (Bar Harbor, ME). Prior to FACS staining cells were subject to a brief (3 minute) red blood cell lysis (RBC lysis buffer, EBiosciences) at room temperature. All FACS staining was performed in FACS buffer (DPBS + 10% FCS + 2mM EDTA). Prior to surface staining FC receptors were blocked with anti-CD16/32 (clone 93) for 30 minutes. Surface staining was performed for 30 minutes in a final volume of 500ul for FACS sorts and 100ul for regular flow cytometry. Surface staining was performed using Live/Dead Yellow (Thermo Fisher) and antibody combinations in accordance with the gating schemes in the figures. Lineage positive cells were identified using pooled APC conjugated anti-CD3, CD19, Ly6G and NK1.1 (clones145-2C11, 1D3, 1A8 and PK136 respectively). Additional cell surface markers used were CD117-PE-Cy7 (ACK2), CD115 PE or BV421 (clone AFS98), Ly6C APC-Cy7 or PerCP-Cy5.5 (clone HK1.4), CD135-PE (clone A2F10.1), CD11c-FITC (clone N418), F4/80 PE-Cy7 (clone BM8), CD11b FITC (clone M1/70) and MHCII-BV605 (clone M1/70). All antibodies were obtained from Biolegend (San Diego, CA). Samples were washed twice in at least 200ul FACS buffer before acquisition. Cells were sorted using a FACS Aria II (BD biosciences) and conventional flow cytometry using an LSRII (BD biosciences). All flow cytometry was performed on live cells.

ChIP and sequencing library preparation

ChIP assays for histone modifications were performed as previously described (Gilfillan et al., 2012). To obtain sufficient numbers of MDPs bone marrow from 10 mice were pooled and sorted. For each ChIP assay 500,000 FACS isolated cells were immediately washed with PBS and resuspended in MNase digestion buffer (50mM Tris pH 8.0, 1mM CaCl2, 0.2% Triton X-100). Cells were then digested in to mononucleosomal fragments using micrococcal nuclease (MNase, Affymetrix, CA). Enzymatic digestion was quenched by addition of 1/10 volume stop buffer (110mM Tris pH 8.0, 55mM EDTA), samples briefly sonicated using a Bioruptor (Diagenode, Belgium) and then adjusted to RIPA buffer conditions by adding an equal volume of 2x RIPA buffer (280mM NaCl, 1.8% Triton x100, 0.2% SDS, 0.2% Sodium Deoxycholate, 5mM EGTA). Immunoprecipitations were performed in a final volume of 100 μl. 2 μg of anti-H3K4me2 (Millipore 07-030) or 2μg of anti-H3K27ac (Diagenode C15410196) were added and incubated overnight with rotation at 4 °C. Antibody-antigen-DNA complexes were recovered by incubating for 2 hours with 10 μl Protein A Dynabeads previously washed in RIPA buffer (10mM Tris pH 8.0, 140mM NaCl, 1mM EDTA, 1% Triton X-100, 0.2% SDS, 0.2% Sodium Deoxycholate). Complexes were then washed 5x in ice cold RIPA buffer and 1x in LiCl wash buffer (10mM Tris pH 8.0, 250mM LiCl, 1mM EDTA, 0.5% Igepal CA-630, 0.5% Sodium deoxycholate,), each wash was performed in 200 μl at 4 °C for 5 minutes with rotation. All above buffers were supplemented with 1x Protease inhibitor Cocktail, 1mM PMSF and 5mM Sodium Butyrate (Sigma). Beads were subject to a final wash in 200 μl ice cold TE buffer (Invitrogen) without protease inhibitors and eluted in 100 μl 1% SDS-TE buffer at 37 °C for 20 minutes. Finally, protein was treated with 2 μl proteinase K (Ambion) for 1h at 55 °C and DNA purified using a ChIP Clean & Concentrate column (Zymo, Irvine, CA) eluting in 30 μl final volume.

ChIP for PU.1 and C/EBPb were performed as previously described (Gosselin et al., 2014). 500,000 sorted cells were washed in PBS and immediately fixated for 9 minutes at room temperature in 1% methanol free formaldehyde in PBS (ThermoFisher Scientific). Fixation was quenched by addition of 1/20 volume 2.625M glycine solution and cells were washed twice with PBS. Cell pellets were then snap frozen in a dry ice/methanol bath and stored at -80oC until needed. Nuclei were enriched by resuspending cell pellets in 10mM HEPES pH 7.9, 85mM KCl, 1mM EDTA, 0.5% Igepal CA-630 and incubating on ice for 10 minutes.

Nuclei were harvested by spinning at 3000 g for 10 minutes, and lysed in 130 μl lysis buffer (10mM Tris pH 8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Sodium Deoxycholate, 0.5% N-lauroylsarcosine). DNA was transferred in to Covaris micro tubes (Covaris, MA) sheared into 150-600bp fragments using a Covaris E220 (14 mins, duty cycle 3%, 100 cycles per burst). Final volume was adjusted to 200 μl and 22 μl 10% Triton X-100 added, cell debris was then cleared by spinning at maximum speed for 5 minutes. 20 μl of protein A dynabeads pre-conjugated to 3ug anti PU.1 or C/EBPb (sc-352x and sc-150x respectively, Santa Cruz Biotechnology, CA) were added to each chromatin preparation and incubated with rotation for 2h at 4 °C. Antibody-antigen-DNA complexes were then washed in 3x WBI (20mM Tris pH 7.4, 150 mM NaCl, 0.1% SDS, 1% Triton X-100, 2mM EDTA,), 3x LiCl WB (10mM Tris Ph 7.4, 250mM LiCl,, 1% Triton X-100, 0.7% Sodium Deoxycholate, 1% Igepal CA-630) and 1x TET (TE, 0.2% Tween-20) buffer. All wassh volumes were 200 μl. DNA was eluted in 1% SDS-TE for 30 minutes at 37 °C, NaCl was added to a final concentration of 300mM and corsslinking reversed overnight at 65 °C. Finally, protein was treated with 2 μl proteinase K (Ambion) for 1h at 55 °C and DNA purified using a ChIP Clean & Concentrate column (Zymo, Irvine, CA) eluting in 30 μl final volume.

ChIP-Seq libraries were prepared using an initial 0.3 - 5 ng DNA using the ThruPlex-FD kit (Rubicon Genomics, MI) in accordance with the manufacturer's guidelines. RNA-Seq libraries were prepared using the Tru-Seq v2 library preparation kit (Illumina, La Jolla, CA) in accordance with the manufacturer's instructions.

RNA-Seq differential expression analysis
Sequencing libraries were sequenced using an Illumina HiSeq at the LIAI sequencing core facility. RNA-Seq libraries were sequenced as paired-end 50 base reads, and ChIP-Seq libraries were sequenced as single-end 50 base reads. Illumina BCL files were converted into FASTQ format using bcl2fastq (v1.8.4, Illumina).

Paired end RNA-Seq reads were mapped to the mouse mm10 reference genome using RNA-STAR v2.3.0 (Dobin et al., 2013) compiled with gene models derived from the Ensembl v73 genome annotation set. Gene counts were quantitated with the same annotation set using featureCounts (Liao et al., 2013). Differential expression analysis was performed using edgeR v3.4.2 (Robinson et al., 2010). Variance calculations were computed using the common dispersion method and sequencing depth differences adjusted using the 'relative log expression' method. Differential expression was computed against all pairwise comparisons and an FDR (Benjamini-Hochberg) corrected P-value threshold of 1e-5 applied for calling statistical significance. RPKM calculations were computed in edgeR using gene lengths derived from featureCounts. Publicly available RNA-Seq data from Levin *et al* were obtained from GEO GSE63340 (Levin *et al.,* 2014) and processed as above.

ChIP-Seq peak calling, differential binding and UCSC genome browser visualization
ChIP-Seq reads were mapped to the mm10 reference genome using Bowtie (v , Langmead et al., 2009). Homer (v4.4, Heinz et al., 2010) was used for ChIP-Seq peak calling.

Enhancer regions were defined by merging biological replicates for H3K4me2 libraries and running the findPeaks algorithm with 'style -histone' against the MNase-treated DNA input control. H3K4me2 profiles defined by each cell type were then aggregated using mergePeaks to give the final H3K4me2 peak set. These peaks were assigned to target genes using HOMER, and enhancers defined as H3K4me2 peaks whose centers were >2.5kb from the nearest annotated transcription start site. Enhancer activity was measured by quantitating H3K4me2 and H3K27ac within these defined enhancer regions. Differences in sequencing depth were accounted for by normalizing to an effective library size of 1×10^7

reads per lane. Transcription factor ChIP-Seq peaks were called using HOMER by running the findPeaks algorithm with the 'style -factor' flag using CH_2O treated input DNA control as background sequence. Transcription factor peaks were called independently for each lane and merged using mergePeaks prior to quantification of tags within peaks using the annotatePeaks command in HOMER. Visualization was performed at the UCSC genome browser using bigWig files generated in HOMER.

ChIP-Seq data for Lavin *et* al (2014) were obtained using GEO accession GSE63339 and processed as above. Human ChIP-Seq data were accessed from GEO accession SRP015328 and mapped to the human hg19 reference genome using the protocols outlined above. The bigWig file of human monocyte DNase-Seq was downloaded from the ENCODE project at (https://www.encodeproject.org/files/ENCFF000TAU/@@download/ENCFF000TAU.bigWig) and visualized directly in the UCSC genome browser.

Microarray data
Human monocyte Klf2 expression was obtained from NCBI GEO Profiles under accession GDS4219, data shown are for probe ID 219371_s_at, which is representative of the three probes available for Klf2 on this array type in this experiment.

Differential enhancer profiling, hierarchical clustering and ChIP-Seq visualization
Differential histone and transcription factor binding were determined using edgeR. To identify patterns of enhancer usage between MDP, cMoP, Ly6C^{hi} and Ly6C^{low} monocytes we employed a very permissive threshold for differential expression, this was to identify any weak signals in the data arising from these closely related cell types. Statistical significance was computed between all pairwise comparisons of cell types using an FDR corrected (Benjamini-Hochberg) p value threshold of 0.01 without the application of a fold-change cutoff. Variance was estimated using the common dispersion estimate by treating all conditions as pseudoreplicates of a single group. The final set of differentially enriched (DE) enhancers represents the union of all differentially bound enhancers determined by H3K4me2 or H3K27ac in any pairwise comparison.

In order to perform hierarchical clustering, matrices of tag counts for H3K4me2 and H3K27ac for the set of DE enhancers were obtained. Each matrix was independently mean centered and variance stabilized (with respect enhancers, row mean / row SD). The normalized matrices were then concatenated and subject to hierarchical clustering using a euclidean distance and complete linkage clustering approach using the base functions in R. Based upon visual inspection, the results the tree was cut into nine clusters, the ordering of these clusters was used to order the final histogram figure (Figure 1c). Within each cluster enhancers were then ranked according to their 'peak score' as determined by HOMER, this ranking was used to order rows within clusters in Figure 1c. All PU.1 peaks overlapping each DE enhancer were extracted and H3K4me2/H3K27ac tag counts calculated +/-1kb from each PU.1 peak center using HOMER's annotatePeaks with option '-ghist'. The final results were plotted using the R package gplots.

UCSC genome browser tracks were visualized using track hubs generated with the HOMER program makeBigWigHub.pl.

Differential PU.1 peak calling between Ly6C^{hi} and Ly6C^{low} Mo
A complete monocyte PU.1 peak set was obtained by concatenating the PU.1 peaks defined in both Ly6C^{hi} and Ly6C^{low} Mo as described in *'ChIP-Seq peak calling, differential binding and UCSC genome browser visualization'*. Tag counts for PU.1 libraries were counted within this peak set for both monocyte populations and differential binding an analysis was performed

in edgeR using the same procedures as for differential histone binding. For PU.1 peaks an FDR corrected (Benjamini-Hochberg) p-value cutoff of 1e-5 and log2 fold change cutoff of 3 were applied.

Enhancer cloning

All specific cloning details for individual sequences are in table S1. The *Nr4a1* 5'UTR and 650 bases of promoter sequence were cloned into the KpnI and SalI sites of a minimal pGL4.10 luc2 reporter vector. This vector, termed pGL4.10-TSS was used as the host vector for all enhancer clones. All enhancer sequences were isolated from genomic DNA by PCR using the primers detailed in Table 1. For PCR using Phusion High-fidelity polymerase (NEB, MA) 10ul 5x HF buffer, 1ul 10mM dNTP, 2.5ul F primer (10uM), 2.5ul R primer (10uM) 2ul mouse genomic DNA (250ng), 1.5ul DMSO, 30ul H2O and 0.5ul Phusion polymerase were used per reaction. PCR was performed using the following protocol at the indicated annealing temperature: - 1x 98oC 30s; 30x 98oC 10s, anneal 30s, 72oC 30s; 72oC 10min. For PCR using LA Taq GC buffer (Clontech, Mountain View, CA) 25ul GC buffer I, 8ul dNTP, 2.5ul F primer (10uM), 2.5ul R primer (10uM), 2ul DNA (250ng), 9.5ul H2O and 0.5ul LA-Taq were used per reaction. PCR was performed as per the manufacturer's instructions using the annealing temperatures in Table 1 and 1-minute extension time. PCR products were purified by gel extraction using Zymoclean Gen DNA recovery columns (Zymo, Irvine, CA). All cloning was performed at 1:4 vector: insert molar ratio using NEB reagents (Ipswich, MA) and clones were validated by PCR sequencing. PCR was performed using a MyCycler (BioRad, Irvine, CA)

Transfection and overexpression experiments

RAW264.7 macrophages were obtained from ATCC and maintained at low passage (<15) in DMEM +(10%FCS, 1% Pen/Strep, 1% L-glut). Transfection studies were performed using 0.3ug pGL4.10 luciferase reporter and 0.2ug cDNA or 0.5ug pGL4.10 luciferase reporter only, in addition 0.04ug pmaxGFP (Lonza, Basel) and 0.01ug pTK Renilla (ThermoFisher, Waltham, MA) luciferase control were added. 0.55ug of plasmid DNA was complexed at 8:1 ratio with lipofectamine LTX as per manufacturer's guidelines and 100ul per well given to RAW.264.7 cells at ~70% confluency and activity measured typically 12-16 hours later using the Dual Luciferase Reporter Assay (Promega, Madison, WI) on Sirius luminometer (Titerkek-Berthold, Pforzheim, Germany).

shRNA virus BMT experiments

PLAT-E packaging cells were grown to 60-70% confluency in DMEM (Gibco, +10%FCS, 1% L-glut, 1 μg/ml puromycin, 10 μg/ml blasticidin) in 10 cm dishes at 37°C in 10% CO2. Pools of three UltramiR (TRANSomiC technologies, Huntsville, AL) retroviral vectors (LMN) for Klf2 (ULTRA-3224750, ULTRA-3224748, ULTRA-3224747), Klf4 (ULTRA-3224770, ULTRA-3224769, ULTRA-3224768) and Non-targeting shRNA control were transfected into PLAT-E cells using JetPrime transfection in accordance with the manufactuer's guidelines. 7 μg total DNA was transfected per 10 cm plate. 1 hour prior to transfection media was replaced with antibiotic free media, cells were incubated overnight. The following were replaced with fresh 10 ml DMEM, without antibiotics. The following morning the retroviral supernatant was harvested and used for transduction of bone marrow stem cells (below). 8 ml fresh DMEM was added to the cells and the last step repeated the following day.

Murine stem cells were isolated from bone marrow by negative selection using the EasySep Mouse Hematopoietic Progenitor Isolation Kit (StemCell technologies, Vancouver, BC) in accordance with manufacturer's guidelines. Cells were resuspended in 2.5ml of STEMSPAN SFEM media (StemCell Technologies, Vancouver) containing 200 ng/ml rmSCF, 40 ng/ml rhIL-6 and 20 ng/ml rmIL-3 in 6 well plates previously coated for 2 hours with 2ml of 25

μg/ml human fibronectin and washed once with PBS. Bone marrow stem cells from one mouse were split between 2 wells and an equal volume (1.25ml) retroviral supernatant was added to each well. Cells were then spun at 850g for 1h at 37°C and incubated for 22h at 37°C in 10% CO_2. A second batch of retrovirus was added the following day and spin transduction repeated. Following the second transduction cells were harvested and transferred via retroorbital injection into lethally irradiated C57BL/6 host mice (2x 600 Rads, 2-3 hours apart). Bone marrow was allowed to reconstitute for 6 weeks prior to analysis.

CRISPR mouse generation

sgRNA sequences were designed using the http://crispr.mit.edu web tool. To minimize off-target effects we only considered sgRNAs with a score >70. For each enhancer deletion two pairs of sgRNAs flanking the target region were designed that were at least 50 base pairs apart. sgRNA templates were ordered as Ultramers from IDT (Coralville, IA) as a chimeric T7 promoter sequence, variable crRNA (see Table S2 for individual sequence information) and invariable tracrRNA. In order to promote robust transcription from the T7 promoter the first two bases of each sgRNA sequence were substituted for guanines. dsDNA templates for each sgRNA were generated by PCR using the Pfu Ultra II polymerase (Agilent Technologies, Santa Clara, CA). PCR reactions were cleaned using ChIP Clean & Concentrate columns (Zymo, Irvine, CA) and used as input for *in vitro* transcription using the MEGAscript T7 kit (Thermo Fisher, Waltham, MA) in accordance with the manufacturer's instructions. RNA was cleaned using the MEGAclear kit (Thermo Fisher) in accordance with the manufacturer's guidelines.

For embryo injections 0.5 day fertilized embryos were collected from 3-4-week-old superovulated C57BL/6 females (Harlan, WI) by injecting 5.0 I.U each of PMSG (Sigma Aldrich) and hCG (Sigma Aldrich). Embryos were transferred into M2 medium (Millipore) and injected into the cytoplasm with sgRNA mixed at 25ng/ul each along with 50ng/ul Cas9 mRNA (GeneArt CRISPR Nuclease mRNA, Thermo Fisher) to give a final 150ng/ul RNA in IDTE buffer. These injected embryos were cultured in an incubator in KSOMaa medium (Zenith) in a humidified atmosphere of 5% CO2 at 37C over night. The embryos were implanted at 2-cell stage into recipient pseudo pregnant ICR female mice. Embryo injections were performed at the University of California, San Diego transgenic core facility.

CRISPR mouse breeding and genotyping

Founder mice were screened for the presence of one or two altered alleles using a PCR strategy that flanks the expected mutation. PCRs were designed such that the wild type product is 2 - 3kb and the respective deletion allele around 1kb shorter. Primer sequences and PCR details are in table S3. PCRs were carried out using the specified kits in accordance with the manufacturer's guidelines, with primer extension and annealing temperatures stated in the table. Founder animals were crossed with C57BL/6 mice obtained from the Jackson laboratory (strain 000664) to obtain heterozygous F1 animals that were interbred via sibling mating to generate *Nr4a1se_2*[-/-], *Nr4a1se_6*[-/-], and *Nr4a1se_9*[-/-] mice. Observed phenotypes were present in homozygous null mice derived from at least three independent founder mice for each strain. For the *Nr4a1se_2*-/- strain we also confirmed by Sanger sequencing that the downstream DNA sequence from the deletion into the Nr4a1 first intron was not modified

LPS challenge

Male mice aged 8-15 weeks old were intraperitoneally injected with $2.5*10^6$ EU/kg (equivalent to 2.5mg/kg) Ultrapure LP-EB (Invivogen) in PBS. Mice were age and sex matched in all experiments. Prior to injection LPS was sonicated for 5 minutes at room

temperature in a bath sonicator (FS20H, Fisher Scientific). Mice were monitored three times daily for the first 72 hours and twice daily thereafter.

<u>B16F10, histology and microscopy quantification</u>
300,000 B16F10 melanoma cells were intravenously injected by tail vein injection into recipient mice as previously described (Hanna *et al* 2015). 18 days following injection mice were sacrificed, lungs were filled with zinc buffered formalin and stored in the same buffer before embedding into paraffin blocks. Sections were cut at 4 μm, adhered to positively charged slides and dried over night. Sections were dewaxed with Slide Brite, rehydrated and stained with hematoxylin (Thermofisher), differentiated with acid alcohol, blued with Scott's water and stained with eosin (Thermofisher). Slides were scanned with ZEISS AxioScan Z1 slide scanner using 10x/0.3NA or 20x/0.8NA objective.

<u>Klf mice</u>
For *Lys2^{Cre} Klf2$^{flox/flox}$* and *Lys2^{Cre} Klf4$^{flox/flox}$* studies bone marrow (femur, tibia and hip bone) and Cre negative littermate controls were harvested, cleaned and shipped overnight on wet ice in BMM (RPMI, 1% Hepes, 1% Anti/Anti (Gibco), 1% BME (Gibco), 1% NEAA (Gibco), 1% Sodium Pyruvate (Gibco), 10% FCS). Bone marrow was harvested by scraping bones clean and immersing in 70% ethanol for 10 seconds before extracting marrow by centrifugation at 5,900g for 15 seconds in a 1.5ml eppendorf tube. Bone marrow was immediately resuspended in room temperature sterile PBS and injected into lethally irradiated recipient mice (2 x 600 rads) by retroorbital injection at a ratio of 3:1 (recipient to donor). Mice were allowed to reconstitute for at least 6 weeks prior to sacrifice and analysis.

Table S1, related to experimental procedures. Details for cloning *Nr4a1* enhancer candidates into PGL4.10 luciferase reporter

Plasmid	Enh_F_primer_seq	Enh_R_primer_seq	PCR protocol	Anneal temp (oC)	Digetion	Com cells
pGL4_Nr4a1_E12	ATATGGATCCCAATGTTGGGTCTCTTTCTCAATTAGTTGC	ATATGTCGACAAATCGCTGTGGTTTGAATGCCA	Phusion HF	72	Bamh1 + Sall	Top10 OneShot
pGL4_Nr4a1_E11	ATATGGATCCGATTGATGTGGGAGGCCAGGGTT	ATATGTCGACGTGTGCACTACCATGTCCAGCATG	Phusion HF	72	Bamh1 + Sall	NEB Stable
pGL4_Nr4a1_E10	ATATGGATCCAACAAAAGCAAAACACTGTTTCATTAG CGG	ATATGTCGACGGGGAGGAACCAAGCTACCCAGGA	Phusion HF	72	Bamh1 + Sall	NEB Stable
pGL4_Nr4a1_E09	ATATGGATCCCTAGACTGGGAGTTAATGACGGTCGTG A	ATATGTCGACAGAAAGATTACCCACAAAATCAAAAACCAGGGGCT	LA Taq GC 1	66	Bamh1 + Sall	Top10 OneShot
pGL4_Nr4a1_E08	ATATGGATCCGTATATGAGTACACTGTAGCTCTTTTCA GACACAC	ATATGTCGACGGACAGCTTAAAGAGACAGGCTGAGAT	Phusion HF	70.4	Bamh1 + Sall	NEB Stable
pGL4_Nr4a1_E07	ATATGGATCCCCCAGAGCACGGGCTAAGGGGT	ATATGTCGACCATCCTGTCCTGAGCAAGCCCTT	LA Taq GC 1	66	Bamh1 + Sall	Top10 OneShot
pGL4_Nr4a1_E06	ATATGGATCCTTCATGAGACATTATACCATCTCACATC T	ATATGTCGACTGAGGGGATTCATCCATGCAGA	LA Taq GC 1	60	Bamh1 + Sall	Top10 OneShot
pGL4_Nr4a1_E05	ATATGGATCCGCTCAGAAGAAAAGACAGTGTACGGT G	ATATGTCGACGTTTGTTTGTTTGTTTTTCGAGACAGGGGTTT CT	LA Taq GC 1	66	Bamh1 + Sall	Top10 OneShot
pGL4_Nr4a1_E04	ATATGGATCCGACAGCGGGGATGGGATAAAGACACCTG	ATATCTCGAGTCTGCAATCCCAGTGTGTCAGGAGAC	Phusion HF	72	Bamh1 + Xhol	NEB Stable
pGL4_Nr4a1_E03	ATATGGATCCCAGGGAGGGCAGTGTGGGCTGAA	ATATGTCGACGTCTAGCTACCTCCATGGAAACTCTGCACC	Phusion HF	72	Bamh1 + Sall	Top10 OneShot
pGL4_Nr4a1_E02	ATATGGATCCCATGGGACCTGGCCAGGTTTCA	ATATGTCGACATTCTCCCTCCATATATACATCTGTTCTATCGAC AG	Phusion HF	72	Bamh1 + Sall	Top10 OneShot
pGL4_Nr4a1_E01	ATATGGATCCGGCTGGCAGCAGAAATCGGGGAA	ATATGTCGACCGACCACGAGGAGGAGGGGGTGTT	Phusion HF	72	Bamh1 + Sall	Top10 OneShot
pGL4_Nr4a1_TSS	ATATGGTACCGAAGGCCAGAGAGTGCCTGTCC	ATATGAGCTCTCCCACTCCCTCGTGGCCG	Phusion HF	68.5	Kpnl + Sacl	Top10 OneShot

sgRNA name	Enhancer region	Score	Guide sequence (genomic)	crRNA sequence (inc G substitution for T7 promoter)
E02_US_1	E02	80	GTGAACTGAACTCCCCACCG	GTGAACTGAACTCCCCACCG
E02_US_2	E02	79	GCGCTGAGATATATGAATGC	GCGCTGAGATATATGAATGC
E02_DS_1	E02	82	GGGCGGGGCGGTTCCTGATT	GGGCGGGGCGGTTCCTGATT
E02_DS_2	E02	78	GCAGCAGGGTCAGCGTGAAC	GCAGCAGGGTCAGCGTGAAC
E06_US_1	E06	88	TGCTTAGGCACGGTAGTCAT	GGCTTAGGCACGGTAGTCAT
E06_US_2	E06	75	TCTGGTCTGGTCACTACAAA	GCTGGTCTGGTCACTACAAA
E06_DS_1	E06	83	GTGATCTAACACACCCCCCT	GTGATCTAACACACCCCCCT
E06_DS_2	E06	80	GGGTTTGGGGCTAGTGTAAT	GGGTTTGGGGCTAGTGTAAT
E09_US_1	E09	73	GGGGTTTGACCTGAGCCATC	GGGGTTTGACCTGAGCCATC
E09_US_2	E09	72	GAGCTTTTGGTGTCTTGACC	GAGCTTTTGGTGTCTTGACC
E09_DS_1	E09	73	GGAGGGGTTAACTAACCCAC	GGAGGGGTTAACTAACCCAC
E09_DS_2	E09	72	GGATCAATAACTACTTGGCT	GGATCAATAACTACTTGGCT
E04_07_US_1	E04_E07	77	TAGCCATCTCCCAGTCAAGC	GAGCCATCTCCCAGTCAAGC
E04_07_US_2	E04_E07	76	ATGGACCCTTACTCCCAAAT	GTGGACCCTTACTCCCAAAT
E04_07_DS_1	E04_E07	94	GAGGTGAAGGGTCCCAATCG	GAGGTGAAGGGTCCCAATCG
E04_07_DS_2	E04_E07	77	CTGCGTTTTAAGCCTTATAA	GTGCGTTTTAAGCCTTATAA

Table S2, related to experimental procedures. crRNA design and sequences. sgRNAs were ordered as ultramers composed of an upstream T7 promoter (green), crRNA (red) and invariant downstream tracrRNA sequence (blue) as shown:

TAATACGACTCACTATAGGTGAACTGAACTCCCCACCGGTTTTAGAGCTAGAAATAGCAAGTTA
AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT

155

Mouse	F primer	R primer	Polymerase kit	anneal (oC)	extend (min)	WT product size	KO product size
Nr4a1se_9	GCATCTCTGCTCCCCACTTT	CAGTAAGCCACCTTGAGCCA	NBE Phusion HF	68	3	2.5kb	1kb
Nr4a1e_6	GGCTCCCAGTGTGACCTTTT	CCTGAACGCCTGAGCTAACA	LA Taq GCI (Clontech)	58	2	1.8kb	500bp
Nr4a1se_2	CTGAGGGCTCCTTATCGGGGA	CTGAATGCCCAAAACGCACC	NBE Phusion HF	68	3	2.6kb	1kb

Table S3, related to experimental procedures. PCR details for enhancer knockout mouse genotyping.

156

Chapter 2, in part, is a reprint of the material as it appears in "*Nxf1 Natural Variant E610G Is a Semi-dominant Suppressor of IAP-Induced RNA Processing Defects*". Concepcion, Dorothy; Ross, Kevin; Hutt, Kasey R.; Yeo, Gene; Hamilton, Bruce A., PLoS Genetics, 2015. The dissertation author was a primary investigator and author of this paper.

Chapter 2, also in part, is a reprint of the material as it appears in "*Nmf9 Encodes a Highly Conserved Protein Important to Neurological Function in Mice and Flies*". Zhang, Shuxiao; Ross, Kevin D.; Seidner, Glen A.; Gorman, Michael R.; Poon, Tiffany H.; Wang, Xiaobo; Keithley, Elizabeth M.; Lee, Patricia N.; Martindale, Mark Q.; Joiner, William J.; Hamilton, Bruce A., PLoS Genetics, 2015. The dissertation author was a primary investigator and author of this paper.

Chapter 2, also in part, is a reprint of the material as it appears in "Deleting an *Nr4a1* Super-Enhancer Subdomain Ablates Ly6C^{low} Monocytes while Preserving Macrophage Gene Function". Thomas, Graham D.; Hanna, Richard N.; Vasudevan, Neelakatan T.; Hamers, Anouk A.; Romanoski, Casey E.; McArdle, Sara; Ross, Kevin D.; Blatchley, Amy; Yoakum, Deborah; Hamilton, Bruce A.; Mikulski, Zbigniew; Jain, Mukesh K.; Glass, Christopher K.; Hedrick, Catherine C., Immunity, 2016. The dissertation author was a co-author of this paper and

provided technical insight towards the generation of CRISPR/Cas9 edited mice.

Chapter 2, also in part is a reprint of the material as it appears in "The lipid elongation enzyme ELOVL2 is a molecular regulator of aging in the retina". Chen, Daniel; Chao, Daniel L.; Rocha, Lorena; Kolar, Matthew; Anh Nguyen Huu,Viet; Krawczyk, Michal; Dasyani, Manish; Wang, Tina; Jafari, Maryam; Jabari, Mary; Ross, Kevin D.; Saghatelian, Alan; Hamilton, Bruce A.; Zhang, Kang; Skowronska-Krawczyk, Dorota, Aging Cell, 2020. The dissertation author was a co-author and co-investigator of this paper, and designed and generated the Elovl2[C234W] mutant mice.

CHAPTER 3: *ANKFN1* VESTIBULAR DEFECTS IN ZEBRAFISH REQUIRE MUTATIONS IN BOTH ANCESTRAL AND DERIVED PARALOGS.

Kevin D. Ross[1], Jie Ren[2], Ruilin Zhang[2], Neil C. Chi[1,2,3,4,5] and Bruce A. Hamilton[1,2,3,4,5,6]

[1] Biomedical Sciences Graduate Program

[2] Department of Medicine

[3] Rebecca and John Moores UCSD Cancer Center

[4] Institute for Engineering in Medicine

[5] Institute for Genomic Medicine

[6] Department of Cellular and Molecular Medicine

University of California San Diego

La Jolla, CA 92093

ABSTRACT

How and to what degree gene duplication events create regulatory innovation, redundancy, or neofunctionalization remain important questions in animal evolution and comparative genetics. *Ankfn1* orthologs are single copy in most invertebrates, partially duplicated in jawed vertebrates, and the original copy lost in mammals. Null mutations in the single mouse homolog have vestibular and neurological abnormalities. Null mutation of the single Drosophila homolog is typically lethal with severe sensorimotor deficits in rare survivors. The functions and potential redundancy of paralogs in species with two copies is not known. Here we define a vestibular role for *Ankfn1* homologs in zebrafish based on simultaneous disruption of each locus. Zebrafish with both paralogs disrupted showed vestibular defects and early lethality from swim bladder inflation failure. One intact copy at either locus was sufficient to prevent major phenotypes. Our results show that vertebrate *Ankfn1* genes are required for vestibular-related functions, with at least partial redundancy between ancestral and derived paralogs.

INTRODUCTION

Ankfn1 is a recently-annotated gene with an interesting genealogy(Zhang et al. 2015). Orthologous genes are recognized by the

160

eponymous ankyrin (ANK) and fibronectin type-III (FN3) motifs and three highly conserved non-motif domains. Orthologs are found in all animal lineages studied to date except urochordates and in some sister groups to animals, including choanoflagellates and filastereans. Some unicellular examples encode an aminoterminal CRIB domain. Some unicellular and nearly all invertebrate homologs include a carboxyterminal Ras association (RA) domain. The orthology group shows a single member per genome outside of vertebrates. In an ancestor to jawed vertebrate, the ancestral gene was incompletely duplicated, with the derived paralog losing the RA domain. In an ancestor to mammals, the ancestral copy was lost. All animal genomes analyzed to date thus have zero (urochordates), one ancestral (all other invertebrates), one derived (mammals), or two (non-mammal vertebrates) *Ankfn1* genes.

Ankfn1 genes were identified independently by three groups using forward genetics in flies and mice. Transposon insertion mutations in *Drosophila* (*wake* alleles) were recovered in a screen for sleep-related phenotypes (Liu et al. 2014). Although originally reported as null, later analyses showed that these were isoform-specific alleles, which prevented expression in some tissues. An RNAi screen to identify genes required for asymmetric cell division in *Drosophila* sensory organ precursor cells

identified the same gene as an essential regulator of Numb segregation during cell division; null mutations (*Banderoula* alleles) had severe developmental consequences and poor viability (Mauri et al. 2014). Banderoula interacted with Discs-large, and inhibition of both genes elicited tumorigenesis in the *Drosophila* brain. We identified a presumptive null allele of the mouse homolog by positional cloning of a neurological mutation (*nmf9*) marked by vestibular-related neurological phenotypes (Zhang et al. 2015). Subsequent alleles made by genome editing confirmed gene identification and showed the functional importance of a non-motif domain including a highly conserved GLYLGYLK peptide sequence, with even a glycine-to-alanine substitution failing to complement the original null. *Nmf9* mice were viable and fertile but had several neurological abnormalities–including defects in circadian onset, fear learning, and vestibular function. We also showed in *Drosophila* that edited mutations in distinct domains had *Banderoula*-like phenotypes. The few flies that survived to adulthood had profound sensorimotor deficits and died prematurely. Flies heterozygous for null mutations had abnormal sleep patterns, broadly consistent with *wake* mutants and demonstrating dosage sensitivity for some phenotypes.

Differences between flies and mice both in domain architecture and in phenotypic severity raise questions about essentiality and conservation of *Ankfn1* function in other animals, especially vertebrate groups that have two paralogous copies. This is relatively common among vertebrate genomes, with approximately 15% of human genes having more than two identifiable orthologs in zebrafish(Howe et al. 2013)**.** Paralogous genes can have variable functional outcomes. Full redundancy can result in one copy decaying into a pseudogene that has lost function relative to the ancestral gene (Mighell et al. 2000). Duplication can also lead to new or divergent functions, through neofunctionalization or subfunctionalization (Force et al. 1999). The facile genetics of the zebrafish *Danio rerio* makes it an ideal system to test *Ankfn1* paralogous gene function in the typical vertebrate arrangement of one ancestral and one derived copy. The ancestral copy with an intact RA domain is on chromosome 24, while the derived copy without RA is on chromosome 12. Assessing null alleles of each homolog independently and in combination should allow direct tests of essentiality, functional redundancy, and possible neofunctionalization after duplication (Ohno 1970).

Here we used CRISPR/Cas9-mediated editing to generate mutations in both *Ankfn1* homologs of *D. rerio* in order to define major

163

functions and degree of genetic redundancy between paralogs in a non-mammal vertebrate. We generated a frameshifting deletion allele of the chromosome 12 homolog, and both a frameshifting insertion allele and an in-frame deletion allele at the GLYLGYLK site of the chromosome 24 copy. We observed mutant progeny among digenic crosses and identified an overt swim bladder phenotype that was predominant in fish with biallelic inactivation at both loci but not in fish with at least one intact copy at either locus. Affected fish did not inflate their swim bladder by 5 days post-fertilization, had severe locomotor and balance deficits, and died prior to adulthood. These results suggest genetic overlap between the ancestral and derived homologs for essential functions, including development of the vestibular system.

RESULTS

Simultaneous editing at the most-constrained sequence in both *Ankfn1* homologs of zebrafish.

Evolutionary constraint analysis of *Ankfn1* homologs across diverse lineages identified strong conservation of ANK, FN3, and three non-motif domains (Zhang et al. 2015). We conducted a multivariate analysis of protein polymorphism (MAPP) (Stone and Sidow 2005) to quantify the extent of physicochemical constraint among 115 homologs of *Ankfn1*

164

(Figure 1A, Table S1). Several of the most constrained positions in Ankfn1 fall outside of the annotated domains that characterize the protein. Of 52 positions in the top 5% of scores, 2 are in the ANK domain (107 residues long), 3 are in FN3 (166 residues), 1 is in the RA domain of ancestral forms (175 residues), and 6 are in the GLYLGYLK peptide (8 residues). This conserved domain is the only region with more than three consecutive positions in the top 5% of scores (Figure 1B). Consistent with this result, mutations of this sequence in mice failed to complement the original *nmf9* allele (Zhang et al. 2015). We therefore targeted the GLYLGYLK peptide site for mutagenesis because it is both highly constrained and previously targeted in flies with a single ancestral homolog and in mice with a single derived homolog, and should therefore be similarly essential to each paralog in zebrafish. We co-injected single-guide RNAs targeting the sequence encoding the GLYLGYLK peptide of each *D. rerio* paralog with Cas9 mRNA to allow simultaneous editing at both loci (Figure 1C). Injected G0 founders were outcrossed to wildtype fish and gDNA was prepared from G1 offspring to screen for germline-transmitted mutations. Among 60-90 injected offspring, we recovered a single frameshifting 11-bp deletion allele on chromosome 12 and both a frame shifting 23-bp insertion and an in-frame 18-bp deletion allele on

chromosome 24. G1 fish from verified founders were raised to sexual

maturity, genotyped from fin clip biopsies, and then intercrossed to

produce animals for behavioral phenotyping (Talbot and Amacher 2014).

Figure 3.1: Constraint and targeted mutagenesis of *Ankfn1* homologs. (A) Multivariate analysis of protein polymorphism (MAPP) scores quantify the degree of physicochemical constraint for 115 homologs of *Ankfn1* in species ranging from fungus to mammals. Colored bars underneath the MAPP Score plot indicate features of *Ankfn1*: ankyrin repeats (orange and red), fibronectin type-III domain (green), conserved GLYLGYLK peptide (pink), variably included C-terminal RA domain (salmon). The horizontal red line indicates the top 5% of scores. (B) The GLYLGYLK motif (pink) exhibits the highest physicochemical constraint of any peptide sequence among 115 *Ankfn1* homologs. (C) Peptide and genomic sequences for *D. rerio* homologs of *Ankfn1*. Reference sequence (wt) shows CRISPR guide RNA sites underlined and PAM sequence boxed. Recovered mutations are shown with deleted base pairs as dashes and insertion of a duplicated sequence at chromosome 24 shown in a box below the caret.

A 115 Homologs

MAPP Score

30
20
10
0

500   1000   1500   2000   2500
Aligned Position

B

MAPP Score

30
20
10
0

bits
4
2
0

P G L Y L G Y L K   L S S S V D Q   R V L V P   K L   P N L P   H K R   D N N V S E E   E L Q

1620   1630   1640   1650   1660   1670
Aligned Position

C

```
Protein   G L Y L G Y L K L S S S V D Q I K V L V P Q
Ch12 wt   GGCCTTTATCTGGGTTACCTGAAACTCAGCAGCTCTGTGGATCAGATCAAGGTTCTGGTTCCTCAG
-11bp allele  GG-----------GTTACCTGAAACTCAGCAGCTCTGTGGATCAGATCAAGGTTCTGGTTCCTCAG


Protein   G L Y L G Y L K L Y T S V D Q I R V L V S H
Ch24 wt   GGTCTGTATCTGGGTTACTTGAAGCTCTACACCTCAGTGGACCAGATCCGAGTCCTGGTGTCACAC
-18bp allele  GGTCTGTATCTGGGTTACT------------------TGGACCAGATCCGAGTCCTGGTGTCACAC
+23bp allele  GGTCTGTATCTGGGTTACTTGAAGCTCTACACCTCAGTGGACCAGATCCGAGTCCTGGTGTCACAC
                                                CAGATCCGAGTCCTGGTGTCACA
```

Highly penetrant major phenotypes require biallelic inactivation of both Ankfn1 homologs.

A series of dihybrid crosses demonstrated strong effects on swim bladder inflation, vestibular function, and survival to adulthood selectively in animals that had both alleles of each *Ankfn1* homolog mutated. An initial cross to assess gross behavioral and physiological phenotypes in mutant offspring showed a minority of larvae with balance and locomotor deficits prior to failure of swim bladder inflation. Larvae whose swim bladders failed to inflate died within two weeks after swim bladder inflation and were not available for subsequent DNA extraction. Larvae from each subsequent cross were phenotyped based on inflation of the swim bladder and genotyped by PCR. Among several cross designs, fish with both genes inactivated displayed abnormal balance, abnormal escape response to touch, and subsequent failure of swim bladder inflation (Figure 2). Double mutant animals failed to maintain a dorsal-up posture (Figure 2).

Genotype ratios of the offspring did not deviate from normal expectations. Among 289 offspring accumulated from four crosses all double mutants, with one exception, failed to inflate their swim bladders within five days post-fertilization (Tables 1-3).
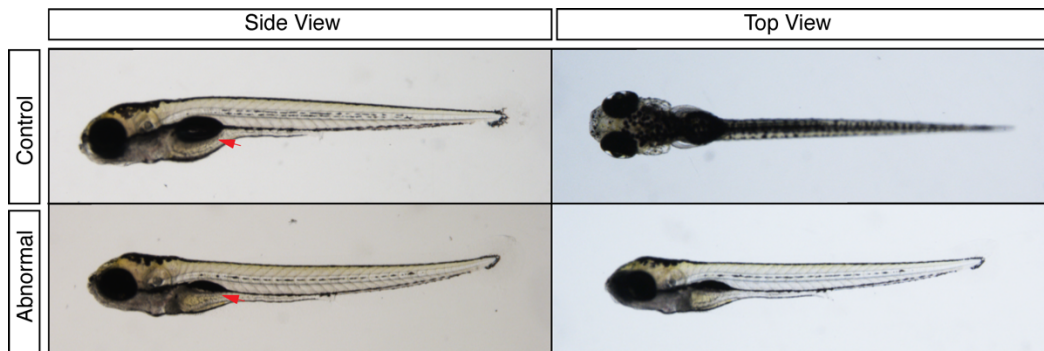
Figure 3.2: Swim bladder and posture defects in progeny of *Chr12/Chr24* mutant zebrafish five days post-fertilization. Swim bladders, indicated by red arrows, in phenotypically normal (top) and abnormal (bottom) zebrafish embryos. Top views (right panels) of normal and abnormal fish demonstrate the posture defect of phenotypically abnormal fish.

All animals that were not double mutants, also with one exception, inflated their swim bladders within five days after fertilization. Both exceptions were re-assessed for genotype and had both target loci resequenced. The asymptomatic double mutant appeared to escape major phenotype despite having both loci mutated (Table 4), but was homozygous for the in-frame Δ18 allele at chromosome 24, which could have residual function. The symptomatic exception was homozygous for the Δ11 allele on chromosome 12, but had no evident mutation at the chromosome 24 locus.

In-frame deletion of the highly conserved peptide is a strong loss-of-function allele.

A summary by genotype of all cross results (Table 4) showed that despite one exceptional animal, an in-frame 18-bp deletion overlapping the GLYLGYLK sequence in the chromosome 24 homolog is a strong loss-of-function allele. This allele deletes the lysine (K) from this sequence and the next 5 residues. Summing all crosses, 12 of 13 progeny homozygous for both chromosome 12 frameshift and chromosome 24 in-frame deletions and all 17 progeny homozygous for chromosome 12 frameshift and trans-heterozygous for chromosome 24 frameshift and in-frame deletions showed phenotypes that were not distinct from the 10 progeny that were homozygous for the frameshift allele at each locus. This complements earlier data in mice showing more-proximal in-frame deletions and a substitution mutation that disrupted the GLYLGYLK motif failed to complement the presumptive null *nmf9* splice donor mutation (Zhang et al. 2015)

DISCUSSION

What happens to paralogous copies after gene duplication across broad time scales is an interesting question, of which Ankfn1 is an interesting example. All invertebrate lineages we have examined have one copy, except for urochordates, which have none. All vertebrate lineages have both an ancestral Ankfn1 homolog (including the RA domain) and a

derived Ankfn1 homolog (lacking the RA domain), except mammals. The

two-homolog arrangement has been conserved across non-mammal

lineages since it arose between the most recent common ancestor of

jawed vertebrates ~465 Mya and the most recent common ancestor of all

vertebrates ~615 Mya (Kumar et al. 2017). The single ancestral gene is

essential in Drosophila, while the single derived gene impacts vestibular

and neurological phenotypes in mice. We used *D. rerio* as a model

vertebrate for a first look at genetic properties of these genes in a

vertebrate with paralogous copies. While evolution acts on a much finer

scale than laboratory phenotypes and single locus deletions may have

subtler or less-penetrant phenotypes than we had power to see, we found

that inactivation of either paralog alone was insufficient to produce major

phenotypes at high penetrance, while biallelic inactivation of both *Danio*

Ankfn1 homologs revealed striking phenotypes.

Our analysis of *Ankfn1* paralogs in zebrafish supports functional

overlap of *Ankfn1* paralogs in normal development of the vestibular

system. We directed mutations to the most highly conserved *Ankfn1*

domain of each gene and recovered both presumptive null frameshift

alleles and, for chromosome 24, an in-frame deletion that destroys the

GLYLGYLK consensus sequence. By design, these mutations mirrored

alleles previously reported for single homologs in mice and flies(Zhang et al. 2015). We found a highly penetrant vestibular-related phenotype in animals homozygous for inactivating mutations in each homolog. We did not see high penetrance of vestibular or other obvious phenotypes in other genotype combinations. The present study is limited in that we did not perform detailed phenotyping of inner ear structures or perform expression level or localization measurements to determine whether disruption of one paralog resulted in increased expression of the other as a form of dosage compensation(Diss et al. 2014). Whether more subtle phenotypes occur with each mutant singly will require further investigation. We had limited power to assess the potential for low-penetrance phenotypes in 3-allele combinations and whether the single affected animal we observed represented a low penetrance of an *Ankfn1*-dependent phenotype or was developmentally abnormal for other reasons.

Few examples of paralogous gene pairs in zebrafish for which disruption of both copies creates a phenotype that differs from knockout of either paralog alone have been published (Kettleborough et al. 2013). Our results provide a striking example, with a gross morphological phenotype arising from disruption of both members of a paralogous gene pair without an overt phenotype in single disruption of either member alone. By

generating single and double mutations, we demonstrated that despite

differences in their domain structure, major phenotypes in swim bladder

inflation and vestibular function require inactivation of both copies,

showing substantial genetic redundancy between paralogous genes,

despite loss of the RA domain in the derived copy and maintenance of

independent paralogs across nearly all lineages during roughly half a

billion years of vertebrate evolution.

MATERIALS AND METHODS

Genome editing.

Mutant fish were generated by injecting a cocktail of sgRNAs and

Cas9 mRNA into one-cell-stage AB strain embryos (Jao, Wente, and

Chen 2013). Ankfn1-homolog sgRNAs were in-vitro-transcribed from

PCR-amplified templates (Table S2). Synthetic Cas9 mRNA was

transcribed from pCS2-nCas9n (gift of Wenbiao Chen, Addgene plasmid #

47929). The injection cocktail contained 200 ng/ul Chr12 sgRNA, 300

ng/ul Chr24 sgRNA, and 300 ng/ul Cas9 mRNA. Founder mutations were

identified by pooling G1 embryos from individual G0 injected fish and

screening for length polymorphisms by PCR. PCR products from putative

mutants were Sanger sequenced to identify specific alleles.

Breeding and epistasis.

G0 founders were outcrossed to wildtype fish of G1 offspring. Animals in cross 1 were G2 offspring from an intercross of G1 mutants. Crosses 2 and 3 were G4 and G5 fish from intercross matings. After phenotyping, single zebrafish larvae were harvested into individual wells in DNA extraction buffer for use in genotyping assays. 1 μl from a 50 μl suspension was used for each 20 μl PCR reaction. Primers for genotyping are given in Table S2.

Behavioral assays.

Balance was assayed by the ability of larvae to rest with their dorsal sides up 1 min after initiating a startle response by tapping Petri dishes containing larvae. To test motor coordination, individual larvae were observed after a startle response, induced by gentle physical stimulation of the tail. Normal larvae rapidly escape, but larvae with vestibular dysfunction do not.

Acknowledgements

Chapter 3, "*Ankfn1* vestibular defects in zebrafish require mutations in both ancestral and derived paralogs" is being prepared for publication and is coauthored with Ren, Jie; Zhang, Ruilin; Chi, Neil and Hamilton, Bruce. The dissertation author was the primary author of this chapter.

Table 3.1: Cross 1: $Chr12^{-11/+}$ $Chr24^{+23/+}$ x $Chr12^{-11/+}$ $Chr24^{-18/+}$

Chr12	Chr24	Unaffected	Affected	Total
-11/-11	+23/-18	0	4	4
-11/-11	+23/+	9	0	9
-11/-11	-18/+	5	0	5
-11/-11	+/+	7	1	8
-11/+	+23/-18	7	0	7
-11/+	+23/+	7	0	7
-11/+	-18/+	14	0	14
-11/+	+/+	7	0	7
+/+	+23/-18	4	0	4
+/+	+23/+	8	0	8
+/+	-18/+	4	0	4
+/+	+/+	5	0	5
	Total	77	5	82

Table 3.2: Cross 2: Chr12$^{-11/+}$ Chr24$^{+23/+}$ x Chr12$^{-11/-11}$ Chr24$^{-18/+}$

Chr12	Chr24	Unaffected	Affected	Total
-11/-11	+23/+23	0	10	10
-11/-11	+23/+	10	0	10
-11/-11	+/+	5	0	5
-11/+	+23/+23	6	0	6
-11/+	+23/+	14	0	14
-11/+	+/+	8	0	8
	Total	43	10	53

Table 3.3: Cross 3. Chr12$^{-11/+}$ Chr24$^{+18/+}$ x Chr12$^{-11/+}$ Chr24$^{-18/+23}$

Chr12	Chr24	Unaffected	Affected	Total
-11/-11	+23/-18	0	13	13
-11/-11	-18/-18	1	12	13
-11/-11	+23/+	10	0	10
-11/-11	-18/+	9	0	9
-11/+	+23/-18	15	0	15
-11/+	-18/-18	15	0	15
-11/+	+23/+	16	0	16
-11/+	-18/+	25	0	25
+/+	+23/-18	14	0	14
+/+	-18/-18	8	0	8
+/+	+23/+	8	0	8
+/+	-18/+	8	0	8
	Total	129	25	154

Table 3.4: Summary by genotype for all crosses.

Chr12	Chr24	Unaffected	Affected	Total
-11/-11	+23/+23	0	10	10
-11/-11	+23/-18	0	17	17
-11/-11	-18/-18	1	12	13
-11/-11	+23/+	29	0	29
-11/-11	-18/+	14	0	14
-11/-11	+/+	12	1	13
-11/+	+23/+23	6	0	6
-11/+	+23/-18	22	0	22
-11/+	-18/-18	15	0	15
-11/+	+23/+	37	0	37
-11/+	-18/+	39	0	39
-11/+	+/+	15	0	15
+/+	+23/-18	18	0	18
+/+	-18/-18	8	0	8
+/+	+23/+	16	0	16
+/+	-18/+	12	0	12
+/+	+/+	5	0	5
	Total	249	40	289

REFERENCES

Diss, Guillaume, Diana Ascencio, Alexander Deluna, and Christian R. Landry. 2014. "Molecular Mechanisms of Paralogous Compensation and the Robustness of Cellular Networks." *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. https://doi.org/10.1002/jez.b.22555.

Force, Allan, Michael Lynch, F. Bryan Pickett, Angel Amores, Yi Lin Yan, and John Postlethwait. 1999. "Preservation of Duplicate Genes by Complementary, Degenerative Mutations." *Genetics*.

Howe, Kerstin, Matthew D. Clark, Carlos F. Torroja, James Torrance, Camille Berthelot, Matthieu Muffato, John E. Collins, et al. 2013. "The Zebrafish Reference Genome Sequence and Its Relationship to the Human Genome." *Nature*. https://doi.org/10.1038/nature12111.

Jao, Li En, Susan R. Wente, and Wenbiao Chen. 2013. "Efficient Multiplex Biallelic Zebrafish Genome Editing Using a CRISPR Nuclease System." *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1308335110.

Kettleborough, Ross N.W., Elisabeth M. Busch-Nentwich, Steven A. Harvey, Christopher M. Dooley, Ewart De Bruijn, Freek Van Eeden, Ian Sealy, et al. 2013. "A Systematic Genome-Wide Analysis of Zebrafish Protein-Coding Gene Function." *Nature*. https://doi.org/10.1038/nature11992.

Kumar, Sudhir, Glen Stecher, Michael Suleski, and S. Blair Hedges. 2017. "TimeTree: A Resource for Timelines, Timetrees, and Divergence Times." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msx116.

Liu, Sha, Angelique Lamaze, Qili Liu, Masashi Tabuchi, Yong Yang, Melissa Fowler, Rajnish Bharadwaj, et al. 2014. "WIDE AWAKE Mediates the Circadian Timing of Sleep Onset." *Neuron*. https://doi.org/10.1016/j.neuron.2014.01.040.

Mauri, Federico, Ilka Reichardt, Jennifer L. Mummery-Widmer, Masakazu

Yamazaki, and Juergen A. Knoblich. 2014. "The Conserved Discs-Large Binding Partner Banderuola Regulates Asymmetric Cell Division in Drosophila." *Current Biology*. https://doi.org/10.1016/j.cub.2014.06.059.

Mighell, A. J., N. R. Smith, P. A. Robinson, and A. F. Markham. 2000. "Vertebrate Pseudogenes." *FEBS Letters*. https://doi.org/10.1016/S0014-5793(00)01199-6.

Ohno, Susumu. 1970. *Evolution by Gene Duplication. Evolution by Gene Duplication*. https://doi.org/10.1007/978-3-642-86659-3.

Stone, Eric A., and Arend Sidow. 2005. "Physicochemical Constraint Violation by Missense Substitutions Mediates Impairment of Protein Function and Disease Severity." *Genome Research*. https://doi.org/10.1101/gr.3804205.

Talbot, Jared Coffin, and Sharon L. Amacher. 2014. "A Streamlined CRISPR Pipeline to Reliably Generate Zebrafish Frameshifting Alleles." *Zebrafish*. https://doi.org/10.1089/zeb.2014.1047.

Zhang, S., K.D. Ross, G.A. Seidner, M.R. Gorman, T.H. Poon, X. Wang, E.M. Keithley, et al. 2015. "Nmf9 Encodes a Highly Conserved Protein Important to Neurological Function in Mice and Flies." *PLoS Genetics* 11 (7). https://doi.org/10.1371/journal.pgen.1005344.

CHAPTER 4: DISCUSSION

Nxf1 **Natural Variant E610G Is a Semi-dominant Suppressor of IAP-Induced RNA Processing Defects**

We demonstrated that a congenic $Nxf1^{CAST}$ allele modifies a diverse range of IAP insertion mutations, including full length and IΔ1 isoforms. IAPs are highly polymorphic, including large structural and single nucleotide variations, and occur in varied genomic contexts, and we found that suppression is limited to a subset of sense-oriented, intronic insertions. Further, we did not find that Nxf1 had a global allele-specific effect on alternative processing.

The suppressive $Nxf1^{CAST}$ allele differs from C57B6/J at two amino acid-coding positions, one of which, E610G, occurs in the highly conserved UBA domain, which mediates interactions with FG repeats of the nuclear pore complex and interactions with other partner proteins. We used CRISPR/Cas9 genome editing to target this position 610 in a B6 strain background, generating a single amino acid mutation, in addition to a pseudo-edited control mutation. The E610G allele suppresses the molecular and neurological phenotypes caused by the *vb* mutation, accounting for the modifier effect of the $Nxf1^{CAST}$ congenic allele. Further,

it replicates the suppression of expression defects at other IAP-inserted loci.

Our Results in Context

The prevailing view of Nxf1 is as an mRNA export factor which licenses transcripts for export through an association with Nxt1, and facilitates nucleocytoplasmic transport via interactions with FG-repeats of the nuclear pore complex (NPC) (Xie and Ren 2019; Scott et al. 2019; Lesbirel and Wilson 2019; Björk and Wieslander 2017). However, mounting evidence, including our work, lends credence to the idea that Nxf1 has additional and/or more nuanced function.

iCLIP experiments of Alyref, Chtop and Nxf1 suggest co-transcriptional loading of these factors during splicing, prior to capping and polyadenylation (Viphakone et al. 2019). In particular, Chtop appears to participate in alternative polyadenylation, and interacts directly with the NTF2 domain of Nxf1, and may be the subject of a differential interaction depending on the residue at the position 610 (Chang et al. 2013). Imaging studies detect Nxf1 at the majority of nuclear pore complexes, which may not be surprising if all complexes are participating in export. However, FRAP data suggests that the turnover time for these signals is much

slower than expected based on the kinetics of mRNA export (Ben-Yishay et al. 2019).

Evidence suggests that Nxf1 can directly impact RNA Pol II dynamics (Chen et al. 2019). RNAi knockdown of Nxf1 in HeLa cells altered global patterns of alternative polyadenylation as determined by 3' UTR sequences in total RNA samples. Additionally, ChIP-seq for RNA Pol II after NXF1 knockdown revealed a shift in Pol II occupancy from transcription start sites towards gene bodies, with associated shifts in nascent transcript levels from those regions. Chen et al. further showed a partially RNAse A-resistant physical association between FLAG-Nxf1 and RNA Pol II. Although this evidence is attractive as a potential model for Nxf1-mediated suppression of IAPs, questions about secondary effects on transcription/export dynamics stemming from Nxf1 knockdown remain when interpreting this data (Chen et al. 2019). Interplay between splicing factors and Nxf1 can also influence the export of specific transcripts. iCLIP experiments suggest co-binding of SR proteins, most potently SRSF3 and SRSF7, to different exonic regions, regulating differential 3' UTR length, and potentially altering the export of specific transcripts (Müller-McNicoll et al. 2016).

There have been additional recent reports of Nxf1 function in response to environmental queues, chemical modifications to RNAs, and to unique RNA isoforms. In response to LPS-induced inflammation, IRAK2 phosphorylates SRSF1, decreasing its affinity for target mRNAs which can be bound by ALYREF and NXF1 for export (Zhou et al. 2017). Interferon regulatory factor 5 (IRF5) complexes with Nxf1 in response to TLR7 ligand stimulation to promote IRF5 transcriptional activity (Fu et al. 2017). Chemical modifications like N6-methyladenosine may subject certain transcripts to differential regulation mediated by Nxf1 and cofactors like SRSF3 (Roundtree et al. 2017; Lesbirel and Wilson 2019). Stable lariat RNA molecules, intronic RNAs which remains following excision, can be transported by NXF1/NXT1 (Talhouarne and Gall 2018). Modifications to interacting partners can regulate the influence of Nxf1 on mRNA export (P. Li et al. 2017). SR proteins have been shown as potentially important regulators of Nxf1-mediated mRNA export in specific cases. Depletion of SRSF1 or disruption of its interaction with Nxf1 in a *Drosophila* model of neurodegeneration caused by a hexanucleotide-repeat expansion alleviates neurotoxicity (Hautbergue et al. 2017)

Future Directions

The molecular mechanism by which Nxf1 suppresses IAP-insertion mutations is still unknown. Allelic variation in Nxf1 may influence the ratio of correctly processed mRNA through several potential mechanisms, such as altering the balance between export and processing of nascent transcripts, or through an effect on the assembly and composition of splicing machinery. Nxf1 fits into a well-characterized network, with ~40 reported interacting proteins that are involved in the maturation and export of mRNPs. Since E610G is well removed from RNA binding domains of Nxf1, we hypothesize that *Nxf1^{E610G}* variant quantitatively alters the interaction between Nxf1 and one or more of its partners.

Variant Specific PPIs May Explain Genetic Suppression

It is becoming increasingly clear that specificity of export and regulation by Nxf1 is conferred through a variety of co-factor proteins. Numerous methods have uncovered Nxf1 binding partners, including immunoprecipitation, yeast-two-hybrid, in-vitro binding, affinity capture, and mass spectrometry. Many biochemical studies of Nxf1 utilize expression systems with epitope-tagged Nxf1 to identify partners and reaction kinetics (Uranishi et al. 2009; Yatherajam, Huang, and Flint 2011; Viphakone et al. 2012; 2019). More recently, proximity-ligation based methods, including BioID and APEX labeling, have shown promise in

identifying interaction partners, and offer a sensitive approach to quantify

interaction differences (Roux, Kim, and Burke 2013; Branon et al. 2018;

Martell et al. 2012; Lam et al. 2014). Already, these methods have been

applied to RNA biology, capturing details about the spatial dynamics of

RNA regulation in the cell (Youn et al. 2018) and specifically to nuclear

transport receptors (Mackmull et al. 2017). These studies recapitulated

some of the interactions between Nxf1 and partner proteins including

Nxt1, and should be applied to the question of differential interactions in

the context of IAP suppression by Nxf1. We have developed lentiviral

TurboID-Nxf1 expression vectors which will allow an assessment of PPI

alterations in the context of Nxf1 allelic variation. We suspect this method

can detect subtle quantitative interactions differences, although it may be

the case that effects must be measured specifically on IAP-inserted

transcripts following enrichment with MS2 tagging or by sequence-specific

probes.

Nxf1 May Alter Transcript Processing Via RNA Pol II Elongation

One specific hypothesis which may be supported by proximity ligation data

is that Nxf1 alters RNA processing by influencing the transcriptional

elongation rate of RNA Pol II. Strong candidate mediators of such a

mechanism would be Thoc1 or Rbm15 (Lindtner et al. 2006; Zolotukhin et

al. 2009). Depletion of Thoc1 in mammalian cells impairs transcriptional elongation and suggests this as a potential hypothesis for the mechanism of suppression (Y. Li et al. 2005; Domínguez-Sánchez et al. 2011). Rbm15 can recruit Nxf1 directly to retroviral RNA transport elements, potentially tethering it to nascent transcripts. The rate of transcriptional elongation can also alter co-transcriptional splice site selection, which could explain the shift in the ratio of wild-type to alternatively spliced transcripts that we observe in *vb* mice with suppressive alleles of Nxf1 (Roberts 1998; Saldi et al. 2016). Epigenetic features such as methylation pattern can impact chromatin accessibility and elongation dynamics (Veloso et al. 2014). The impact of an IAP insertion on splice-choice may depend on specific features of the inserted intron or genomic locus (Singh and Padgett 2009). SR proteins should be considered in any such analysis, given the evidence that they may mediate interactions between Nxf1 and specific transcripts, although this would be a surprising mechanism to explain suppression since SR protein-Nxf1 interactions have not been mapped to the UBA domain.

***Nmf9* Encodes a Highly Conserved Protein Important to Neurological Function in Mice and Flies, and Vestibular Function in Zebrafish**

Our investigations of *Nmf9/Ankfn1* are among the first reports characterizing its functional role. Starting from a vestibular mutant identified through an ENU-mutation screen, we mapped a G-to-A transition mutation to a predicted gene, *Ankfn1*, and additional sequence on chromosome 11, which we refer to as *Nmf9*. Expression of *Nmf9* is restricted to the nervous system, and to specific structures and regions which suggested potential behavioral tests to characterize its biological function. *Nmf9* is expressed in the inner ear, amygdala, suprachiasmatic nuclei, lateral septum, piriform cortex, olfactory bulb and epithelium, and ventromedial nuclei, and mutants had sexually dimorphic vestibular phenotypes and alterations in fear learning, and disordered circadian rhythms, but did not have other strong phenotypes attributable to its expression pattern.

Nmf9 is highly conserved across metazoans, and is characterized by two ankyrin repeats, a fibronectin type III domain, a variably included Ras-association domain, and three highly conserved but unannotated domains which include a nearly-invariant GLYLGYLK peptide sequence in the second of these putative domains. We targeted the sequence encoding this peptide using CRISPR/Cas9, both to confirm the positional cloning of the *nmf9* mutation, and to demonstrate the importance of this

conserved domain. Predicted pathogenic CRISPR alleles of *nmf9*, including a single amino acid glycine-to-alanine substitution of the first conserved residue in this sequence failed to compliment the original *nmf9* mutant allele across a range of behavioral measures.

We also conducted editing experiments in the *Drosophila* homolog of *Nmf9* by targeting the first ankyrin repeat, the FN3 domain, and the GLYLGYLK peptide. Null and deleterious mutations in flies resulted in severe locomotor defects and reduced viability. Heterozygotes had abnormal, sexually dimorphic sleep patterns. It was unclear whether the severe outcomes in flies relative to mice are attributable to the presence of the C-terminal Ras association domain.

In part to address differential function of this variably-included RA domain, and in order to address the evolutionary consequence of its loss in mammalian lineages, we conducted gene editing experiments in the zebrafish paralogs. A duplication event in fish yielded an ancestral copy which contains the RA domain, and a derived copy which does not. We again targeted the conserved GLYLGYLK peptide, following an expanded analysis of physicochemical properties in 115 homologs which confirmed it as the most constrained sequence. Interestingly, we observed an overt swim bladder phenotype, but only when both paralogs were disrupted.

Affected fish did not inflate their swims bladders by 5 days post-fertilization and died within two weeks.

Our Results in Context

Around 1/5 of protein coding genes are still uncharacterized, representing an obstacle to a full understanding of biological systems important to health and disease (Wood et al. 2019). Our work is the first report of the function of *Nmf9/Ankfn1* in mice and fish, and resolves a conflict in prior results from *Drosophila* (Liu et al. 2014; Mauri et al. 2014) In our hands, homozygous mutations of three different domains of *Ankfn1* were mostly unviable, with the few animals that survived to eclosure dying shortly thereafter. This is consistent with the Mauri et al. report, where they deleted the entire *bnd* coding region and noted predominant pupal-lethality, with significant locomotor defects in surviving animals. In contrast, Liu et al. report sleep analysis data from apparently viable homozygous mutants for two P-element alleles of WIDE AWAKE and an additional deletion allele which did not span into the *Ankfn1*-homologous sequence of the gene locus. Further, in a direct comparison of sleep patterns between heterozygous *wake* mutants and our engineered *Ankfn1* mutations in our lab, we observed significant, sexually dimorphic abnormalities which differed from those of *wake* mutants. The timing of

192

these publications has meant that although the Liu et al. paper appears to describe an alternative transcript from this locus, it has garnered considerable attention despite this discordant result. The unique evolutionary history of *Nmf9/Ankfn1*, including domain and total gene loss in some lineages and duplication in others, means there is significant value in performing comparative genetic studies across diverse organisms. Ours is a rare example of such a study, and sets the stage for a further exploration of functiona evolution of this gene (Lynch and Katju 2004; Conant and Wolfe 2008). Gene duplications are a topic of sustained debate, and aside from *in-silico* modeling and computational analysis of homologous sequences, it is difficult to design experiments that can distinguish between the different models for and outcomes of gene evolution after duplication. We have laid the foundation for such studies by generating mutant alleles of *Ankfn1* in three different model organisms, one of which, zebrafish, expresses both the ancestral and the derived copies of this gene. The lack of an observed phenotype in single mutants suggests that there is functional overlap between the paralogs with respect to swim bladder inflation.

Shortcomings of the Study

The swim bladder inflation phenotype was easy to identify, and suggests a role for *Ankfn1* in the developing vestibular system of zebrafish (Kwak et al. 2006). Indeed, we suspected that such a phenotype might be the fish analog to the circling behavior that we observed in the most severely effected *nmf9* mutant mice. In adult zebrafish, the swim bladder is linked to the inner ear through a series of bones which improve acoustic response (Popper, Platt, and Edds 1992). These comparisons are provocative, but we did not perform deeper phenotyping to more clearly delineate the phenotypic similarities and differences between fish, mice, and flies. We are able to conclude that there is functional overlap between the zebrafish *Ankfn1* paralogs, but we cannot rule out that we may be blind to subfunctionalization which would be revealed by more in-depth phenotyping. Additionally, replicating this finding in additional clutches will resolve the two mutant fish which contradict the main finding, and give us a stronger estimate of penetrance.

The conservation analysis is limited by the availability of high-quality sequences, and by biases in the representation of different taxons. We originally identified several putative conserved domains, an expanded analysis of full length sequences does not clarify the boundaries between these domains, nor do we know what specific functions are encoded by

these regions. Manual curation may capture additional homologs with sequences related to the GLYLGYLK peptide, but common alignment algorithms struggle to include these sequences as part of a coherent set.

Future Directions

Zebrafish as a model system has the potential to distinguish between some of the duplicate fate hypotheses. A clearer functional characterization of the zebrafish paralogs compared to the single gene orthologs in mice and flies could differentiate DDC versus another mode of subfunctionalization (Ohno 1970; Force et al. 1999; Lynch and Force 2000; Lynch and Katju 2004; Conant and Wolfe 2008). At present, we don't know whether the two paralogs actually encode unique functions, although we suspect this to be the case given the difference in severity of phenotypes between flies and mice and the absence of the RA domain in the derived paralog. Zebrafish exhibit specific and replicable locomotor behaviors in early development which we could test in to distinguish subtle phenotypic differences between mutations in the ancestral and derived copies (Brustein et al. 2003; Kwak et al. 2006). Beyond behavior, we can also test hearing in adult fish and examine physical structures of the vestibular system (Whitfield et al. 2002; 1996; Whitfield 2002).

Although there are insights into the molecular mechanism of *Ankfn1* in *Drosophila,* nothing is known about the mode of action in mice and fish by which *Nmf9*/*Ankfn1* influences neurological function and development. It is therefore worth taking an unbiased approach to address this gap in our knowledge. RNA-seq of single and double mutants would let us determine whether dosage balance applies to these duplicates in the context of gene evolution, and can determine if Ankfn1 regulates other targets (Papp, Pál, and Hurst 2003; Birchler and Veitia 2012). It would also be informative to identify protein interactions partners of *Ankfn1*, which could be accomplished using some of the methods described previously, such as making an epitope-tagged version of *Nmf9* (which we have designed and produced and using gene editing in mice) to perform Co-IP experiments, or may tuning an expression system to attempt a proximity ligation assay in murine cells.

REFERENCES

Ben-Yishay, Rakefet, Amir Mor, Amit Shraga, Asaf Ashkenazy-Titelman, Noa Kinor, Avital Schwed-Gross, Avi Jacob, et al. 2019. "Imaging within Single NPCs Reveals NXF1's Role in MRNA Export on the Cytoplasmic Side of the Pore." *The Journal of Cell Biology*. https://doi.org/10.1083/jcb.201901127.

Birchler, James A., and Reiner A. Veitia. 2012. "Gene Balance Hypothesis: Connecting Issues of Dosage Sensitivity across Biological Disciplines." *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1207726109.

Björk, Petra, and Lars Wieslander. 2017. "Integration of MRNP Formation and Export." *Cellular and Molecular Life Sciences*. https://doi.org/10.1007/s00018-017-2503-3.

Branon, Tess C., Justin A. Bosch, Ariana D. Sanchez, Namrata D. Udeshi, Tanya Svinkina, Steven A. Carr, Jessica L. Feldman, Norbert Perrimon, and Alice Y. Ting. 2018. "Efficient Proximity Labeling in Living Cells and Organisms with TurboID." *Nature Biotechnology*. https://doi.org/10.1038/nbt.4201.

Brustein, Edna, Louis Saint-Amant, Robert R. Buss, Mabel Chong, Jonathan R. McDearmid, and Pierre Drapeau. 2003. "Steps during the Development of the Zebrafish Locomotor Network." In *Journal of Physiology Paris*. https://doi.org/10.1016/j.jphysparis.2003.10.009.

Chang, Chung Te, Guillaume M. Hautbergue, Matthew J. Walsh, Nicolas Viphakone, Thamar B. Van Dijk, Sjaak Philipsen, and Stuart A. Wilson. 2013. "Chtop Is a Component of the Dynamic TREX MRNA Export Complex." *EMBO Journal*. https://doi.org/10.1038/emboj.2012.342.

Chen, Suli, Ruijia Wang, Dinghai Zheng, Heng Zhang, Xingya Chang, Ke Wang, Wencheng Li, Jing Fan, Bin Tian, and Hong Cheng. 2019. "The MRNA Export Receptor NXF1 Coordinates Transcriptional Dynamics, Alternative Polyadenylation, and MRNA Export." *Molecular*

Cell. https://doi.org/10.1016/j.molcel.2019.01.026.

Conant, Gavin C., and Kenneth H. Wolfe. 2008. "Turning a Hobby into a Job: How Duplicated Genes Find New Functions." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg2482.

Domínguez-Sánchez, María S., Sonia Barroso, Belén Gómez-González, Rosa Luna, and Andrés Aguilera. 2011. "Genome Instability and Transcription Elongation Impairment in Human Cells Depleted of THO/TREX." *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1002386.

Force, Allan, Michael Lynch, F. Bryan Pickett, Angel Amores, Yi Lin Yan, and John Postlethwait. 1999. "Preservation of Duplicate Genes by Complementary, Degenerative Mutations." *Genetics*.

Fu, Bishi, Mengmeng Zhao, Lingyan Wang, Girish Patil, Jennifer A. Smith, Ignacio J. Juncadella, Ljiljana Zuvela-Jelaska, Martin E. Dorf, and Shitao Li. 2017. "RNAi Screen and Proteomics Reveal NXF1 as a Novel Regulator of IRF5 Signaling." *Scientific Reports*. https://doi.org/10.1038/s41598-017-02857-z.

Hautbergue, Guillaume M., Lydia M. Castelli, Laura Ferraiuolo, Alvaro Sanchez-Martinez, Johnathan Cooper-Knock, Adrian Higginbottom, Ya Hui Lin, et al. 2017. "SRSF1-Dependent Nuclear Export Inhibition of C9ORF72 Repeat Transcripts Prevents Neurodegeneration and Associated Motor Deficits." *Nature Communications*. https://doi.org/10.1038/ncomms16063.

Kwak, Su Jin, Shruti Vemaraju, Stephen J. Moorman, David Zeddies, Arthur N. Popper, and Bruce B. Riley. 2006. "Zebrafish Pax5 Regulates Development of the Utricular Macula and Vestibular Function." *Developmental Dynamics*. https://doi.org/10.1002/dvdy.20961.

Lam, Stephanie S., Jeffrey D. Martell, Kimberli J. Kamer, Thomas J. Deerinck, Mark H. Ellisman, Vamsi K. Mootha, and Alice Y. Ting. 2014. "Directed Evolution of APEX2 for Electron Microscopy and Proximity Labeling." *Nature Methods*.

https://doi.org/10.1038/nmeth.3179.

Lesbirel, Simon, and Stuart A. Wilson. 2019. "The m 6 A-methylase Complex and MRNA Export." *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. https://doi.org/10.1016/j.bbagrm.2018.09.008.

Li, Ping, Maria Stumpf, Rolf Müller, Ludwig Eichinger, Gernot Glöckner, and Angelika A. Noegel. 2017. "The Function of the Inner Nuclear Envelope Protein SUN1 in MRNA Export Is Regulated by Phosphorylation." *Scientific Reports*. https://doi.org/10.1038/s41598-017-08837-7.

Li, Y., X. Wang, X. Zhang, and D. W. Goodrich. 2005. "Human HHpr1/P84/Thoc1 Regulates Transcriptional Elongation and Physically Links RNA Polymerase II and RNA Processing Factors." *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.25.10.4023-4033.2005.

Lindtner, Susan, Andrei S. Zolotukhin, Hiroaki Uranishi, Jenifer Bear, Viraj Kulkarni, Sergey Smulevitch, Martina Samiotaki, George Panayotou, Barbara K. Felber, and George N. Pavlakis. 2006. "RNA-Binding Motif Protein 15 Binds to the RNA Transport Element RTE and Provides a Direct Link to the NXF1 Export Pathway." *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M608745200.

Liu, Sha, Angelique Lamaze, Qili Liu, Masashi Tabuchi, Yong Yang, Melissa Fowler, Rajnish Bharadwaj, et al. 2014. "WIDE AWAKE Mediates the Circadian Timing of Sleep Onset." *Neuron*. https://doi.org/10.1016/j.neuron.2014.01.040.

Lynch, Michael, and Allan Force. 2000. "The Probability of Duplicate Gene Preservation by Subfunctionalization." *Genetics*.

Lynch, Michael, and Vaishali Katju. 2004. "The Altered Evolutionary Trajectories of Gene Duplicates." *Trends in Genetics*. https://doi.org/10.1016/j.tig.2004.09.001.

Mackmull, Marie-Therese, Bernd Klaus, Ivonne Heinze, Manopriya

Chokkalingam, Andreas Beyer, Robert B Russell, Alessandro Ori, and Martin Beck. 2017. "Landscape of Nuclear Transport Receptor Cargo Specificity." *Molecular Systems Biology*. https://doi.org/10.15252/msb.20177608.

Martell, Jeffrey D., Thomas J. Deerinck, Yasemin Sancak, Thomas L. Poulos, Vamsi K. Mootha, Gina E. Sosinsky, Mark H. Ellisman, and Alice Y. Ting. 2012. "Engineered Ascorbate Peroxidase as a Genetically Encoded Reporter for Electron Microscopy." *Nature Biotechnology*. https://doi.org/10.1038/nbt.2375.

Mauri, Federico, Ilka Reichardt, Jennifer L. Mummery-Widmer, Masakazu Yamazaki, and Juergen A. Knoblich. 2014. "The Conserved Discs-Large Binding Partner Banderuola Regulates Asymmetric Cell Division in Drosophila." *Current Biology*. https://doi.org/10.1016/j.cub.2014.06.059.

Müller-McNicoll, Michaela, Valentina Botti, Antonio M. de Jesus Domingues, Holger Brandl, Oliver D. Schwich, Michaela C. Steiner, Tomaz Curk, Ina Poser, Kathi Zarnack, and Karla M. Neugebauer. 2016. "SR Proteins Are NXF1 Adaptors That Link Alternative RNA Processing to MRNA Export." *Genes and Development*. https://doi.org/10.1101/gad.276477.115.

Ohno, Susumu. 1970. *Evolution by Gene Duplication. Evolution by Gene Duplication*. https://doi.org/10.1007/978-3-642-86659-3.

Papp, Balázs, Csaba Pál, and Laurence D. Hurst. 2003. "Dosage Sensitivity and the Evolution of Gene Families in Yeast." *Nature*. https://doi.org/10.1038/nature01771.

Popper, Arthur N., Christopher Platt, and Peggy L. Edds. 1992. "Evolution of the Vertebrate Inner Ear: An Overview of Ideas." In *The Evolutionary Biology of Hearing*. https://doi.org/10.1007/978-1-4612-2784-7_4.

Roberts, G. 1998. "Co-Transcriptional Commitment to Alternative Splice Site Selection." *Nucleic Acids Research*. https://doi.org/10.1093/nar/26.24.5568.

Roundtree, Ian A., Guan Zheng Luo, Zijie Zhang, Xiao Wang, Tao Zhou, Yiquang Cui, Jiahao Sha, et al. 2017. "YTHDC1 Mediates Nuclear Export of N6-Methyladenosine Methylated MRNAs." *ELife*. https://doi.org/10.7554/eLife.31311.

Roux, Kyle J., Dae In Kim, and Brian Burke. 2013. "BioID: A Screen for Protein-Protein Interactions." *Current Protocols in Protein Science*. https://doi.org/10.1002/0471140864.ps1923s74.

Saldi, Tassa, Michael A. Cortazar, Ryan M. Sheridan, and David L. Bentley. 2016. "Coupling of RNA Polymerase II Transcription Elongation with Pre-MRNA Splicing." *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2016.04.017.

Scott, Daniel D., L. Carolina Aguilar, Mathew Kramar, and Marlene Oeffinger. 2019. "It's Not the Destination, It's the Journey: Heterogeneity in MRNA Export Mechanisms." In *Advances in Experimental Medicine and Biology*. https://doi.org/10.1007/978-3-030-31434-7_2.

Singh, Jarnail, and Richard A. Padgett. 2009. "Rates of in Situ Transcription and Splicing in Large Human Genes." *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.1666.

Talhouarne, Gaëlle J.S., and Joseph G. Gall. 2018. "Lariat Intronic RNAs in the Cytoplasm of Vertebrate Cells." *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1808816115.

Uranishi, Hiroaki, Andrei S. Zolotukhin, Susan Lindtner, Soren Warming, Gen Mu Zhang, Jenifer Bear, Neal G. Copeland, Nancy A. Jenkins, George N. Pavlakis, and Barbara K. Felber. 2009. "The RNA-Binding Motif Protein 15B (RBM15B/OTT3) Acts as Cofactor of the Nuclear Export Receptor NXF1." *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M109.040113.

Veloso, Artur, Killeen S. Kirkconnell, Brian Magnuson, Benjamin Biewen, Michelle T. Paulsen, Thomas E. Wilson, and Mats Ljungman. 2014. "Rate of Elongation by RNA Polymerase II Is Associated with Specific

Gene Features and Epigenetic Modifications." *Genome Research*. https://doi.org/10.1101/gr.171405.113.

Viphakone, Nicolas, Guillaume M. Hautbergue, Matthew Walsh, Chung Te Chang, Arthur Holland, Eric G. Folco, Robin Reed, and Stuart A. Wilson. 2012. "TREX Exposes the RNA-Binding Domain of Nxf1 to Enable MRNA Export." *Nature Communications*. https://doi.org/10.1038/ncomms2005.

Viphakone, Nicolas, Ian Sudbery, Llywelyn Griffith, Catherine G. Heath, David Sims, and Stuart A. Wilson. 2019. "Co-Transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome." *Molecular Cell*. https://doi.org/10.1016/j.molcel.2019.04.034.

Whitfield, Tanya T. 2002. "Zebrafish as a Model for Hearing and Deafness." *Journal of Neurobiology*. https://doi.org/10.1002/neu.10123.

Whitfield, Tanya T., Michael Granato, Fredericus J.M. Van Eeden, Ursula Schach, Michael Brand, Makoto Furutani-Seiki, Pascal Haffter, et al. 1996. "Mutations Affecting Development of the Zebrafish Inner Ear and Lateral Line." *Development*.

Whitfield, Tanya T., Bruce B. Riley, Ming Yung Chiang, and Bryan Phillips. 2002. "Development of the Zebrafish Inner Ear." *Developmental Dynamics*. https://doi.org/10.1002/dvdy.10073.

Wood, Valerie, Antonia Lock, Midori A. Harris, Kim Rutherford, Jürg Bähler, and Stephen G. Oliver. 2019. "Hidden in Plain Sight: What Remains to Be Discovered in the Eukaryotic Proteome?" *Open Biology*. https://doi.org/10.1098/rsob.180241.

Xie, Yihu, and Yi Ren. 2019. "Mechanisms of Nuclear MRNA Export: A Structural Perspective." *Traffic*. https://doi.org/10.1111/tra.12691.

Yatherajam, G., W. Huang, and S. J. Flint. 2011. "Export of Adenoviral Late MRNA from the Nucleus Requires the Nxf1/Tap Export Receptor." *Journal of Virology*. https://doi.org/10.1128/jvi.02108-10.

Youn, Ji Young, Wade H. Dunham, Seo Jung Hong, James D.R. Knight, Mikhail Bashkurov, Ginny I. Chen, Halil Bagci, et al. 2018. "High-Density Proximity Mapping Reveals the Subcellular Organization of MRNA-Associated Granules and Bodies." *Molecular Cell*. https://doi.org/10.1016/j.molcel.2017.12.020.

Zhou, Hao, Katarzyna Bulek, Xiao Li, Tomasz Herjan, Minjia Yu, Wen Qian, Han Wang, et al. 2017. "IRAK2 Directs Stimulus-Dependent Nuclear Export of Inflammatory MRNAs." *ELife*. https://doi.org/10.7554/eLife.29630.

Zolotukhin, Andrei S., Hiroaki Uranishi, Susan Lindtner, Jenifer Bear, George N. Pavlakis, and Barbara K. Felber. 2009. "Nuclear Export Factor RBM15 Facilitates the Access of DBP5 to MRNA." *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkp782.