Applications Of Genome Editing Tools In Drug Discovery And Basic Research

Inauguraldissertation

zur Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Guillermo Rodrigo Villaseñor Molina

Aus Chihuahua, Mexiko

Basel, 2016

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von:

Prof. Dr. Marc Bühler

(Fakultätsverantwortlicher/ Dissertationsleiter)

Prof. Dr. René Ketting

(Korreferent, extern)

Basel, den 8. Dezember 2015

Prof. Dr. Jörg Schibler

Dekan

TABLE OF CONTENTS

Summary

Since the discovery of the DNA double helix, major advances in biology have been; the development of recombinant DNA technology in the 1970s, methods to amplify DNA and gene targeting technology in the late 1980s. In organisms such as yeast and mice, the ability to accurately add or delete genetic information transformed biology, allowing an unmatched level of precision in studies of gene function. But, the ability to easily and specifically edit the genetic material of other cells and organisms remained impossible until recently for molecular biologists. The recent advent of programmable nucleases has dramatically changed the efficiency and speed of genome manipulation in several model organisms including cultured cells, as well as whole animals and plants. These tools opened up a powerful technique for biology research now called “genome editing” or “genome engineering” (Carroll, 2011; Hsu et al., 2014; Kim and Kim, 2014).

In the first half of my doctoral studies, I developed genome-editing strategies to discover drug targets for a rare genetic disease called Friedreich´s Ataxia. Friedreich’s Ataxia (FRDA) is a neurodegenerative disease caused by deficiency of the mitochondrial protein frataxin (FXN) (Campuzano et al., 1997). This deficiency results from an expansion of a trinucleotide GAA repeat in the first intron of the *FXN* gene (Campuzano et al., 1996; Durr et al., 1996). Therapeutics that reactivate *FXN* gene expression are expected to be beneficial to FRDA patients (Gottesfeld, 2007). However, high-throughput screening (HTS) for *FXN* activators has so far met with limited success because current cellular models do not accurately assess endogenous *FXN* gene regulation. Here I used genome-editing technologies to generate a cellular model in which a luciferase reporter is introduced into the endogenous *FXN* locus. Using this system in a

high-throughput genomic screen, we discovered novel inhibitors of FXN-luciferase expression. I confirmed that reducing expression of one of these inhibitors, PRKD1, led to an increase in *FXN* expression in FRDA patient fibroblasts (Villasenor et al., 2015). We then used reprogramming technologies to create a disease-relevant situation and test small molecules that specifically modulate PRKD1. We found that WA-21-JO19, a chemical inhibitor of PRKD1, increases *FXN* expression levels in iPSC-derived FRDA patient neurons. This approach, developed at the interface between academic and pharmaceutical research, demonstrates how a combination of genome editing, cellular reprogramming, and high-throughput biology can generate an effective novel drug discovery platform.

In the second part of my doctoral work, we developed an interface between genome editing and proteomics to isolate native protein complexes produced from their natural genomic contexts. In many biological processes, proteins act as members of protein complexes. Understanding the molecular composition of protein complexes is a key task towards explaining their function in the cell. Conventional affinity purification followed by mass spectrometry analysis is a broadly applicable method to decipher molecular interaction networks and infer protein function. However, traditional affinity purification methods are limited by a number of factors such as antibody specificity and are sensitive to perturbations induced by overexpressed target proteins. Here, we combined genome editing with tandem affinity purification to circumvent current limitations. I uncovered subunits and interactions among well-characterized complexes and report the isolation of novel Mettl3-binding partners. The multi-protein complex composed of two active methyltransferases Mettl3 and Mettl14 mediates methylation of adenosines at position N6 on RNA molecules (Bokar et al., 1994; Bokar et al., 1997; Liu et al., 2014). N6-methyladenosine is the most abundant internal modification in eukaryotic mRNA and is often found on introns, which implies that methylation occurs co-transcriptionally (Fu et al., 2014). My work identified a set of nuclear RNA binding proteins, which specifically interact with

the Mettl3-Mettl14 complex. We are currently testing the ability of these factors to function as “recruiters” of the Mettl3-Mettl14 complex to nascent mRNAs in the cell nucleus.

In summary, our approach solidly establishes how a combination of genome editing and proteomics can simplify explorations of protein complexes as well as the study of post-translational modifications. In addition, this approach opens up new opportunities to study native protein complexes in a wide variety of cells and model organisms and will likely enable the systematic investigation of mammalian proteome function.

1 Introduction

1.1 A brief research history of genome editing with programmable nucleases

A major goal in genetics is to connect genotypes with phenotypes. Species with experimentally manipulable genomes are essential to investigate the role of genes in biology and disease. The ability to modify a complex genome with high precision transformed biology since the 1980s (Capecchi, 2001; Evans, 2001; Smithies, 2001). The underlying technology is known as gene targeting and relies on the homologous recombination (HR) machinery found in all eukaryotes. In gene targeting, an exogenously introduced DNA fragment replaces an endogenous piece of DNA by homologous recombination. The procedure for gene replacement was pioneered in baker´s yeast, *Saccharomyces cerevisiae*, more than 35 years ago (Orr-Weaver et al., 1981; Scherer and Davis, 1979). In the mid 1980s, gene targeting was demonstrated in human and mouse embryonic stem (ES) cells (Doetschman et al., 1987; Mansour et al., 1988; Smithies et al., 1985; Thomas et al., 1986), which enabled the production of transgenic mice for studying gene function and for creating models of human genetic diseases (Kuehn et al., 1987; Wu and Melton, 1993). Similar to the process in yeast, gene targeting in mice depends on homologous recombination between the donor and the target site. In addition, positive selection for the specific integration of the transgene must be applied against the more common products of random integration (Capecchi, 2005). This is achieved by double selection strategies yielding the desired replacements. The pluripotent nature of mouse ES cells allows spreading of the transgene into all cell lineages after injection into early embryos (Evans and Kaufman, 1981; Zijlstra et al., 1989). An important development in the use of recombination-based methods for gene manipulation in yeast and mammals involved the use of polymerase chain reaction (PCR), rather than restriction enzymes, to generate DNA fragments

with sufficiently long homologous sequences (Baudin et al., 1993; Lubahn et al., 1993).

Advances in gene targeting have made this technique routine in yeast and nearly routine in mice. The ability to specifically add or delete genetic information in yeast enabled an unmatched level of precision in studies of eukaryotic gene function leading to breakthrough discoveries in molecular biology. In addition, thousands of transgenic mice and ES cell lines with precise genomic modifications have been created. Their characterization has increased our knowledge of mammalian physiology and the pathogenesis of numerous human diseases.

In both yeast and mouse cells, the frequency of homologous recombination events between donor and target site is inherently low – on the order of one in every 10^3 to 10^9 cells (Bahler et al., 1998; Thomas et al., 1986). Applications of gene targeting in other experimental organisms, where embryonic stem cells are not available, have been hindered by the low frequency of HR and the need for positive selection in cell culture before incorporation into whole organisms. The challenge in extending gene targeting to other model organisms can be viewed largely as one of increasing the frequency of homologous recombination (Carroll, 2011).

What limits the frequency of homologous recombination in gene targeting? Experiments in yeast and mammalian cells demonstrated clearly that manipulations of the donor DNA have modest effects, but activation of the chromosomal target with a double-strand break (DSB) boosts the frequency of HR by several orders of magnitude. These experiments were inspired by the discovery that natural recombination events, such as meiotic crossing over and mating-type switching in yeast, are initiated by DSBs. Pioneering experiments in yeast and mammalian cells demonstrated that induction of a single DSB can dramatically increase the frequency of HR at the cleavage site (Choulika et al., 1995; Rouet et al., 1994; Rudin and Haber, 1988). Both approaches made use

of yeast DNA endonucleases that specifically cut a previously inserted recognition site in the genome. Further studies revealed the power of a DSB in stimulating gene targeting to levels that would be useful in other experimental systems (Rong and Golic, 2000; Smih et al., 1995; Taghian and Nickoloff, 1997).

I-SceI and HO are yeast homing endonucleases that were used to induce a DSB and increase the efficiency of gene targeting. A key feature of these enzymes is that they create DSBs at recognition sites that are 14 to 40 base-pairs (bp) long (Chevalier and Stoddard, 2001). These enzymes provided important information on the efficiency and mechanisms of DSB repair, but they were limited in their utility as programmable gene-targeting reagents. The main reason was that their recognition sites had to be introduced in the genome by a low-frequency process before they could mediate high-efficiency homologous recombination (Carroll, 2011). Therefore, several investigators used protein engineering to modify homing endonucleases to recognize target sites in mammalian genes (Belfort and Bonocora, 2014; Chevalier et al., 2002).

To harness the stimulatory power of DSBs in facilitating gene targeting, methods for creating site-specific DSBs were required. Several approaches were developed such as modified triplex-forming oligonucleotides (Kuan and Glazer, 2004), modified polyamides (Dervan and Edelson, 2003; Wurtz and Dervan, 2000), modified peptide-nucleic acids (Kaihatsu et al., 2004), modified homing endonucleases (Chevalier et al., 2002), and programmable nucleases (Kim and Kim, 2014). In this doctoral thesis, I will focus on the recent progress made with programmable nucleases to edit eukaryotic genomes.

1.2 Overview of existing genome editing technologies

The currently existing and most utilized programmable nucleases for genome editing include zinc-finger nucleases (ZFNs), transcription activator-like

effector nucleases (TALENs) and RNA-guided nucleases (RGNs) derived from the adaptive immune defense system of bacteria termed “clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas)” CRISPR-Cas system (Figure 1B&C). In the next chapters, I will review the development and applications of programmable nucleases for genome editing from a historical perspective.

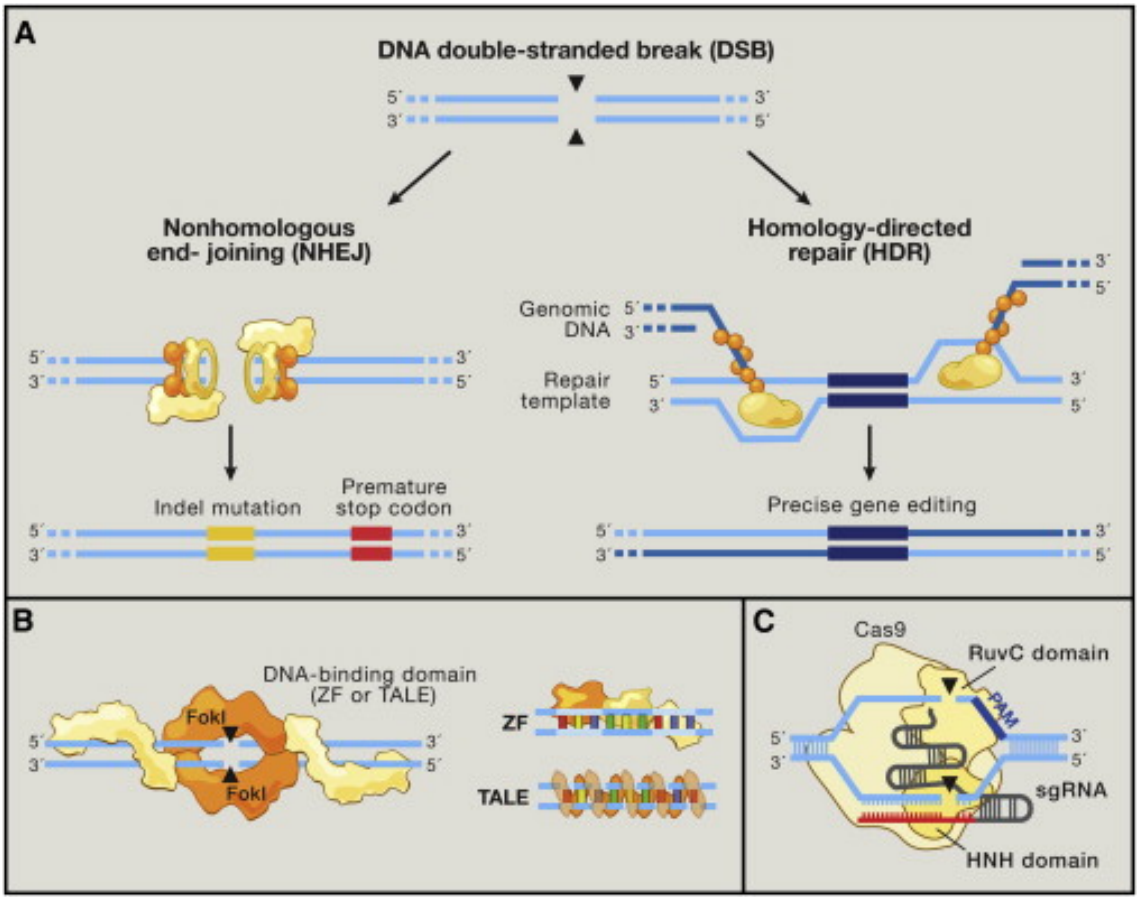


Figure 1. Genome Editing Technologies Exploit Endogenous DNA Repair Machinery

(**A**) DNA double-strand breaks (DSBs) are typically repaired by non-homologous end-joining (NHEJ) or homology-directed repair (HDR). In the error-prone NHEJ pathway, Ku heterodimers bind to DSB ends and serve as a molecular scaffold for associated repair proteins. Indels are introduced when the complementary strands undergo end resection and misaligned repair due to microhomology, eventually leading to frameshift mutations and gene knockout. Alternatively, Rad51 proteins may bind DSB ends during the initial phase of HDR, recruiting accessory factors that direct genomic recombination with homology arms on an

exogenous repair template. Bypassing the matching sister chromatid facilitates the introduction of precise gene modifications. (**B**) Zinc finger (ZF) proteins and transcription activator-like effectors (TALEs) are naturally occurring DNA-binding domains that can be modularly assembled to target specific sequences. ZF and TALE domains each recognize 3 and 1 bp of DNA, respectively. Such DNA-binding proteins can be fused to the FokI endonuclease to generate programmable site-specific nucleases. (**C**) The Cas9 nuclease from the microbial CRISPR adaptive immune system is localized to specific DNA sequences via the guide sequence on its guide RNA (red), directly base-pairing with the DNA target. Binding of a protospacer-adjacent motif (PAM, blue) downstream of the target locus helps to direct Cas9-mediated DSBs. Image taken from (Hsu et al., 2014).

1.2.1 Zinc-finger Nucleases (ZFNs)

In 1996, Chandrasegaran and colleagues developed the first zinc-finger nucleases (ZFN), originally termed chimeric restriction enzymes (Kim et al., 1996). They hypothesized that the modular structure of the FokI type II restriction endonuclease might allow the creation of hybrid endonucleases with novel sequence specificities by linking other DNA-binding proteins to the cleavage domain. The first ZFNs consisted of the non sequence-specific cleavage domain of the FokI endonuclease fused to zinc-finger DNA-binding domain of transcription factors with known consensus sequences (Kim et al., 1996; Kim et al., 1998). Because of their modular structure, FokI endonucleases offered an attractive framework for designing chimeric restriction enzymes with tailor-made sequence specificity (Durai et al., 2005).

Zinc-finger proteins of the Cys_2His_2-like fold group are found in the DNA-binding domains of the most abundant family of eukaryotic transcription factors (Vaquerizas 2009) (Figure 2A). The crystal structure of the transcription factor Zif268 (EGR1) bound to DNA greatly stimulated research into engineered zinc-finger arrays for ZFNs (Durai et al., 2005). This structure revealed that the DNA-binding domain of EGR1 consists of three zinc-finger modules of the Cys_2His_2 type. Each zinc-finger module consists of 30 amino acids, folds into a ββα configuration, and coordinates one Zn^{2+} ion using two cysteine and two histidine residues (Pavletich and Pabo, 1991). Two critical features of the structure made

Cys_2His_2 zinc-fingers of particular interest for the development of ZFNs: the first is that each zinc-finger module independently binds three base-pairs of DNA. The second is that each nucleotide in the major groove of the DNA target site seemed to be contacted by a single amino-acid side chain of the α-helix (Porteus and Carroll, 2005). From these features, researchers concluded that DNA-binding domains with novel specificities could be designed by altering the number of fingers and the nature of critical amino-acid residues that contact DNA directly (Smith et al., 2000).

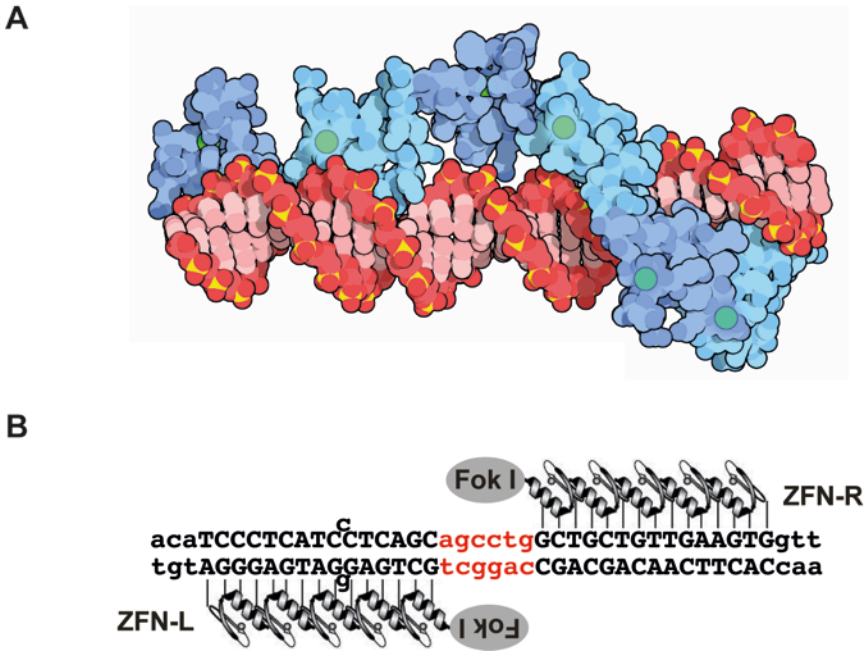


Figure 2. Zinc-Finger Nucleases (ZFNs)

(**A**) Crystal structure of the Cys_2His_2-like zinc-finger domain of transcription factor TFIIIA from frog (Nolte et al., 1998). PDB entry 1tf6, shown here, includes 6 of the zinc-fingers (blue) bound to a long stretch of DNA (red). Zinc atoms are shown in green. (**B**) Schematic showing a pair of zinc-finger nucleases targeting a piece of DNA (*FXN* locus). Each ZFN contains the cleavage domain of FokI linked to an array of five zinc-fingers that have been designed to specifically recognize sequences (black uppercase letters) flanking the cleavage site (red letters) in intron 4 of the *FXN* gene.

Although not recognized initially (Kim et al., 1996), ZFNs cuts DNA more efficient when dimers of the FokI cleavage domain are formed (Bitinaite et al.,

1998; Smith et al., 2000). Further improvements lead to the development of a paired heterodimeric nuclease design strategy that increased cleavage specificity while minimizing off-target activity (Bibikova et al., 2001; Miller et al., 2007). The most optimal configuration of a ZFN pair cleaves DNA targets containing two 9−12 bp binding sites separated by a 5−6 bp spacer (Handel et al., 2009; Shimizu et al., 2011; Urnov et al., 2005) (Figure 2B). The requirement of FokI dimerization is important for nuclease specificity because cleavage should not occur at single ZFN binding sites. As a result, the overall 18−24 bp sequence specificity of a ZFN pair should be sufficient to pick out a unique target within a complex eukaryotic genome (Shimizu et al., 2011).

The first genomic locus successfully targeted with designed ZFNs was the yellow gene of the fruit fly *D. melanogaster*. Bibikova et al. demonstrated targeted mutagenesis (Bibikova et al., 2002) and targeted gene replacement (Bibikova et al., 2003) at the yellow locus in somatic cells and in the germline. Since then, ZFN pairs have been used to target numerous genes in a wide variety of organisms including human cells (Carroll, 2011; Urnov et al., 2005). However, the initial enthusiasm for the broad application of ZFNs as genome editing tools was hampered by difficulties in their design (Cornu et al., 2008; Ramirez et al., 2008).

The most basic design strategy, modular assembly, optimizes individual zinc-fingers against target triplet DNA sequences and then links them together to target a larger sequence. But, this assembly method suffers from a high failure rate when assembled in an array (Bae et al., 2003; Ramirez et al., 2008). Zinc-fingers failed to be treated as independent modules in several design strategies, mainly because zinc-finger domains exhibit context-dependent binding preferences (Kim and Kim, 2014). This made the selection of functional ZFNs a labour-intensive and time-consuming screening process (Maeder et al., 2008; Sander et al., 2011b). Consequently, the available options for academic researchers interested in using ZFN technology were limited. Before other genome editing tools became widely available, researchers had to purchase

customized ZFNs through the Sigma-Aldrich CompoZr service. The downside was the cost of these proteins and the legal restrictions when using the ZFN technology, which greatly limited the scale and scope of projects that could be performed by academic institutions.

1.2.2 Transcription activator-like effector nucleases (TALENs)

Since the development of the first engineered zinc-finger nuclease, researchers spent about 15 years optimizing ZFNs, trying to make their design easier and cheaper. In 2009, two seminal studies reported that transcription activator-like effectors (TALEs) rely on a previously unknown way to bind DNA that is much simpler than that of zinc-finger proteins (Boch et al., 2009; Moscou and Bogdanove, 2009). TALEs were first discovered in plant pathogenic bacteria of the genus *Xanthomonas* and are virulence factors that are translocated into rice crops via a Type III bacterial secretion system. These factors that act as transcriptional activators in the plant cell nucleus, where they directly bind host DNA and activate genes that contribute to disease or turn on defence (Boch and Bonas, 2010). Target specificity depends on an effector-variable number of typically 34 amino-acid repeats in the TALE's central domain (Figure 3A). Boch et al. and Moscou and Bogdanove independently discovered that a pair of adjacent amino-acid residues at positions 12 and 13 in each repeat, the 'repeat-variable di-residue' (RVD), is crucial to specifically recognize the DNA target site (Boch et al., 2009; Moscou and Bogdanove, 2009). The code is simple: one RVD binds to one nucleotide of the DNA target. Four of the most commonly occurring RVDs in natural TALEs preferentially associate with one of the four DNA bases (Moscou and Bogdanove, 2009) (Figure 3A).

The lessons learned from ZFN design greatly facilitated the development of transcription activator-like effector nucleases (TALENs), fusions of transcription activator-like effectors (TALE) to the FokI nuclease domain (Figure 3B & C). Two years after breaking the code for DNA binding specificity of

TALEs, subsequent studies revealed the potential of TALENs for genome editing (Cermak et al., 2011; Miller et al., 2011). The use of TALENs showed superior genome editing efficiencies over ZFNs in a variety of cellular and model organisms including human pluripotent stem cells, rats, worms and zebrafish (Hockemeyer et al., 2011; Huang et al., 2011; Sander et al., 2011a; Tesson et al., 2011; Wood et al., 2011).

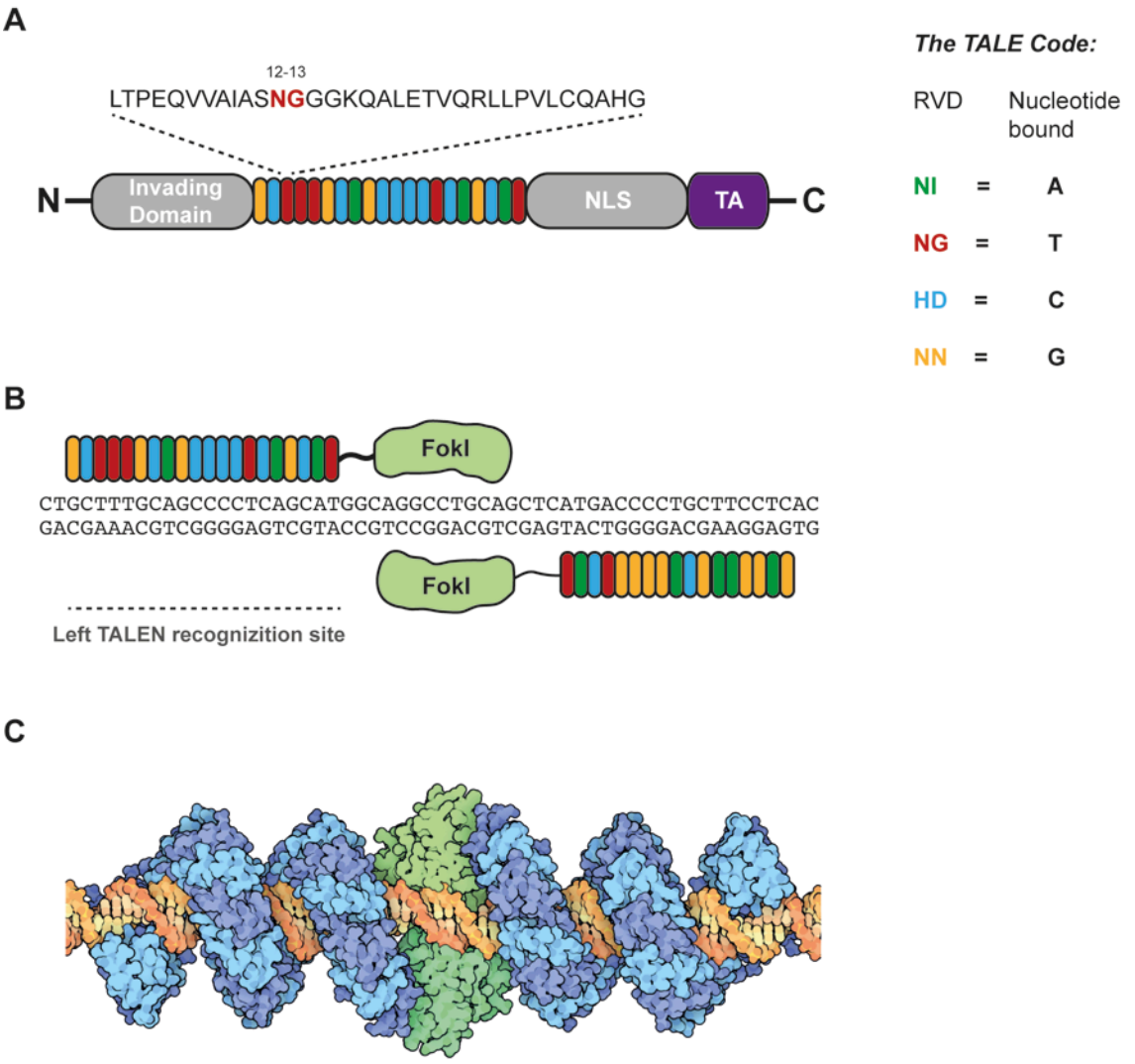


Figure 3. TAL Effector And TALEN Structure

(**A**) Structure of a naturally occurring TAL effector. A consensus repeat sequence is shown with the repeat-variable di-residue (RVD) marked in red. The sequence of RVDs determines the target nucleotide sequence. The four most common RVDs are shown with their most frequently associated nucleotide (right).

(**B**) Structure of a TALEN. Two monomeric TALENs are required to bind the target site to enable FokI (green) to dimerize and cleave DNA. NLS, nuclear localization signal(s); TA, transcriptional activation domain. (**C**) Crystal structure of a TALEN pair bound to DNA (orange). The FokI nuclease domain (PDB entry 1fok, shown here in green) is fused to one end of a TAL effector (blue).

What made scientists switch so rapidly from ZFNs to TALENs? The ease of use made TALENs an attractive alternative for genome editing (Baker, 2012). Unlike ZFNs, the DNA-binding domains of TALEs (repeat monomer with RVD) can be assembled in a modular fashion and are easier to customize than those of zinc-finger proteins. Current protocols allow the design of TALEN pairs in less than 5 days, reducing costs and manpower (Cermak et al., 2011; Sanjana et al., 2012). In addition, only a few restrictions (see section 1.4.1) have to be taken into account when designing synthetic TALEs (Cermak et al., 2011). This offers a larger flexibility in the choice of DNA target sequences, which is a great advantage over the ZFN technology.

Since their invention, TALENs and synthetic TALEs have been used in a variety of targeted genome engineering applications. In chapter 1.5, I will discuss how genome editing technologies are enabling a broad range of applications from research to biotechnology and medicine.

1.2.3 Clustered regularly interspaced short palindromic repeat (CRISPR)–Cas9 system

In 2013, a genome editing tool referred to as CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats/ CRISPR-associated-9 (Cas9) system) transformed biology research and spread through laboratories faster than any genome editing tool before. Many researchers and scientific journals believe that this “powerful gene-editing technology is the biggest game changer to hit biology since PCR” (Ledford, 2015). In this chapter, I will briefly review how

the CRISPR-Cas9 system developed to a powerful and popular genome editing tool.

The acronym “CRISPR” was invented by Mojica et al. and Jansen et al. and used to reflect the characteristic features of a family of repetitive sequences commonly found in genomes of prokaryotic organisms (Jansen et al., 2002). Genomic analysis of microbial genomes suggested that CRISPR repeats are widespread among bacteria and archaea (Mojica et al., 2000). These findings stimulated interest in such microbial repeats and whether those repeats have a biological function in prokaryotes. Subsequent studies identified conserved CRISPR-associated (Cas) genes, which are typically located next to the repeat elements and encode Cas proteins (Haft et al., 2005; Jansen et al., 2002; Makarova et al., 2011a). More systematic analysis of CRISPR elements and Cas proteins lead to our current view of the CRISPR-Cas system as an adaptive immunity system in bacteria and archaea that uses short RNA fragments to destroy foreign nucleic acids (Barrangou et al., 2007; Bolotin et al., 2005; Brouns et al., 2008; Marraffini and Sontheimer, 2010; Mojica et al., 2005; Pourcel et al., 2005).

The CRISPR defence system involves three key steps: (i) acquisition and integration of new targeting sequences (spacers) from invading viruses or plasmid DNA into the CRISPR locus, (ii) expression and processing of short guiding CRISPR RNAs (crRNAs) consisting of spacer-repeat units, and (iii) cleavage of nucleic acids (most commonly DNA) complementary to the spacer (Westra et al., 2012) (Figure 4).

Three classes of CRISPR systems have been described so far (Type I, II and III). The most highly conserved Cas proteins, Cas1 and Cas2, are present in all three CRISPR basic types (Westra et al., 2012). To process the crRNAs and cleave target nucleotides sequences, Type I and Type III systems require multiple effector proteins acting as a complex. In contrast, Type II CRISPR

systems utilize a single effector enzyme, Cas9, to cleave double-stranded DNA (Makarova et al., 2011b; Westra et al., 2012) (Figure 4).

During the adaptation stage, short pieces of DNA homologous to virus or plasmid sequences are integrated into the CRISPR loci (spacers, sized ~30-bp) (Barrangou et al., 2007). Invading DNA is recognized most probably through the protospacer-adjacent motif (PAM) either downstream (Type I) or upstream (Type II) of the protospacer (Deveau et al., 2008). New spacer sequences are integrated at the leader end of the CRISPR array, a process termed “polarized addition” (Horvath et al., 2008; Pourcel et al., 2005). Cas1 and Cas2 seem to be involved in spacer integration into the CRISPR locus, but the process of spacer integration is still not fully understood (Nunez et al., 2014; Westra et al., 2012).

Transcription of the CRISPR array gives rise to a pre-crRNA molecule (Brouns et al., 2008), which is subsequently cleaved in the repeat sequence by an endo-ribonuclease in Type I and Type III systems (often a Cas6 homolog), and by RNase III and Cas9 in Type II systems (Deltcheva et al., 2011) (Figure 4).

A study by Charpentier and colleagues investigating the processing of crRNA in type II CRISPR systems elucidated a crucial component for RNA-guided targeting of foreign nucleotides that could be bacteriophages or plasmids – a short trans-activating crRNA (tracrRNA). The authors could show that the tracrRNA acts with crRNA as a RNA hybrid, which is then used by the Cas9 enzyme to promote the maturation of crRNAs (Deltcheva et al., 2011). Genetic studies further determined that Cas9 is the only enzyme required to confer immunity against bacteriophages and plasmids (Garneau et al., 2010), suggesting that a few components are essential for reconstituting the type II CRISPR system. Around the same time, Siksnys and coworkers demonstrated that components of the type II CRISPR-Cas system could be transferred across distant bacteria species to protect against exogenous plasmids and phage infection (Sapranauskas et al., 2011).

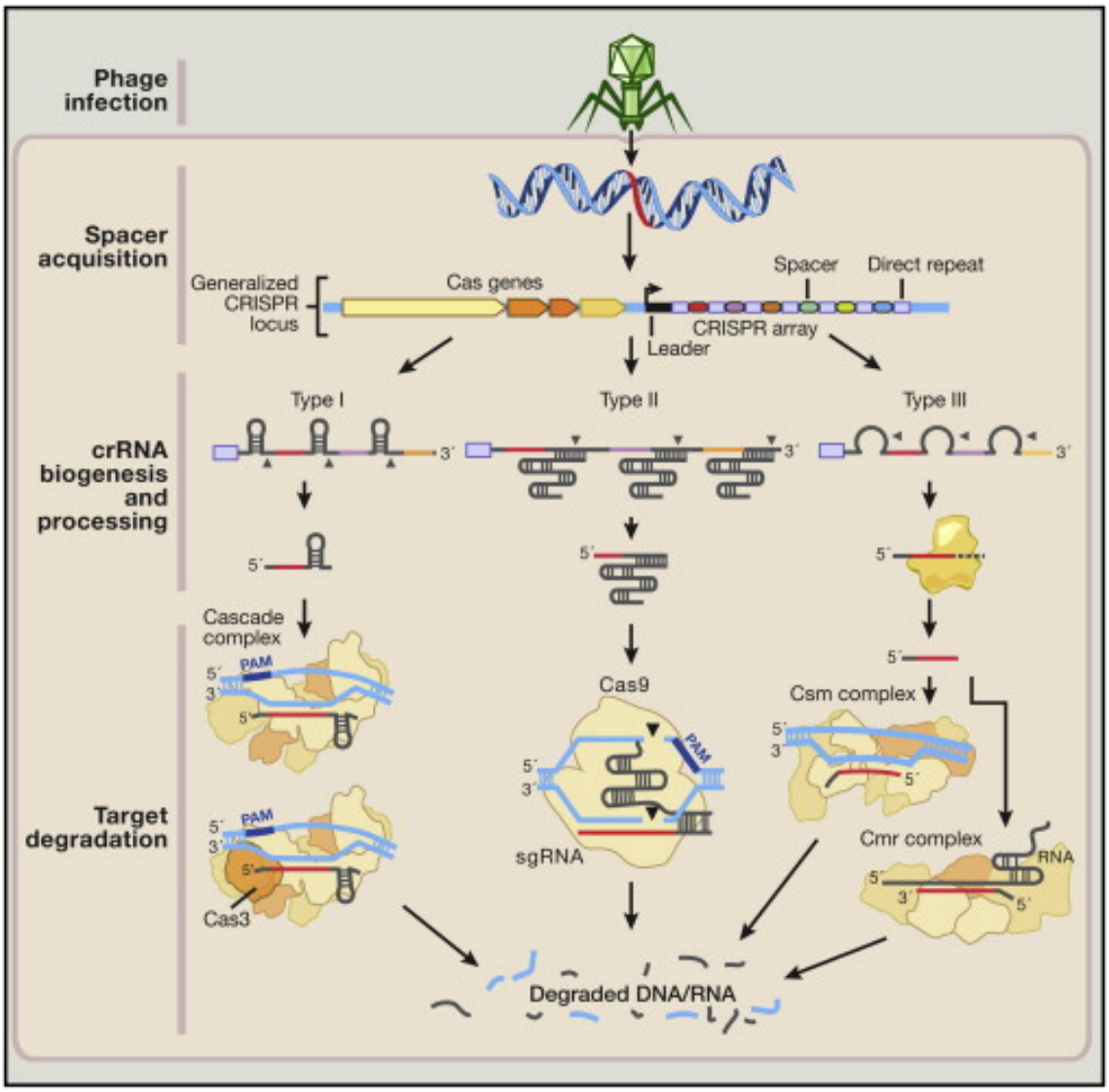


Figure 4. Natural Mechanisms of Microbial CRISPR Systems In Adaptive Immunity

Following invasion of the cell by foreign genetic elements from bacteriophages or plasmids (step 1: phage infection), certain CRISPR-associated (Cas) enzymes acquire spacers from the exogenous protospacer sequences and install them into the CRISPR locus within the prokaryotic genome (step 2: spacer acquisition). These spacers are segregated between direct repeats that allow the CRISPR system to mediate self and nonself recognition. The CRISPR array is a noncoding RNA transcript that is enzymatically maturated through distinct pathways that are unique to each type of CRISPR system (step 3: crRNA biogenesis and processing).

In types I and III CRISPR, the pre-crRNA transcript is cleaved within the repeats by CRISPR-associated ribonucleases, releasing multiple small crRNAs. Type III crRNA intermediates are further processed at the 3′ end by yet-to-be-identified RNases to produce the fully mature transcript. In type II CRISPR, an associated *trans*-activating CRISPR RNA (tracrRNA) hybridizes with the direct repeats, forming an RNA

duplex that is cleaved and processed by endogenous RNase III and other unknown nucleases. Maturated crRNAs from type I and III CRISPR systems are then loaded onto effector protein complexes for target recognition and degradation.

In type II systems, crRNA-tracrRNA hybrids complex with Cas9 to mediate interference. Both type I and III CRISPR systems use multiprotein interference modules to facilitate target recognition. In type I CRISPR, the Cascade complex is loaded with a crRNA molecule, constituting a catalytically inert surveillance complex that recognizes target DNA. The Cas3 nuclease is then recruited to the Cascade-bound R loop, mediating target degradation. In type III CRISPR, crRNAs associate either with Csm or Cmr complexes that bind and cleave DNA and RNA substrates, respectively. In contrast, the type II system requires only the Cas9 nuclease to degrade DNA matching its dual guide RNA consisting of a crRNA-tracrRNA hybrid. Figure taken with permission from (Hsu et al., 2014).

Considering the vast attention programmable nucleases for genome editing obtained in late 2011 (Baker, 2012) and the importance of these tools for basic research as well as biomedical applications, a race to harness the potential of Cas9 for genome editing began. By 2012, several biochemical analyses showed that Cas9 is a DNA endonuclease guided by two RNAs, a hybrid between the activating tracrRNA and the targeting crRNA. A key study by Jinek et al. revealed the potential of Cas9 for eukaryotic genome editing (Jinek et al., 2012). The authors could reprogram Cas9 with a single chimeric RNA, so called guide RNA (sgRNA or gRNA), to cleave DNA (Figure 5A). The guide RNA can be engineered to direct Cas9 site-specific cleavage of almost any DNA locus, making the Cas9 RNA-guided system easy to use, “efficient, versatile, and programmable by changing the DNA target-binding sequence in the guide chimeric RNA” (Jinek et al., 2012).

In 2013, a wave of studies demonstrated how the Cas9 system could be used to accomplish highly efficient genome editing in mammalian cells and zebrafish (Cho et al., 2013; Cong et al., 2013; Hwang et al., 2013; Mali et al., 2013b). Since these pioneering studies, the scientific community has rapidly adopted the CRISPR-Cas9 technology. The fast spread of Cas9 into research laboratories around the world was greatly facilitated by three additional aspects: i) the accessibility of expression constructs for nuclear-localized Cas9 and

variant forms of guide RNAs through open-source plasmid distributors such as Addgene, ii) a number of online user forums, and iii) online software tools to design guide RNAs for the use with the CRISPR-Cas9 system.

1.3 Shared mechanistic features of programmable nucleases

One of the most harmful DNA lesions that cells encounter are DNA double strand breaks. Unrepaired or error-prone repaired DSBs can potentially lead to cell death or oncogenic mutations (Khanna and Jackson, 2001). To protect cells from DSBs, high fidelity repair of DNA damage evolved in every cell to maintain genomic integrity (Chapman et al., 2012; Moynahan and Jasin, 2010). Homologous directed repair (HDR) is essentially a 'copy and paste' mechanism, which uses an intact homologous segment of DNA as a template to copy and replace the damaged DNA across the break (Figure 1A). This mechanism is the most accurate form of DSB repair active in most eukaryotic cells. An alternative pathway besides HDR is non-homologous end joining (NHEJ), which joins DSB-ends without regard for homology. NHEJ is an error-prone DNA repair pathway often resulting in small, localized deletions and/or insertions at the break that can cause gene disruption (Lieber, 2010).

A common action of programmable nucleases such as ZFNs, TALENs and Cas9 is the generation of a DSB at a target genomic locus (Hsu et al., 2014; Joung and Sander, 2013; Urnov et al., 2010). Upon cleavage by programmable nucleases, the target locus typically undergoes one of two major pathways for DNA damage repair: the error-prone NHEJ or the high-fidelity HDR pathway, both of which can be used to achieve a desired editing outcome (Figure 1A). The phase of the cell cycle largely governs the choice of pathway to repair a DSB using NHEJ or HDR. NHEJ dominates DNA repair during G1, S and G2 phases (Karanam et al., 2012).

In the absence of a repair template, DSBs are joined through NHEJ, which generates variable insertion/deletion (indel) mutations at the break (Lieber, 2010). NHEJ can be harnessed to mediate gene knockouts, as indels occurring within a coding exon can lead to frameshift mutations and premature stop codons (Ran et al., 2013). Additionally, multiple DSBs can be exploited to facilitate larger deletions in the genome (Cong et al., 2013; Flemr et al., 2013).

HDR is an alternative major DNA repair pathway present in most eukaryotic cells. HDR typically occurs at lower and more variable frequencies than NHEJ (Karanam et al., 2012). Nonetheless, it can be utilized to generate exact and defined modifications at a target locus in the presence of an exogenously introduced repair template (Carroll, 2011; Kim and Kim, 2014; Ran et al., 2013). The repair template can either be in the form of conventional double-stranded DNA (dsDNA) targeting donors with homology arms flanking the insertion sequence, or single-stranded DNA oligonucleotides (ssODNs). The latter provides an effective and simple method for making small edits in the genome, such as the introduction of single-nucleotide mutations for probing causal genetic variations, insertion of small DNA sequences to tag genes or create conditional alleles (Chen et al., 2011; Flemr and Buhler, 2015). HDR is active only in dividing cells, where it is restricted to S phase when DNA replication is highest and sister chromatids are available to serve as repair templates (Heyer et al., 2010; Karanam et al., 2012). HDR-mediated genome editing can vary widely depending on the cell cycle state, cell type, as well as the genomic locus and repair template.

1.4 Improving genome editing technologies

1.4.1 Limitations of existing genome editing technologies

Genome editing leads to permanent modifications within the genome. Targeting Specificity and activity are of particular concern when designing programmable nucleases (Kim and Kim, 2014). Other key features that need consideration when using DNA targeting nucleases are delivery and genome editing outcomes (Cox et al., 2015; Hsu et al., 2014). Here, I will examine the most important aspects concerning design and targeting specificity of ZFNs, TALENs and Cas9 nucleases.

ZFNs

Design. Compared to other programmable nucleases, the construction of ZFNs with high activity and high targeting specificity remains challenging. Methods for modular assembly of ZFNs that account for context dependence between neighbouring zinc-finger modules are laborious and often fail to produce high-quality ZFNs. In fact, off-target DNA cleavage events of self-made ZFNs are believed to cause cytotoxicity (Cornu et al., 2008). Alternatively, custom-made ZFN services (Sigma-Aldrich) use a proprietary archive of zinc-finger modules that yields ZFNs of higher quality (Kim and Kim, 2014). However, the service is expensive for most academic institutions making ZFNs the least attractive genome editing tool.

Targeting specificity. A general feature of FokI-domain containing nucleases is their modular structure, which is composed of two domains: a DNA-binding domain and the FokI nuclease domain. Furthermore, FokI-domain containing nucleases must bind as pairs on the target site to cleave DNA (Bitinaite et al., 1998). This feature can yield high specificity and is often

explained making the following approximation: if DNA were a random polymer of four nucleotides, a 16-bp recognition sequence would be cleaved every 4^{16} or 4.3×10^9 base-pairs, a number that is greater than the size of the human haploid genome (3.2×10^9). Target sequences of ZFNs are usually 18 – 30-bp in length (excluding spacers) and theoretically unique in a complex genome (Figure 2B). Nevertheless, the unpredictable context-dependent neighbouring effects of zinc-finger modules frequently result in poor ZFN cleavage activity and targeting specificity (Cornu et al., 2008; Ramirez et al., 2008). In contrast to ZFNs, TALENs and Cas9 are less cytotoxic, which facilitates the manipulation of single cells to create animals with edited genomes (Hsu et al., 2014; Kim and Kim, 2014).

TALENs

Design. Similar to ZFNs, TALENs are FokI-domain containing nucleases composed of two domains: a DNA-binding domain and the FokI-nuclease domain. TALE-nucleases (TALENs) use a different class of DNA-binding domain, known as TALEs, which were discovered in plant pathogenic bacteria. Natural TALEs consist of an array of 34 amino-acid long repeats (Figure 3A). Each repeat recognizes a single DNA base-pair through the adjacent amino-acid residues at positions 12 and 13, the 'repeat-variable di-residue' (RVD). Four different RVD modules are most commonly used to bind the four DNA nucleotides and even methylated cytosine (Kim and Kim, 2014; Valton et al., 2012). The natural occurrence of TALEs and the one-to-one match between the four RVDs and the four DNA bases makes it easy to design novel TALENs to target almost any DNA sequence. Structural and computational studies showed that nearly all TALE binding sites observed in nature require a thymidine at the 5´ end of the target site (Mak et al., 2013; Moscou and Bogdanove, 2009). Currently, this seems to be the most important parameter in the design of TALENs (Reyon et al., 2012).

Several cloning methods have been developed for the assembly of custom-made TALE arrays (Briggs et al., 2012; Cermak et al., 2011; Reyon et al., 2012; Sanjana et al., 2012; Schmid-Burgk et al., 2013). A TALE array for genome engineering applications typically consists of 15 – 20 RVDs (Figure 3B). The construction of several arrays can be a time-consuming and laborious process. Therefore, improved cloning methods have been developed to construct large libraries of TALENs targeting genes and non-coding sequences in the human genome (Kim et al., 2013a; Kim et al., 2013b) allowing the application of TALENs in functional genomics screens.

Targeting specificity. TALENs can be designed to target almost any given DNA sequence, which is a crucial advantage over ZFNs, Cas9 and other known Cas nucleases. Target sequences of TALEN pairs are typically 30 – 40 bp in length, excluding spacers. In contrast to ZFNs and Cas9, TALENs are believed to have less widespread off-target activity (Kim and Kim, 2014; Miller et al., 2015). However, the principles that govern TALEN-specificity remain poorly understood and are currently and active area of research. A few studies addressing these principles used *in vitro* techniques or *in vivo* methods to investigate TALEN off-target activity. Both approaches have their own pros and cons. But, they are beginning to reveal additional factors that affect target specificity and cleavage activity such as design strategy, chromatin structure and locus accessibility (Guilinger et al., 2014; Miller et al., 2015; Veres et al., 2014).

Cas9

Design. Streptococcus pyogenes Cas9 (SpCas9) is the most broadly used RNA-guided nuclease to date. SpCas9 can be directed to target DNA loci either with a pair of crRNA and tracrRNA (Cong et al., 2013) or with a single RNA-hybrid guide sgRNA (Cho et al., 2013; Cong et al., 2013; Jinek et al., 2012; Jinek et al., 2013; Mali et al., 2013b). Researchers rapidly adopted this

technology because of its simple preparation and high efficiency. The only component of the system that needs to be designed to target a particular DNA locus is the crRNA or sgRNA, which contains 20-nt sequence complementarity to match a DNA target site (Figure 5A). Both sequences can be cloned into plasmids under the control of a RNA polymerase III promoter for efficient transcription in cells (Cong et al., 2013; Ran et al., 2013). Unlike TALENs or ZFNs, the Cas9 enzyme does not require *de novo* design and *in vitro* synthesis. Several laboratories have made plasmids encoding SpCas9 for expression in a variety of species and cell types available to the scientific community. In addition, some guide RNA design software is freely available online such as: http://crispr.mit.edu or http://www.e-crisp.org/E-CRISP/. In summary, these features represent a crucial advantage of RNA-guided nucleases like Cas9 over ZFNs and TALENs.

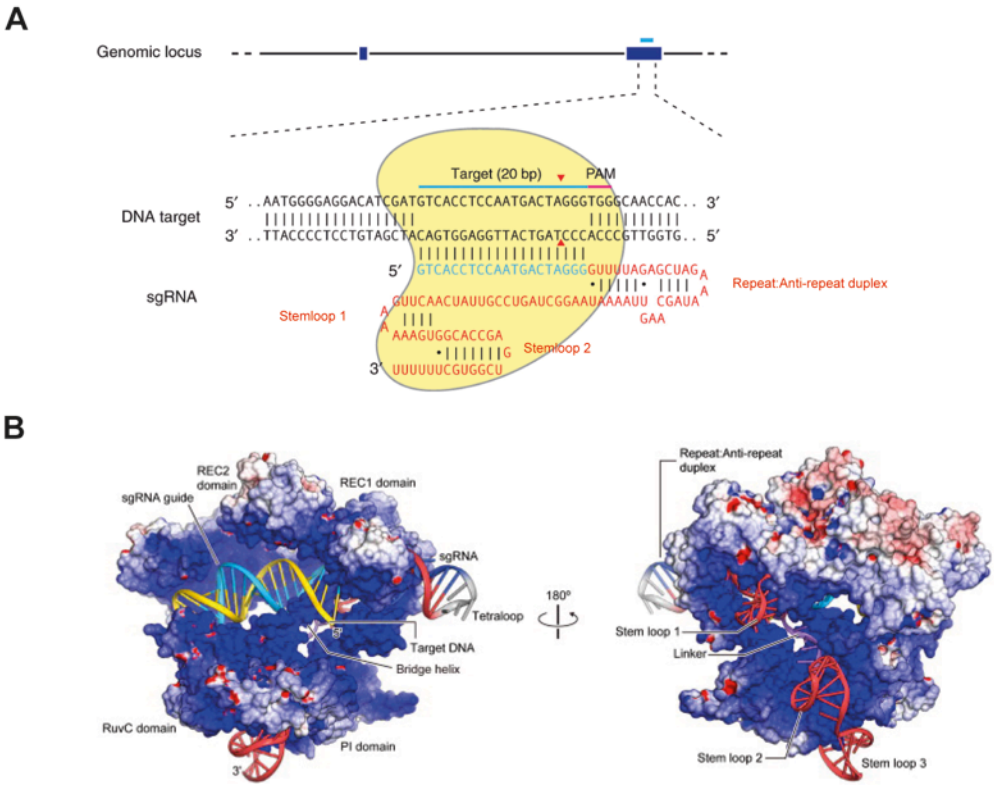


Figure 5. Overall Structure of the Cas9-sgRNA-DNA Ternary Complex

(**A**) The Cas9 nuclease from *S. pyogenes* (in yellow) is targeted to genomic DNA (shown for example is the human EMX1 locus) by an sgRNA consisting of a 20-nt guide sequence (blue) and a scaffold (red). The guide sequence pairs with the DNA target (blue bar on top strand), directly upstream of a requisite 5′-NGG adjacent motif (PAM; pink). Cas9 mediates a DSB ~3 bp upstream of the PAM (red triangle). Image was taken with permission from (Ran et al., 2013). (**B**) Electrostatic surface potential of Cas9. The HNH domain is omitted for clarity. Image taken with permission from (Nishimasu et al., 2014).

Targeting specificity. A critical feature for target recognition by Cas9 and other RNA-guided nucleases is the PAM, which flanks the 3´end of the DNA target site (Figure 5A). Potential target sites lacking a PAM are unlikely to be interrogated for DNA cleavage by the Cas9 enzyme (Hsu et al., 2013; Pattanayak et al., 2013). Biochemical studies propose that sgRNA-loaded Cas9 uses PAM recognition to identify potential target sites while scanning DNA (Gasiunas et al., 2012; Sternberg et al., 2014). Loading of the RNA guide induces large conformational changes, leading to the conversion of Cas9 from an inactive into a structurally activated DNA endonuclease (Jinek et al., 2014). Binding to a correct target sequence results in an additional structural rearrangement (Figure 5B). This activates both nuclease domains of Cas9 to engage for DNA cleavage (Jinek et al., 2014; Nishimasu et al., 2014). Collectively, these investigations highlight the importance of the PAM and sequences adjacent to this motif for target recognition and cleavage.

In addition, the PAM sequence is specific to each Cas9 ortholog (Hsu et al., 2014). SpCas9 targets DNA sites flanked by 5´– NGG and with a much lower efficiency sites flanked by 5´– NAG (Hsu et al., 2013). The complexity of the PAM sequence determines the overall frequency of potential target sites in a given genome. For example, the 5´– NGG of SpCas9 allows it to target every 8-bp on average within the human genome (Cong et al., 2013). Current evidence suggests that Cas9 technology can cause widespread off-target effects and genome alterations (Cho et al., 2014; Frock et al., 2015; Fu et al., 2013; Hsu et al., 2013; Kuscu et al., 2014; Pattanayak et al., 2013). However, these studies also suggest that potential off-target effects and possible genome alterations

following Cas9 treatment could be overcome by rational sgRNA design and optimal enzymatic concentration. A few studies have demonstrated that choosing unique target sites in the genome minimize or completely avoid off-target effects (Cho et al., 2014; Kim et al., 2013a). Unlike ZFNs and TALENs, Cas9 enzyme binds its targets mostly by simple Watson-Crick base-pairing rules. Considering the importance of the PAM and sequences adjacent to this motif, these features will allow a more reliable prediction of potential off-target sites using sequence homology algorithms. Finally, enzymatic concentration and duration of Cas9 expression are likely further factors that modulate target specificity and require careful investigation in the future.

1.4.2 Enhancing homologous recombination in genome editing

Site-specific genetic insertions using programmable nucleases are greatly limited by the low rates of homologous directed repair. Besides homology directed repair (HDR), a second major pathway for the repair of DNA breaks is the error-prone NHEJ pathway. Unrepaired DSBs can result in genomic instability leading to apoptosis or senescence. Therefore, cells evolved a less accurate form of DSB repair, the NHEJ repair pathway, in which the broken DNA ends are processed and re-ligated without the need of a repair template. As discussed in the previous chapter, NHEJ is generally believed to be the dominant cellular repair pathway (Karanam et al., 2012).

Researchers have designed several strategies to shift the balance from NHEJ to HDR and increase the frequency of recombination. In principle, DSB repair pathways could be manipulated either genetically or chemically to favour HDR over NHEJ. Genetic approaches to modulate the HDR repair machinery have been pioneered in plants without the involvement of programmable nucleases. Early studies overexpressed key recombination genes from Escherichia coli (RecA and RuvC) in plants without effects on gene targeting (Weinthal et al., 2010). More recent approaches applying protein fusions

consisting of site-specific TALENs fused to several recombination proteins such as Rad51, Ledgf, Znhit1, Nabp2 and Nbn failed to enhance targeted homologous recombination in mammalian cells (Flemr M., unpublished results). Transient inhibition of NHEJ pathway has recently been regarded as an alternative way to favour HDR and boost precise genetic insertions into mammalian genomes. A study by Kühn and colleagues published earlier this year demonstrated that genetic suppression of key enzymes involved in NHEJ such as Ku70, Ku80 or DNA ligase IV greatly stimulates the efficiency of HDR in mammalian cells (Chu et al., 2015). Along the same lines, chemical inhibition of DNA ligase IV, with the small molecule SCR7, showed increased efficiency of HDR-mediated genome editing in cultured cells and mice (Chu et al., 2015; Maruyama et al., 2015). However, chemical inhibition of NHEJ seems to be toxic even at low concentrations in cultured cells. Therefore, the development of more specific and less toxic “NHEJ inhibitors” would further extend the use of this approach. Since NHEJ and HDR pathways are evolutionary conserved across eukaryotes, it is likely that suppression of NHEJ is applicable to other model organisms to enhance HDR-mediated genome editing.

1.4.3 Enriching nuclease activity and homologous recombination events

Although TALEN and Cas9 technology greatly enhanced the success rate of genome editing, the search for genome-edited cells is time-consuming and requires laborious screening procedures. In addition, genome editing strongly depends on efficient delivery of programmable nucleases into the cell nucleus. Delivery of site-specific programmable nucleases can be achieved in cultured cells by chemical or biological transfection methods. Chemical transfection methods are more widely used in research laboratories because of their ease of use and low costs. However, transfection efficiency of these methods depends on several factors and varies greatly across different cell types (Kim and Eberwine, 2010). To further improve the technology and circumvent poor

delivery of nuclease encoding-plasmids into cells, systems that allow enrichment of genome-edited cells were required.

The most successful strategies to enrich for genome-edited cells use reporter plasmids that monitor NHEJ or HDR events as a result of nuclease activity. Kim and colleagues developed reporter constructs for monitoring NHEJ and enrich for mutant cells by flow-cytometry or antibiotic selection (Kim et al., 2011; Kim et al., 2014b; Ramakrishna et al., 2014). The first generation of these reporters contained the nuclease target sequence and expressed green fluorescent protein (GFP) only after a cut was made and the break was repaired by NHEJ (Kim et al., 2011). Later reporter constructs using the same strategy were combined with antibiotic resistance genes instead of GFP (Kim et al., 2014b; Ramakrishna et al., 2014).

Our laboratory recently established a similar approach to monitor HDR of nuclease-induced DSBs, which in turn allow the enrichment of genome-edited cells. Flemr and Bühler “adopted the architecture of an existing HR-dependent β-galactosidase reporter (Wefers et al., 2013) to create recombination reporter plasmids (pRR-EGFP, pRR-Puro) that contain a prematurely terminated N-terminal part of the EGFP or puromycin coding sequence followed by a multiple cloning site (MCS) and a full-length enhanced GFP (EGFP) or puromycin resistance coding sequence, respectively, with a stop codon in place of a start codon" (Flemr and Buhler, 2015). Upon insertion of a nuclease target sequence into the MCS and co-transfection of this reporter plasmid with the respective nuclease-encoding constructs, the nuclease-induced DSB in the plasmid is anticipated to be repaired via recombination of the N-terminal part with the homologous portion of the full-length sequence. This results in functional EGFP or puromycin resistance, reporting on successful transfection, nuclease activity, and an effective HDR pathway (Figure 6). Finally, this approach allows the efficient generation of conditional knockouts and gene tagged cell lines in a single step.

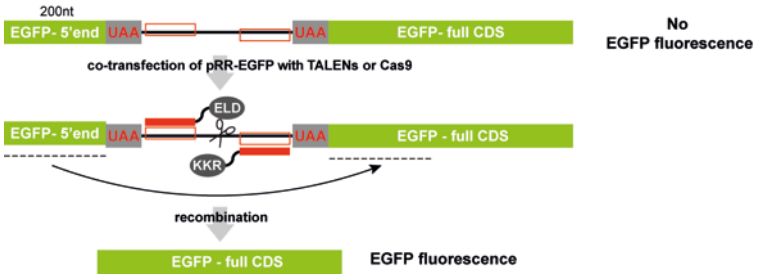


Figure 6. Schematic Demonstration of TALEN-mediated pRR-EGFP Recombination

Red lined rectangles depict TALEN recognition sites. ELD, KKR, TALENs with obligate heterodimeric FokI nuclease domain. Image adapted with permission from (Flemr and Buhler, 2015).

1.5 Applications of genome editing technologies

1.5.1 Applications in life science research

The recent advances in genome editing technologies based on ZFNs, TALENs and the RNA-guided endonuclease Cas9 are enabling powerful applications in life science research, biotechnology and medicine. Excellent review articles on this topic have been published elsewhere offering a great resource of information for the interested reader (Cox et al., 2015; Hilton and Gersbach, 2015; Hsu et al., 2014; Joung and Sander, 2013; Kim and Kim, 2014). In this chapter, I will briefly review the current and potential future applications of programmable nucleases in life science research and medicine.

The ease of use of TALENs and Cas9 has enabled the rapid generation of cellular and animal models to study human disease (Chen et al., 2013b; Niu et al., 2014; Schwank et al., 2013; Wang et al., 2013; Wu et al., 2013). We pioneered the use of ZFNs in drug discovery by introducing a luciferase reporter into an endogenous gene to identify potential drug targets that activate transcription at that particular disease locus (Villasenor et al., 2015). This approach could be applied to screen drugs to treat several human monogenic

diseases. Other laboratories and our own group have combined genome-editing tools with proteomic or functional genomic approaches to map the protein-protein, protein-DNA or protein-RNA interactions of nearly any endogenous protein (Dalvai et al., 2015)(Villaseñor et al., unpublished results, Tuck et al., unpublished results). Furthermore, the simple preparation of sgRNAs allows application of Cas9 in large-scale unbiased gene disruption screens (functional genomic screens, Figure 7F) to elucidate gene function (Shalem et al., 2014; Wang et al., 2014a). In contrast to RNA interference (RNAi) screens, functional genomics screens using Cas9 or TALENs enable the interrogation of non-coding parts of mammalian genomes.

The ability to engineer ZFs, TALEs or reprogram Cas9 to bind a particular DNA site has unlocked several powerful applications. ZFs, TALEs and RGNs can be joined to a wide variety of effector domains such as nucleases, transcription effectors, and epigenetic modifying enzymes to carry out site-specific modifications near their DNA binding site (Hilton and Gersbach, 2015; Hsu et al., 2014). Early proof-of-concept studies demonstrated the ability of ZFs to tether transcription activating or repressing domains to DNA and induce transcriptional changes at the target site (Beerli et al., 2000a; Beerli et al., 2000b; Beerli et al., 1998). These and other studies encouraged applying similar strategies to modulate endogenous gene expression using TALEs or Cas9. Subsequent studies showed that synthetic TALEs fused to transcription activators or repressors could robustly modulate transcription at endogenous chromatin target sites (Zhang et al., 2011). Mutation of the catalytic residues of both nuclease domains, the HNH and RuvC-like domains, converts Cas9 to an inactive enzyme (dCas9), but preserves its ability to specifically bind DNA target sites (Jinek et al., 2012). Similar to ZFs or TALENs, dCas9 can then be fused to transcription activating or repressing domains to modulate transcription at the target locus (Gilbert et al., 2013; Konermann et al., 2013; Konermann et al., 2015; Maeder et al., 2013; Mali et al., 2013a; Perez-Pinera et al., 2013) (Figure 7G). Most of these approaches require tethering at multiple target sites and

achieve moderate changes in gene expression at target loci. Konermann and colleagues recently developed an elegant strategy to overcome these difficulties (Konermann et al., 2015). They identified two regions of the sgRNA that can be adapted with short RNA sequences, so called aptamers, which in turn attract an RNA-binding protein. The RNA-binding protein can then be fused to the transcription-activation domains of mammalian transcription factors. The authors termed this system the synergistic activation mediator (SAM), and demonstrate that it activates endogenous human genes more efficiently than activated by the dCas9-activator fusion protein (Konermann et al., 2015). Furthermore, ZFs, TALEs and dCas9 can be fused to chromatin modifiers, such as histone modifying enzymes and DNA modifying enzymes, to alter epigenetic marks near the target site (Hilton and Gersbach, 2015). These “designed epigenetic effectors” can artificially install or remove epigenetic marks at specific target loci and could serve as a powerful tool to probe the causal effects of epigenetic modifications on gene expression for instance.

“Genomes are more than linear sequences. In vivo they exist as elaborate physical structures, and their functional properties are strongly determined by their cellular organization” (Misteli, 2007). To study the spatio-temporal organisation and dynamics of genomes researchers require robust methods to visualize DNA in living cells. Pioneering studies have developed live imaging techniques by fusing TALEs or dCas9 to fluorescent proteins to visualize and record the dynamics of single or multiple chromatin sites in mammalian cells (Chen et al., 2013a; Ma et al., 2015; Miyanari et al., 2013) (Figure 7I).

Taken together, advances in genome engineering technologies enabled researchers to easily edit or modulate DNA sequences in a variety of organisms. These powerful tools will serve to systematically interrogate mammalian genome function and gain fundamental insights of human biology and disease.

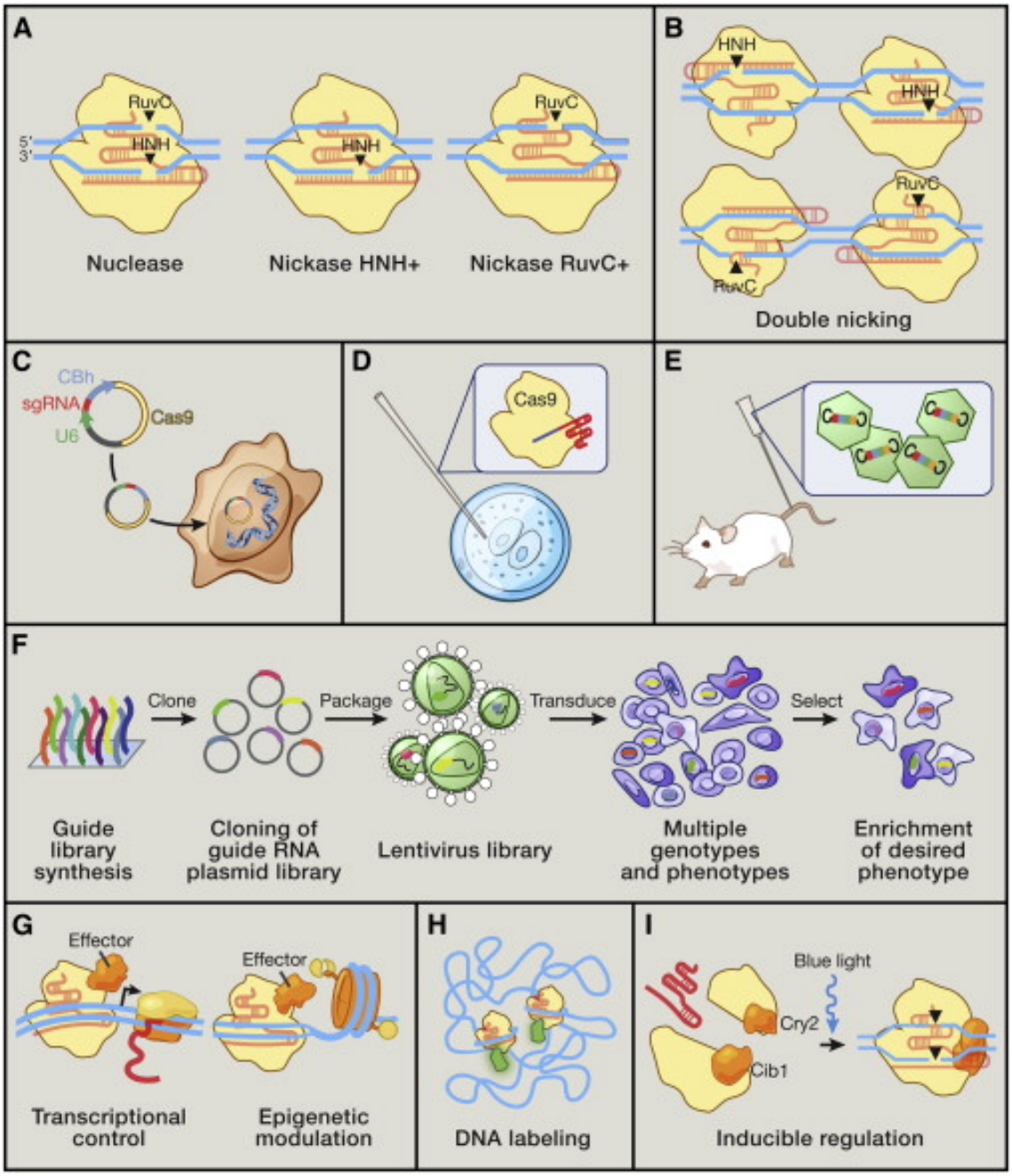


Figure 7. Applications of Cas9 as a Genome Engineering Platform

(**A**) The Cas9 nuclease cleaves DNA via its RuvC and HNH nuclease domains, each of which nicks a DNA strand to generate blunt-end DSBs. Either catalytic domain can be inactivated to generate nickase mutants that cause single-strand DNA breaks. (**B**) Two Cas9 nickase complexes with appropriately spaced target sites can mimic targeted DSBs via cooperative nicks, doubling the length of target recognition without sacrificing cleavage efficiency. (**C**) Expression plasmids encoding the Cas9 gene and a short sgRNA cassette driven by the U6 RNA polymerase III promoter can be directly transfected into cell lines of interest.

(**D**) Purified Cas9 protein and in vitro transcribed sgRNA can be microinjected into fertilized zygotes for rapid generation of transgenic animal models. (**E**) For somatic genetic modification, high-titer viral vectors encoding CRISPR reagents can be transduced into tissues or cells of interest. (**F**) Genome-scale functional screening can be facilitated by mass synthesis and delivery of guide RNA libraries. (**G**) Catalytically dead Cas9 (dCas9) can be converted into a general DNA-binding domain and fused to functional effectors such as transcriptional activators or epigenetic enzymes. The modularity of targeting and flexible choice of functional domains enable rapid expansion of the Cas9 toolbox. (**H**) Cas9 coupled to fluorescent reporters facilitates live imaging of DNA loci for illuminating the dynamics of genome architecture. (**I**) Reconstituting split fragments of Cas9 via chemical or optical induction of heterodimer domains, such as the cib1/cry2 system from Arabidopsis, confers temporal control of dynamic cellular processes. Figure taken with permission from (Hsu et al., 2014).

1.5.2 Applications in medicine

A tantalizing application for programmable nucleases is the potential to correct disease-causing genetic mutations in affected tissues or cells and treat human diseases with unmet medical need. This particular application has raised tremendous hope and excitement (Gaj et al., 2013; Ledford, 2015). Therapeutic genome editing was pioneered using ZFNs to correct disease-causing mutations in human cells (Chu et al., 2015; Genovese et al., 2014; Li et al., 2011; Li et al., 2015; Perez et al., 2008; Urnov et al., 2005). But the later development of TALENs and CRISPR-Cas9 dramatically increased proof-of-principle studies demonstrating the potential of genome editing for therapeutic purposes (Bloom et al., 2013; Ding et al., 2014; Hu et al., 2014; Kennedy et al., 2015; Kennedy et al., 2014; Liang et al., 2015; Lin et al., 2014; Long et al., 2014; Mahiny et al., 2015; Ousterout et al., 2013; Schwank et al., 2013; Wu et al., 2013; Ye et al., 2014; Yin et al., 2014).

Examples of applications of genome editing to therapeutic models

Disease type	Nuclease platform	Therapeutic strategy	Reference(s)
Hemophilia B	ZFN	HDR-mediated insertion of correct gene sequence	Li et al., 2011
HIV	ZFN and CRISPR	NHEJ-mediated inactivation of CCR5	Perez et al., 2008, Holt et al., 2010, Tebas et al., 2014, Ye et al., 2014
Duchenne muscular dystrophy (DMD)	CRISPR and TALEN	NHEJ-mediated removal of stop codon, and HDR-mediated gene correction	Ousterout et al., 2013, Long et al., 2014
Friedreich´s ataxia	ZFN	NHEJ-mediated excision of expanded repeats	Li et al., 2015
Hepatitis B virus (HBV)	TALEN and CRISPR	NHEJ-mediated depletion of viral DNA	Bloom et al., 2013, Lin et al. 2014
SCID	ZFN	HDR-mediated insertion of correct gene sequence	Genovese et al., 2014
Cataracts	CRISPR	HDR-mediated correction of mutation in mouse zygote	Wu et al., 2013
Cystic fibrosis	CRISPR	HDR-mediated correction of CFTR in intestinal stem cell organoid	Schwank et al., 2013
Hereditary tyrosinemia	CRISPR	HDR-mediated correction of mutation in liver	Yin et al., 2014

Table 1. Examples of Genome Editing to Therapeutic Models

Therapeutic genome editing involves a number of strategies including correction or inactivation of disease-causing mutations, introduction of protective mutations, addition of therapeutic transgenes and disruption of viral DNA (Cox et al., 2015) (Table 1). Some of these approaches have been successfully applied to a number of diseases at the preclinical level as well as in a phase 1 clinical trial. However, there are still major challenges and questions that need to be addressed before the technology can move to the clinic (Cox et al., 2015). The most important are believed to be the following:

i) The specificity of each programmable nuclease.
ii) The different efficiencies of DSB repair pathway. NHEJ is more active than HDR challenging strategies that require gene correction or gene insertion.
iii) The challenge to efficiently deliver the nuclease into the cell or tissue.
iv) The uncertainty of programmable nucleases causing cellular or systemic immune responses.

Although the genome editing field is still in its infancy, it has the potential to revolutionize medicine (Ledford, 2015). Scientists have seen other promising

technologies such as RNAi and gene therapy that prompted great excitement, concern and then disappointment when major problems arose (Kim and Rossi, 2007; Thomas et al., 2003). Therapeutic genome editing faces similar problems to those of gene therapy and therapeutic RNAi (Cox et al., 2015). Therefore, therapeutic genome editing might profit from recent advances in these fields and lessons learned from past difficulties. Important questions concerning safety and efficacy are now the focus of current investigations to translate this technology to the clinic.

1.6 Aim of this Thesis

During the course of this work I applied genome editing tools in four different projects:

Project I: Application of genome editing for drug discovery

In the first half of my doctoral studies, I developed genome-editing strategies to discover drug targets for a rare genetic disease. I generated cell-based assays using ZFNs for high-throughput genomic screens to uncover novel strategies to treat Friedreich's Ataxia (FRDA). In collaboration with Novartis, we performed high-throughput RNAi screens and employed iPSC technology to generate neurons from FRDA-patient derived cells. Our innovative proof-of-concept study led to the identification of potential drug targets (Villasenor et al., 2015)(Villaseñor et al., unpublished results).

Project II: Endogenous in-vivo biotinylation system for mapping protein networks and protein-DNA interactions of RNAi factors in embryonic stem cells

In the second part of this thesis, I will explain how I combined more recent genome editing tools (TALENs and CRISPR/Cas9 system) with affinity purification approaches in mouse stem cells. I took advantage of this approach to study protein networks of RNAi factors keeping endogenous protein levels intact and to identify a potential protein variant of Argonaute-1 made by alternative splicing (Villaseñor et al., unpublished results).

Project III: Functional analysis of Ago1 splicing isoform

Argonaute (Ago) proteins form the functional core of the RNA-induced silencing complexes that mediate RNA silencing in eukaryotes. In mammals, four proteins of the Ago-clade (Ago1-4) are ubiquitously expressed and bind two classes of small RNAs (sRNAs); microRNAs (miRNAs) and small-interfering RNAs (siRNAs). In this project, I will describe the discovery and functional characterization of a novel splice variant of mouse *Ago1*.

Project IV: Novel insights in mammalian m6A RNA methylation

In the last project presented in this thesis, I will describe how I used the above-mentioned techniques to uncover novel binding partners of the m6A methyltransferase complex, which are now under further investigation (Knuckles, Villaseñor et al., unpublished results).

The results for each project will be presented in a separate chapter. Additionally, I will give a short introduction to the related biology.

2 Results

2.1 Project I: Application of genome editing for drug discovery

Results published in:

- Villaseñor R*, Miraglia L*, Romero A, Tu B, Punga T, Knuckles P, Duss S, Orth T, Bühler M. Genome-Engineering Tools to Establish Accurate Reporter Cell Lines That Enable Identification of Therapeutic Strategies to Treat Friedreich's Ataxia. ***J Biomol Screen.*** 2015 Jul;20(6):760-7. doi: 10.1177/1087057114568071. * equal contribution
- Villaseñor R, Manos P, Berenshteyn F, Littlefield C, Lacoste A, Bühler M. ***Unpublished results.***

2.1.1 Introduction

Clinical features and pathology of Friedreich's Ataxia

Frataxin is a highly conserved protein that is ubiquitously expressed and plays a crucial role in many aspects of mitochondrial iron metabolism, respiratory control, and resistance to oxidative stress. Deficiency of frataxin causes Friedreich's Ataxia (FRDA), an autosomal recessive degenerative disease that primarily affects the nervous system and the heart. FRDA is the most common cause of inherited recessive ataxias across Europe and several epidemiological studies have estimated the prevalence of FRDA as 2-3 cases per 100'000 people in Caucasian populations (Campuzano et al., 1996; Schulz et al., 2009). Unlike the small number of homozygous FRDA patients, the number of heterozygous mutation carriers has been estimated to be as high as 1:60 and they are clinically normal. Intriguingly, FRDA is only found in individuals of European, North African, Middle Eastern, or Indian origin (Labuda et al., 2000). Clinical hallmarks of FRDA are progressive gait and limb ataxia, loss of posture and vibratory sense, leg muscle weakness, vision loss as well as non-neurological signs such as thickening of heart muscle (i.e. cardiomyopathy) and diabetes (Schulz et al., 2009). The first symptoms usually appear around puberty, but age of onset can vary from infancy (2-3 years) to adulthood (after

25 years). Progressive loss of coordination and muscle weakness often result in the inability to walk 10-15 years after disease onset and ultimately makes the patient wheelchair-bound. Two-thirds of patients show cardiac symptoms, which further contributes to disability and premature death. Indeed, cardiomyopathy is the most frequent cause of death among FRDA patients (Durr et al., 1996).

The GAA-triplet repeat expansion

Friedreich's Ataxia belongs to the human trinucleotide repeat expansion diseases (TREDs). Trinucleotide repeat expansion underlies at least 17 neurological diseases, which can be generally categorized into two classes: the first, includes Huntington's disease and many spinocerebellar ataxias, and is characterized by exonic (CAG)n repeat expansions encoding polyQ tracts. The second class has its repeats in non-coding sequences and is typically characterized by large and variable repeat expansions, which result in multiple tissue dysfunctions. Friedreich's Ataxia belongs to the second class of TREDs.

Large tracts of trinucleotides can cause disease in several ways: by affecting gene expression, by producing toxic RNA species, or by altering the function of the resultant protein. The genetic basis of the frataxin protein (FXN) deficiency lies in the structure of the *FXN* gene, which is encoded on human chromosome 9 (Chamberlain et al., 1988). The majority of FRDA patients (95%) have a pathogenic hyperexpansion of a trinucleotide GAA repeat in the first intron of the FXN gene resulting in reduced protein levels (Pandolfo, 2009). Generally, healthy individuals have up to 38 GAA repeats, whereas FRDA patients have approximately 70 to more than 1000 GAA triplets (Figure 8A), most commonly 600 – 900 GAA triplets on both alleles of the *FXN* gene (Durr et al., 1996; Montermini et al., 1997). Similar to other TREDs, the copy number of the GAA repeats defines the disease. Thus, the larger the number of GAA repeats, the earlier the onset of the disease and the quicker the decline of the patient. As a functional consequence of GAA repeat hyperexpansion, FRDA

patients have a marked deficiency of *FXN* mRNA and thus FXN protein levels, thought to be due to reduced *FXN* gene expression (Campuzano et al., 1997; Campuzano et al., 1996) (Figure 8B).

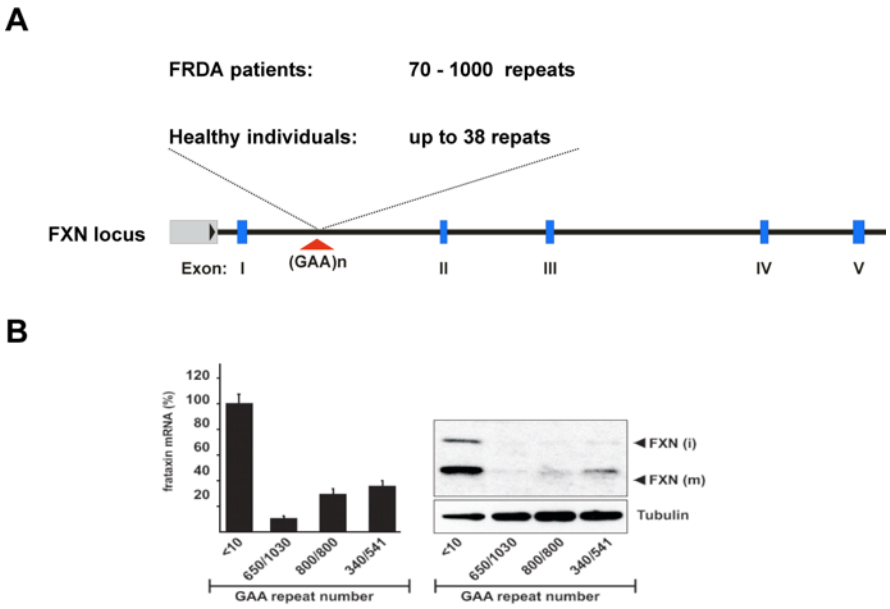


Figure 8. FXN Gene Structure And Consequences of GAA Triplet Repeat Expansion

(A) Schematic view of the human *FXN* gene (ENSG00000165060). Exons and untranslated regions are depicted as blue and gray boxes, respectively. The red arrowhead indicates localization of the GAA repeats. **(B)** FXN mRNA levels (left) and FXN protein accumulation (right) in FRDA patient cell lines. The numbers indicate established GAA repeat numbers in two alleles in three different cell lines. Healthy individual cells display less than 10 GAA repeats (<10). FXN (i) and FXN (m) designate the intermediate and mature forms of the FXN protein, respectively. Figures shown in (B) were kindly provided by Dr. Tanel Punga (Uppsala, Sweden).

FXN gene silencing mechanism

Two molecular models have been proposed to explain silencing of *FXN* gene expression at the level of transcription. At first, mainly based on *in vitro* experiments, it was shown that expanded GAA repeats adopt unusual, non-B form DNA structures (such as DNA triplexes or “sticky DNA”) that impede the progress of RNA-polymerase II and thus reduce transcription (Bidichandani et

al., 1998; Sakamoto et al., 1999). Secondly, it was suggested that expanded GAA repeats induce formation of heterochromatin, which would render the *FXN* gene transcriptionally inactive (Saveliev et al., 2003).

FXN gene silencing correlates with lysine hypoacetylation of histones H3 and H4 and methylation of histone H3 at lysine 9 (H3K9) residues in cells with expanded GAA repeats (Figure 9)(Herman et al., 2006). These particular histone modifications are considered to be hallmark features of heterochromatin. Of these modifications, the methylation of H3K9 (H3K9me) is the most studied and best understood heterochromatic mark in various organisms. It has been established that H3K9 methylation is essential for recruitment of HP1 (Heterochromatin Protein 1) family members, which in turn mediate chromatin condensation (Bühler and Moazed, 2007). A study in mice showed the involvement of the HP1 proteins in chromatin condensation around the expanded genomic GAA repeats. In addition, the same study demonstrated the repressive role of the expanded GAA repeats on a neighbouring reporter gene. Conclusively, it was proposed that the expanded GAA repeats silence nearby genes by position effect variegation (PEV) mechanism (Saveliev et al., 2003).

Consistent with previous reports, Punga and Bühler could demonstrate that H3K9 methylation accumulates in the proximity of the GAA repeats in patient-derived cell lines (Herman et al., 2006; Punga and Buhler, 2010). It was also confirmed that the *FXN* gene is silenced on the transcriptional level. However, the authors could show that expanded GAA repeat tracts affect transcription elongation rather than initiation. Most importantly, reduction of H3K9 methylation by a specific drug (BIX-01294) did not enhance *FXN* gene expression (Punga and Bühler, 2010). Therefore, H3K9 methylation plays, if at all, a redundant role in *FXN* silencing and it appears that the non-B form DNA structure formed by the GAA repeat constitutes a major obstacle for the transcription machinery (Figure 9). One interesting question that remains to be addressed in the future is how GAA repeats trigger H3K9 methylation and what prevents them from spreading into the *FXN* promoter region.

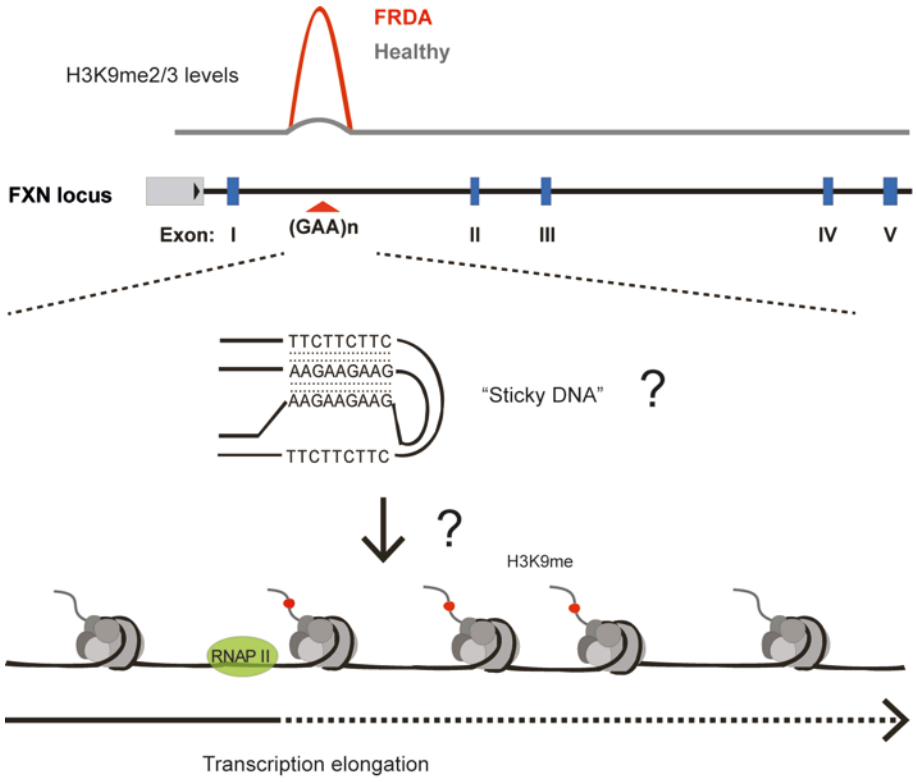


Figure 9. *FXN* Gene Silencing Mechanism

Long intronic GAA repeats can induce the methylation of histone H3 at lysine 9 (H3K9me) at the *FXN* locus in FRDA cells. This epigenetic modification is a hallmark of heterochromatin and it has therefore generally been assumed that the *FXN* locus is assembled into a chromatin structure that hinders transcription. However, our findings suggest that H3K9me is restricted to the GAA repeat tract and that initiation of transcription is not affected. The long GAA repeats - and probably their adopted chromatin status - seem to be the major obstacle to the transcription machinery impeding transcription elongation in FRDA cells (Punga and Bühler, 2010). How GAA repeats and the non-B DNA form adopted by the repeats trigger H3K9 methylation and what prevents it from spreading into the promoter region remains unknown.

2.1.2 Manuscript I / see Appendix

As an aberrant non-B DNA conformation adopted by the long GAA repeats appears to constitute the major obstacle for RNA-polymerase II (RNAPII) in transcribing the *FXN* gene (Figure 9), we believe that strategic chemical design and high-throughput screening to identify compounds with the potential to promote transcription elongation through expanded GAA repeat

tracts should become a major focus in the quest for novel therapeutic strategies to treat FRDA. In addition, the regulation of the native *FXN* gene remains poorly understood. The identification of regulators of *FXN* gene expression would benefit the search for potential therapeutic targets for FRDA patients. Before this study, no reliable and robust system existed for cell-based drug discovery, hampering high-throughput screens for genes or low molecular weight (LMW) enhancers of frataxin expression.

During the course of my master thesis and PhD studies, I generated reporter cell lines for simple detection of endogenous human *FXN* gene expression in its natural genomic context. To achieve this aim, I generated a reporter cell line of human origin that is compatible with high-throughput biology and enables accurate monitoring of endogenous *FXN* gene expression. I chose a zinc-finger nuclease (ZFN)-mediated genome-editing approach – in 2011 the only available genome editing tool – to tag the 3' end of the endogenous *FXN* gene with a firefly luciferase (FL) reporter gene (Lombardo et al., 2007; Urnov et al., 2005). In contrast to previous work (Lufino et al., 2013; Martelli et al., 2012a), the established HEK293T-FF2AP cell line allowed me to screen for modulators of *FXN* expression in its natural genomic context and in a high-throughput format. The FF2AP cell line is easy to transfect and hence well suited for RNA interference (RNAi) or cDNA overexpression screens.

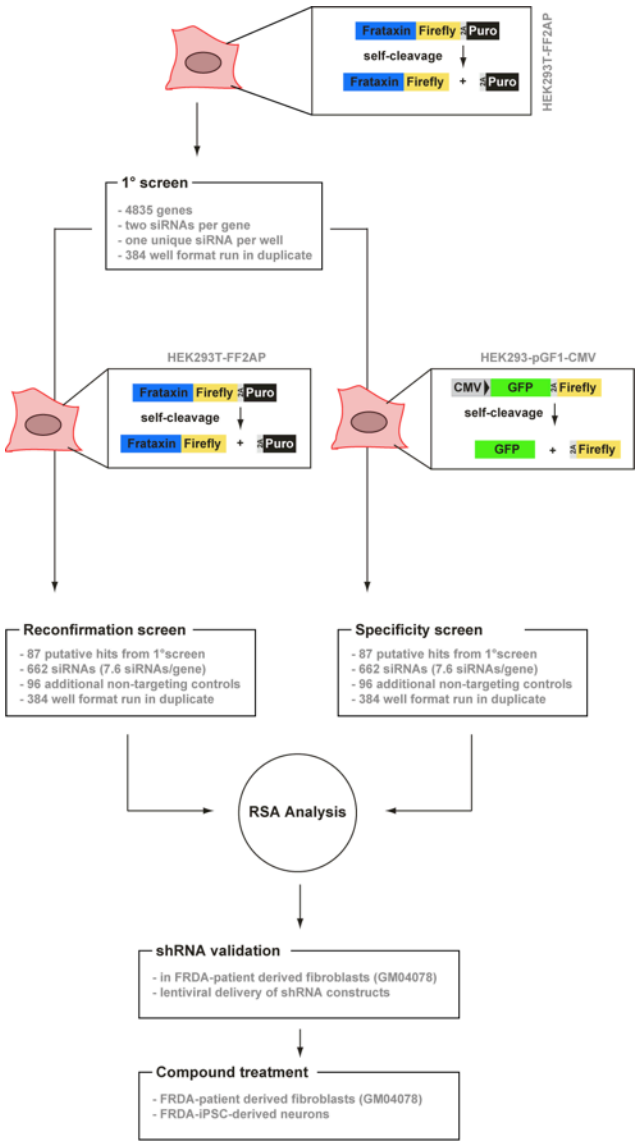


Figure 10. Workflow of siRNA Screen With The HEK293T-FF2AP cell line

For details see Materials & Methods of published manuscript found in the appendix.

To identify potential repressors of *FXN* gene expression or regulators of FXN protein stability, we collaborated with Loren Miraglia and Dr. Anthony Orth (GNF, San Diego) to run an RNAi screen targeting 4,835 human genes in the HEK293T-FF2AP cell line, using luciferase signal amplification as the readout (Figure 10). Using the HEK293T-FF2AP cell line, we performed a reconfirmation

screen of the top primary screen hits. In parallel, we screened positive-hit siRNAs for non-FXN-specific activators of luciferase activity with an HEK293 cell line stably expressing firefly luciferase from a CMV promoter (HEK293-pGF1-CMV). This allowed us to filter out false-positive hits and uncover novel regulators of *FXN* gene expression (Figure 10).

To determine if the primary hits could increase *FXN* gene expression also in cells containing expanded GAA repeats, I performed secondary hit validation experiments in FRDA fibroblasts using lentiviral shRNA knockdown followed by quantitative real-time PCR (qRT-PCR). I could identify two genes, *PRKD1* and *SBF1*, whose knockdown leads to *FXN* gene reactivation. The roughly 1.5-fold increase in *FXN* mRNA levels after *PRKD1* or *SBF1* knockdown is quite modest and hence FXN protein increase was not reliably detectable by Western blot. However, this proof-of-concept study demonstrates the enormous power of using genome editing to easily monitor changes of endogenous *FXN* expression, but also urges us to expand our screens to genome-wide siRNA libraries or compound libraries to identify genes or drugs that lead to more drastic changes in *FXN* gene expression.

Generation of patient-derived reporter cell line

I set out to introduce the FL gene into a *FXN* allele with an expanded GAA tract in intron #1 with a similar strategy as described in section 2.1.2 (Figure 11A). For this I used primary skin fibroblasts from a clinically affected individual (Coriell Cell Repositories; GM04078) homozygous for the GAA expansion (GAA_{541}/GAA_{420}). After immortalization of GM04078 cells with BMI1 and hTERT (hereafter referred as FRDA4078iBT), these cells were co-transfected with plasmids encoding the frataxin-ZFNs and Donor-2 (Figure 11A). Donor-2 was chosen because of its autonomous neomycin (Neo) selection cassette (Figure 11B). This strategy permits clonal selection in G418 containing medium independently of the expression status of the *FXN* gene. Therefore, we

selected single cell colonies after G418 selection and screened for correct and unique genomic integration events by Southern Blot analysis (Figure 11C). One positive clone (FRDA4078iBT-FF) was further characterized for GAA repeat stability and FL activity. Genomic PCR assays with primers that flank the GAA repeat tract (Figure 11D, bottom) revealed that repeat length in parental (GM04078), FRDA4078iBT, and FRDA4078iBT-FF cells remained unaffected (Figure 11D). Moreover, luciferase activity was observed in FRDA4078iBT-FF cells but not in the parental FRDA4078iBT cell line (Figure 11E). The expression of the endogenous FXN-FL fusion protein was absent in FRDA4078iBT-FF cells (data not shown). Notably, this FRDA patient-derived cell line showed a relatively low firefly luciferase activity, possibly reflecting the silenced state of the FXN locus (Figure 11E).

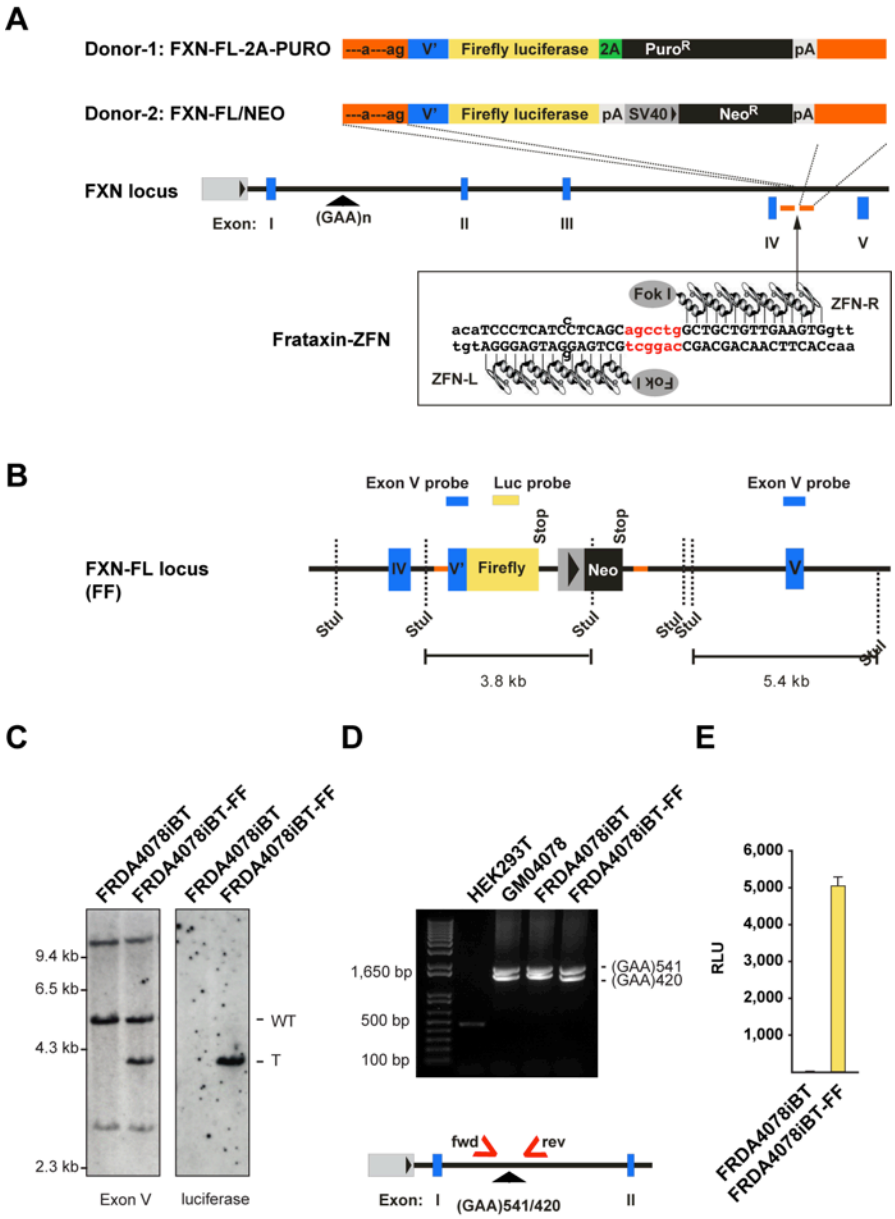


Figure 11. Generation of the FRDA4078iBT-FF Cell Line

(**A**) Scheme of the ZFN-mediated genome editing strategy to tag the 3´end of the protein-coding region of the endogenous FXN gene with a firefly luciferase (FL) reporter gene. Each frataxin-ZFN contains the cleavage domain of FokI linked to an array of five zinc fingers that have been designed to specifically recognize sequences (black uppercase letters) flanking the cleavage site (red letters) in intron 4 of the FXN gene. The double-strand break induced by the frataxin-ZFN pair permits site-specific integration of a transgenic donor DNA fragment (Donor-1 or -2) encoding the last FXN exon fused to the open reading frame of firefly luciferase (V'-firefly luciferase). Both Donor constructs are flanked by regions homologous to the insertion site (orange) to enable homology-directed repair. To ensure correct splicing of the FXN mRNA

after successful genome editing, the left homology arms were designed to introduce functional branchpoint and 3' splice sites (---a---ag). Donor-1 encodes puromycin N-acetyl-transferase (Puro) preceded by a 2A self-cleaving peptide linked to the V'-firefly luciferase fusion. Using Donor-1, puromycin resistance is only expected after correct integration into the FXN locus and transcription thereof. Donor-2 contains an autonomous neomycin (Neo) selection cassette driven by a SV40 promoter. This makes neomycin resistance independent of the transcriptional status of the FXN locus but might increase the rate of false-positive integration events. The location of the GAA triplet repeat is indicated by a black arrowhead. pA, cleavage and polyadenylation site (**B**) Blow up of the ZFN-targeted integration site after successful integration of the Donor-2 construct. StuI restriction sites and expected DNA fragment sizes after StuI digestion is depicted. Probes used for Southern Blot are indicated (FXN-exon V: blue boxes; Firefly luciferase: yellow box). (**C**) Southern Blot with StuI digested genomic DNA from the cell lines indicated was probed with Exon V probe, stripped and re-probed with luciferase probe. Single luciferase signal confirms unique integration event. Exon V and luciferase signals of the expected size (T) demonstrate correct integration of the Donor-2 construct at the FXN locus in FRDA4078iBT-FF cells. (**D**) Bottom: Clinically affected patient is homozygous for the GAA expansion in the FXN gene with alleles of approximately 541 and 420 repeats. PCR primers to assess GAA repeat length are depicted by red arrows. GAA repeat length in the cell lines indicated was assessed by genomic PCR using fwd and rev primers indicated in scheme. PCR product size = 457+3N base-pairs (N= number of GAA triplets). HEK293T cells serve as control (<10 GAA triplets). GM04078, patient-derived primary fibroblasts; FRDA4078iBT, GM04078 fibroblasts immortalized with BMI1 and hTERT; FRDA4078iBT-FF, FRDA4078iBT cells with the edited FXN locus (FXN-Firefly luciferase, FF). (**E**) Firefly luciferase activity assayed from non-targeted FRDA4078iBT and FRDA4078iBT-FF cells. RLU, relative light units normalized to total protein. Error bars represent standard deviation.

2.1.3 Use of PRKD1 chemical inhibitors in patient-derived lymphoblasts

The RNAi screen performed in section 2.1.1 demonstrates the potential of our high throughput screening (HTS) system to reveal prospective drug targets for the development of FRDA treatments. Similar to 2-aminobenzamides that inhibit histone deacetylases (HDACs) (Chou et al., 2008; Herman et al., 2006; Rai et al., 2008) compounds with an antagonistic effect on the newly identified *FXN* repressors might be used to alleviate *FXN* expression in FRDA cells. To identify low molecular weight inhibitors of PRKD1, we searched the Novartis compound archive. A previous HTS campaign from the Novartis archive identified 2,6-naphthyridine as a dual PKC/PRKD inhibitor, which was further developed into a potent prototype pan-PRKD inhibitor (Figure 12A, compound

13c “Cpd13c”) (Meredith et al., 2010a). Compound 13c has 1000-fold selectivity versus PKC isoforms (Meredith et al., 2010b) and can be administered orally, which makes it a good candidate for drug development. As observed with shRNA-mediated suppression (Villasenor et al., 2015), chemical inhibition of PRKD1 with compound 13c in FRDA4078iBT-FF patient-derived fibroblasts provoked up to 5-fold elevation of luciferase activity (Figure 12B), suggesting that PRKD1 is a potent negative regulator of endogenous *FXN* expression.

Next, I determined the effect of compound 13c on *FXN* mRNA and protein expression in FRDA4078iBT-FF cells and in patient-derived lymphoblasts. To this end, I treated FRDA4078iBT-FF cells with compound 13c for 24 hours and measured levels of mature FXN protein and *FXN* mRNA by western blot and qRT-PCR, respectively. *FXN* mRNA, when normalized to *RPL13A* expression, and mature FXN protein levels remained constant in treated versus non-treated FRDA4078iBT-FF cells (Figures 12C & 12D). Moreover, I tested the ability of compound 13c in restoring *FXN* expression in another patient-derived cell type with longer GAA repeat expansions. I decided to work with a FRDA patient-derived lymphoblastoid cell line (Coriell Cell Repositories), which has between 650 and 1030 GAA triplets on both *FXN* alleles (“FRDA”, GM15850). As a control, a lymphoblastoid cell line from a healthy sibling (“Healthy”, GM15851) of one of the FRDA patients (GM15850) was used. I treated both cell lines with compound 13c for 24 hours and measured *FXN* mRNA levels by qRT-PCR. Similarly to the effects observed in FRDA4078iBT-FF cells, *FXN* mRNA expression remained constant in treated versus non-treated healthy and FRDA cells (Figure 12E).

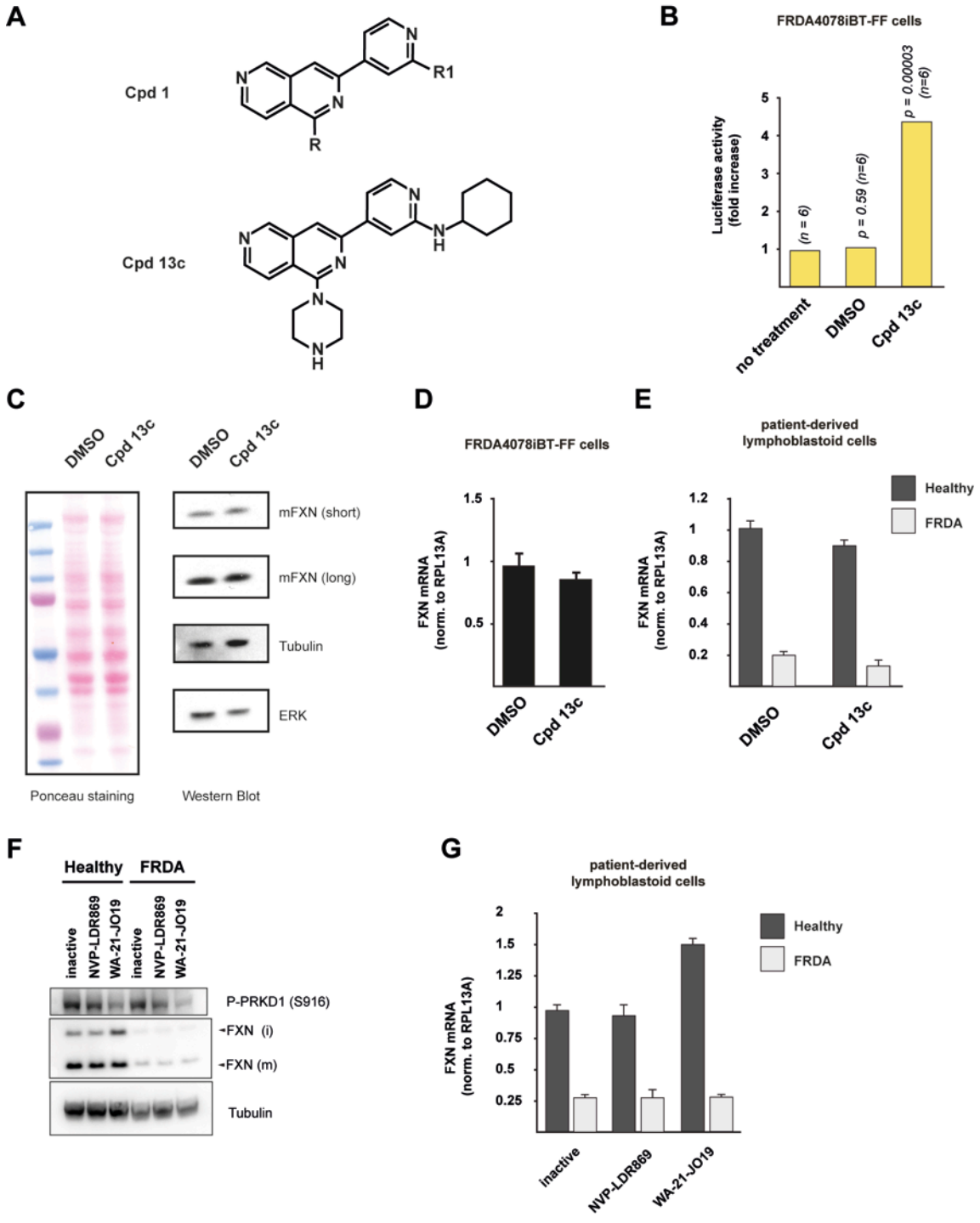


Figure 12. Chemical Inhibition of PRKD1 In Patient-derived Cells

(**A**) The molecular structures of dual PKC/PRKD inhibitor and compound 13c (Cpd 13c) are depicted. (**B**) Firefly luciferase activity assayed from drug-treated FRDA4078iBT-FF cells with DMSO (control) or compound 13c (Cpd13c). RLU, relative light units normalized to total protein. P values were determined using the Student's t test in 6 biological replicates. (**C**) Right: Western Blot showing unchanged expression

of the mature FXN protein (~15 kDa) in cells treated with DMSO (control) and compound 13c. Blot was probed with an antibody recognizing FXN. Antibodies against Tubulin or ERK were used to control sample loading. Left: Ponceau staining used as loading control. Effect of compound 13c on endogenous *FXN* mRNA levels in FRDA patient derived fibroblasts (**D**) and lymphoblasts (**E**). FRDA patient-derived lymphoblastoid cell line has between 650/1030 GAA triplets on both FXN alleles ("FRDA"). A lymphoblastoid cell line from a healthy sibling ("Healthy") with no expanded GAA repeats was used as control. Cells were treated with 1 μM of PRKD1 inhibitor (Cpd13c) for 24h in experiments shown in B-E. (**F-G**) FRDA lymphoblasts (10^7 cells) were treated with 5 μM PRKD1 inhibitor (NVP-LDR869 or WA-21-JO19) for 48 hours. Cells treated with 5 μM of inactive compound NVP-LEG653 (inactive) serve as control. (**F**) Western blot showing FXN protein levels and PRKD1 auto-phosphorylation of drug-treated cells. (**G**) FXN mRNA levels were determined by real-time PCR, normalized to *RPL13A* mRNA levels and shown relative to the inactive compound. Data represent the average plus standard deviation of 3 biological replicates.

Rational design and modification of pan-PRKD inhibitor compound 13c led to a more potent and specific PRKD1 inhibitor (compound WA-21-JO19; L. Monovich, personal communication). In addition, high throughput screening against recombinant PRKD1 identified 3,5-diarylpyrazole as a novel kinase inhibitor scaffold with moderate activity. Rational design and modification of the 3,5-diarylpyrazole scaffold led to a more potent selective, and orally bioavailable PRKD1 inhibitor (compound NVP-LDR869)(Gamber et al., 2011). Next, I treated healthy and FRDA lymphoblasts with PRKD1 inhibitor WA-21-JO19 or NVP-LDR869 to test their effect on *FXN* gene expression. Cells treated with an inactive compound NVP-LEG653, which is closely related chemically to WA-21-JO19 but lacks activity on PRKD1, served as control. Only WA-21-JO19 reduced PRKD1 auto-phosphorylation at Serine 916 (S916) in healthy and FRDA lymphoblasts, showing the efficacy of WA-21-JO19 in blocking PRKD1 activity (Figure 12F). Nevertheless, mature FXN protein levels remained constant in FRDA lymphoblasts as measured by Western blot (Figure 12F). Moreover, *FXN* mRNA expression remained unchanged in WA-21-JO19 treated FRDA patient-derived lymphoblasts (Figure 12G). Intriguingly, in healthy cells, WA-21-JO19 increased the intermediate form of FXN protein and mRNA levels by roughly 50 percent (Figure 12G).

In summary, our hypothesis that chemical inhibition of PRKD1 restores *FXN* gene expression in FRDA cells lacks convincing evidence. However, the effect of PRKD1 inhibition, particularly using WA-21-JO19, could be significant in more disease-relevant cell types such as neurons or cardiomyocytes. To test this idea we set out to generate neurons from patient-derived iPSCs.

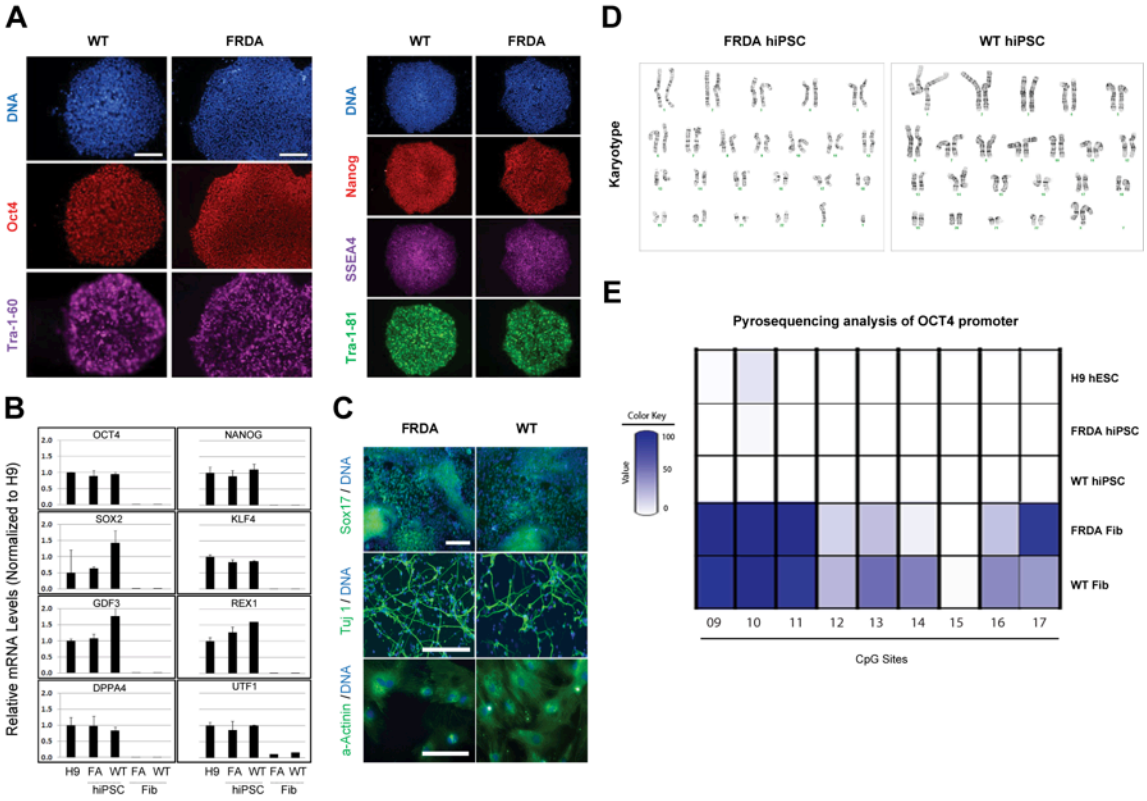


Figure 13. Characterization of FRDA- and WT-SeV Derived HiPSCs

(**A**) Immunostaining showing expression of pluripotency markers in hiPSCs derived from unaffected- (WT) and FRDA-patient (FRDA) fibroblasts. Scale bar represents 200 µm. (**B**) Quantitative real-time PCR analysis for expression of pluripotency genes in FRDA (FA) and WT- hiPSCs and parental fibroblasts. Samples were normalized to internal housekeeping genes GAPDH and presented relative to H9 hESC. Analysis was performed in biological and technical triplicates; error bars represent standard error. (**C**) Immunostaining showing expression of lineage-specific markers Sox17 (definitive endoderm), β-III Tubulin (Tuj1, neuronal, ectodermal) and alpha-actinin (mesoderm) following hiPSC differentiation. (**D**) FRDA- and WT hiPSCs have a normal karyotype and proper parental fingerprint. (**E**) Heat-map representation of pyrosequencing analysis of the OCT4 promoter in FRDA- and WT-hiPSCs compared with their respective parental fibroblasts. H9 hESC was included as a positive control.

2.1.4 Use of iPSCs to model FRDA

We completed our drug discovery platform described above by designing disease-relevant cellular models, which enable the validation of primary hits and discovery of chemical modulators of these targets. In collaboration with Dr. Arnaud Lacoste´s group at NIBR (Cambridge, USA), we first reprogrammed dermal fibroblasts from healthy, unaffected individuals (wildtype WT, GM08399) and from patients affected by FRDA (FRDA, GM04078) into induced pluripotent stem cells (iPSCs). Sendai virus particles encoding OCT4, SOX2, KLF4, and cMYC (OSKM) were used to minimize the risk that random integration of reprogramming transgenes would affect FXN pathways. We obtained iPSCs from both WT and FRDA patients and all iPSC lines expressed similar pluripotency cell markers (Figure 13A and 13B). The iPSC lines could also differentiate in representatives of all three germ layers, including neurons and cardiomyocytes (Figures 13C), two cell types that are affected in FRDA patients. In addition, WT and FRDA iPSCs contained a normal karyotype and proper parental fingerprint compared with parental fibroblasts (Figure 13D and 13E).

Both WT and FRDA iPSCs also formed teratomas when injected into mice (Figure 14). Importantly, the iPSC lines obtained from WT and FRDA patients exhibited a similar ability to differentiate into Pax6-expressing neural progenitors and β-III Tubulin-expressing neurons (Figure 15A and 15B). These results confirm previous reports that genetic alterations causing FRDA do not affect reprogramming or differentiation processes (Hick et al., 2013; Ku et al., 2010). Finally, FRDA neuronal cells expressed lower levels of *FXN* mRNA and protein than WT neurons, (Figure 15C and 15D), indicating that the FRDA molecular phenotype is conserved during reprogramming.

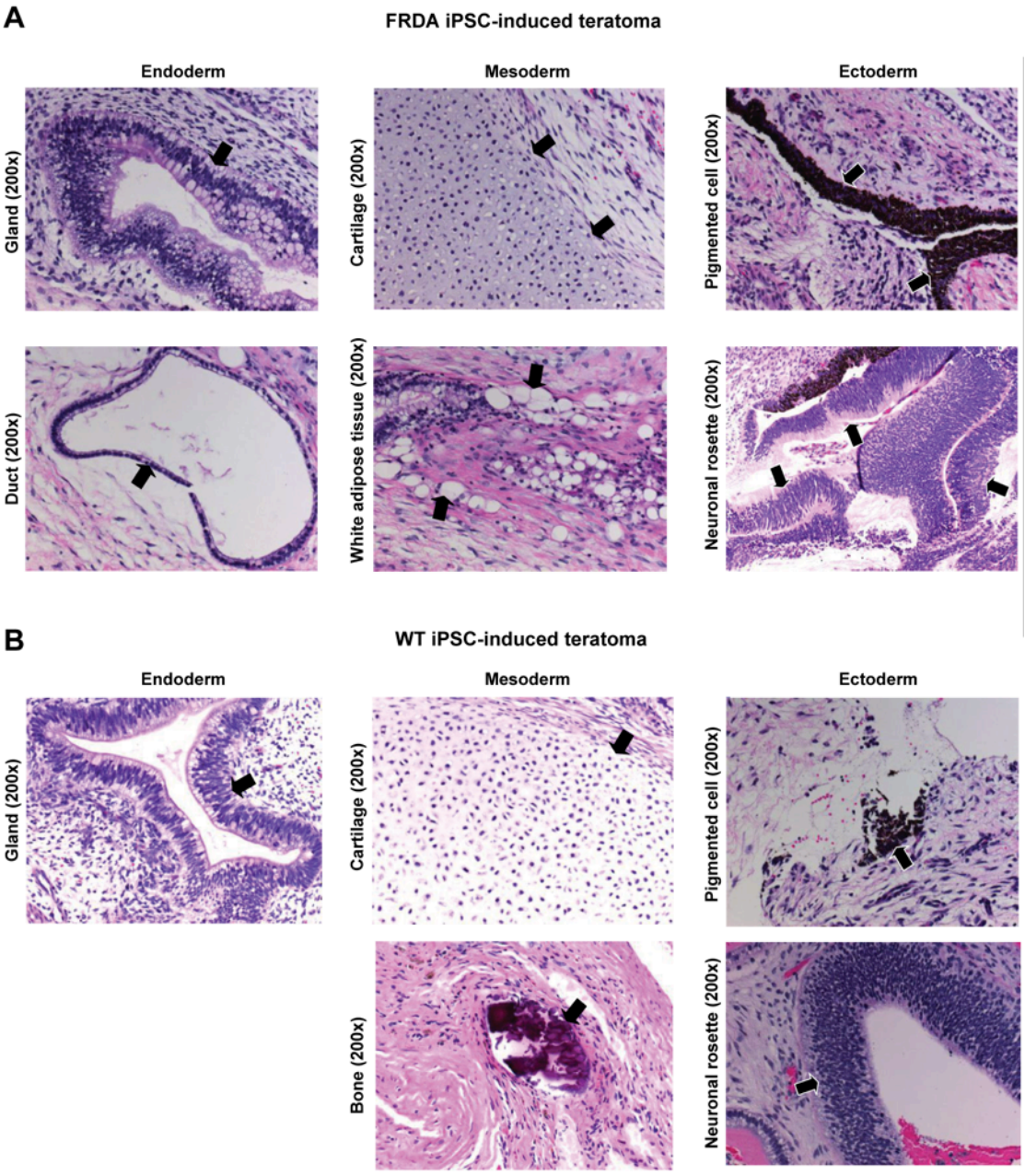


Figure 14. Teratoma Formation Assay

Representatives of all three germ layers can be identified in teratomas derived from both (**A**) FRDA-hiPSC and (**B**) WT-hiPSC.

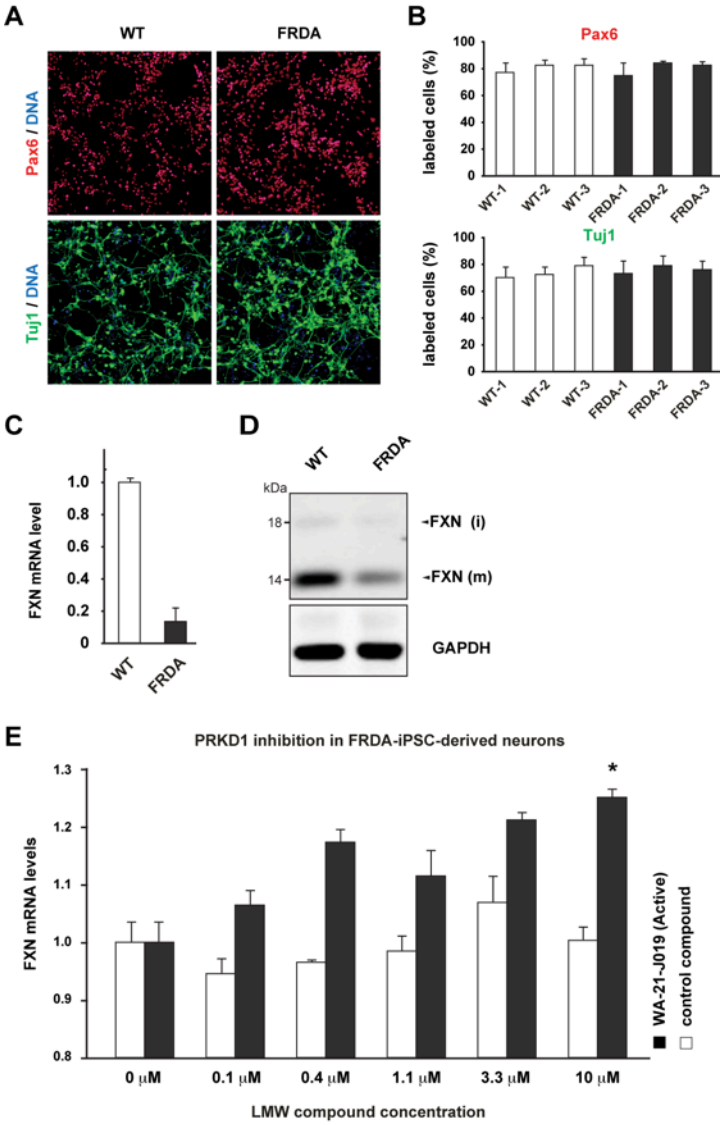


Figure 15. Studies In FRDA-iPSC-derived Neurons Validate PRKD1 as a Druggable Target And Identify a Chemical Modulator of FXN Expression

(**A**) Immunostaining showing expression of neuronal markers. Both WT and FRDA iPSCs can be differentiated into neural progenitors expressing Pax6 and neurons expressing β-III Tubulin (Tuj1). (**B**) Triplicate cell counts of three independent WT (WT1-3) and FRDA (FRDA1-3) cultures show that WT and FRDA iPSCs can be differentiated into Pax6-positive neural progenitors or β-III Tubulin (Tuj1)-positive neurons with similar efficiencies. (**C-D**) The FRDA molecular phenotype is conserved during reprogramming. (**C**) Quantitative real-time PCR analysis for frataxin (*FXN*) expression in WT versus FRDA neuronal cells. Samples were normalized to the internal housekeeping gene GAPDH. Analysis was performed in biological and technical triplicates; error bars represent standard error. (**D**) Western blot showing protein levels of frataxin (FXN) and the housekeeping protein GAPDH in WT versus FRDA neuronal cells. (**E**) Exposure of FRDA-iPSC-derived neurons to WA-21-JO19, a chemical inhibitor of

PRKD1, leads to an increase in FXN expression while a control compound (NVP-LEG253) has no effect. Neuron cultures were treated for 3 weeks with five different concentrations of PRKD1 inhibitor WA-21-JO19, or an inactive analog of WA-21-JO19. The experiment was run in biological duplicates and no compound toxicity was observed at the concentrations used. FXN mRNA levels were determined by real-time RT-PCR, normalized to RPL13A mRNA levels. *P = 0.03.

Subsequently, we assessed the effect of low molecular weight inhibitors of PRKD1 on *FXN* expression in WT and FRDA neurons. We treated FRDA-iPSC-derived neurons with either compound WA-21-JO19 or a control compound. As observed with shRNA-mediated suppression in FRDA fibroblasts, chemical inhibition of PRKD1 by WA-21-JO19 resulted in a significant increase in endogenous *FXN* gene expression in FRDA-iPSC-derived neurons (Figure 15E). WA-21-JO19 treatment led to a 25% increase in *FXN* mRNA expression in FRDA neurons after 3 weeks of treatment. Importantly, an inhibitor against an unrelated protein kinase (control compound) had no significant effect on *FXN* expression in FRDA patient neurons, confirming that WA-21-JO19 activity is due to an inhibitory effect on PRKD1 (Figure 15E).

In summary, these results validate PRKD1 as a modest negative regulator of endogenous FXN expression. They also demonstrate that PRKD1 is a druggable target in patient- and disease-relevant cell types such as neurons.

2.2 Project II: Endogenous in-vivo biotinylation system for mapping protein networks and protein-DNA interactions of RNAi factors in embryonic stem cell

Unpublished results:

- Villaseñor R, Knuckles P, Daniel H, Burger L, Seebacher J, Bühler M.

The RNA interference (RNAi) pathway consists of evolutionary conserved enzymes indispensable for gene regulation mediated via small non-coding RNAs (Figure 16A). MicroRNAs (miRNAs) are a class of small non-coding RNAs that control gene expression through regulation of messenger RNA (mRNA) stability and translational repression (Figure 16B). MiRNAs are genomically encoded in miRNA genes or introns of protein coding genes and are transcribed by RNA-polymerase II to primary miRNA (pri-miRNA) transcripts. A typical pri-miRNA transcript is capped, polyadenylated and is composed of a ~33-nt long stem loop with two flanking segments (Lee et al., 2002). Central to the production of miRNAs are members of the ribonuclease III family of double stranded RNA (dsRNA) processing enzymes Drosha and Dicer. Drosha carries out the first processing step in the nucleus in complex with other cofactors (microprocessor complex) for efficient and accurate processing of pri-miRNA (Gregory et al., 2004; Lee et al., 2003). In mammals, the main cofactor of the microprocessor complex is the DiGeorge syndrome critical region 8 (Dgcr8) protein (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Landthaler et al., 2004). The first processing step of pri-miRNAs results in the formation of a ~70-nt long precursor or pre-miRNA. Pre-miRNAs are then exported to the cytoplasm to undergo a second step of processing (Figure 16A).

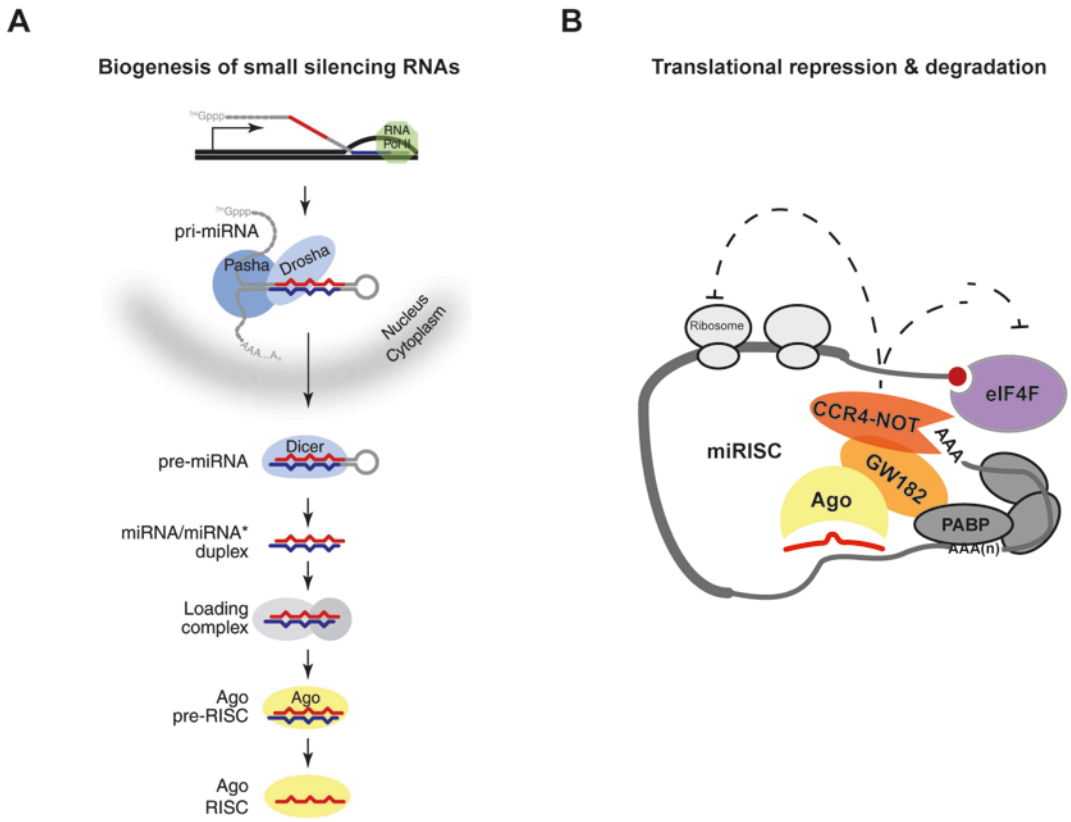


Figure 16. The Mammalian RNA Interference Pathway

(**A**) In animals, pri-miRNAs are transcribed by RNA-polymerase II (or, rarely, by RNA-polymerase III), processed in the nucleus into pre-miRNAs by the RNase III enzyme, Drosha, and then exported to the cytoplasm where the pre-miRNAs are converted to miRNA/miRNA* duplexes by a second RNase III enzyme, Dicer. miRNAs are then loaded into Argonaute effector protein complexes to silence target mRNAs. Image adapted with permission from (Matranga and Zamore, 2007) (**B**) Translational repression: the miRISC inhibits translation initiation by interfering with eIF4F-cap recognition and 40S small ribosomal subunit recruitment or by antagonizing 60S subunit joining and preventing 80S ribosomal complex formation. The miRISC might inhibit translation at post-initiation steps by inhibiting ribosome elongation. mRNA decay: the miRISC interacts with the CCR4–NOT and PAN2–PAN3 deadenylase complexes to facilitate deadenylation of the poly(A) tail (indicated by AAA(n)).

Dicer, a cytoplasmic RNAse III protein of ~216 kDa in size, mediates cytoplasmic processing of pre-miRNAs (Bernstein et al., 2001). This processing step produces a ~22-nt long dsRNA duplex. Human DICER interacts with TARBP2 (TAR RNA-binding protein 2) (Chendrimada et al., 2005; Haase et al.,

2005) and PRKRA (Protein kinase, interferon-inducible double-stranded RNA-dependent activator) (Lee et al., 2006). Both proteins are closely related proteins that seemingly facilitate cleavage of pre-miRNAs (Kok et al., 2007) and are believed to be involved in the transfer of the mature miRNA to the downstream RNA Induced Silencing Complex (RISC) (Tomari et al., 2004a; Tomari et al., 2004b). Once processed, both strands of the dsRNA intermediate are separated and one strand, the guide strand, is incorporated into effector protein complexes to recognize target mRNAs (Martinez et al., 2002) (Figure 16A and 16B).

A key member of RISC, the main RNAi effector protein complex in mammals, is Argonaute (Ago). Argonaute proteins are enzymes with the ability to bind small non-coding RNAs and guide RISC to target mRNAs (Ketting, 2011; Meister, 2013). Agos belong to an evolutionary conserved protein family indispensable for gene regulation mediated via small non-coding RNAs (Swarts et al., 2014). Structural studies of Ago proteins revealed a bi-lobed architecture composed of four globular domains (N, PAZ, MID and PIWI) connected through two structured linker domains (L1 and L2). In mammals, four Ago proteins (Ago1-4) are ubiquitously expressed and bind primarily two classes of small RNAs (sRNAs), miRNAs and siRNAs. sRNA binding occurs through the conserved MID and PAZ domains. The PAZ domain binds the 3´end of the small RNA. Whereas the MID domain anchors the 5´ end of the sRNA by providing a binding pocket in which the 5´-terminal base engages in stacking interactions with a conserved tyrosine (Tyr529). In addition, several hydrogen bonds coordinate correct 5´-end binding (Elkayam et al., 2012; Schirle and MacRae, 2012). The PIWI domain is responsible for the endonucleolytic activity and structurally resembles RNase H enzymes (Parker et al., 2004; Song et al., 2004; Yuan et al., 2005). In mammals, only Ago2 has endonucleolytic activity and is therefore the only Ago protein capable of cleaving mRNA targets (Meister et al., 2004). The remaining Ago homologs (Ago1, Ago3, and Ago4) facilitate translational repression through Ago interacting proteins (Meister, 2013).

The incorporated guide strand in RISC directs the sequence-specific cleavage of complementary target mRNAs, a process termed RNA interference or RNAi (Fire et al., 1998; Hammond et al., 2000; Martinez et al., 2002). Cleavage only occurs when full complementarity between guide RNA and RNA target is achieved. In addition, cleavage-independent silencing mechanisms exist to inhibit mRNA function. This is especially important when the involved Ago protein has no endonucleolytic activity (as shown for Ago 1, 3, and 4), or when miRNAs only have restricted complementarity with their targets.

In fact, most mammalian miRNAs do not guide cleavage of target mRNAs by RNAi-like mechanisms. In contrast, mammalian miRNAs predominantly repress gene expression on the level of translation or affect mRNA stability by guiding cellular decay processes, including mRNA deadenylation and decapping (Figure 16B). Argonautes interact with members of the GW182 protein family (GW proteins) to mediate post-transcriptional gene silencing via miRNAs (Eulalio et al., 2008b; Meister et al., 2005). Members of the GW182 protein family are characterized by the presence of multiple Gly-Trp (GW) repeats. Three paralogues exists in vertebrates, Tnrc6A/GW182 (Trinucleotide repeat-containing gene 6A protein), Tnrc6B and Tnrc6C. GW proteins coordinate most downstream steps in gene silencing through recruitment of additional protein effector complexes to target mRNAs (Meister, 2013). Recent studies have revealed that GW proteins induce mRNA deadenylation by recruiting two major cellular deadenylase complexes, PAN2-PAN3 and CCR4-NOT (Braun et al., 2011; Chekulaeva et al., 2009; Chekulaeva et al., 2011; Fabian et al., 2009; Zipprich et al., 2009).

In mammals, core components of the RNAi pathway have extensively been studied over the last 15 years. Biochemical studies provided insights into protein networks of most members of the pathway (Chendrimada et al., 2005; Gregory et al., 2004; Haase et al., 2005; Hock et al., 2007; Landthaler et al., 2008; Meister et al., 2005). Most studies made significant progress in mapping the mammalian RNAi protein network by combining protein affinity purification

with mass spectrometry (MS). However, these studies were performed in cancer cell lines in most of the cases using over-expression of the protein of interest. Ectopic expression of a protein beyond endogenous levels may result in artifactual protein complexes and increase nonspecific/non-physiological protein – protein interactions. In addition, traditional affinity purification methods are limited by use of antibodies against the protein of interest. The lack of specific antibodies against many proteins represents a significant caveat in current protocols. Therefore, we combined genome editing with *in vivo* biotinylation to dissect protein–protein and protein–DNA interactions of protein complexes in mouse embryonic stem (mES) cells keeping endogenous protein levels intact (Figure 17). Finally, I validated our MS-approach to map protein interaction networks by studying well-characterized factors of the RNAi pathway.

2.2.1 Endogenous tagging of genes encoding RNAi factors

We recently established an efficient protocol for genome editing in mES cells using TALENs (Flemr and Buhler, 2015). We used BirA-expressing cells (RosaB) as a parental cell line for endogenous gene tagging of several RNAi factors with the FLAG-AviTag sequence, which is a well-established tagging system for tandem affinity purification of protein complexes in mES cells (Kim et al., 2009). The FLAG-AviTag (FLAvi) tag contains the Flag epitope and a short peptide tag “AviTag” that serves as a substrate for *in-vivo* biotinylation mediated by the *Escherichia coli* biotin ligase BirA. Our genome editing approach, using TALENs and the CRISPR/Cas9 system, allows insertion of the FLAvi tag at the 5’- or 3’-end of the target gene in BirA expressing mES cells (Figure 17). For this study, my colleague Dr. Philip Knuckles and I set out to endogenously tag the following RNAi factors, individually, in RosaB cells: Dicer, Dgcr8 and Ago1.

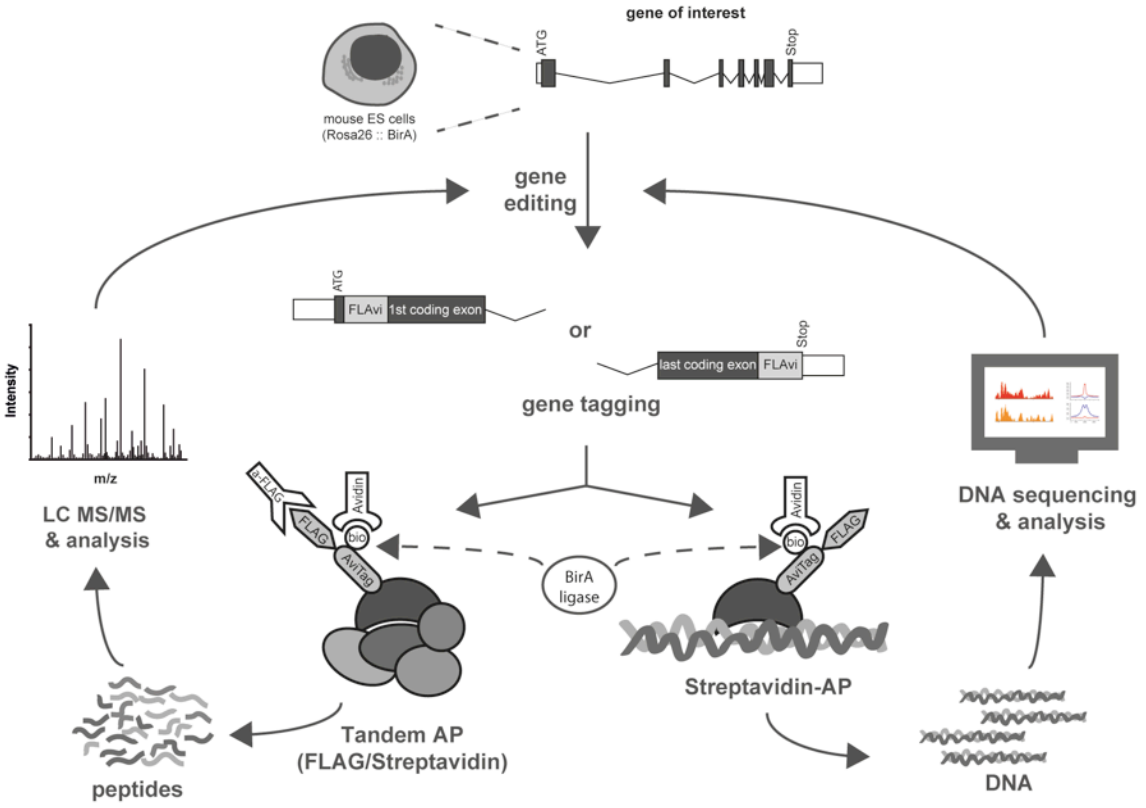


Figure 17. Combination of Genome Editing With In Vivo Biotinylation To Dissect Protein-Protein And Protein-DNA Interactions of Protein Complexes In Mouse Embryonic Stem Cells

To tag the endogenous Dicer protein, I designed a pair of TALENs (*Dicer1*-TALENs) to cut at the first coding exon of the mouse *Dicer1* gene in RosaB cells (Figure 18A & 18B). Both plasmids encoding for the *Dicer1*-TALENs were co-transfected with a recombination reporter (pRR-EGFP), and a single-stranded oligodeoxynucleotides (ssODN) carrying the FLAvi-tag sequence. The homology regions against the first coding exon of *Dicer1* gene in the ssODN serve as template for the HDR machinery to repair the cut mediated by TALENs (Figure 18B). Therefore, homology directed repair allows insertion of the FLAG-AviTag into the TALEN target site. In addition, I used the pRR-EGFP recombination reporter, which is designed to produce an EGFP signal upon NHEJ repair of the *Dicer1*-TALENs cut, to enrich for cells with a TALEN-modified genomic target site (see Methods). Using this approach, I obtained 9 of

40 positive clones (22.5%) by PCR genotyping (Figure 18C & 18E). Furthermore, I confirmed the expression of endogenous FLAvi-Dicer1 protein in three clones by Western blot (Figure 18D). Importantly, unique peptides matching to mouse Dicer1 protein were detected after streptavidin affinity purifications followed by MS analysis in two clonal cell lines (Figure 18F & 18G), demonstrating that our approach generates endogenously tagged FLAvi-Dicer1 expressing cells.

We applied a different strategy to tag endogenous mouse *Dgcr8* and *Ago1* genes. To tag the endogenous *Dgcr8* or *Ago1* locus, we designed a guide RNA (gRNA) to target the first coding exon of the mouse *Ago1* gene (Figure 19A) or the mouse *Dgcr8* gene (Figure 19B), respectively. The gRNA was subcloned into the SpCas9-2A-mCherry plasmid (Cas9-2C, see Methods section). The resulting vector encoding for the gRNA and Cas9 enzyme (gR-Cas9-2C) was co-transfected with an ssODN carrying the FLAG-AviTag sequence and homology regions against the first coding exon of the target gene. For the *Dgcr8* gene tagging, 3 of 24 positive clones (12.5%) were identified by PCR (Figure 19C). We confirmed the expression of FLAG-AviTag-Dgcr8 protein in one of three tested clones by Western blot (Figure 19C). For mouse *Ago1* gene tagging, we obtained 25 of 28 positive clones (89.3%) by PCR genotyping (Figure 19C). Western blot analysis confirmed FLAG-AviTag-Ago1 protein expression in two clones (Figure 19C). Furthermore, FLAG immunoprecipitations followed by MS analysis confirmed the expression of endogenously tagged Ago1 protein (Figure 19D).

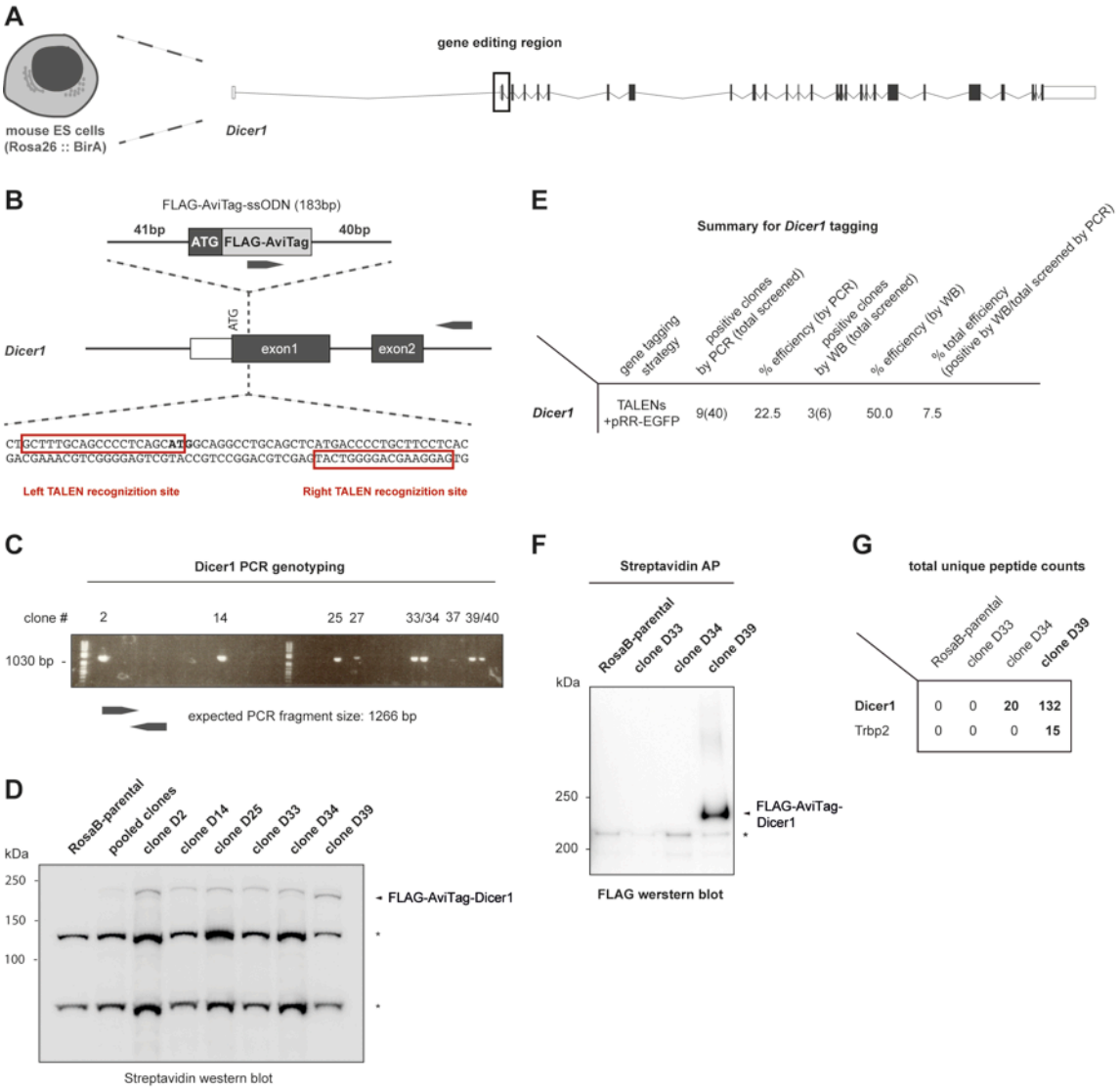


Figure 18. Endogenous Dicer1 Tagging With FLAG-AviTag In RosaB Cells

(**A**) Illustration of mouse *Dicer1* locus and TALENs target site (box). (**B**) Schematic overview of TALEN-assisted integration of an ssODN carrying a FLAG-AviTag sequence downstream of the Dicer1 start codon (bold letters). Gray arrows indicate positions of genotyping primers. (**C**) PCR genotyping of FLAG-AviTag insertions in pRR-EGFP selected clones with a forward primer binding in the AviTag sequence and a primer downstream the integration site, where the expected PCR product size is 1266-bp for the allele containing the FLAG-AviTag sequence. (**D**) Western blot analysis of RosaB (FLAG-AviTag-Dicer1) clones with anti-streptavidin conjugate. (**E**) Summary statistics for *Dicer1* gene tagging. (**F-G**) Detection of FLAG-AviTag-Dicer1 protein in gene edited clones with streptavidin-AP followed by anti-FLAG Western blot (**F**) and mass spectrometry analysis (**G**).

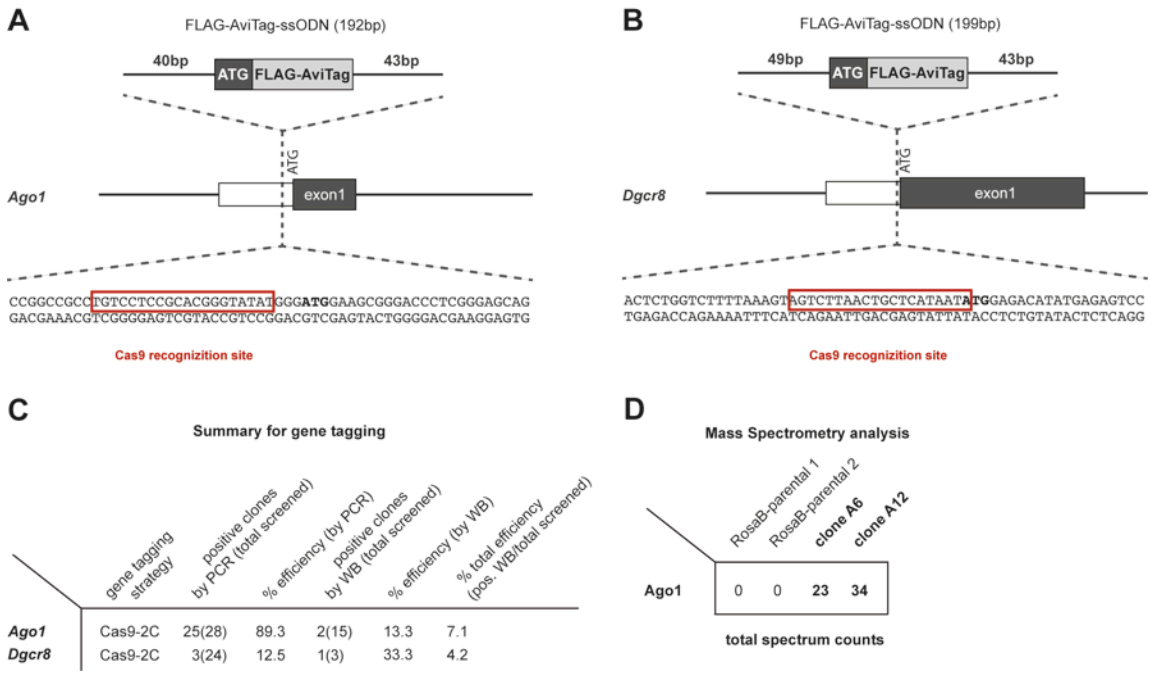


Figure 19. Endogenous Ago1 and Dgcr8 Tagging With FLAG-AviTag In RosaB Cells

(**A-B**) Schematic overview of Cas9-assisted integration of an ssODN carrying a FLAG-AviTag sequence upstream of the Ago1 and Dgcr8 start codon (bold letters), respectively. Gray arrows indicate positions of genotyping primers. (**C**) Summary statistics for *Ago1 and Dgcr8* gene tagging. (**D**) Detection of FLAG-AviTag-Ago1 protein in gene edited clones with streptavidin-AP followed by mass spectrometry analysis.

Taken together, I established and applied a genome editing approach to tag endogenous *Dicer1* gene in mouse ES cells. Moreover, endogenous Dicer1 protein is efficiently enriched after streptavidin affinity purifications allowing the faithful identification of Dicer1 protein binding partners. Finally, the results obtained for *Ago1* and *Dgcr8* gene editing establish the Cas9-system as an efficient alternative method to tag endogenous genes of interest where TALENs are not available.

2.2.2 Mapping protein-protein interactions of RNAi factors in mES cells

To test the robustness of affinity purification protocols, I performed single and tandem affinity purification of Dicer and Ago1 protein complexes for MS analysis. One-step streptavidin purification of Dicer protein complexes was efficient for enrichment and detection of Dicer and its known high-confidence interaction partner Trbp2 (Figure 18F & 18G). However, this approach also yielded high signal for endogenous biotinylated mitochondrial protein contaminants that interfere with the detection of lowly abundant interaction partners (data not shown). High quantities of contaminants were also observed in one-step FLAG immunoaffinity purifications of Dicer1 and Ago1 (data not shown). These observations represented a significant disadvantage of one-step streptavidin and FLAG immunoaffinity purifications. This prompted me to explore tandem affinity purification of protein complexes for accurate detection of protein – protein interactions (Figure 20A). This approach minimized background signals of contaminants in RosaB control cells and lacked detection of endogenously biotinylated proteins (Figure 20B).

After optimization of the tandem affinity purification (TAP) protocol, I conducted TAP-MS experiments using the FLAvi-tagged Dicer, Dgcr8 and Ago1 cells plus control cell line RosaB. To account for RNA-independent and direct physical interactions of Ago1, RNase A treated samples were included to MS analysis. Dr. Lukas Burger and I statistically analysed the MS data using the SAINT algorithm (Choi et al., 2011). SAINT determines a probability of interaction for each possible bait-prey pair. Figures 20C-E display, for each bait protein separately, the average fold-change between bait and control samples versus the SAINT probability of interaction (poi). Proteins with a probability of interaction score (poi-score) of at least 0.8 were defined as “potential interactors”. Previously reported interactors are marked in red.

MS analysis of Dgcr8 protein complexes confirmed the high-confidence interaction with the second component of the microprocessor complex, Drosha

(Figure 20C). Moreover, SAINT analysis further confirmed a specific but weak interaction to Dhx9 and Srpk1 (Figure 20C). Dhx9 (DEAH box protein 9), also known as ATP-dependent RNA helicase A (RHA), is an RNA helicase implicated in a number of cellular processes involving alteration of RNA secondary structure (Fuller-Pace, 2006). Human DHX9 has been reported to directly interact with DGCR8 (Shiohama et al., 2007). In contrast, Srpk1 is a serine/arginine-rich protein-specific kinase that specifically phosphorylates splicing factors (Gonçalves 2014). Interestingly, the interaction to human DGCR8 has been recently revealed in IP-MS experiments of human SRPK1 (Varjosalo et al., 2013).

In Dicer1 purifications, all core subunits of the miRNA-loading complex (miRLC), which is composed of Dicer1, Ago1, Ago2 and Tarbp2, were identified (Figure 20D) (Gregory et al., 2005; Maniataki and Mourelatos, 2005). We also found other reported binding partners like Dhx9 (Robb and Rana, 2007). However, this interaction shows less probability in our analysis (Figure 20D). Finally, Ago1 co-purified all three Tnrc6 paralogs as well as Ago2 and Hspa8 (Figure 20E). Ago1 interacts with Tnrc6 proteins (GW proteins) to mediate post-transcriptional gene silencing (Meister, 2013). Furthermore, we detect Ago1 binding to Heat shock 70 kDa protein 8 (Hspa8). The Hspa8/Hsp90 chaperone machinery mediates a conformational opening of Ago proteins, so that Ago can bind to stiff dsRNA, thus facilitating efficient loading of small RNAs (Iwasaki 2010, Johnston 2010 (Iwasaki et al., 2010; Johnston et al., 2010). Importantly, most interactions remained constant upon RNase A treatments suggesting direct physical interactions of these proteins to Ago1 (Figure 20F). In agreement with previous results (Hock et al., 2007; Landthaler et al., 2008), a weak association of Ago1 to Dicer was detected (Figure 20E).

In summary, our combined genome editing approach with tandem affinity purification coupled to MS analysis exemplifies a highly versatile and efficient method for identification of high confidence protein – protein interactions in mouse ES cells.

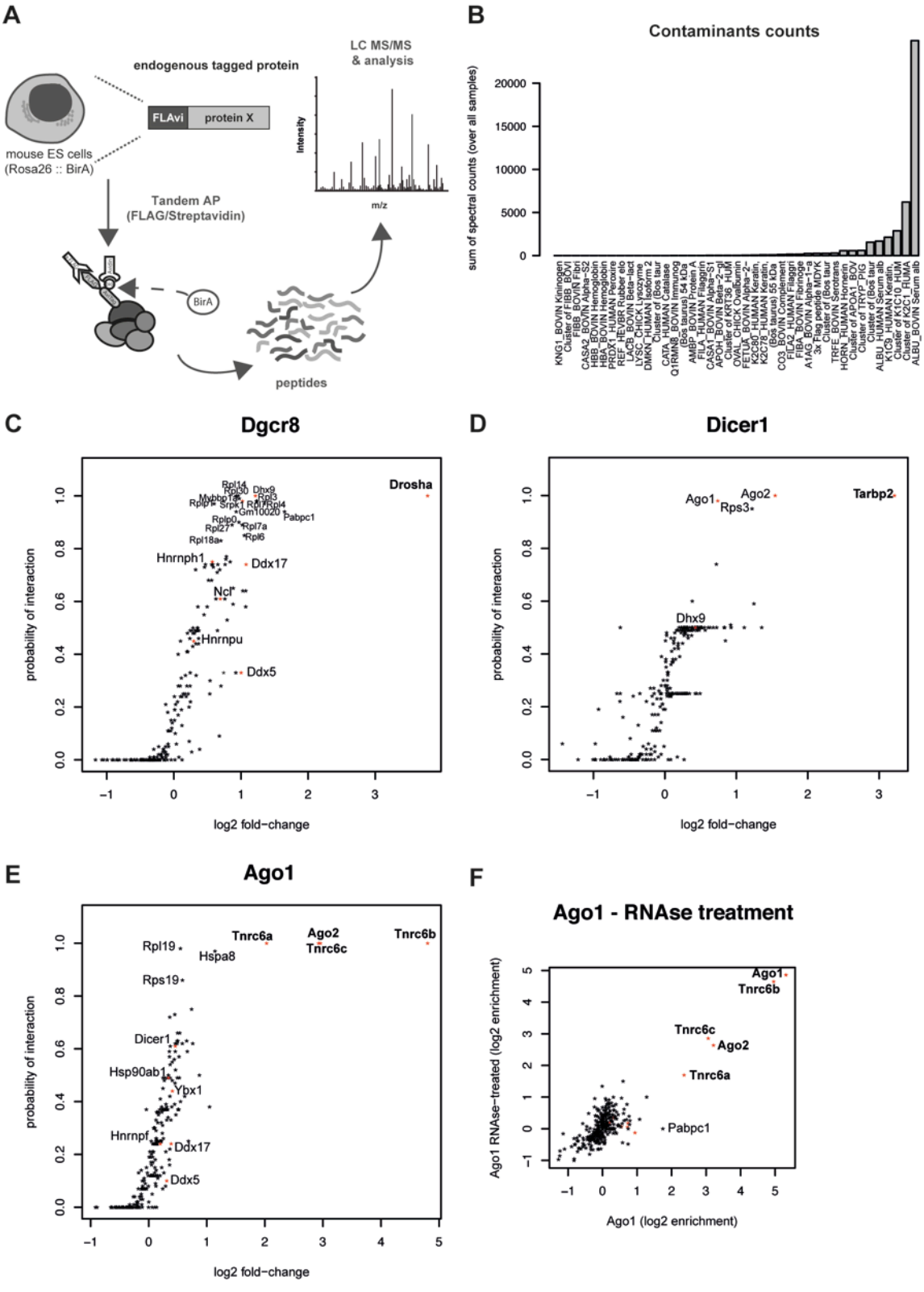


Figure 20. Protein – Protein Interaction Networks of RNAi Factors In Mouse ES Cells

(**A**) Schematic of endogenous gene tagging coupled to tandem affinity purification mass-spectrometry (TAP-MS) analysis of protein complexes. (**B**) Graph showing sum of spectral counts over all samples for contaminant proteins. (**C-E**) Graphs displaying, for each bait protein separately, the average fold-change between bait and control samples versus the SAINT probability of interaction (poi). Previously reported interactors are shown as red asterisks. (**F**) Proteins associated with endogenously tagged Ago1 in TAP-MS samples, compared to RNAse-treated counterparts. All MS-data shown here was averaged from four biological replicates.

2.3 Project III: Functional analysis of Ago1 splicing isoform

Unpublished results:

- Villaseñor R, Flemr M, Hotz HR, Daniel H, Gut H, Seebacher J, Bühler M.

One of the most surprising initial findings after sequencing the human genome was the total number of protein encoding genes (Lander et al., 2001). First studies found between 20,000 and 40,000 genes in the human genome (International Human Genome Sequencing, 2004; Lander et al., 2001), which was a stunning result for scientists. It is only about double the number of genes in a fruit fly or worm. Eleven years after, analysis of the complete human genome sequence led to the identification of approximately 20,687 protein-coding genes (Consortium, 2012), although the annotation still continues to be refined.

A second mystery of life has recently been tackled with the first drafts of the human proteome (Kim et al., 2014a; Uhlen et al., 2015). According to these studies, the human body is made up of at least 17,294 different proteins. Further findings suggest, for instance, a core set of 10,000 to 12,000 proteins identified in most cell types and tissues. In addition, many tissues are characterised by the presence of specific proteins. In contrast to humans, a total of 7,349 proteins have been found in the mouse organism (Geiger et al., 2013). Today, it is becoming clear that there is not a direct correlation between the complexity of an organism and its gene and maybe even protein count. Nevertheless, it remains elusive how human complexity is achieved at the molecular level.

Alternative splicing (AS) is one of the many crucial processes that mediate gene regulation in metazoans (Chen and Manley, 2009; Raj and Blencowe, 2015). During AS of precursor mRNA (pre-mRNA), different combinations of 5 ´ and 3 ´ splice site pairs are selected, resulting in the

generation of diverse mRNA variants. In theory, AS produces multiple proteins from a single gene and is thought to increase the number of proteins extensively over the number of genes in a genome. This in turn, could be required to generate complexity of higher eukaryotes including humans (Blencowe, 2006). It is estimated that 95% of all human genes undergo AS with alternative exons often encoding for disordered regions important for protein-protein interactions (Ellis et al., 2012). Recently, it has been shown that alternative splicing remodels the protein-protein interaction network of alternative spliced genes (Buljan et al., 2012; Ellis et al., 2012; Irimia et al., 2014). Rewiring the interaction network of protein complexes in a given tissue could possibly create further complexity in higher eukaryotes.

Protein variants of RNAi factors made by alternative splicing have remained undiscovered. Recently, an oocyte-specific isoform of mouse Dicer1, a core component of the RNAi pathway essential for processing of small RNAs, was shown to evolved as a consequence of a specific retrotransposon insertion, and to be essential for oocyte function in mice (Flemr et al., 2013). This peculiar example prompted us to search for alternative spliced isoforms of other RNAi factors in mouse cells.

2.3.1 Detection of mouse alternative splice variant of Ago1 (Masva1)

Dr. Matyas Flemr discovered a novel splice variant of mouse Ago1, hereafter referred as Masva1, in cDNA of mouse ES cells and embryonic fibroblasts (NIH3T3). Sanger sequencing of Masva1 cDNA revealed that two alternative splicing events occur between coding exons 9 & 13 of the mouse *Ago1* gene (Figure 21A & 21B). The first AS event causes a +1 frame shift on exon 10. A second AS event on exon 12 restores the original frame of *Ago1* by shifting the frame to -1 (Figure 21B bottom). Both alternative splicing events result in a new stretch of 106 amino acids, which corresponds to most of the MID domain of Ago1 protein (amino acids 422 – 528, Figure 21C). The new

stretch of amino acids “replaces” the original MID domain of Ago1 and creates a new domain (hereafter termed novel MID domain: NMID) with probably entirely new structural properties (Figure 21D). However, *in silico* analysis done by Heinz Gut (FMI) suggests that NMID unlikely folds into structured domain and does not match to any annotated domains in structural databases (data not shown).

To gain first insights into the biological function of Masva1, I set out to test Masva1´s ability to bind small RNAs and determine the spatio-temporal expression levels of *Masva1* gene transcript in mice.

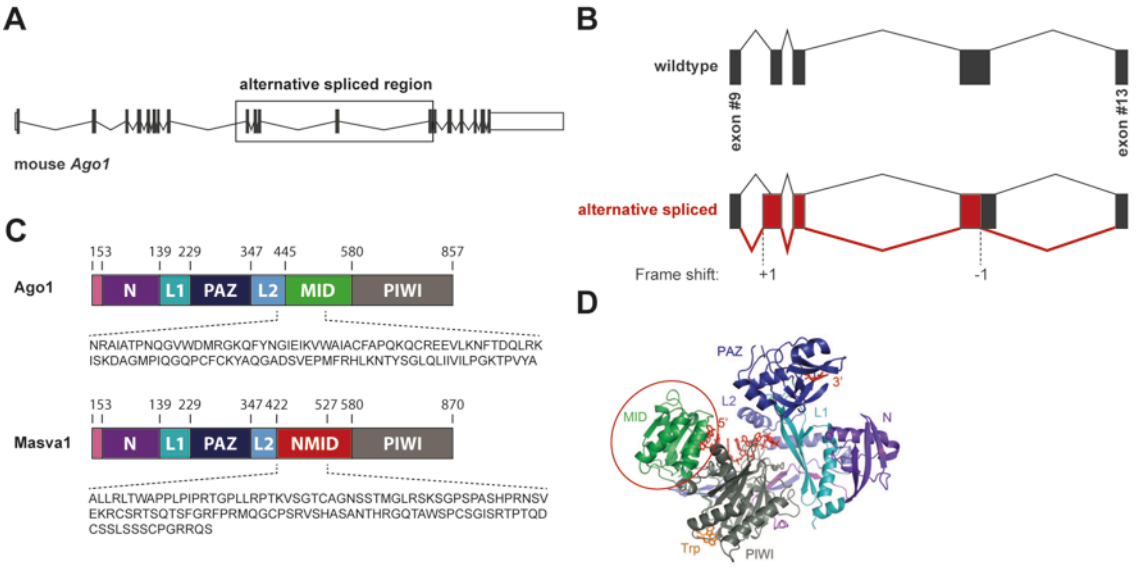
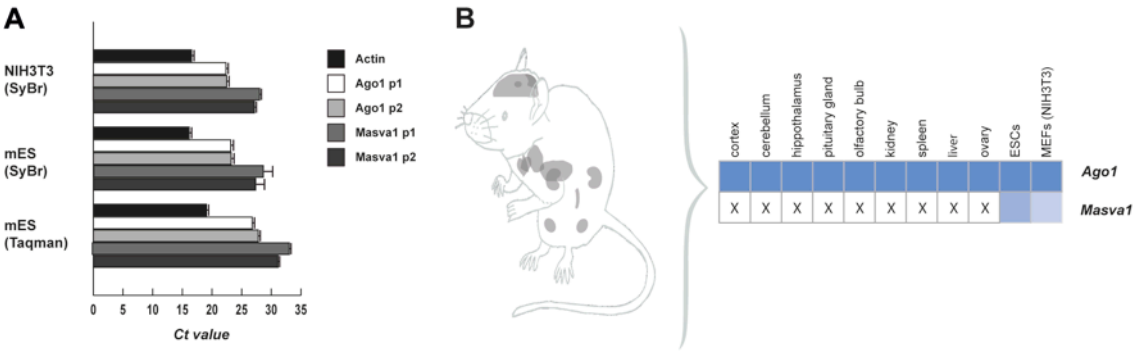


Figure 21. Identification of a Mouse Alternative Splice Variant of Ago1

(**A**) Scheme of mouse Ago1 locus and protein structure. Exons and untranslated regions are shown as grey and white boxes, respectively. The black box indicates the alternative spliced region. (**B-C**) Two alternative splicing events between exon #9 and #13 of the *Ago1* gene result in a new stretch of 106 amino acids, which corresponds to most of the MID domain of Ago1 protein (amino acids 422 – 528). (**D**) This stretch of amino acids possibly creates a novel domain with entirely new structural properties. The crystal structure of HsAgo2 (Schirle and MacRae, 2012) is shown here for comparison.

2.3.2 *Spatio-temporal expression of Masva1 transcript*

Alternative splicing of *Ago1* seems only to occur in mouse. Sequence alignment across 37 eutherian mammals did not show evolutionary conservation of the 5´alternative splice site, suggesting that expression of *Masva1* is restricted to *Mus musculus* (data not shown). I used two specific primer pairs against *Masva1* or *Ago1* to measure mRNA levels by SYBR-qRT-PCR. I measured 4 to 5 higher threshold cycles (Ct) of *Masva1* in comparison to *Ago1* mRNA in mES cells and NIH3T3 cells (Figure 22A). Similarly, qRT-PCR using two specific taqman-probes against *Masva1* or *Ago1* transcript showed 3 to 6 higher Ct values of *Masva1* in comparison to *Ago1* mRNA in mES cells (Figure 22A). These results suggest that *Masva1* is at least ~10-fold lower expressed than *Ago1* in mES and NIH3T3 cells.



Reference	SRA	Origin	library	fastq	bigWig	*Ago1* reads	*Masva1* reads
Grote et al.	SRR639080	E12.5	SE	yes	yes	yes	n.d.
Seisenberger et al.	ERR192339	E16.5m	SE	yes	yes	yes	n.d.
	ERR192340	E11.5	SE	yes	yes	yes	n.d.
	ERR192341	E13.5m	SE	yes	yes	yes	n.d.
	ERR192342	E13.5f	SE	yes	yes	yes	n.d.
	ERR192343	E16.5f	SE	yes	yes	yes	n.d.
Shen et al.	GSM723775	MEF	PE	no	yes	yes	n.d.
	GSM723776	mESC	PE	no	yes	yes	n.d.
	GSM850906	brain-E14.5	PE	no	yes	yes	n.d.
	GSM723773	lung	PE	no	yes	yes	n.d.
	GSM723768	cerebellum	PE	no	yes	yes	n.d.
Smallwood et al.	ERR034795	d10-oocytes	SE	yes	yes	yes	n.d.
	ERR034797	FG-GV_oocyte	SE	yes	yes	yes	n.d.

Figure 22. Spatio-Temporal Expression of Masva1 Transcript

(**A-B**) Quantitative real-time PCR measurement of *Ago1* and *Masva1* mRNA expression in ES and NIH3T3 cells (A), as well as nine different mouse tissues (B). Computational survey of 13 published RNA-

sequencing datasets to detect expression of *Masva1* gene in several murine tissues and developmental stages (**C-D**).

Next, I focused on finding out where *Masva1* is most abundantly expressed in murine tissues to understand its physiological function. To determine the expression profile of *Masva1* in the adult mouse, total RNA was isolated from six different adult mouse tissues to prepare cDNA and measure *Masva1* and *Ago1* mRNA levels. *Masva1* expression failed to be detected as measured by qRT-PCR using two specific primer sets (Figure 22B). In contrast, *Ago1* expression was faithfully detected in all tissues tested.

Publicly available RNA sequencing (RNA-seq) datasets of different mouse tissues and developmental stages represent a rich source for evaluating expression of genes and finding alternative spliced transcripts. I examined several transcriptome-wide RNA-seq datasets (Grote et al., 2013; Seisenberger et al., 2012; Shen et al., 2012; Smallwood et al., 2011) to identify expression of the *Masva1* alternative spliced exon (Figure 22B). To do this, raw data was downloaded, processed and mapped against the mouse genome (Figure 22C). Aligned reads were visualized using the UCSD genome browser (http://genome.ucsc.edu). My computational survey of 13 RNA-seq datasets failed to reveal expression of *Masva1* gene in several murine tissues and developmental stages other than mES and NIH3T3 cells (Figure 22D).

2.3.3 Sub-cellular localization of Masva1

Ago1 is a predominantly cytoplasmic protein and has previously been shown to localize in cytoplasmic sites of mRNA decay known as processing-bodies (P-bodies) (Liu et al., 2005; Sen and Blau, 2005). To determine if this feature is conserved in Masva1, I examined Ago1 and Masva1 sub-cellular distribution in NIH3T3 cells by immunofluorescence (IF) analysis.

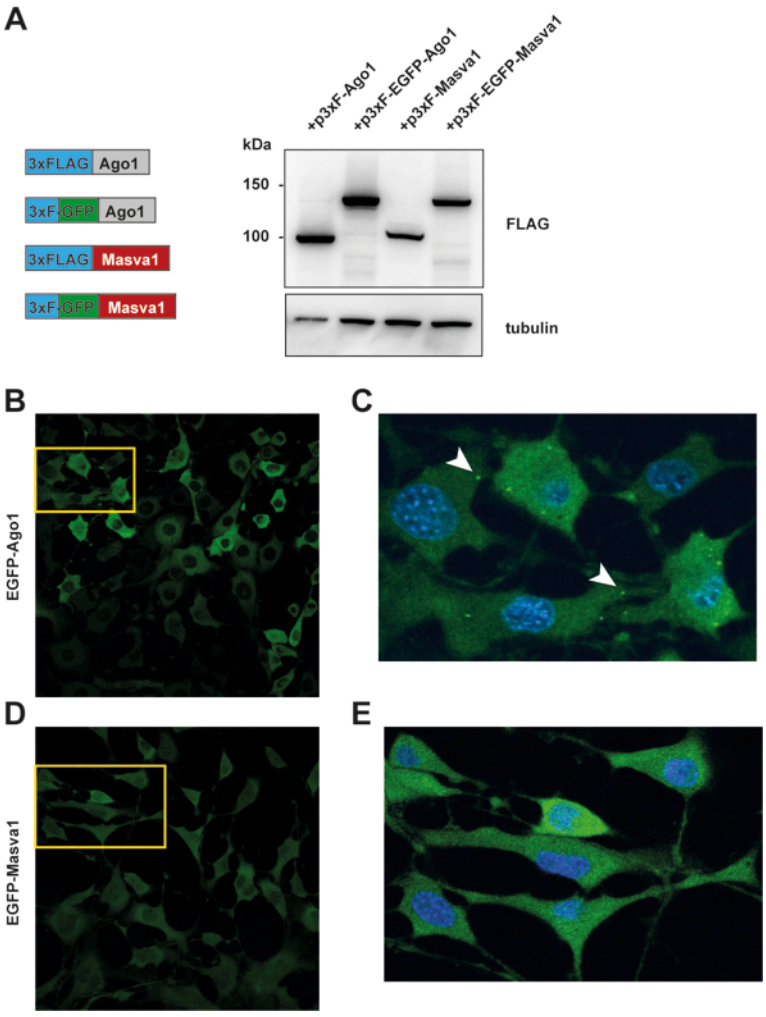


Figure 23. Cellular Distribution of Masva1 In NIH3T3 Cells

(**A**) Schematic of mammalian expression constructs and Western blot showing expression of tagged Ago fusions. (**B**) Image showing the cellular distribution of Ago1 and Masva1 protein in NIH3T3 cells. (**C**) Blow-up of yellow box shown in image B. EGFP-Ago1 shows a predominant cytoplasmic localization. White arrows indicate P-bodies. (**D**) EGFP-Masva1 shows a diffused cellular localization with no formation of cytoplasmic foci. (**E**) Blow-up of yellow box shown in image D. DNA stained with DAPI.

First, I generated plasmids encoding for tagged fusions of Ago1 and Masva1 (Figure 23A). Stable overexpression of exogenous 3xFLAG-EGFP-tagged Ago1 (EGFP-Ago1) or Masva1 (EGFP-Masva1) in NIH3T3 cells was established (data not shown). Then, I determine the sub-cellular distribution of GFP-tagged Ago1 by IF analysis (Figure 23B). In most cells, Ago1 was found in

the cytoplasm and accumulated in cytoplasmic foci that resembled P-bodies (Figure 23B & 23C). In contrast, GFP-tagged Masva1 was dispersed throughout cytoplasm and nucleus, although the signal appeared more prominent in the cytoplasm when whole cell distribution was observed (Figure 23D). Interestingly, the formation of Masva1 cytoplasmic foci was absent in most cells visualized (Figure 23E).

2.3.4 Profiling of Masva1 associated small RNAs

Many small RNA classes possess specific sequence biases at the 5´-end. Based on structural evidence, it has become clear that the Argonaute´s MID domain senses the 5´-nucleotide of small RNAs. A rigid loop in the MID domain of human Ago2 allows specific contacts to a 5-terminal uridine or adenine (Elkayam et al., 2012; Schirle and MacRae, 2012; Schirle et al., 2014). Small RNAs with Gs or Cs at the 5´-end bind to human Ago2 with low affinities. Similar small-RNA-binding strategies have been found for plant proteins (Frank et al., 2012). Therefore, it is appealing to speculate that Masva1 would have different sRNA-binding properties due to its novel MID domain (NMID). Understanding the small RNA binding characteristics of Masva1 is crucial to elucidate its biological function.

To profile Masva1-bound small RNAs, I generated NIH3T3 cell lines stably expressing 3xFLAG-EGFP tagged Ago1 and Masva1 proteins (see Section 2.3.3). Futhermore, 3xFLAG-Ago1 and 3xFLAG-Masva1 fusions were transiently over-expressed in NIH3T3 cells (Figure 24A). I purified total RNA from FLAG-immunoprecipitations (FLAG-IP) of stably or transiently expressed proteins and prepared libraries for small RNA sequencing (Figure 24B & 24C). Computational analysis of Masva1- and Ago1-bound small RNAs revealed a high correlation between all samples sequenced (Figure 24D). Moreover, Masva1 binds miRNAs (Figure 24D) and other sRNAs (data not shown). I observed similar sequence biases at the 5´-end of Masva1-bound miRNAs

compared to Ago1-associated miRNAs. In comparison to Ago1, however, Masva1 seemed to bind fewer miRNAs as shown by total read count, suggesting that Masva1 binds small RNAs with a lower efficiency than Ago1 (Figure 24E). To assess differential binding between Masva1 and Ago1, Dr. Hans-Rudolf Hotz and I performed additional computational analysis. Some miRNAs associated to Ago1 were less efficiently bound by Masva1 (Figure 24F). Interestingly, a group of miRNAs was strongly reduced in Masva1 samples (Mirg and miR410). These miRNAs originate from a cluster of about 40 miRNAs in the imprinted *Dlk1–Dio3* locus on mouse chromosome 12qF1. In contrast, miR330 and miR196b preferentially associated with Masva1 (Figure 24F).

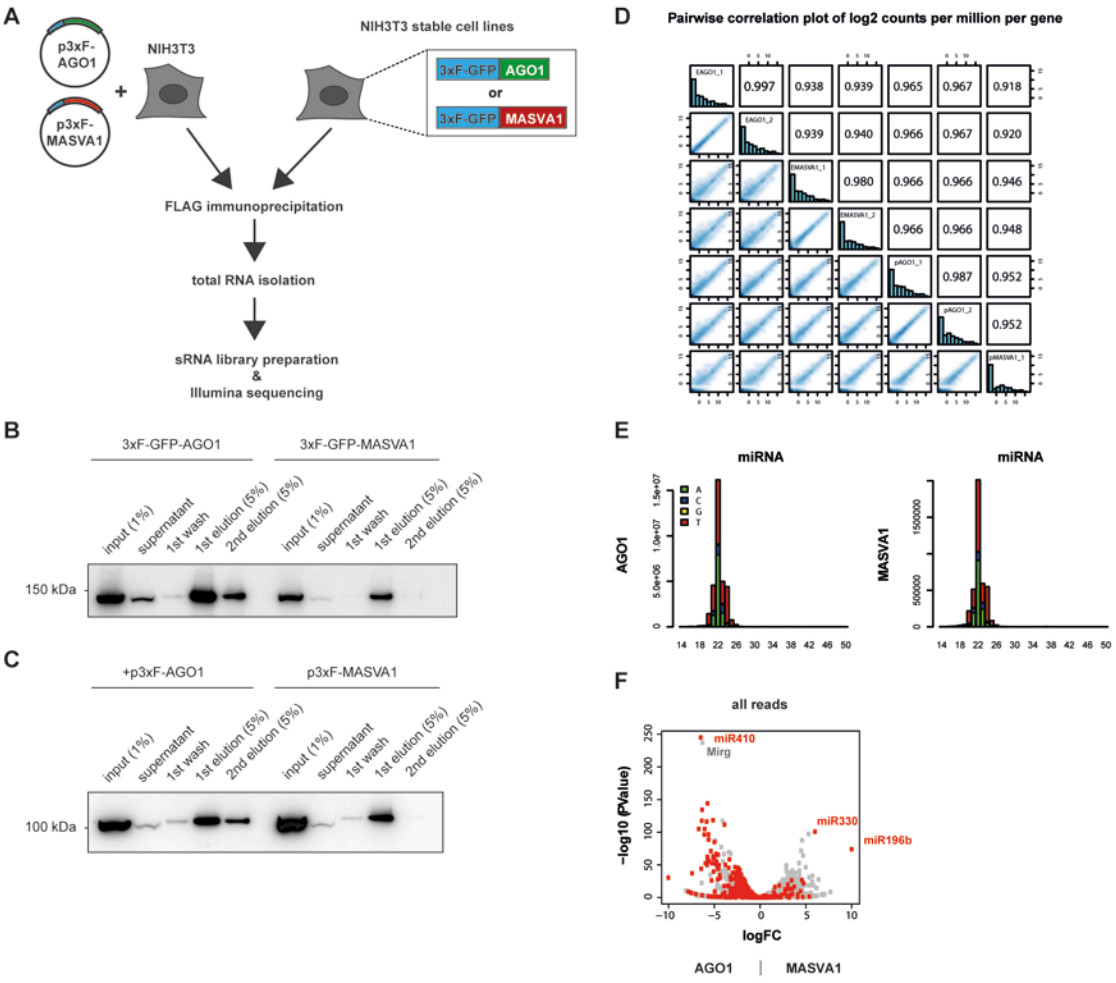


Figure 24. Purification of Small RNAs Associated To Ago Proteins and Their Sequencing

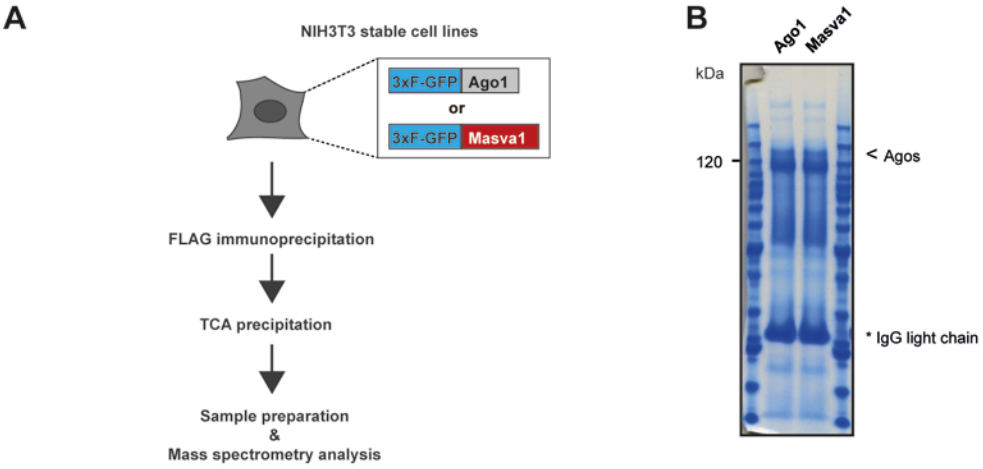
(**A**) Schematic describing experimental workflow. (**B-C**) FLAG-immunoprecipitations of Ago-protein complexes. Western blot showing enrichment of Ago proteins after purification. (**D**) Graph showing pairwise correlation plot of log2 counts per million per gene of reads mapping to small RNAs. Isolated sRNAs were sequenced and the corresponding reads were aligned to the *Mus musculus* genome (GRCm38/ mm10). (**E**) miRNA expression levels were calculated and normalized using mouse miRBase (http://www.mirbase.org/) as a reference. Graphs showing miRNA profile of Ago1 and Masva1, respectively. (**F**) Vulcano plot showing differential binding of small RNAs to Ago1 and Masva1. Reads containing the word "miRNA" in their genome annotation are marked in red.

2.3.5 Analysis of protein-protein interaction network of Masva1

To gain additional insights into the biological function of Masva1 from another experimental angle, I purified Masva1 protein complexes to map its protein-protein interactions. To determine the protein interaction network of Masva1, the generated NIH3T3 cell line stably expressing 3xFLAG-EGFP tagged Masva1 was used for FLAG-IP experiments followed by MS analysis (Figure 25A). Importantly, IP-MS analysis was also performed from NIH3T3 cells stably expressing 3xFLAG-EGFP-Ago1 (Figure 25B). The results obtained in Ago1 IP-MS experiments served as a positive control. Interestingly, the most striking difference between Masva1 and Ago1 IPs was the absence of Tnrc6B-peptides in Masva1 pull-downs (Figure 25C). Furthermore, unique peptides matching to other Ago proteins were not detected in both Masva1 IPs, suggesting that Masva1 is not able to interact with other Ago proteins. However, Masva1 co-purified the Hspa8/Hsp90 chaperone complex, which suggests a similar mechanism of small RNA loading as Ago proteins (Iwasaki et al., 2010; Johnston et al., 2010). In contrast, I could observe unique peptides for Ago2, Tnrc6B, and Hsp proteins in both Ago1 pull-down experiments (Figure 25C).

Taken together, Masva1 protein seems to lack the ability to interact with other Ago proteins and more importantly with Tnrc6B. Although limitations in peptide detection cannot be ruled out in these experiments, the results above predict that Masva1 lacks the ability to mediate gene silencing through

recruitment of the Ccr4-Not effector complex to targeted mRNAs. This hypothesis remains to be tested in future experiments.



C

Identified Proteins	Accession Number	MW	Peptide count		Peptides (protein coverage %)	
			Ago1	Masva1	*Lanthaler et al.*	*Hoeck et al.*
Cluster of Protein argonaute-1	AGO1_MOUSE	97 kDa	131	45		
Protein argonaute-1	AGO1_MOUSE	97 kDa	130	45		
Protein argonaute-2	AGO2_MOUSE	97 kDa	28	0		
Protein argonaute-1	AGO1_MOUSE	97 kDa	11	0		
Putative protein Masva1		98 kDa	0	10		
Heat shock protein HSP 90-beta	HS90B_MOUSE	83 kDa	63	52	10 (18)	
Heat shock protein HSP 90-alpha	HS90A_MOUSE	85 kDa	49	22	15 (11)	
Heat shock cognate 71 kDa protein	HSP7C_MOUSE	71 kDa	35	50	34 (21)	
Trinucleotide repeat-containing gene 6B protein	TNR6B_MOUSE	192 kDa	**27**	**0**	21 (26)	
Elongation factor 1-alpha 1	EF1A1_MOUSE	50 kDa	24	26	24 (8)	
Coatomer subunit alpha	COPA_MOUSE	138 kDa	19	3		6 (6)
DNA damage-binding protein 1	DDB1_MOUSE	127 kDa	12	3		4 (3)
Stress-70 protein, mitochondrial	GRP75_MOUSE	73 kDa	7	6	12 (7)	
Polyadenylate-binding protein 1	PABP1_MOUSE	71 kDa	3	2	23 (14)	
Lupus La protein homolog	LA_MOUSE	48 kDa	1	3	19 (7)	
Nuclease-sensitive element-binding protein 1	YBOX1_MOUSE	36 kDa	3	1	28 (6)	
Eukaryotic translation initiation factor 3 subunit D	EIF3D_MOUSE	64 kDa	2	1		0 (1)
Elongation factor 2	EF2_MOUSE	95 kDa	2	1	10 (4)	
Insulin-like growth factor 2 mRNA-binding protein 3	IF2B3_MOUSE	64 kDa	2	0	12 (7)	
Eukaryotic translation initiation factor 4B	IF4B_MOUSE	69 kDa	4	4		0 (1)
RuvB-like 1	RUVB1_MOUSE	50 kDa	0	6		4 (2)
Heterogeneous nuclear ribonucleoprotein U	HNRPU_MOUSE	88 kDa	2	3	19 (11)	4 (5)
Probable ATP-dependent RNA helicase DHX36	DHX36_MOUSE	114 kDa	0	3		1 (1)
Heterogeneous nuclear ribonucleoprotein K	HNRPK_MOUSE	51 kDa	2	1	7 (3)	
Heterogeneous nuclear ribonucleoprotein Q	HNRPQ_MOUSE	70 kDa	0	2	*	
Heterogeneous nuclear ribonucleoprotein F	HNRPF_MOUSE	46 kDa	2	0	*	1 (3)
Heterogeneous nuclear ribonucleoprotein M	HNRPM_MOUSE	78 kDa	1	0	13 (8)	

Figure 25. Masva1 Interaction Partners Identified By Mass Spectrometry Analysis

(**A**) Outline describing experimental workflow. (**B**) Coomassie stained gel showing FLAG-immunoprecipitations of Ago-protein complexes. (**C**) Summary of two independent IP-MS experiments showing interaction partners of Masva1 in comparison to Ago1 in NIH3T3 cells. Peptide count and protein coverage for each identified protein in other studies are shown. Masva1 proteo-specific peptides matching to the NMID domain can be detected.

2.3.6 *Proteomic analysis in Ago1 gene tagged mouse ES cells*

Expression of *Masva1* mRNA was only detected in NIH3T3 mouse embryonic fibroblasts and mouse ES cells. Over-expression experiments in NIH3T3 cells demonstrate that *Masva1* can be translated to a functional protein with the ability of binding small RNAs (Results section 2.3.4). To prove the existence of the endogenous Masva1 protein variant, I aimed to detect Masva1 protein in biochemical purifications of endogenous Ago1 protein complexes by using mouse ES cell lines expressing endogenously tagged *Ago1* generated in Section 2.2.1 (Figure 26A).

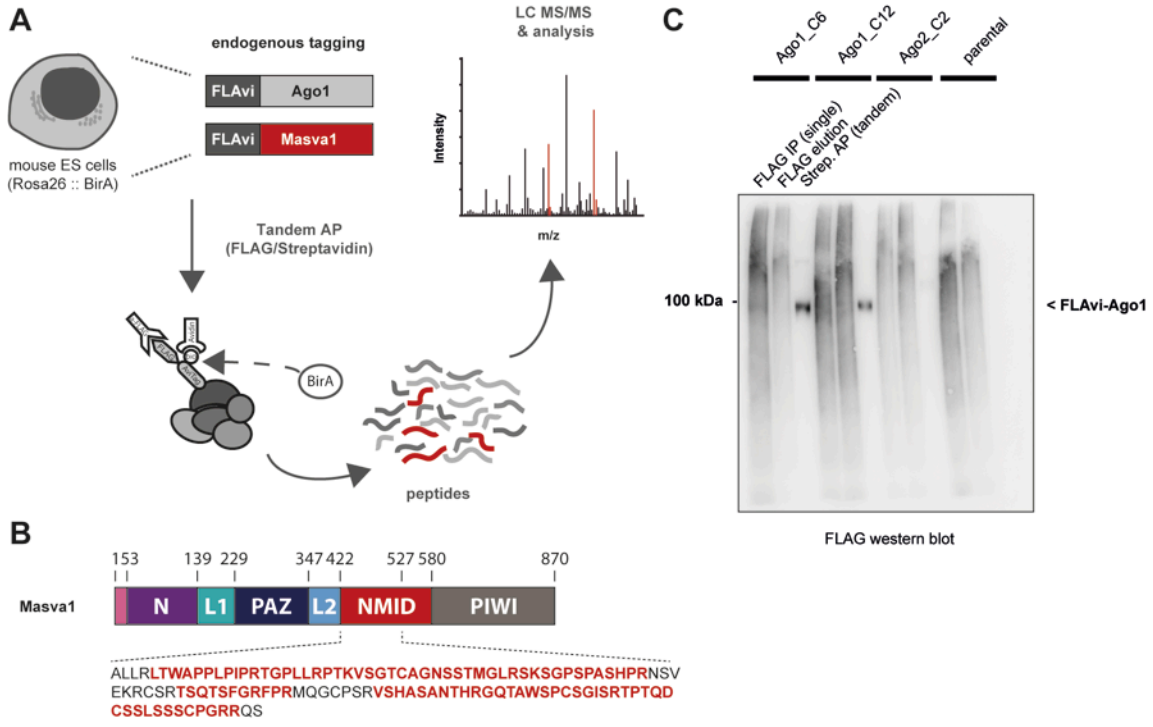


Figure 26. Proteomic Search of Endogenous Masva1 Protein In Mouse ES Cells

(**A**) Schematic of endogenous gene tagging coupled to tandem affinity-purification mass-spectrometry (TAP-MS) analysis of tagged Ago1 protein complexes. (**B**) Scheme of Masva1 protein showing the proteo-specific peptides matching to the NMID domain, marked in red. (**C**) Western blot showing FLAG-immunoprecipitations of Ago-protein complexes. Two endogenously tagged Ago1 cell lines were used for single or tandem affinity purification followed by mass-spectrometry analysis.

The amino acid sequence of the NMID domain in Masva1 represents the only difference between Masva1 and Ago1. In over-expression experiments, I could detect unique peptides matching to the NMID domain of Masva1 by MS analysis (Figure 26B). Next, I performed tandem affinity purifications in two endogenously tagged *Ago1* cell lines and searched for unique peptides corresponding to Ago1 protein or to the NMID domain of Masva1 by MS analysis (Figure 26C). Detection of peptides exclusively matching to Masva1 failed in ten independent IP-MS experiments, suggesting that Masva1 protein is extremely low abundant in mES cells. In contrast, unique peptides corresponding to Ago1 were found in all TAP-MS experiments.

2.3.7 Summary of results

In summary, the findings presented in chapter 2.3 suggest that Masva1 might be expressed predominantly during early developmental stages such as mouse ES cells and embryonic fibroblasts. The immunofluorescence results indicate that Ago1 and Masva1 have different sub-cellular localization patterns when overexpressed in mouse fibroblasts, which may reflect differences in their biological function. I could show that Masva1 has the ability to bind small RNAs although with a much lower binding efficiency than Ago1. Furthermore, the NMID domain allows binding of most miRNAs found in murine fibroblasts with the exception of imprinted miRNAs. How structural changes in the NMID domain allow binding of some miRNAs but not others remains to be addressed in future experiments. In addition, Masva1 protein seems to lack the ability to interact with other Ago proteins and more importantly with Tnrc6B. Although limitations in peptide detection cannot be ruled out in these experiments, the results above predict that Masva1 lacks the ability to mediate gene silencing through recruitment of the Ccr4-Not effector complex to targeted mRNAs. This hypothesis remains to be tested in future experiments. Finally, I applied genome editing to tag endogenous *Ago1* gene and performed TAP-MS to identify

endogenous Masva1 protein. However, peptides matching to the NMID domain of Masva1 were absent in ten independent TAP-MS experiments.

2.4 Project IV: Novel insights in mammalian m6A RNA methylation

Unpublished results:

- Villaseñor R, Knuckles P, Daniel H, Burger L, Seebacher J, Bühler M.

Methylation of adenosines at position N6 on RNA molecules is the most abundant internal modification in eukaryotic mRNA. Besides mRNA, this RNA modification is found on small nucleolar RNA (snoRNA), ribosomal RNA (rRNA), transfer RNA (tRNA) and long non-coding RNA (lncRNA), such as XIST (X-inactive specific transcript) and MALAT1 (Dominissini et al., 2012; Meyer et al., 2012). N6-Methyladenosine (m6A) was initially discovered in the 1970´s in polyadenylated RNA fractions isolated from mammalian cells (Desrosiers et al., 1974). But the initial attention on m6A in mRNA was abandoned owing to concerns that m6A may have come from contamination of known sources of methylated RNAs (rRNAs and snoRNAs) often found in polyadenylated RNA isolations (Meyer and Jaffrey, 2014). After four decades of latency, the interest in m6A was revived by development of novel high-throughput and genome-wide m6A detection methods (MeRIP-seq). Two pioneering studies mapped the transcriptome-wide distribution of m6A methylation sites on human and murine transcriptomes (Dominissini et al., 2012; Meyer et al., 2012). The resulting maps have shown that about 25% of all mRNAs are m6A methylated and that the modification is enriched near the stop codons, in 3' untranslated regions (3' UTRs) and within internal long exons. Furthermore, the m6A-methylated regions have often a consensus sequence of RRACH (where R denotes A or G, and H represents A, C or U).

Discoveries from biochemical and genetic studies further enhanced the interest on m6A (Bokar et al., 1997; Hongay and Orr-Weaver, 2011; Jia et al.,

2008). Recently, m6A has been linked to a plethora of physiological processes, ranging from obesity, synaptic signalling and cancer to circadian periods, sperm development, and stem cell differentiation (Fu et al., 2014). Despite important recent advances, the precise function of m6A and how mRNA methylation is regulated remain poorly understood.

Mettl3 (Methyl-transferase-like 3) is an enzyme that promotes m6A methylation of RNAs (Figure 27A). Purification of the enzyme showed substrate specificity towards GAC and AAC sequences in single-stranded RNA (Bokar et al., 1994). Cloning of the human *METTL3* gene revealed that it has a classic SAM-binding methyltransferase domain, which is consistent with its function (Bokar et al., 1997). Additionally, homologues in plants (MTA), *Saccharomyces cerevisiae* (inducer of meiosis 4; Ime4) and *Drosophila melanogaster* (Ime4) have also been identified (Clancy et al., 2002; Hongay and Orr-Weaver, 2011). Genetic ablation of *Mettl3* results in early embryonic lethality in mice and impaired oogenesis in flies (Geula et al., 2015; Hongay and Orr-Weaver, 2011).

Biochemical studies on human METTL3 revealed interactions to another N6-methyltransferase, known as METTL14 (Methyl-transferase-like 14) in mammals (Figure 27). In addition to METTL14, Wilms tumour 1-associated protein (WTAP) was found to associate to the human METTL3-METTL14 complex (Liu et al., 2014; Ping et al., 2014; Wang et al., 2014c). Previously, human WTAP has been shown to interact with RNA binding proteins in the nucleus and bind nuclear lncRNAs. Another recent study expanded the interaction network of the METTL3-METTL14 complex. Schwartz and colleagues previously found the protein VIRILIZER to interact with METTL14 in human cells. Knockdown experiments of VIRILIZER showed that it is required in human cells for cellular mRNA methylation (Schwartz et al., 2014). Interestingly, knockdown of human WTAP lowered global m6A levels and decreased the amount of RNA bound to METTL3, suggesting that WTAP may recruit the METTL3-METTL14 enzymatic complex to substrate RNAs (Liu et al., 2014; Ping et al., 2014; Schwartz et al., 2014).

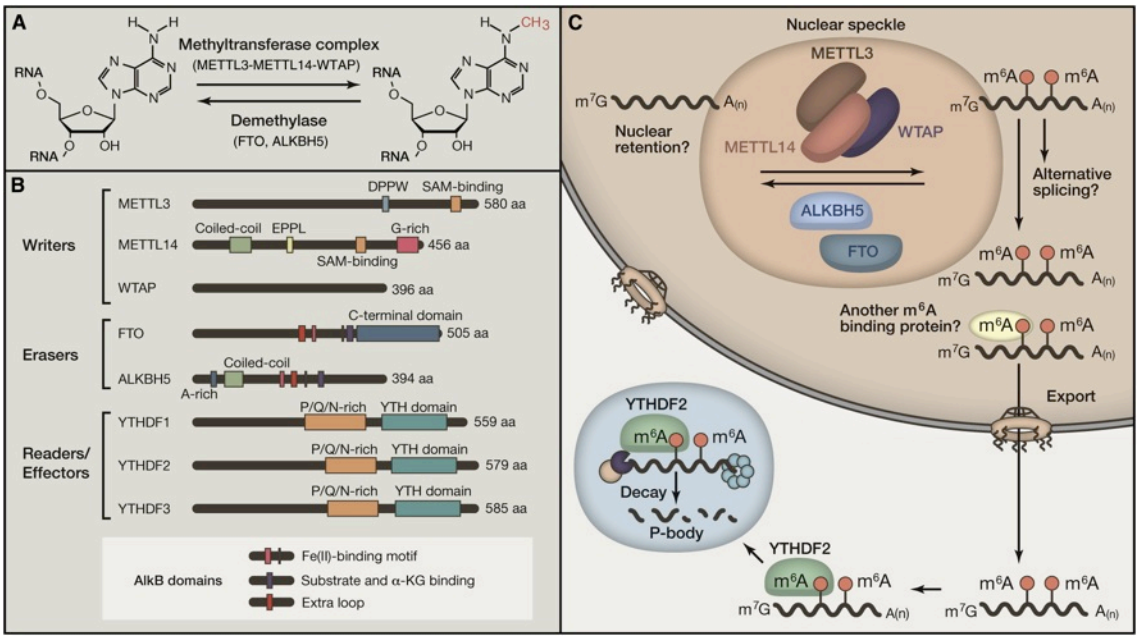


Figure 27. The m6A Pathway and Its Cellular Functions

(**A**) N6-methyladenosine methylation and demethylation reactions. Methyltransferase complex containing METTL3, METTL14, and WTAP catalyzes m6A methylation, whereas FTO and ALKBH5, the demethylases, catalyze oxidative demethylation of m6A. (**B**) Domain structures of writer, eraser, and reader/effector proteins in human. METTL3, METTL14, and WTAP are components of the methyltransferase complex. METTL3 and METTL4 have a SAM-binding domain required for m6A methylation, whereas WTAP contains no characteristic domain. Eraser proteins, FTO and ALKBH5 demethylases, have an AlkB domain in common. Compared to ALKBH5, FTO has an additional C-terminal domain. YTHDF1,2,3 containing a YTH RNA-binding domain are effector proteins. The P/Q/N-rich domain is known to be important for the localization of YTHDF2 to P body (Fu et al., 2014). (**C**) Proposed model of the cellular function of m6A on mRNA. Reversible methylation/demethylation is thought to occur in nuclear speckles where the enzymes are concentrated. Methylation may affect the export and splicing of mRNAs in the nucleus. Exported methylated mRNAs are recognized by YTHDF2 in the cytoplasm and then localize to P bodies, where mRNA decay factors are enriched. Image taken with permission from (Lee et al., 2014).

m6A is likely to have a role in mediating RNA-protein interactions. In principle, methylation of adenosine could either block or induce RNA-protein interactions (Figure 27C). To date, several m6A-binding proteins have been identified from mammalian cellular extracts using RNA pull-down approaches, which are followed by mass spectrometry (Dominissini et al., 2012; Schwartz et

al., 2014; Wang et al., 2014c). These include the mammalian proteins YTH domain-containing family 1 (YTHDF1), YTHDF2, YTHDF3, YTHDC1 and YTHDC2 each of which contains an YTH RNA-binding domain (Figure 27B).

To date, only two cytoplasmic m6A readers have been characterized in human cells – namely YTHDF1 and YTHDF2. He and co-workers showed that the cytoplasmic YTHDF1 protein promotes the translation of m6A methylated mRNAs in human cells (Wang et al., 2015). Transcriptome-wide protein-RNA interaction mapping indicated that around 5,000 mRNA are targets of YTHDF1. Ribosome profiling revealed a significant decrease in the translation efficiency of YTHDF1 targets following YTHDF1 knockdown or the reduction of m6A levels. Furthermore, YTHDF1 associates with translation initiation factors and subunits of the ribosome. Therefore, it is believed that YTHDF1-dependent translation could be promoted by delivering mRNAs to the translation machinery and by enhancing translation initiation (Wang et al., 2015).

Both YTHDF1 and YTHDF2 locate in the cytoplasm**.** YTHDF1 and YTHDF2 seem to share about 50% common target transcripts. However, it has been previously reported that m6A reader protein YTHDF2 decreases the stability of its m6A-modified target mRNAs (Wang et al., 2014b), which seems to contradict the findings about the translation-promotion role of m6A. A possible explanation for this conundrum suggests the following: YTHDF1 and YTHDF2 could have differences in timing of their binding to common RNA targets. Support for this hypothesis showed that YTHDF1 binds to m6A methylated mRNA transcripts earlier than YTHDF2, which would be consistent with the expectation that translation of most of these mRNAs should occur before degradation under normal growth conditions (Wang et al., 2015).

Recently m6A has been identified as a novel regulator of miRNA processing. Human METTL3 mediates methylation of pri-miRNAs, seemingly facilitating their recognition and processing by the microprocessor complex (Alarcon et al., 2015). To study the link between N6-adenosinemethylation and

miRNA biogenesis, and more importantly how METTL3 binds its substrates, we endogenously tagged *Mettl3* using CRISPR/Cas9 technology in mouse ES cells. By TAP-MS I found novel protein interaction partners. We currently are characterizing the function of these interaction partners by measuring the effect their depletion has on Mettl3 binding to chromatin and enzymatic activity.

2.4.1 Mapping protein-protein interactions of Mettl3 in mouse ES cells

To study the protein interaction network of Mettl3, Dr. Philip Knuckles and I set out to endogenously tag Mettl3 in RosaB (BirA-expressing) mouse ES cells. To tag the endogenous *Mettl3* gene, we designed a gRNAs to target the first coding exon of the mouse *Mettl3* gene (Figure 28A). A vector encoding for the Mettl3-gRNA and Cas9 enzyme (gR-Cas9-2C) was co-transfected with an ssODN carrying the FLAG-AviTag sequence and homology regions against the first coding exon of *Mettl3* gene. We obtained 15 of 96 positive clones (15.6% efficiency) by PCR genotyping analysis. Moreover, we confirmed the expression of FLAvi-Mettl3 protein in six out of eight tested clones by western blot (Figure 28B).

Next, I conducted tandem affinity purification of Mettl3 protein complexes using two cell lines expressing endogenously tagged FLAvi-Mettl3. The same tandem affinity purification was done in non-tagged RosaB cells that served as control samples. RNase A treated samples were included to the MS analysis to account for RNA-dependent and RNA-independent direct physical interactions of Mettl3. All TAP-samples were subjected to MS analysis and the obtained data was statistically analysed using the SAINT algorithm (as previously described in Section 2.2.2).

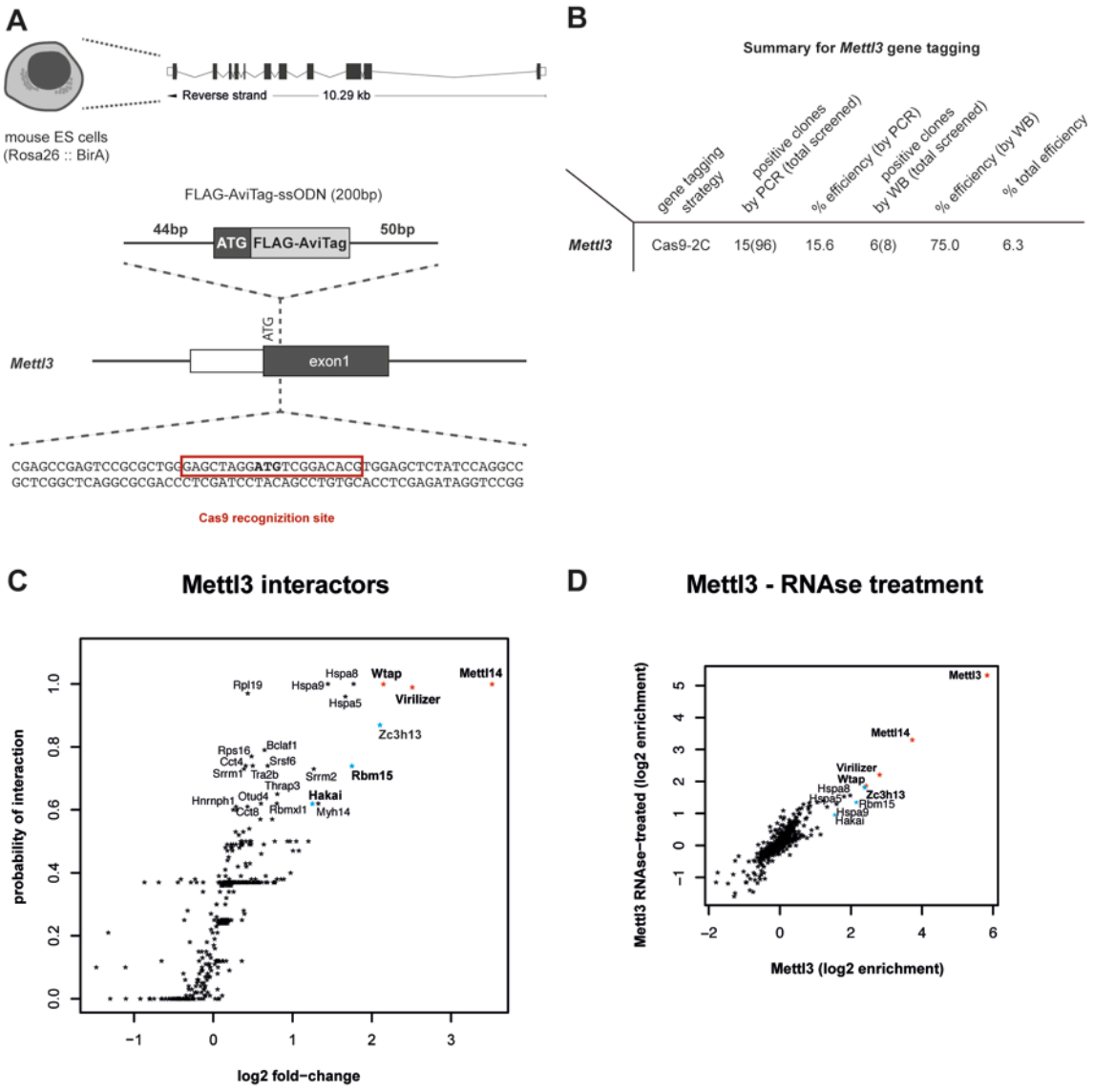


Figure 28. Protein-Protein Interaction Network of Mettl3 In Mouse ES Cells

(**A**) Schematic overview of Cas9-assisted integration of an ssODN carrying a FLAG-AviTag sequence downstream of the Mettl3 start codon (bold letters). (**B**) Summary statistics for Cas9-assisted *Mettl3* gene tagging. (**C**) Graphs displaying the average fold-change between Mettl3-IP and control samples versus the SAINT probability of interaction (poi). Previously reported interactors are shown as red asterisks. Blue asterisks denote RNA-binding proteins previously identified by Horiuchi et al. (Horiuchi et al., 2013). (**F**) Proteins associated with endogenously tagged Mettl3 in TAP-MS samples, compared to RNAse-treated counterparts. All MS-data shown here was averaged from four biological replicates.

Previously identified partners of Mettl3, namely Mettl14, Wtap and Virilizer (alias KIAA1429), were found among the proteins with the highest enrichment and probability of interaction in all TAP purifications (Figure 28C). In addition to these interactors, we identified novel binding partners of the Mettl3-Mettl14 complex that were not detected in previous MS-experiments using overexpression of Mettl3. SAINT analysis revealed a high-probability interaction to the chaperone proteins Hspa5, Hspa8 and Hspa9. These interactions seem highly specific to Mettl3, as we do not observe similar interactions to other bait proteins (Figure R11). Our analysis further revealed high-confidence interactions of Mettl3 with other RNA-binding proteins and splicing factors. Among those interactions, I observed the strongest log2 fold-change (2.2) and a probability of interaction score of ~0.9 for Zinc finger CCCH domain-containing protein 13 (Zc3h13) – a poorly characterized CCCH-type zinc finger protein in eukaryotes with the potential to bind RNA transcripts. Interestingly, another RNA-binding protein known to bind RNA through its conserved RNA recognition motifs (RRM) domains, called Rbm15 (RNA binding motif 15), shows a strong enrichment and high poi-score in all TAP-MS experiments. In contrast, other observed interactions had lower poi-scores and enrichments over control samples, suggesting that these factors are weak or indirect interactors of the Mettl3-Mettl14 complex (Figure 28C). Most of the weak Mettl3-interacting proteins are RNA-binding proteins, which participate in the control of pre-mRNA splicing. These results are noteworthy because they could provide further evidence for the hypothesis that m6A is set co-transcriptionally on nascent mRNAs, which I will further discuss chapter 3.4.

In addition, I treated Mettl3-TAP samples with RNase A to account for RNA-dependent and RNA-independent direct physical interactions of Mettl3. Most high-probability interactions remained constant upon RNase A treatment, suggesting direct physical interactions of these proteins to Mettl3 (Figure 28D).

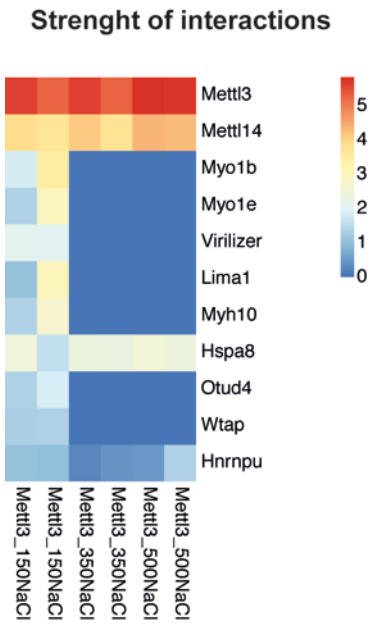


Figure 29. The Mouse N6-Methyladenosine Transferase Complex Consists of a Highly Stable Mettl3-Mettl14 Complex

Heatmap showing enrichment of the most highly detected proteins in all samples that show at least 2-fold enrichment in the 150-mM NaCl condition. Most interactions detected at 150-mM NaCl are lost under more stringent affinity purification conditions (350 – 500-mM NaCl).

To test the strength of the identified interactions, I performed TAP-MS of Mettl3 protein complexes under increasing salt concentrations (150, 350 and 500-mM NaCl). The parental cell line, which lacks expression of endogenous tagged proteins, was treated under exact same conditions as the Mettl3-TAP samples. We calculated the enrichments over TAP-MS control samples using the spectral counts for each detected protein after MS-analysis. The heatmap shown in Figure 29 displays the enrichment of the most highly detected proteins in all samples that show at least 2-fold enrichment in the 150-mM salt condition. Most of the observed interactions in Figure 28D are lost under more stringent affinity purification conditions (Figure 29). Remarkably, the Mettl3 to Mettl14 interaction resisted all three salt concentrations, suggesting that Mettl3 acts primarily in a complex with Mettl14 *in vivo* and might function synergistically. Other previously reported components of the Mettl3-Mettl14 complex such as

Wtap and Virilizer are lost under high salt concentrations (350 and 500-mM NaCl). Surprisingly, Hspa8 is the only interaction partner besides Mettl14, which seems to remain bound to Mettl3 under high stringency conditions (Figure 29). The biological meaning of the strong Mettl3-Hspa8 interaction awaits further investigation.

Taken together, we applied genome editing to tag endogenous *Mettl3* gene and identify novel protein binding partners of the Mettl3-Mettl14 protein complex. Our combined genome editing approach with tandem affinity purification coupled to MS analysis led to the discovery of high-confidence and novel protein binding partners of the m6A methyltransferase complex such as chaperone proteins Hspa5, Hspa8 and Hspa9 and the RNA-binding proteins Zc3h13 and Rbm15. In current follow-up experiments, we are assessing the biological significance of these findings.

3 Discussion & Conclusion

Recent advances in genome editing technologies are having a remarkable impact in the life sciences. During the course of my doctoral studies I experienced the development of more advanced genome editing tools such as TALENs and the CRISPR/Cas9 system. The first project presented in this doctoral thesis applied ZFN technology – at that time the only available tool for genome editing. However, this technology has severe limitations compared to more recent programmable nucleases (see Introduction). Therefore, I switched from ZFNs to TALENs and Cas9. These technologies were successfully applied in subsequent projects presented in this thesis, which are now becoming standard molecular techniques in biology research. Here, I will discuss the results and open questions for each project in a separate chapter.

3.1 Project I: Application of genome editing for drug discovery

In recent years, a significant body of literature has proposed that in theory, genome-editing technologies, cellular reprogramming, and high-throughput biology each have the potential to revolutionize biomedicine and pharmaceutical pipelines. This study establishes that these emerging biotechnologies can be effectively combined to perform as proposed, exemplifying a new paradigm for drug discovery.

3.1.1 Genome-engineering tools to establish accurate reporter cell lines that enable identification of therapeutic strategies to treat Friedreich's Ataxia.

Since the expanded GAA repeats in FRDA reside in the first intron of the *FXN* gene and thus do not alter the amino acid sequence of the protein, gene

reactivation would be of therapeutic benefit. Efforts to find small molecules that reverse *FXN* gene silencing are thus of medical importance. Several studies have demonstrated that benzamide-based histone deacetylase inhibitors (HDACi) and synthetic PPAR-γ agonist (e.g. A-PAF) have the ability to moderately increase FXN protein levels in non-affected and FRDA patient-derived cells as well as in FRDA mouse models (Chou et al., 2008; Herman et al., 2006; Marmolino et al., 2009; Rai et al., 2008). However, it is not clear how this compounds act on *FXN* gene expression.

One important application for HEK293T-FF2AP cell line is its use in high-throughput genomic screens for the discovery of drug targets that activate *FXN* gene expression or stabilize FXN protein turnover. The HEK293T-FF2AP cell line established in this study is the first luciferase reporter-based cellular model, which allows an accurate and effortless assessment of endogenous frataxin gene regulation and is compatible with high-throughput biology. HEK293T-FF2AP cells are easy to transfect and can be expanded to a very large scale at a relatively low cost. Thus, this system should be compatible with the technical and financial constraints of most academic and industrial screening platforms.

Using the HEK293T-FF2AP cell line in a pilot high-throughput genomic screen, we uncovered novel negative regulators of *FXN* expression, demonstrating the importance of monitoring gene expression from the endogenous *FXN* locus. Our results showing that *FXN* expression can be increased even in the presence of an expanded repeat tract also demonstrate that screening for general regulators of *FXN* transcription, irrespective of GAA repeat length, is an appropriate strategy. Thus, our approach opens up new opportunities for target discovery and we anticipate that high-throughput screening of genome-wide siRNA or cDNA libraries will reveal additional potential drug targets in the near future.

Furthermore, the genome engineering tools designed in this study will serve as a powerful resource to the FRDA community. The ZFNs and donor

constructs developed in this study will enable others to knock in any given sequence tag into the *FXN* locus in many different cell types, facilitating the dissection of molecular pathways involved in FRDA.

Generation of patient-derived reporter cell line

High-throughput screening for *FXN* gene activators or low molecular weight compounds has so far met with limited success because current cellular models may not accurately assess endogenous *FXN* gene regulation under disease conditions. To facilitate the identification of molecules that enhance *FXN* gene expression and therefore could alleviate FXN deficiency, the cause behind FRDA, cell lines that fully reflect the *in vivo* disease situation are essential tools for drug discovery. Although *FXN* knockout mice and cell lines are important tools for understanding the role of frataxin in cell physiology and disease etiology and progression (Martelli et al., 2012a), these resources are not appropriate for discovery of *FXN* gene activators. For that purpose, several groups have developed reporter cell lines harbouring a reporter gene separated by an artificial intron containing extended GAA repeat tracts to mimic the genetic defect observed in FRDA (Puccio, 2009). Another group recently developed a HeLa-based reporter cell line with a randomly integrated BAC containing the human *FXN* locus and 350-nt long GAA repeats (Lufino et al., 2013). Because the FXN-FL fusion is expressed from BAC clones, this system does not monitor *FXN* expression in its native chromosomal context. Consequently, none of these cell lines fully recapitulate the molecular defects of an endogenous, silenced *FXN* gene. Therefore, we set out to develop a FRDA patient-derived reporter cell line that entirely reflects the complexity of the repressed *FXN* gene in the *in vivo* disease scenario in FRDA.

My efforts lead to the generation of the FRDA4078iBT-FF cell line. Several lines of evidence indicated the suitability of this cell line for basic research and HTS applications. In contrast to the HEK293T-FF2AP cell line, where other

integrations of the reporter gene were observed, only a single specific integration at the *FXN* gene locus was detected by southern hybridizations. Furthermore, the FRDA4078iBT-FF line contains extended GAA repeats tracts on both alleles. More importantly, the cells show stable luciferase activity that responds to treatments of one specific chemical inhibitor of PRKD1. However, more experiments performed to further validate the FRDA4078iBT-FF cell line failed to detect the low abundant FXN-FL fusion protein (data not shown). Treatment of these cells with PRKD1 chemical inhibitor compound 13c led to a 5-fold increase of luciferase activity that failed to translate to a similar increase of FXN protein.

In conclusion, I believe the FRDA4078iBT-FF cell line is erroneous and caution is needed if these cells are employed in future experiments. Moreover, the *FXN*-ZFNs used in this study were less efficient in FRDA patient-derived fibroblasts preventing *FXN* gene editing in those cells. Therefore, modern genome editing tools like TALENs and CRISPR/Cas9 might offer additional possibilities to edit the *FXN* locus in FRDA patient-derived cells with expanded GAA repeats. As shown in later projects, TALENs or the CRISPR/Cas9 system combined with recombination reporters can be used to efficiently edit mouse gene loci. I am confident that a similar approach can be applied in human primary cells, iPSCs or stem cells to edit disease loci such as the human *FXN* gene.

Use of PRKD1 chemical inhibitors in patient-derived lymphoblasts

To assess the effect of PRKD1 chemical inhibition in reactivation of *FXN* expression, we tested several low molecular weight inhibitors of PRKD1 in patient-derived cells. *FXN* mRNA and protein levels remained constant after 24h-treatments using compound 13c in FRDA fibroblasts (GAA_{541}/GAA_{420}) and FRDA lymphoblasts (GAA_{1030}/GAA_{650}). Similarly, WA-21-JO19 failed to reactivate *FXN* expression in FRDA lymphoblasts after 48h treatment.

Remarkably, I noticed a modest increase in *FXN* mRNA and FXN intermediate protein expression after 48h WA-21-JO19-treatment of lymphoblasts derived from healthy individuals. However, these results still need further verification to evaluate their statistical significance.

What could be the mechanisms of action for PRKD1 inhibitors? PRKD1 belongs to the serine/threonine protein kinase D family, which consists of three isoforms playing a role in growth factor signalling and in stress-induced signalling (Steinberg, 2012). In cardiac myocytes, PRKD1 phosphorylates HDAC5 resulting in nuclear export of HDAC5 and de-repression of downstream target genes. HDAC5 suppresses stress-dependent remodelling of the heart via association with the myocyte enhancer factor 2 (MEF2) transcription factor (Fielitz et al., 2008). MEF2 gene transcription in the myocyte, in turn, alters myocyte growth, contraction, calcium handling, metabolism and promotes cardiac hypertrophy (Kim et al., 2008).

In other cells, the MEF2 family of transcription factors are bound to their target promoters and repressed by a variety of HDACs (Potthoff and Olson, 2007). Other transcription factors might be repressed by HDACs similarly to MEF2 in cardiac cells. It remains to be tested if PRKD1 inhibition might lead to MEF2-dependent transcription of frataxin. In addition, several studies have demonstrated that benzamide-based HDAC inhibitors promote moderate increase of FXN protein levels in non-affected and FRDA patient-derived cells as well as in FRDA mouse models (Chou et al., 2008; Herman et al., 2006; Rai et al., 2008). However, it is not clear how these compounds act on *FXN* gene expression. A possible mechanism of action for benzamide-based HDACi could be the activation of MEF2-dependent transcription leading to increased *FXN* expression. Alternatively, other transcription factors repressed by HDACs might be involved in *FXN* transcription. Therefore, the identification of regulators of *FXN* gene expression would benefit the search for potential therapeutic targets for FRDA patients. The established HEK293T-FF2AP cell line also enables high-

throughput screening of cDNA libraries to decipher the transcriptional regulatory network of *FXN* gene expression.

Use of iPSCs to model FRDA

In section 2.1.4, I described the use of human iPSCs to model FRDA and the testing of a PRKD1 inhibitor in FRDA neurons. The outcome of this work is two-fold: we found that PRKD1 is a druggable target in the patient- and disease-relevant cell type, and we found 2,6-naphthyridines as candidate compounds for drug development to treat FRDA. Importantly, WA-21-JO19 treatment led to a 25% increase in *FXN* mRNA expression in FRDA neurons after 3 weeks of treatment. Therefore, I speculate that the effect of PRKD1 chemical inhibition might be dependent on GAA repeat length and duration of treatment. It remains to be verified if a 25% increase in *FXN* expression is therapeutically sufficient for at least a subpopulation of FRDA patients as some FRDA patients exhibit FXN levels only 20-30% lower than asymptomatic carriers (Martelli et al., 2012b; Miranda et al., 2002). In addition, neuron viability was not affected by exposure to WA-21-JO19, even at 10 μM. Therefore, it may be possible to increase WA-21-JO19 efficacy or the efficacy of its structure-activity relationship (SAR) derivatives by increasing treatment duration and/or compound concentration. Oral availability, high PRKD1 selectivity, and low toxicity thus make WA-21-JO19 an attractive candidate for drug development.

In conclusion, this multifaceted approach, developed collaboratively by academic and industrial research teams, allowed us to identify a means to counteract a disease-causing phenotype in a patient-derived cellular context, highlighting the potential of this discovery paradigm.

3.2 Project II: Endogenous in-vivo biotinylation system for mapping protein networks and protein-DNA interactions of RNAi factors in embryonic stem cells

We have recently developed reagents and protocols for *in vitro* genome engineering such as endogenous gene tagging in mouse ES cells (Flemr and Buhler, 2015). Additionally, we established Cas9-constructs and protocols for efficient editing of gene loci where TALENs are not available (Knuckles, Villaseñor et al., unpublished). This study establishes that endogenous gene tagging and *in vivo* biotinylation can be successfully combined to map the protein interaction network of virtually any protein of interest. This method exemplifies a new paradigm for mass-spectrometry based proteomics of protein complexes. Unlike traditional affinity purification approaches to study protein-protein interactions, the technique presented here allows detection of most physical interactions for a bait protein. Importantly, this approach circumvents the need for specific antibodies against the protein of interest, retains endogenous expression levels of the gene of interest, and allows reverse genetic experiments in isogenic mouse ES cells. In addition, the relatively small size of the FLAG-AviTag is likely to lack interference with the native function of the target protein. However, some proteins lose functionality when tags are fused to them, representing a foremost disadvantage of this approach. Therefore, tagging of endogenous proteins requires additional downstream analysis such as functional testing of the generated fusions using biochemical methods and/or genomic approaches.

To assess the competence of our method to map protein-protein interactions, we established several mouse ES cell lines expressing endogenously tagged RNAi factors. Both gene-tagging approaches using TALENs or the CRISPR/Cas9 system showed similar efficiencies (Figures 18 & 19). It remains to be investigated why just a few and not all positive clones

identified by PCR genotyping express tagged proteins. A possible explanation is the efficiency and accuracy of the underlying mechanism to repair a DSB using an ssODN as a template. Microhomology-mediated end-joining (MMEJ) is an additional DSB repair mechanism active in G1 that uses microhomologous sequences (5–25 bp) for error-prone end-joining (McVey and Lee, 2008). Depending on the cell cycle stage where cells are targeted for genome editing, repair of induced DSBs might be achieved using the error-prone MMEJ pathway.

The combined genome editing approach with tandem affinity purification coupled to MS analysis yielded high confidence bait-prey interactions for all tagged RNAi factors. I could identify most of the previously reported protein-binding partners of Dgcr8, Dicer1 and Ago1 regardless of their predominant cellular localization (Figure 20). For all three bait purifications, I observe novel interaction partners with a probability of interaction score greater than 0.8. However, the novel interactors are mostly ribosomal proteins, which are difficult to reconcile with a function in the canonical miRNA biogenesis pathway. Further validation experiments are needed to assess the functional role of these interactors in the miRNA pathway.

Interestingly, our analysis failed to detect a known interaction partner of Ago1 in all TAP purifications, the Moloney leukemia virus 10 protein (Mov10). Much of the existing literature describes Mov10 as a cytoplasmic protein that associates with Ago complexes in P-bodies (Chendrimada et al., 2007; Meister et al., 2005). A recent proteomics study confirmed the inability of human MOV10 to interact with AGO1 (Gregersen et al., 2014). Instead, AGO2 seems to be a direct binding partner of MOV10, suggesting additional functions of AGO2 in other pathways. This and my results suggest that Ago proteins can have differential binding partners. To provide evidence for this hypothesis TAP-MS experiments of Ago2 would be needed. Finally, my results indicate that Ago1s main function is miRNA-mediated gene silencing in mammals. This function

would be consistent with the role of Ago1 in *Drosophila* showing an astonishing evolutionary conservation across animal kingdoms (Figure 30).

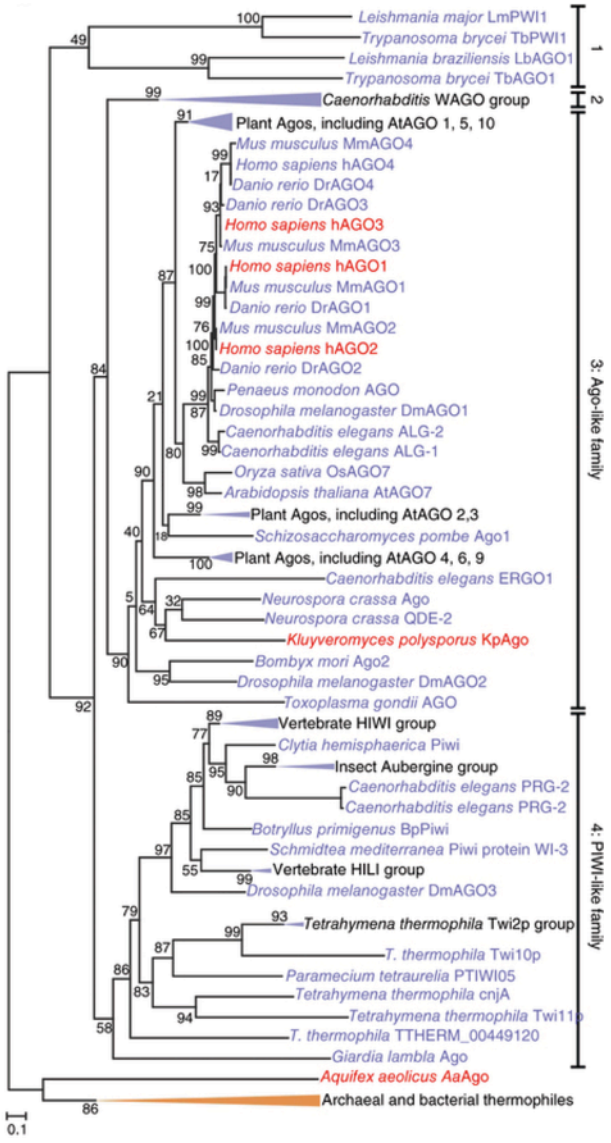


Figure 30. Phylogenetic Analysis of a Representative Set of 177 Eukaryotic Agos

1, *Trypanosoma* Ago family; 2, WAGO family. Ago sequence alignment and uncollapsed phylogenetic tree are in found in (Swarts et al., 2014). Picture adapted from (Swarts et al., 2014). Copyright © 2014, Rights managed by Nature Publishing Group.

In summary, the methodology presented in this chapter is highly reproducible and versatile, making it ideal to study interaction networks of almost any protein in mouse ES cells.

3.3 Project III: Functional analysis of Ago1 splicing isoform

Alternative splicing (AS) has the potential to remodel the protein interaction network of alternative spliced genes (Ellis et al., 2012; Irimia et al., 2014). Until now, alternative spliced variants of RNAi genes have remained undiscovered. In this project, we discovered an alternative spliced variant of the mouse *Ago1* gene, which we termed *Masva1*. Masva1 protein acquired through two AS events a novel domain (NMID) that replaces the MID domain of Ago1. *Masva1* transcript appears to be specific for *Mus musculus*, since the alternative splice sites and amino acid sequence of NMID are not evolutionary conserved.

To assess the biological relevance of *Masva1*, I first determined its spatio-temporal expression in different mouse organs and cells. My results described in section 2.3 showed that *Masva1* mRNA is only expressed in mouse ES and NIH3T3 cells, suggesting that *Masva1* is predominantly expressed during early murine development. Additionally, *Masva1* is ~10-fold less abundant than *Ago1* in these cells. However, I cannot rule out the expression of *Masva1* in other mouse tissues that were not tested – possibly at higher levels than *Ago1*. Therefore, it will be crucial for future follow-up experiments to determine the exact spatio-temporal expression of *Masva1* in other mouse tissues and developmental stages.

The MID domain of Ago proteins is crucial for binding the 5´-end of small RNAs (Meister, 2013). Surprisingly, Masva1 is able to bind small RNAs through its NMID domain. Masva1 showed similar sequence preferences at the 5´-end of its bound miRNAs in comparison to Ago1-associated miRNAs. Modelling of the 5´-binding pocket of both proteins (Heinz Gut, FMI) partly explains why Masva1 retained the ability to specifically bind miRNAs (Figure 31). A conserved tyrosine residue (Tyr529) is important to stabilize the binding of the first base of the small RNA (Elkayam et al., 2012; Schirle and MacRae, 2012). Although Masva1 lacks this particular tyrosine residue, it seems to replace it by a glutamine residue

(Gln540) to bind the first base of small RNAs. Therefore, it is reasonable to speculate that NMID domain of Masva1 folds into a structured domain partly resembling the three-dimensional fold of Argonaute MID domains.

Moreover, results from small RNA profiling experiments suggest that Masva1 binds small RNAs less efficient than Ago1 (Figure 16E). This qualitative observation could partially be explained by the replacement of the aromatic Tyr529 to the non-aromatic Gln540 residue. Gln540 could in theory stabilize the first base-pair of the sRNA via hydrophobic interactions and/or hydrogen bond formation. However, this effect would be less efficient than stabilization via electrostatic interactions between aromatic residues (Figure 31). Lastly, it is more likely that the unstructured parts of the NMID domain have a negative effect on the overall fold of Masva1 leading to reduced sRNA binding efficiencies.

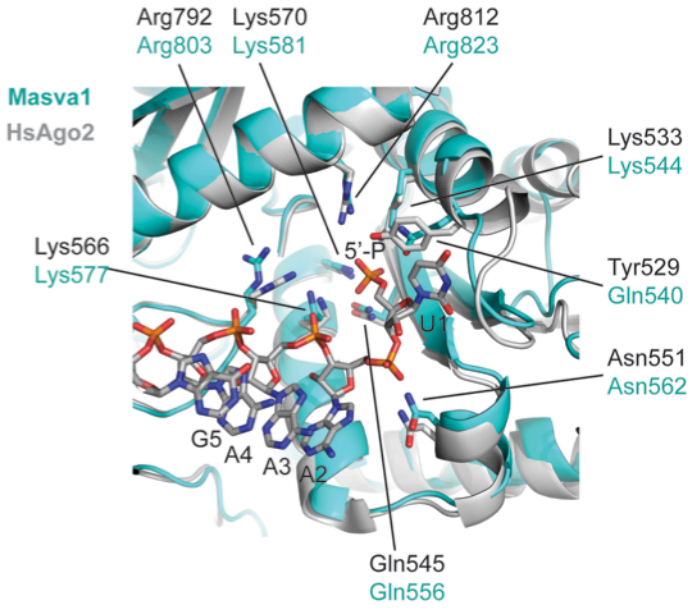


Figure 31. Detailed View Onto The 5´Binding Site of Human Ago2 In Complex With Mir-20a

The crystal structure of human Ago2 [(Elkayam et al., 2012), PDB ID 4F3T] is displayed as grey cartoon model with selected residues involved in miRNA binding and the first five nucleotides on the 5' side of microRNA-20a shown as sticks (grey, atom colors). A homology model of the Masva1 is superimposed and displayed in cyan with selected residues predicted to interact with miRNA as sticks in cyan (atom colors).

In contrast to Ago1, Masva1 lacks the ability to localize to cytoplasmic P-bodies (section 2.3.3). In addition, Tnrc6B was absent in Masva1 affinity purifications, suggesting that Masva1 lost the ability to interact with GW proteins (section 2.3.5). As discussed above, Masva1 binds small RNAs in a similar fashion as Ago1. These results raise the question whether Masva1 is able to mediate gene silencing through recruitment of effector complexes to target transcripts. The Ccr4-Not complex is recruited to target mRNAs through its interaction with GW proteins. The absence of Tnrc6B and possibly other Tnrc6 proteins in Masva1 complexes, suggests that Masva1 lacks the ability to recruit deadenylase complexes to target mRNAs. Moreover, Masva1 showed a diffused cellular localization without accumulation in cytoplasmic foci. But, the formation of P-bodies seems to be a consequence, rather than the cause, of miRNA-mediated gene silencing (Eulalio et al., 2008a). Therefore, I assume that Masva1 is not involved in miRNA-mediated gene silencing mechanisms. This assumption could be tested in tethering experiments of Masva1 to reporter mRNAs.

The discovery of *Masva1* opens up the exciting possibility to remodel the interaction network of Ago1 to gain additional functions. In the following paragraph, I will briefly describe four aspects to consider if somebody continues this line of research. First, *Masva1* mRNA quantities are almost below detection levels in mouse ES and NIH3T3 cells. Bioinformatics search across several RNA-seq datasets from different mouse tissues and developmental stages could not reveal expression of *Masva1*. Second, Masva1 protein is below detection in mouse ES cells. Third, AS leading to the generation of *Masva1* seems to be restricted to *Mus musculus,* raising concerns about its general biological significance. Fourth, the functional redundancy between Argonaute proteins further complicates functional studies of individual members. The biological role of mammalian Ago1 still remains unclear, since genetic ablation of the *Ago1* gene *in vivo* and *in cellulo* does not lead to striking observable phenotypes. Therefore, it is questionable to what extent genetic ablation studies of *Masva1*

might provide clear insights into its biological role. However, some findings presented in this project are intriguing and deserve further attention. I believe that future efforts should be put primarily in identifying the tissue or biological scenario (e.g. disease) where Masva1 is most highly expressed to understand its function.

What could be the biological function of Masva1, if any, in the cell? Masva1 could possibly act as a storage protein to protect miRNAs and other sRNAs from degradation. To function as proposed, Masva1 protein levels could be lower than the protein levels of endogenous Argonaute proteins. This hypothesis also foresees a mechanism of miRNA exchange between Masva1 and Ago proteins, which could be facilitated by heat-shock proteins. Additionally, signalling pathways and tissue distribution might control the abundance of Masva1 in mice. Future experiments should focus on finding the place and time of endogenous Masva1 protein expression as well as unbiased approaches to uncover its biological role.

3.4 Project IV: Novel insights in mammalian m6A RNA methylation

In this project, I performed TAP-MS analysis of endogenous Mettl3 complexes purified from murine ES cells. Previously identified partners of Mettl3, namely Mettl14, Wtap and Virilizer, were found among the proteins with the highest enrichment and probability of interaction in all TAP-MS experiments. More importantly, our proteomic approach allowed us to uncover novel interaction partners of Mettl3 that are now under further investigation. In this chapter, I will discuss how the identified Mettl3 binding partners could contribute to mammalian N6-adenosine methylation of RNA transcripts.

Heat shock proteins are commonly observed contaminants in affinity purification experiments presumably because of the presence of unfolded polypeptides during the purification procedure (Gingras et al., 2007). Heat shock proteins perform chaperone function by stabilizing new proteins to ensure correct folding or by helping to refold proteins that were damaged during cell stress (Saibil, 2013). I could observe a significant interaction of Hspa5, Hspa8 and Hspa9 with Mettl3, suggesting a role for chaperone proteins in stabilizing the Mettl3-Mettl14 complex. What could be the function for protein chaperones in the m6A RNA-methylation pathway? Although caution is required in the interpretation of these findings, chaperones play important roles in other pathways. In the RNAi pathway for example, the Hspa8/Hsp90 complex facilitates the binding of “unloaded” Argonaute proteins to duplex miRNAs (Iwasaki et al., 2010; Johnston et al., 2010). Since empty Ago proteins are actively destroyed, chaperones are thought to protect unloaded Argonautes from degradation (Martinez and Gregory, 2013). It might be possible that similar functions for chaperones operate in m6A pathway. For instance, chaperone proteins might help stabilize the interaction of the Mettl3-Mettl14 complex to other proteins. This might be required to control the enzymatic activity of both

methytransferases and protect the complex from degradation. *In vitro* methylation assays of RNA substrates in the presence and absence of chaperone proteins could provide important insights to understand the role of chaperones in m6A methylation.

The RNA binding factors identified as Mettl3 interactors could be involved in recruitment of the methyltransferase complex to certain RRACH sites on mRNAs. In fact, why and how the Mettl3-Mettl14 complex methylates certain mRNAs and not others remains elusive. Most m6A sites are located in the vicinity of the stop codon. Additionally, a subset of mRNAs contains m6A residues in their 5'-untranslated regions (Dominissini et al., 2012; Meyer et al., 2012). Analysis of the MeRIP-seq generated peaks using motif-discovery algorithms identified the enrichment of RRACH consensus motifs within these peaks. However, the majority of RRACH sites are not methylated in mammalian mRNAs, which begs the question of how asymmetric distribution of m6A sites is achieved.

Mettl3 and Mettl14 proteins lack obvious RNA-binding domains (Figure 27B). Therefore, it remains unclear how both enzymes specifically recognize their substrates. Recent studies have pointed towards Wtap as the RNA-binding component of the methyltransferase complex (Ping et al., 2014; Schwartz et al., 2014). Schwartz and colleagues recently demonstrated that Wtap is crucial for the methylation around the stop codon of certain mRNAs (Schwartz et al., 2014). However, Wtap also lacks RNA-binding domains, suggesting that other Mettl3 interaction partners achieve binding to mRNA targets. Another recent study identified Rbm15, Zc3h13, Hakai and Virilizer as binding partners of human WTAP (Horiuchi et al., 2013). These proteins seem to form a complex that is required for cell cycle and alternative splicing regulation. But, the authors could not provide further mechanistic insights into this potential complex, and more importantly, link their findings to the m6A methylation pathway. In this project, I was able to show that Mettl3 specifically interacts with nuclear RNA binding proteins such as Zc3h13, Rbm15 and Hakai (Figure 28). Therefore it is

tempting to speculate that these RNA binding proteins might function as “recruiters” of the Mettl3-Mettl14 complex to target nascent mRNAs in the cell nucleus (Figure 32). In addition, Wtap might act as a scaffold protein to anchor Mettl3 recruiters. In the next paragraphs, I will briefly discuss about potential functions for Zc3h13, Rbm15 and Hakai, which are currently the subject of further investigations in our group.

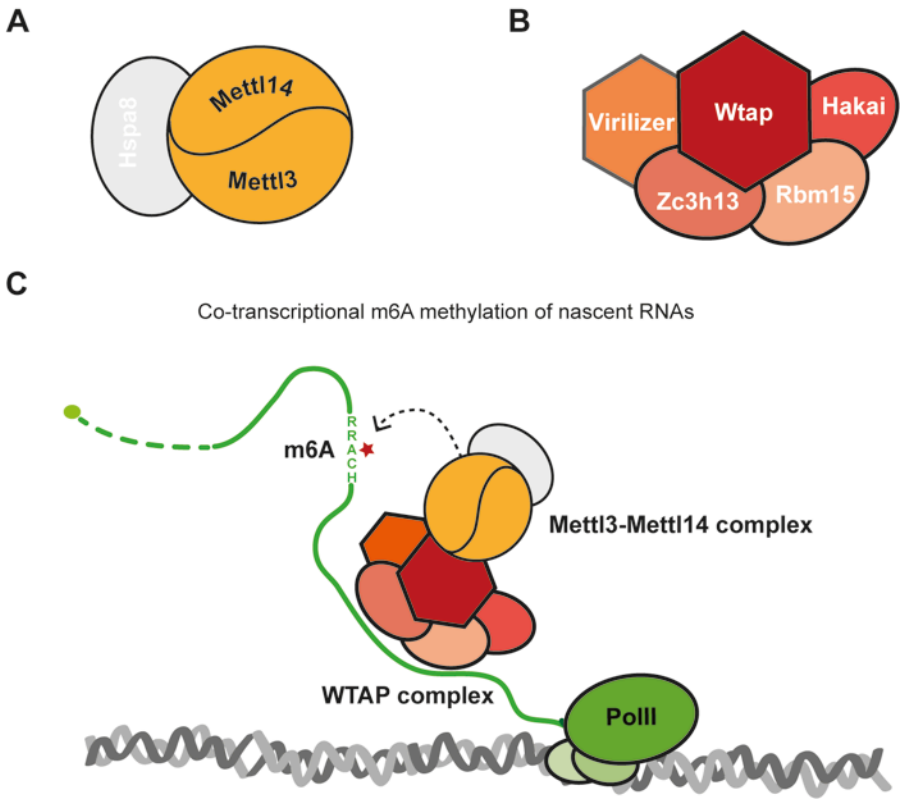


Figure 32. Speculative Model For Co-Transcriptional Methylation of Nascent RNAs

(A) Schematic representation of the Mettl3-Mettl14 N6-methyladenosine transferase complex (yellow). The biological role of Hspa8 (gray) is not clear. (**B**) Scheme showing the multi-protein Wtap complex consisting of Wtap, Virilizer and the RNA-binding proteins Hakai, Rbm15 and Zc3h13. (**C**) Cartoon showing a speculative model for the co-transcriptional methylation of nascent RNAs by the Mettl3-Mettl14 complex. m6A methylation site (red star) on RRACH consensus sequence (where R denotes A or G, and H represents A, C or U).

Zc3h13 is a poorly characterized and evolutionary conserved CCCH-type zinc-finger protein. Proteins of the CCCH-type zinc-finger family have the ability to bind single-stranded RNA through their zinc-finger domains. The *Trypanosoma brucei* Zc3h13 protein has been reported to play various roles in RNA metabolism (Ouna et al., 2012). In addition, a recent genome-wide association study identified the *Zc3h13* gene as a schizophrenia risk locus (Oldmeadow et al., 2014). We are currently testing whether genetic ablation of Zc3h13 affects recruitment of Mettl3 to target transcripts and reduces m6A levels on mRNAs. Additionally, transcriptome-wide protein-RNA association mapping will be crucial to determine which transcripts Zc3h13 binds. This might enhance our understanding of how the Mettl3-Mettl14 complex targets only certain mRNAs in mammals.

Rbm15 is a member of the Split ends (Spen) protein family of transcriptional repressors originally discovered in *Drosophila melanogaster* in the late 1990s. Every family member contains three predicted RNA recognition motifs (RRMs) in the N-terminal region of the protein and a conserved Spen paralog and ortholog C-terminal domain (SPOC). Repression of transcription by Spen proteins can occur through interactions with other repressors, by recruitment of proteins involved in histone deacetylation, or through sequestration of transcriptional activators. The human Spen family proteins Sharp, Rbm15/Ott1, and Rbm15B/Ott3 share the structural domain architecture but show distinct functional properties. The SPOC domain of the human homolog of the fly Spen protein, named SHARP (SMRT/HDAC1 associated repressor protein), has been shown to mediate the interaction with the silencing mediator for retinoid and thyroid-hormone repressor protein (SMRT/NCoR) as well as with the histone deacetylase HDAC1 (Ariyoshi and Schwabe, 2003; Mikami et al., 2014; Shi et al., 2001). In contrast, the SPOC domain of Rbm15, a primarily nuclear protein, has been reported to interact with the general mRNA export factor Nxf1 (Lindtner et al., 2006). Similarly to Rbm15, Rbm15B/Ott3 binds Nxf1 via its C-terminal region and shares with Rbm15 the association with

the splicing factor compartment and the nuclear envelope as well as the binding to other export factors such as Aly/REF (Uranishi et al., 2009). Nevertheless, genetic ablation of Rbm15 in mice results in embryonic lethality, indicating that Ott3 cannot compensate for the Rbm15 loss, which supports the notion that these proteins, in addition to sharing similar activities, likely have distinct biological roles (Raffel et al., 2007; Uranishi et al., 2009).

Rbm15 was recently identified in a genetic screen as a factor involved in Xist-mediated gene silencing. Interestingly, the same study also identified additional proteins such as Wtap and Virilizer, which interact with the Mettl3-Mettl14 complex. Additionally, a recent proteomic study by Guttmann and colleagues identified Rbm15 as the second most abundant Xist-interacting protein (McHugh et al., 2015). This study shows direct evidence that Rbm15 binds lncRNAs such as Xist and Malat1, which are known to be m6A methylated (Dominissini et al., 2012).

What is the role of Rbm15 in the m6A methyltransferase complex? The proteomic study by Guttmann and colleagues identified Sharp as the major Xist-interacting protein (McHugh et al., 2015). Sharp knockdown leads to de-repression of genes silenced by Xist, suggesting that Sharp recruits repressor complexes to promote Xist-mediated gene silencing. In contrast, knockdown of Rbm15 did not reactivate genes silenced by Xist (McHugh et al., 2015). A possible explanation for this observation could be that Rbm15 is not involved in the recruitment of repressor complexes to Xist target loci. Alternatively, Rbm15 could recruit the Mettl3-Mettl14 complex to Xist and possibly other transcripts. Binding to RNA might be achieved through the three predicted RNA recognition motifs (RRMs) in the N-terminus of Rbm15. mRNA binding might occur in a sequence specific manner and/or through recognition of secondary structures. The C-terminal SPOC domain might serve as an interaction surface to recruit the Mettl3-Mettl14 complex to certain mRNAs. Transcriptome-wide protein-RNA interaction mapping will be essential to assess which transcripts Rbm15 binds and whether these substrates are shared with Zc3h13.

What is then the function of m6A methylation of nuclear transcripts? The m6A mark could have a role in mediating RNA-protein interactions. Apart from this purpose, m6A could also affect the local secondary structure of mRNAs. Chang and colleagues lately reported a biochemical approach that allows a global view of RNA secondary structures in living cells. Using this method they could show that m6A impacts RNA structure, favouring the transition from paired to unpaired RNA (Spitale et al., 2015). It remains to be tested if deletion of m6A sites affects the secondary structure of methylated transcripts like Xist and whether this has an impact on their function, localization or stability.

The E3 ubiquitin-protein ligase Hakai (Hakai, which means “destruction” in Japanese) also known as Casitas B-lineage lymphoma-transforming sequence-like protein 1 (Cbll1) is an enzyme that in humans is encoded by the *CBLL1* gene. Hakai was originally discovered as a RING finger domain-containing E3 ubiquitin-ligase targeting the E-cadherin complex, a major component of adherens junctions (Pece and Gutkind, 2002). Hakai binds to the cytoplasmic domain of E-cadherin and mediates its ubiquitination, endocytosis and lysosomal degradation by the proteasome. Hakai contains RING, zinc-finger and proline-rich domains, and interacts with E-cadherin in a tyrosine phosphorylation-dependent manner (Fujita et al., 2002).

In search of downstream effectors of Hakai that function independently of binding to E-cadherin, PSF (polypyrimidine tract-binding protein-associated splicing factor) was identified as a novel Hakai-interacting protein (Figueroa et al., 2009). PSF is a nuclear protein implicated in transcription, DNA binding, unwinding, and repair, as well as pre-mRNA splicing and RNA editing. PSF co-localizes with Hakai in the nucleus (Figueroa et al., 2009). Another recent study identified Wtap as a Hakai-interaction partner in human cancer cells (Horiuchi et al., 2013). Hakai co-localizes with Wtap in the nucleus, raising the possibility that Hakai could play a role in nuclear RNA processing through its association with PSF or Wtap.

What could be the function of Hakai in the nucleus? The crystal structure of Hakai revealed that it forms an atypical, zinc-coordinated homodimer by utilizing residues from the phosphotyrosine-binding domain of two Hakai monomers. Hakai dimerization allows the formation of a phosphotyrosine-binding pocket that recognizes specific phosphorylated tyrosines and flanking acidic amino acids of Src kinase substrates (Mukherjee et al., 2012). Interestingly, Hakai purification experiments revealed a strong interaction to Wtap. The use of a Hakai mutant lacking a small part of the RING finger domain (amino-acids 125-133), which is probably required to form a functional phosphotyrosine-binding pocket, abolished interaction with all Hakai binding partners including Wtap (Horiuchi et al., 2013). Thus, it is appealing to speculate that Hakai binds to Wtap in a tyrosine phosphorylation-dependent manner and mediates ubiquitination of Wtap.

Could Wtap be a Src substrate and be phosphorylated upon Src signalling? Src, and related Src family kinases, are involved in numerous functions and signalling pathways. Recently, two proteomic studies identified nuclear RNA binding proteins as bona fide Src kinase substrates (Ferrando et al., 2012; Sirvent et al., 2012). However, components of the Mettl3-Mettl14 complex were not identified as Src substrates. But, the data suggests that Src over-expression might have a profound effect on mRNA splicing and stability *in vivo*. In conclusion, the role of Hakai in the nuclear m6A-transferase complex remains elusive.

Although m6A is generally believed as a posttranscriptional modification, the actual timing and localization of its deposition on RNA transcripts is not clear (Fu et al., 2014; Meyer and Jaffrey, 2014). mRNA processing events like capping, splicing, cleavage and poly-adenylation occur primarily co-transcriptionally and because m6A modifications were found within introns as well as exons (Dominissini et al., 2012), it is likely that the Mettl3-Mettl14 complex methylates nascent RNAs co-transcriptionally. Consistent with this idea

are three additional lines of evidence:

i) Primary miRNA transcripts and lncRNAs are m6A-methylated (Alarcon et al., 2015; Dominissini et al., 2012; Meyer et al., 2012)

ii) Mettl3 co-precipitates components of the splicing machinery and other nuclear proteins (results presented in this doctoral thesis).

iii) Mettl3 localizes to nucleus and associates with chromatin. Genome wide Mettl3 is enriched at genes known to be m6A-methylated as well as at miRNAs containing loci suggesting the mark is installed co-transcriptionally (results obtained by P. Knuckles during the writing process of this doctoral thesis).

We currently are characterizing the function of the novel interaction partners by measuring the effect their depletion has on Mettl3 binding to chromatin and activity. Finally, the results obtained from these experiments will shed light into the mechanisms of Mettl3 recruitment to target RNAs.

4 Methods

4.1 Methods Project I

See experimental procedures in the published manuscript found in the appendix. Methods not covered in the manuscript are found here.

Cell culture

Dermal fibroblasts from patients affected by Friedreich's Ataxia (FRDA, GM04078) and transformed B-lymphocytes from FRDA-affected (GM15850) and unaffected individuals (GM15851) were obtained from Coriell Cell Repositories (Camden, N.J.). H9 hESCs were obtained from WiCell (Madison, Wis.). GM04078 cells were immortalized by lentiviral delivery of BMI1 and hTERT as described previously (Villasenor et al., 2015), resulting in the cell line FRDA-4078iBT. GM04078 and FRDA-4078iBT human dermal fibroblasts (HDF) were maintained in MEM-α (Life Technologies, Grand Island, N.Y.) media containing 20% heat-inactivated FBS (Geminibio, West Sacramento, Calif.) and 1X L-glutamine (Life Technologies) or in Minimal Essential Medium Eagle (Sigma, St. Luis, Mo.) supplemented with 10% FCS (Sigma) and 2 mM L-glutamine at 37°C in 5% CO_2. Transformed B-lymphocytes were propagated in RPMI1640 medium supplemented with 15% FCS and 2mM L-glutamine at 37°C in 5% CO2. HEK293T cells were cultured in Dulbecco´s Modified Eagle Medium (DMEM) supplemented with 10% FCS (Sigma), 100 U ml^{-1} penicillin, 100 µg ml^{-1} streptomycin and 2 mM L-glutamine at 37°C in 5% CO_2. Human induced pluripotent stem cells (hiPSC) and human embryonic stem cells (hESC) were maintained in mTeSR (Stem Cell Technologies, B.C., Canada) on Growth-Factor-Reduced Matrigel (BD, San Jose, Calif.) culture conditions and were passaged every 4-6 days using dispase (Stem Cell Technologies). Karyotype and fingerprint analyses were performed by Cell Line Genetics (Madison, Wis.).

Genome editing

Plasmids encoding FXN-ZFN-L, FXN-ZFN-R, and donor constructs (sequences available as supplementary material) were co-transfected into FRDA-4078iBT human dermal fibroblasts by Amaxa Nucleofection (Lonza). In brief, cells were passaged 2-3 days before nucleofection. Prior to nucleofection, cells were trypsinized and counted. 1-2 x 10^6 viable cells were then centrifuged, and the supernatant was removed completely. Cells were re-suspended in 100 µl Nucleofector Solution R, mixed with three non-linearized plasmids (two ZFN encoding plasmids + Donor 1 or Donor 2) in a 1:10 ZFN / Donor molar ratio in 10 µl nuclease-free water, and nucleofected using program X-01 (Amaxa, Kit-R). Following nucleofection, sterile plastic pipettes were used to transfer the cells to warm MEM media in 10 cm tissue culture dishes that were already prepared with MEM media. After transfer, cells were cultured in standard conditions for 3 days prior selection with puromycin or G418 (neomycin). After 1-2 weeks of drug selection colonies start to form. Single puromycin-resistant or G418-resistant clones were picked and expanded for further analysis. Sequences of plasmids used in this study are provided as text files. Following clonal isolation and expansion, individual clones were screened by southern blotting and by luciferase assays.

Southern blotting

Southern hybridizations were performed using the Digoxigenin-system (DIG, Roche) for chemiluminescent detection of DNA samples. Genomic DNA was extracted from parental and gene edited cells (FRDA-4078iBT and FRDA-4078iBT cells) using the QIAGEN DNeasy Blood & Tissue Kit (QIAGEN, 69504). 10 µg of extracted genomic DNA were incubated with StuI at 37°C for 16 h. DNA was concentrated after digestion by standard ethanol precipitation followed by resuspension in 20 µl TE-buffer. Digested DNA was separated on a 0.7% agarose gel overnight (~16h) at 20 V. After complete separation of DNA fragments, DNA was denatured by gently shaking the gel in denaturing solution

(2x 15min). Gel was neutralized by gently shaking in neutralization solution (2x 15min). DNA was transferred to a nylon membrane (Roche) by capillary action using absorbent paper to soak transfer solution (20x SSC) through the gel and the membrane (for a detailed description of the protocol see reference (Southern, 2006)). After over-night transfer procedure, DNA was cross-linked to the membrane by exposure to short-wavelength ultraviolet light. The dry membrane was pre-hybridize in 20 ml DIG Easy Hyb solution (Roche, 11796895001) in hybridization bottles for 1-2 h at 42°C. The purified DIG-labeled DNA probe, prepared by PCR with DIG-labeled dNTPs (Roche, 11835289910) was diluted in 10 ml DIG Easy Hyb solution to a final concentration of 25 ng/ml and denatured by heating 68°C for 10 minutes just before use (probe solution can be re-used several times). The pre-hybridization solution was replaced by the freshly denatured probe solution. Hybridization was performed over-night at 42°C in a roller oven. After hybridization, probe solution was removed and membrane was washed twice in 2x wash solution at room temperature. Before proceeding with chemiluminescent DIG detection, membrane was washed twice for 15 minutes each at 68°C in pre-warmed 0.5x wash solution. Next, membrane was washed for 5 min in 50 ml washing buffer by gentle shaking. The membrane was incubated then for 30 min in 50 ml blocking buffer with gentle shaking. The blocking buffer was replaced by 20 ml fresh blocking buffer containing 1 µl of anti-DIG-AP conjugate (1:20000; Roche, 11093274910). The membrane was washed 3 times for 15 minutes each in washing buffer at RT. Then the membrane was equilibrated for 2 min in 20 ml of AP-buffer, before incubation with chemiluminescence substrate CDP-Star (Roche, 11685627001) diluted 1:100 in AP-buffer for 5 minutes. The membrane was exposed to an X-ray film for 20 minutes. To strip for re-hybridization, the membrane was washed twice for 10 minutes in 0.2N NaOH 0.1%SDS at 37-42°C and then briefly washed again in 2x SSC. The dry membrane can be stored at 4°C.

Solutions	**Components**
Denaturation solution	0.4N NaOH, 0.6M NaCl
Neutralization solution	0.5M Tris-HCl pH 7.5, 1.5M NaCl
20x SSC	3M NaCl, 0,3M Na-Citrate, pH 7 (autoclave)
DIG Easy Hyb	make from Hyb Granules (Roche, 11796895001)
2x wash solution	2x SSC, 0.1% SDS
0.5x wash solution	0.5x SSC, 0.1% SDS
MS-buffer	100mM maleic acid, 150mM NaCl; pH 7.5 adjusted with NaOH
Washing buffer	MS-buffer + 0.3% Tween-20
Blocking buffer	MS-buffer with 1% (w/v) Roche blocking reagent (11096176001) and autoclave
AP-buffer	100mM Tris-HCl pH 9.5, 100mM NaCl

Table 2. Solutions And Buffers For Southern Blotting

Luciferase assays

FRDA-4078iBT human dermal fibroblasts were washed with PBS and trypsinized before preparation of cell lysates. The trypzined cells were washed at least twice with PBS to remove residual trypsin. Cell pellets were re-suspended in 100 µl of 1x passive lysis buffer (PLB, Promega Dual-Luciferase Reporter Assay System E1910) supplemented with 1x cOmplete Protease Inhibitor Cocktail (Roche) and cell lysates were transferred to Eppendorf tubes. Tubes were incubated at 25°C for 15 min and rocked several at 1000 rpm to ensure complete lysis. Cell lysates were centrifuged at maximum speed in an Eppendorf microcentrifuge for 10 min at 4°C. Luciferase measurements were performed with 10 µl of lysate supernatants in triplicate. All measurements were performed on a Centro LB 960 luminometer (Berthold Technologies, Germany). Bradford assays were carried out to determine total protein concentration, which

was used for normalization. Cell lysate can be stored at minus 80°C for later use.

Immunoflorescent staining

hiPSCs were seeded on a matrigel-coated black-clear bottom imaging plate (BD). Two days after plating, cells were washed once with PBS and fixed in 4% formaldehyde (USB Corp, Cleveland, Ohio) for 20 min. Cells were permeabilized and incubated for 30 min in blocking buffer containing 3% Donkey Serum (Jackson Labs, Bar Harbor, Maine) and 0.1% Triton (Sigma) in PBS. Cells were stained with primary antibodies Oct4/Nanog (Abcam, Cambridge, Mass./Santa Cruz Biotech, Santa Cruz, Calif.) and dye-conjugated antibodies SSEA-4 and Tra-1-81/1-60 (BD) overnight at 4˚C in blocking buffer. Cells were washed in 1% Tween (Sigma) in PBS and stained with secondary antibodies for 1 h at room temperature. Images were captured using a Zeiss Confocal Microscope (Thornwood, N.Y.).

Pyrosequencing

Cells were collected and outsourced to EpigenDX (Worcester, Mass.). Pyrosequencing analysis was performed according to standard procedures with primers that were developed for the CpG sites at positions (-50; +96) from the start codon of OCT4 (Assay ID: ADS510) gene.

Real-time RT-PCR

For quantitative RT-PCR (qRT-PCR) experiments, total RNA was extracted from FRDA-4078iBT, and transformed B-lymphocytes cells with the Absolutely RNA Microprep Kit (Stratagene). A 1-µg aliquot of total RNA was reverse transcribed with AffinityScript enzyme (Stratagene) using random oligo dT primers according to the manufacturer's instructions. qRT-PCR was performed on a CFX96 Real-Time PCR System (Bio-Rad) using the SsoAdvanced SYBR Green Supermix (Bio-Rad, #172-5264). Relative RNA

levels were calculated from C_T values according to the ΔC_T method and normalized to RPL13A mRNA levels. Primer sequences are provided in Table 3.

Reprogrammed cells were washed once and total RNA isolated using TRIzol (Life Technologies). After chloroform extraction, RNA was purified using a RNeasy Plus Kit (Qiagen, Germantown, Md.). RNA purity and concentration were verified using an Aligent 2100 Bioanalyzer (Santa Clara, Calif.). In a two-step procedure, 1.5 μg of total RNA input was first used to synthesize cDNA with random primers with a High-Capacity cDNA Reverse Transcription Kit (Life Technologies) according to the manufacturer's instructions. Q-PCR analysis was performed on a ViiA-7 Real Time PCR System (Life Technologies) using Fast Advanced SYBR Green PCR master mix (Life Technologies). PCR conditions were as follows: 95˚C – 5 min; (60˚C – 30 s; x 40 cycles); 74˚C – 10 min.

Sendai virus reprogramming and isolation

Sendai virus (SeV) particles encoding OCT4, SOX2, KLF4 and cMYC (OSKM) were purchased from Life Technologies and SeV-reprogramming was performed as follows: patient and unaffected HDFs were seeded on Matrigel-coated 24-well plates at a density of 25-50 x 10^3 per well. Cells were transduced with SeV particles encoding the 4-factors (MOI; O=10, SKM=7.5) overnight in fibroblast media containing 10 μg/ml polybrene (Sigma). The medium was changed to mTeSR 3 days post-infection. On days 5 or 6, cells were dissociated using accutase (Life Technologies) and plated onto Matrigel-coated 35-mm dishes seeded with mitomycin-C-treated human neonatal fibroblast feeders (GlobalStem, Rockville, Md.). Emerging hiPSC colonies were observed 20-30 days post-infection, selected and transferred to matrigel and mTeSR culture conditions.

Tri-layer differentiation and teratoma formation

Endoderm: Accutase was used for cell dissociation to single cells, which were then plated on matrigel in mTeSR with rock-inhibitor Y27632 (Calbiochem, Darmstadt, Germany) at a density of 100 x 10^3 per cm^2. Three days post-plating,

the medium was changed to RPMI containing 0.5% FBS (Thermo, Waltham, Mass.) and 100 ng/ml Activin A (R&D, Minneapolis, Minn.). Cells were washed once, replenished with fresh media daily, fixed on day 5-6, and stained with Sox17 (Abcam) for analysis.

Mesoderm: hiPSC/hESC cultures were grown to confluence. Using dispase (Stem Cell Technologies), cells were collected and plated into non-adherent 6-well culture dishes for embryoid body (EB) formation in media containing BMP4, G-CSF (R&D) and hSCF, hFlt3, and IL-3/6 (Peprotech, Rocky Hill, N.J.). Media were changed every 3 days. Cells were collected or re-plated on MG-coated 96-well plates for fixation at day 18 and then stained with α-Actinin (Abcam) for analysis.

Ectoderm and neural derivatives: Undifferentiated hESC/hiPSC cultures were first exposed to dual inhibition of SMAD signaling, which has been shown to convert hESC to neural progenitors (Chambers et al., 2009). For neuronal differentiation, neural progenitors were expanded in DMEM/F-12 with B27 and N2 supplements (Gibco), 10 ng/ml human epidermal growth factor (hEGF) and 10 ng/ml human basic fibroblast growth factor (hbFGF) (Gibco). Neural progenitors were finally differentiated into neurons in neurobasal medium (Gibco) supplemented with B27, 10 ng/ml human neurotrophin-3 (hNT-3) (R & D), and 10 ng/ml human brain-derived neurotrophic factor (hBDNF) (R & D) for at least 4 weeks. Antibodies against Pax6 (DSHB) and β-III Tubulin (Tuj1, Covance) were used to identify neural progenitors and neurons, respectively. In order to quantify the efficiency of differentiation into Pax6-expressing neural progenitors, immunostained cultures were generated and the percentage of Pax6-positive and DAPI-positive cells determined in 10 randomly selected microscopic fields. In order to quantify the efficiency of differentiation into β-III Tubulin-expressing neurons, a similar approach was used with cultures previously treated with cytosine-1-β-D-arabinofuranoside to prevent the expansion of neural progenitors.

Teratoma formation was outsourced to Applied Stem Cells (Menlo Park, Calif.).

Western blotting

Proteins were extracted for 30 min on ice, the lysates were centrifuged at 16,000 × g for 20 min at 4°C, and protein concentration in the supernatant was determined using the Bio-Rad protein assay. For western blotting, 20 µg of protein were resolved on SDS-PAGE. Alternatively, frataxin protein samples were separated on NuPAGE-Novex Bis-Tris 4-12% gradient gels (Invitrogen) in MES buffer at 200 V for 40 min. Semi-dry transfer on polyvinylidene fluoride (PVDF) membrane was performed at 15 V for 15 min, then immediately shifted to 25 V for 20 min. The membranes were blocked for 30 min 1% non-fat dry milk in TBS-0.05% Tween 20 (TBST), and stained with primary antibodies at 4°C overnight. The primary antibodies used for western blotting were mouse anti-FXN (1:500; Life Technologies, clone 18A5DB1), mouse anti-FLAG (1:2,000; Sigma clone M2), rabbit-anti-PRKD1-S916 (1:1,000; Cell Signaling Technology, Antibody #2051), and rat-anti-tubulin (1:5,000; Abcam clone YL1/2). Signal was detected with corresponding HRP-conjugated secondary antibodies and Immobilon Western Chemiluminiscent HRP Substrate (Millipore). For streptavidin staining, membranes were blocked after transfer in 2% BSA in TBST and incubated with streptavidin-HRP (1:25,000, Sigma) for 30 min at room temperature, followed by signal development as above.

Table 3. qRT-PCR and PCR primers used in project I

qRT-PCR primers

Transcript	Forward	Reverse
FXN	ACAAGCAGACGCCAAACAAGCA	ACCCAGTTTTTCCCAGTCCAGTCA
PRKD1	ACGTCCGCGAGATGGCTTGC	TCACTGGCCGCTTTCACCAGC
RPL13A	CTGCCCTTCCTCCATTGTT	CTTTCAAGCAACTTCGGGAG
OCT4	CCTCACTTCACTGCACTGTA	CAGGTTTTCTTTCCCTAGCT
NANOG	TGAACCTCAGCTACAAACAG	TGGTGGTAGGAAGAGTAAG
SOX2	CCCAGCAGACTTCACATGT	CCTCCCATTTCCCTCGTTTT
KLF4	GATGAACTGACCAGGCACTA	GTGGGTCATATCCACTGTCT
GDF3	AAATGTTTGTGTTGCGGTCA	TCTGGCACAGGTGTCTTCAG
REX1	TCGCTGAGCTGAAACAAATG	CCCTTCTTGAAGGTTTACAC
DPPA4	GGAGCCGCCTGCCCTGGAAAATTC	TTTTTCCTGATATTCTATTCCCAT
UTF1	CCGTCGCTGAACACCGCCCTGCTG	CGCGCTGCCCAGAATGAAGCCCAC
SeV	GGATCACTAGGTGATATCGAGC	ACCAGACAAGAGTTTAAGAGATATGTA
SeV-OCT4	CCCGAAAGAGAAAGCGAACCAG	AATGTATCGAAGGTGCTCAA
SeV-SOX2	ATGCACCGCTACGACGTGAGCGC	AATGTATCGAAGGTGCTCAA
SeV-KLF4	TTCCTGCATGCCAGAGGAGCCC	AATGTATCGAAGGTGCTCAA
SeV-cMyc	TAACTGACTAGCAGGCTTGTCG	TCCACATACAGTCCTGGATGATGATG
GAPDH	CCTTCATTGACCTCAACTAC	GGAAGGCCATGCCAGTGAGC
B-Actin	TGAAGTGTGACGTGGACATC	GGAGGAGCAATGATCTTGAT

PCR primers

Transcript	Forward	Reverse
GAA repeats	GGGATTGGTTGCCAGTGCTTAAAAGTT	GATCTAAGGACCATCATGGCCACACTTGCC

Table 3. qRT-PCR and PCR Primers Used In Project I

4.2 Methods Project II

Cell Culture And Transfection

Mouse embryonic fibroblast cell line NIH3T3 from was obtained from ATCC (clone CRL-1658). NIH3T3 cells were cultured in Dulbecco´s Modified Eagle Medium (DMEM) supplemented with 10% FBS (Sigma), 100 U ml^{-1} penicillin, 100 µg ml^{-1} streptomycin and 2 mM L-glutamine at 37°C in 5% CO_2. Purified plasmids were transfected into murine NIH3T3 cells and treated with blasticidin 48 hr post-trasfection for selection. All transfections were carried out using Lipofectamine 3000 reagent (Invitro- gen) at 3 ml/1 mg DNA ratio in OptiMEM medium (Invitrogen).

Mammalian Expression Vector Cloning

To construct mammalian expression vectors, parental plasmid pEF1a-Blasticidin (pEF1a_B) was a gift from Dr. Matyas Flemr. A gBLOCK encoding for the 3xFLAG-AviTag sequence was cloned into the HindIII site of pEF1a_B. Integrated DNA Technologies synthesized the gBLOCK as linear dsDNA and its sequence is listed in Table 3. The resulting plasmid pEF1a-3xFLAG-AviTag was cut with XbaI to introduce mouse Ago1, Masva1, EGFP-Ago1, and EGFP-Masva1 coding sequences by In-Fusion cloning (In-Fusion HD Cloning Plus Kit, Clontech Laboratories). All primers used for In-Fusion cloning are listed in Table 4. Plasmids containing coding sequences were obtained from MF. All plasmids were verified by sequencing.

Fluorescence Microscopy

Living cells expressing EGFP-Ago1 fusions were imaged in FlouroBrite DMEM medium (Invitrogen) on an Olympus iX-53 inverted micro- scope using the Olympus U-FGFP filter cube.

Argonaute Immunoprecipitation

Immunoprecipitation of FLAG-Ago1 and FLAG-Masva1 was performed as described previously (de la Mata 2015) with Anti-FLAG M2 magnetic beads (Sigma, M8823). Briefly, beads were washed twice with TBS buffer (50 mM Tris HCl, 150 mM NaCl, pH 7.4). For each IP, two 15-cm plates with cells grown to 80-90% confluency were used. Cells were washed once with cold PBS, pelleted and lyzed in 500 µl of lysis buffer [50 mM Tris–HCl pH 7.5, 150mM NaCl, 1% (v/v) TRITON X-100, 1 mM EDTA, containing protease inhibitors (cOmplete, EDTA-free Protease Inhibitor Cocktail, Roche) and RNAse inhibitor (Invitrogen)]. The lysates were incubated 30 min on ice, cleared by centrifugation at 16,000g for 10 min at 4°C and mixed with the washed beads. After 2 h of rotation at 4°C, the beads were washed three times with TBS buffer. As a control for the IPs, parental NIH3T3 cells were used. FLAG-Ago expression and immunoprecipitation efficiency were determined by western blot. RNA was extracted by adding Trizol reagent (Invitrogen) directly on the beads or FLAG-peptide elutions.

Generation Of Small RNA Libraries For High-Throughput Sequencing

RNA purity and concentration were verified using an Aligent 2100 Bioanalyzer (Santa Clara, Calif.). ~17–30 nt small RNAs were PAGE purified. ALL small RNA libraries were prepared using Illumina TruSeq Small RNA kits (Cat # RS-200-0024).

Small RNA Data Analysis

Individual reads were assigned to their sample based on the first four nucleotides containing the barcode. The 3′ adaptor was removed by aligning it to the read allowing one or two mismatches in prefix alignments of at least seven or ten bases, respectively. Low-complexity reads were filtered out based on their dinucleotide entropy (removing <1% of the reads). All reads that were shorter than 14 nucleotides were removed. Alignments to the *Mus musculus* genome

(GRCm38/ mm10) were performed using the FMI galaxy pipeline. miRNA expression levels were calculated and normalized using mouse miRBase (http://www.mirbase.org/) as a reference.

Homology Modeling

The amino acid sequence of the *M. musculus* Masva1 was submitted to the HHPRED server for homology detection and structure prediction (Soding et al., 2005). The structure of *H. sapiens* Ago2 in complex with miR-20a (PDB ID 4F3T) was among the top hits of similar structures yielding a very high score (Score=1647, E-value=7.2e-182, 75% sequence identity). The HHPRED alignment between HsAgo2 and Masva1 was then sent to the HHPRED MODELLER pipeline for homology modelling of the mouse protein using the human Ago2 structure as template. Structural figures were prepared using PyMOL (www.pymol.org).

Table 4. qRT-PCR primers and taqman probes used in project II

qRT-PCR primers and Taqman probes		
Transcript	**Forward**	**Reverse**
Ago1 p1	AGCCGCCTGATGAAGAATG	TTGGTTGGGCGTAGCAAT
Ago1 p2	ATTTGGAATCAAAGTGAAGGATGAC	TCCCATTGTAGAACTGTTTCCC
Masva1 p1	AGCCGCCTGATGAAGAATG	CCAGGTGAGCCTCAGTAGA
Masva1 p2	TAGCTACAACCTGGATCCCTAC	GGTTCTAGGGATAGGCAGAGG
Taqman Ago1 probe1	5HEX/TGGCAATGA/ZEN/GAGGGTTGACTTTGAGG/3IABkFQ	
Ago1 probe1 primers	GCCTGTGTATGATGGAAAGAAGA	GCC ACT TGA TGG AGA CCT TAA A
Taqman Ago1 probe2	5HEX/AGTATGGCG/ZEN/GCCGGAACCG/3IABkFQ	
Ago1 probe2 primers	AGGATGACATGACGGAGGT	TCCCATTGTAGAACTGTTTCCC
Taqman Masva1 probe1	56-FAM/CAGTATGGC/ZEN/GGCCGGGCTC/3IABkFQ	
Masva1 probe1 primers	TTTGGAATCAAAGTGAAGGATGAC	GCAATGGCCCGGTTCTA

Table 4. qRT-PCR Primers and Taqman Probes Used In Project II

4.3 Methods Projects III-IV

Similar methods were used for projects II – IV. Therefore, those methods are covered together in this section.

TALEN Assembly

TALENs were assembled using the Golden Gate TALEN cloning kit (Cermak et al., 2011) (Addgene 1000000024) and previously described acceptor vectors SV40-ELD and SV40-KKR carrying obligate heterodimeric FokI nuclease domains (Flemr et al., 2013).

The following repeat variable diresidues were assembled for individual TALENs:
Dicer-Nterm_left_ELD (targeted sequence GCTTTGCAGCCCCTCAGCAT)
NN HD NG NG NG NN HD NI NN HD HD HD HD NG HD NI NN HD NI NG
Dicer-Nterm_right_KKR (targeted sequence GAGGAAGCAGGGGTCAT)
NN NI NN NN NI NI NN HD NI NN NN NN NN NG HD NI NG

SpCas9-2A-mCherry Cloning

To generate the SpCas9-2A-mCherry construct, we mutated the BbsI recognition site on pX330-U6-Chimeric_BB-CBh-hSpCas9 doing mutagenesis with In-Fusion® HD Cloning Plus Kit (Clontech Laboratories). pX330-U6-Chimeric_BB-CBh-hSpCas9 was a gift from Feng Zhang (Addgene plasmid # 42230). To perform mutagenesis with In-Fusion, PCR primers were designed so that they have a 15 bp overlap with each other at their 5' ends and incorporate a BsaI recognition site. pX330-U6-Chimeric_BB-CBh-hSpCas9 was PCR amplified (CloneAmpTM HiFi PCR Premix, Clontech Laboratories) using the forward primer CGGAGACCGAGGTCT CTGTTTTAGAGCTAGAAATAGCAAG and the reverse primer AGACCTCGGTCTCCG GTGTTTCGTCCTTTCCACAAGATA following the manufacturer´s instructions

(In-Fusion Mutagenesis Protocol). The amplicon was treated with the In-Fusion HD enzyme premix and transformed into DH5alpha competent cells. The resulting intermediate plasmid was digested with EcoRI and the full mCherry coding sequence was then PCR amplified using the forward primers GCCGGCCAGGCAAAAAAGAAAAAGctcgagATGgtgagcaagggcgagg and the reverse primer GATCAGCGAGCTCTAGGAATTtatcgattaCTTGTACAGCTCGTCCATG, and integrated using In-Fusion cloning kit resulting in a Cas9-XhoI-mCherry fusion plasmid. The 2A self cleavage peptide was obtained by annealing two primers: tcgagGGCAGCGGAGAgggcagaggaagtcttctaacatgcggtgacgtggaggagaatcccggcc & tcgaggccgggattctcctccacgtcaccgcatgttagaagacttcctctgcccTCTCCGCTGCCc, and then cloned into the XhoI site to obtain pX330-U6-Chimeric_BsaI-CBh-hSpCas9-2A-mCherry, termed SpCas9-2A-mCherry for convenience.

To clone individual Cas9-sgRNA sequences into SpCas9-2A-mCherry plasmid, the corresponding oligos (listed in Table 3) were annealed in annealing buffer (10 mM Tris-HCl [pH 7.5], 0.1 M NaCl, 1 mM EDTA) and cloned into BsaI cut SpCas9-2A-mCherry.

Recombination Reporter Cloning

Cloning of pRR-EGFP recombination reporter has been previously described (Flemr and Buhler, 2015). To clone individual TALEN-recognition and Cas9-recognisition sequences into pRR-EGFP recombination reporter, the corresponding oligos (listed in Table 3) were annealed in annealing buffer (10 mM Tris-HCl [pH 7.5], 0.1 M NaCl, 1 mM EDTA) and cloned into SacI/AatII cut pRR-EGFP.

Cell Culture and Transfection

Mouse embryonic stem cells were cultured on gelatin-coated dishes in mES medium (DMEM (Gibco 21969-035), supplemented with 15% fetal bovine serum (Biosera), 1 x non-essential amino acids (Gibco), 1 mM sodium pyruvate

(Gibco), 2 mM L-glutamine (Gibco), 0.1 mM 2-mercaptoethanol (Sigma), 50 mg/ml penicillin, 80 mg/ml streptomycin and home-made LIF conditioned medium) at 37°C in 5% CO_2.

For endogenous gene tagging using TALENs, Rosa26:BirA-V5-expressing cells (RosaB; (Flemr and Buhler, 2015)) were transfected with 400 ng of each TALEN, 200 ng pRR-EGFP-Target_site, and 1 mg Target_Gene-FLAG-AViTag ssODN. The ssODNs were synthesized as Ultramers by Integrated DNA Technologies and their sequences are listed in Table 5.

For endogenous gene tagging using SpCas9-2A-mCherry, Rosa26:BirA-V5-expressing cells (RosaB) were transfected with 500 ng of SpCas9-sgRNA-2A-mCherry, 500 ng pRR-EGFP-Target_site, and 1 mg Target_Gene-FLAG-AViTag ssODN. The ssODNs were synthesized as Ultramers by Integrated DNA Technologies and their sequences are listed in Table 5.

All transfections were carried out using Lipofectamine 3000 reagent (Invitro- gen) at 3 ml/1 mg DNA ratio in OptiMEM medium (Invitrogen).

Cell Sorting

Gene edited cells were trypsinized 48-72 hr post-transfection, re-suspended in PBS, and filtered through a 45-μm filter. EGFP-positive cells were sorted on a BD FACSAria III cell sorter (Becton Dickinson) using a 488-nm laser and a 530/30-nm filter. mCherry-positive cells were sorted using a 561-nm laser and a 610/20nm filter. EGFP/mCherry-positive and EGFP cells were collected as single cells into gelatin-coated 96-well plates containing pre-warmed mES medium for clonal expansion.

ES Cell Clone Selection and PCR Genotyping

Sorted EGFP-positive cells transfected with pRR-EGFP were sorted as single cells into gelatin-coated, mES medium containing 96-well plates for clonal expansion. After 7 days of culture, individual clones were picked and expanded

in 96-well plates for 2 more days. Cells were then washed with PBS and lysed directly in 96-well plates with 20 ml lysis buffer (10 mM Tris-HCl [pH 8.0], 0.5 mM EDTA, 0.5% Triton X-100, and 0.5 mg/ml Proteinase K) for 2 hr at 55°C in a humidified box. Proteinase K was inactivated by incubating the lysates at 95°C for 10 min and 0.5 ml of the lysate was directly used for PCR. Genotyping was done in a single 35 cycle run in 6-µl reactions using QIAGEN Fast Cycling PCR Kit (QFC). FLAG-AviTag integrations into targeted genes were screened with universal primers FLAvi_F or FLAvi_R binding on the integration and a locus specific primer binding 0.5 - 1 kb distant of the target site. The region spanning FLAG-AviTag integrations was amplified with locus specific primers yielding PCR amplicons of ~ 500 bp. Primer sequences are listed in Table 5.

Tandem Affinity Purification Of Protein Complexes

A detailed bench protocol can be found in the appendix.

Mass Spectrometry Data

Data can be found in the attached SD card.

Table 4. guide RNAs and ssODN used in project III & IV

guide RNAs

Target	Gene ID	guide RNA (forward primer)	guide RNA (reverse primer)
Ago1	ENSMUSG00000041530	caccgTGTCCTCCGCACGGGTATAT	aaacATATACCCGTGCGGAGGACAc
Dgcr8	ENSMUSG00000022718	caccgAGTCTTAACTGCTCATAATA	aaacTATTATGAGCAGTTAAGACTc
Mettl3	ENSMUSG00000022160	caccgGAGCTAGGATGTCGGACACG	aaacCGTGTCCGACATCCTAGCTCc

ssODN

Target	sequence
Ago1	cctcagtctcccggccgcctgtcctccgcacgggtatatgtcCCACCATGGACTACAAAGACGATGACGATAAAGGTGATGGGTCCGGCCTGAACGACATCTTCGAGGCTCAGAAAATCGAATGGCACGAAGGCGCGCCGAGCTCGAGGGAAGCGGGACCCTCGGGAGCAGgtaagggccctcgggggactgg
Dgcr8	ATGTAAGGTGGTTTAAAACTCTGGTCTTTTAAAGTAGTCTTAACTGCTCCCACCATGGACTACAAAGACGATGACGATAAAGGTGATGGGTCCGGCCTGAACGACATCTTCGAGGCTCAGAAAATCGAATGGCACGAAGGCGCGCCGAGCTCGAGGGAGACATATGAGAGTCCCTCTCCTCTCCCGCGTGAGCCCGCAG
Mettl3	GCTAGTCCGCCCGCCCTGATTCGAGCCGAGTCCGCGCTGGGAGCCCACCATGGACTACAAAGACGATGACGATAAAGGTGATGGGTCCGGCCTGAACGACATCTTCGAGGCTCAGAAAATCGAATGGCACGAAGGCGCGCCGAGCTCGAGGTCGGACACGTGGAGCTCTATCCAGCCCATAAGAAACAGCTGGACTCGCT

Table 5. Guide RNAs and ssODNs Used In Project III & IV

Acknowledgements

I am humbled and excited knowing that every piece of this work has been shaped by my colleagues – all of them great scientists to whom I owe my deepest gratitude.

First and foremost, thank you to my principal supervisor and advisor, Marc Bühler. I remember meeting him in a conference in Vienna where I ask him for an opportunity to join his lab as a summer student. I found him to be a vibrant young scientist, excited about everything around RNA and heterochromatin. His encouragement and enthusiastic support from the initial to the final level enabled me to gain a deeper understanding of science in general. Thank you so much Marc, it was a great pleasure to conduct my doctoral thesis under your guidance. Second, thank you to all people with whom I directly collaborated: the first project would not have been possible without the careful groundwork of Tanel Punga. He has made available his support in a number of ways and always provided constructive suggestions during the project. I specially thank Loren Miraglia and Tony Orth for setting up several screens and for their easy-going collaborative spirit. Tony further supported my applications for a postdoctoral training with a beautiful letter of recommendation that reminded me about what science should be about – thank you Tony. Arnaud Lacoste and his team were crucial for the generation of human iPSCs and drug-treatments in neurons – thank you. For the last projects, I am enormously grateful to my colleagues Matyas Flemr and Philip Knuckles who were essential for these projects, they were always generous and provided their best scientific advice to me. I am indebted to the FMI technology platforms, in particular the protein analysis platform for processing the mass spec samples – thank you Daniel Hess and Jan Seebacher. I am grateful to Lukas Burger and Hans-Rudolf Hotz from the bioinformatics facility who provided indispensable support in data-analysis. Thank you to Heinz Gut for structure modelling and, furthermore, I

thank Hubertus Kohler from the flow cytometry facility for his continuous help in sorting cells for genome editing.

Furthermore, I am grateful to all the people who helped me to solve the small and big problems that one encounters during the scientific day life and shared reagents: I would specially like to thank Yukiko Shimada, Nathalie Laschet, Veronika Ostapcuk, Alex Tuck, Stephan Duss, Lauren Monovich, Arnaud Krebs, Tuncay Baubec, Manuel de la Mata, Uli Rass and Jörg Betschinger. My warm thanks go to all present and past members from the Bühler lab for the nice working atmosphere and inspiring time. I thank several present and past members from the Friedrich Miescher Institute (not mentioned here) for providing an outstanding research environment. Marc Bühler and Mike Abanto proofread this thesis and gave helpful editorial suggestions. Finally I thank the other members of my thesis committee, Helge Grosshans and René Ketting, for taking the time to fulfil their scientific advisory role. René had to come for the meetings from Mainz to Basel – thank you René for taking the time.

G. Rodrigo Villaseñor Molina

Basel, November 2015

References

Alarcon, C.R., Lee, H., Goodarzi, H., Halberg, N., and Tavazoie, S.F. (2015). N6-methyladenosine marks primary microRNAs for processing. Nature *519*, 482-485.

Ariyoshi, M., and Schwabe, J.W. (2003). A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling. Genes Dev *17*, 1909-1920.

Bae, K.H., Kwon, Y.D., Shin, H.C., Hwang, M.S., Ryu, E.H., Park, K.S., Yang, H.Y., Lee, D.K., Lee, Y., Park, J.*, et al.* (2003). Human zinc fingers as building blocks in the construction of artificial transcription factors. Nat Biotechnol *21*, 275-280.

Bahler, J., Wu, J.Q., Longtine, M.S., Shah, N.G., McKenzie, A., 3rd, Steever, A.B., Wach, A., Philippsen, P., and Pringle, J.R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. Yeast *14*, 943-951.

Baker, M. (2012). Gene-editing nucleases. Nat Methods *9*, 23-26.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F., and Cullin, C. (1993). A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. Nucleic Acids Res *21*, 3329-3330.

Beerli, R.R., Dreier, B., and Barbas, C.F., 3rd (2000a). Positive and negative regulation of endogenous genes by designed transcription factors. Proc Natl Acad Sci U S A *97*, 1495-1500.

Beerli, R.R., Schopfer, U., Dreier, B., and Barbas, C.F., 3rd (2000b). Chemically regulated zinc finger transcription factors. J Biol Chem *275*, 32617-32627.

Beerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F., 3rd (1998). Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc Natl Acad Sci U S A *95*, 14628-14633.

Belfort, M., and Bonocora, R.P. (2014). Homing endonucleases: from genetic anomalies to programmable genomic clippers. Methods in molecular biology *1123*, 1-26.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Bibikova, M., Beumer, K., Trautman, J.K., and Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. Science *300*, 764.

Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G., and Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol *21*, 289-297.

Bibikova, M., Golic, M., Golic, K.G., and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics *161*, 1169-1175.

Bidichandani, S.I., Ashizawa, T., and Patel, P.I. (1998). The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. Am J Hum Genet *62*, 111-121.

Bitinaite, J., Wah, D.A., Aggarwal, A.K., and Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A *95*, 10570-10575.

Blencowe, B.J. (2006). Alternative splicing: new insights from global analyses. Cell *126*, 37-47.

Bloom, K., Ely, A., Mussolino, C., Cathomen, T., and Arbuthnot, P. (2013). Inactivation of hepatitis B virus replication in cultured cells and in vivo with engineered transcription activator-like effector nucleases. Molecular therapy : the journal of the American Society of Gene Therapy *21*, 1889-1897.

Boch, J., and Bonas, U. (2010). Xanthomonas AvrBs3 family-type III effectors: discovery and function. Annual review of phytopathology *48*, 419-436.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509-1512.

Bokar, J.A., Rath-Shambaugh, M.E., Ludwiczak, R., Narayan, P., and Rottman, F. (1994). Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. J Biol Chem *269*, 17697-17704.

Bokar, J.A., Shambaugh, M.E., Polayes, D., Matera, A.G., and Rottman, F.M. (1997). Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. Rna *3*, 1233-1247.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551-2561.

Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol Cell *44*, 120-133.

Briggs, A.W., Rios, X., Chari, R., Yang, L., Zhang, F., Mali, P., and Church, G.M. (2012). Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. Nucleic Acids Res *40*, e117.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

Bühler, M., and Moazed, D. (2007). Transcription and RNAi in heterochromatic gene silencing. Nat Struct Mol Biol *14*, 1041-1048.

Buljan, M., Chalancon, G., Eustermann, S., Wagner, G.P., Fuxreiter, M., Bateman, A., and Babu, M.M. (2012). Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. Mol Cell *46*, 871-883.

Campuzano, V., Montermini, L., Lutz, Y., Cova, L., Hindelang, C., Jiralerspong, S., Trottier, Y., Kish, S.J., Faucheux, B., Trouillas, P.*, et al.* (1997). Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. Hum Mol Genet *6*, 1771-1780.

Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A.*, et al.* (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science *271*, 1423-1427.

Capecchi, M.R. (2001). Generating mice with targeted mutations. Nature medicine *7*, 1086-1090.

Capecchi, M.R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat Rev Genet *6*, 507-512.

Carroll, D. (2011). Genome engineering with zinc-finger nucleases. Genetics *188*, 773-782.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res *39*, e82.

Chamberlain, S., Shaw, J., Rowland, A., Wallis, J., South, S., Nakamura, Y., von Gabain, A., Farrall, M., and Williamson, R. (1988). Mapping of mutation causing Friedreich's ataxia to human chromosome 9. Nature *334*, 248-250.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol *27*, 275-280.

Chapman, J.R., Taylor, M.R., and Boulton, S.J. (2012). Playing the end game: DNA double-strand break repair pathway choice. Mol Cell *47*, 497-510.

Chekulaeva, M., Filipowicz, W., and Parker, R. (2009). Multiple independent domains of dGW182 function in miRNA-mediated repression in Drosophila. Rna *15*, 794-803.

Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat Struct Mol Biol *18*, 1218-1226.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S.*, et al.* (2013a). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479-1491.

Chen, C., Fenk, L.A., and de Bono, M. (2013b). Efficient genome editing in Caenorhabditis elegans by CRISPR-targeted homologous recombination. Nucleic Acids Res *41*, e193.

Chen, F., Pruett-Miller, S.M., Huang, Y., Gjoka, M., Duda, K., Taunton, J., Collingwood, T.N., Frodin, M., and Davis, G.D. (2011). High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat Methods *8*, 753-755.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol *10*, 741-754.

Chendrimada, T.P., Finn, K.J., Ji, X., Baillat, D., Gregory, R.I., Liebhaber, S.A., Pasquinelli, A.E., and Shiekhattar, R. (2007). MicroRNA silencing through RISC recruitment of eIF6. Nature *447*, 823-828.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436*, 740-744.

Chevalier, B.S., Kortemme, T., Chadsey, M.S., Baker, D., Monnat, R.J., and Stoddard, B.L. (2002). Design, activity, and structure of a highly specific artificial endonuclease. Mol Cell *10*, 895-905.

Chevalier, B.S., and Stoddard, B.L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res *29*, 3757-3774.

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol *31*, 230-232.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome research *24*, 132-141.

Choi, H., Larsen, B., Lin, Z.Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.C., and Nesvizhskii, A.I. (2011). SAINT: probabilistic scoring of affinity purification-mass spectrometry data. Nat Methods *8*, 70-73.

Chou, C.J., Herman, D., and Gottesfeld, J.M. (2008). Pimelic diphenylamide 106 is a slow, tight-binding inhibitor of class I histone deacetylases. J Biol Chem *283*, 35402-35409.

Choulika, A., Perrin, A., Dujon, B., and Nicolas, J.F. (1995). Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. Mol Cell Biol *15*, 1968-1973.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kuhn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol *33*, 543-548.

Clancy, M.J., Shambaugh, M.E., Timpte, C.S., and Bokar, J.A. (2002). Induction of sporulation in Saccharomyces cerevisiae leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. Nucleic Acids Res *30*, 4509-4518.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A*., et al.* (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science.

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Cornu, T.I., Thibodeau-Beganny, S., Guhl, E., Alwin, S., Eichtinger, M., Joung, J.K., and Cathomen, T. (2008). DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. Molecular therapy : the journal of the American Society of Gene Therapy *16*, 352-358.

Cox, D.B., Platt, R.J., and Zhang, F. (2015). Therapeutic genome editing: prospects and challenges. Nature medicine *21*, 121-131.

Dalvai, M., Loehr, J., Jacquet, K., Huard, C.C., Roques, C., Herst, P., Cote, J., and Doyon, Y. (2015). A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. Cell Rep.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231-235.

Dervan, P.B., and Edelson, B.S. (2003). Recognition of the DNA minor groove by pyrrole-imidazole polyamides. Current opinion in structural biology *13*, 284-299.

Desrosiers, R., Friderici, K., and Rottman, F. (1974). Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. Proc Natl Acad Sci U S A *71*, 3971-3975.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. Journal of bacteriology *190*, 1390-1400.

Ding, Q., Strong, A., Patel, K.M., Ng, S.L., Gosis, B.S., Regan, S.N., Cowan, C.A., Rader, D.J., and Musunuru, K. (2014). Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing. Circulation research *115*, 488-492.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature *330*, 576-578.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M.*, et al.* (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201-206.

Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M.H., and Chandrasegaran, S. (2005). Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucleic Acids Res *33*, 5978-5990.

Durr, A., Cossee, M., Agid, Y., Campuzano, V., Mignard, C., Penet, C., Mandel, J.L., Brice, A., and Koenig, M. (1996). Clinical and genetic abnormalities in patients with Friedreich's ataxia. The New England journal of medicine *335*, 1169-1175.

Elkayam, E., Kuhn, C.D., Tocilj, A., Haase, A.D., Greene, E.M., Hannon, G.J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. Cell *150*, 100-110.

Ellis, J.D., Barrios-Rodiles, M., Colak, R., Irimia, M., Kim, T., Calarco, J.A., Wang, X., Pan, Q., O'Hanlon, D., Kim, P.M.*, et al.* (2012). Tissue-specific alternative splicing remodels protein-protein interaction networks. Mol Cell *46*, 884-892.

Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008a). Getting to the root of miRNA-mediated gene silencing. Cell *132*, 9-14.

Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008b). GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. Nat Struct Mol Biol *15*, 346-353.

Evans, M.J. (2001). The cultural mouse. Nature medicine *7*, 1081-1083.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A*., et al.* (2009). Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. Mol Cell *35*, 868-880.

Ferrando, I.M., Chaerkady, R., Zhong, J., Molina, H., Jacob, H.K., Herbst-Robinson, K., Dancy, B.M., Katju, V., Bose, R., Zhang, J*., et al.* (2012). Identification of targets of c-Src tyrosine kinase by chemical complementation and phosphoproteomics. Mol Cell Proteomics *11*, 355-369.

Fielitz, J., Kim, M.S., Shelton, J.M., Qi, X., Hill, J.A., Richardson, J.A., Bassel-Duby, R., and Olson, E.N. (2008). Requirement of protein kinase D1 for pathological cardiac remodeling. Proc Natl Acad Sci U S A *105*, 3059-3063.

Figueroa, A., Kotani, H., Toda, Y., Mazan-Mamczarz, K., Mueller, E.C., Otto, A., Disch, L., Norman, M., Ramdasi, R.M., Keshtgar, M*., et al.* (2009). Novel roles of hakai in cell proliferation and oncogenesis. Mol Biol Cell *20*, 3533-3542.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Flemr, M., and Buhler, M. (2015). Single-Step Generation of Conditional Knockout Mouse Embryonic Stem Cells. Cell Rep *12*, 709-716.

Flemr, M., Malik, R., Franke, V., Nejepinska, J., Sedlacek, R., Vlahovicek, K., and Svoboda, P. (2013). A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. Cell *155*, 807-816.

Frank, F., Hauver, J., Sonenberg, N., and Nagar, B. (2012). Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. The EMBO journal *31*, 3588-3595.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat Biotechnol *33*, 179-186.

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m(6)A RNA methylation. Nat Rev Genet *15*, 293-306.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol *31*, 822-826.

Fujita, Y., Krause, G., Scheffner, M., Zechner, D., Leddy, H.E., Behrens, J., Sommer, T., and Birchmeier, W. (2002). Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex. Nat Cell Biol *4*, 222-231.

Fuller-Pace, F.V. (2006). DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. Nucleic Acids Res *34*, 4206-4215.

Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in biotechnology *31*, 397-405.

Gamber, G.G., Meredith, E., Zhu, Q., Yan, W., Rao, C., Capparelli, M., Burgis, R., Enyedy, I., Zhang, J.H., Soldermann, N.*, et al.* (2011). 3,5-diarylazoles as novel and selective inhibitors of protein kinase D. Bioorganic & medicinal chemistry letters *21*, 1447-1451.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67-71.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A *109*, E2579-2586.

Geiger, T., Velic, A., Macek, B., Lundberg, E., Kampf, C., Nagaraj, N., Uhlen, M., Cox, J., and Mann, M. (2013). Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse. Mol Cell Proteomics *12*, 1709-1722.

Genovese, P., Schiroli, G., Escobar, G., Di Tomaso, T., Firrito, C., Calabria, A., Moi, D., Mazzieri, R., Bonini, C., Holmes, M.C.*, et al.* (2014). Targeted genome editing in human repopulating haematopoietic stem cells. Nature *510*, 235-240.

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A.A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y.S.*, et al.* (2015). Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. Science *347*, 1002-1006.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A.*, et al.* (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442-451.

Gingras, A.C., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol *8*, 645-654.

Gottesfeld, J.M. (2007). Small molecules affecting transcription in Friedreich ataxia. Pharmacol Ther *116*, 236-248.

Gregersen, L.H., Schueler, M., Munschauer, M., Mastrobuoni, G., Chen, W., Kempa, S., Dieterich, C., and Landthaler, M. (2014). MOV10 Is a 5' to 3' RNA helicase contributing to UPF1 mRNA target degradation by translocation along 3' UTRs. Mol Cell *54*, 573-585.

Gregory, R.I., Chendrimada, T.P., Cooch, N., and Shiekhattar, R. (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell *123*, 631-640.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235-240.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Wahrisch, S., Beisaw, A., Macura, K., Blass, G., Kellis, M., Werber, M.*, et al.* (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev Cell *24*, 206-214.

Guilinger, J.P., Pattanayak, V., Reyon, D., Tsai, S.Q., Sander, J.D., Joung, J.K., and Liu, D.R. (2014). Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. Nat Methods *11*, 429-435.

Haase, A.D., Jaskiewicz, L., Zhang, H., Laine, S., Sack, R., Gatignol, A., and Filipowicz, W. (2005). TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. EMBO Rep *6*, 961-967.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS computational biology *1*, e60.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293-296.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev *18*, 3016-3027.

Handel, E.M., Alwin, S., and Cathomen, T. (2009). Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. Molecular therapy : the journal of the American Society of Gene Therapy *17*, 104-111.

Herman, D., Jenssen, K., Burnett, R., Soragni, E., Perlman, S.L., and Gottesfeld, J.M. (2006). Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. Nat Chem Biol *2*, 551-558.

Heyer, W.D., Ehmsen, K.T., and Liu, J. (2010). Regulation of homologous recombination in eukaryotes. Annu Rev Genet *44*, 113-139.

Hick, A., Wattenhofer-Donze, M., Chintawar, S., Tropel, P., Simard, J.P., Vaucamps, N., Gall, D., Lambot, L., Andre, C., Reutenauer, L.*, et al.* (2013). Neurons and cardiomyocytes derived from induced pluripotent stem cells as a model for mitochondrial defects in Friedreich's ataxia. Disease models & mechanisms *6*, 608-621.

Hilton, I.B., and Gersbach, C.A. (2015). Enabling functional genomics with genome engineering. Genome research *25*, 1442-1455.

Hock, J., Weinmann, L., Ender, C., Rudel, S., Kremmer, E., Raabe, M., Urlaub, H., and Meister, G. (2007). Proteomic and functional analysis of Argonaute-containing mRNA-protein complexes in human cells. EMBO Rep *8*, 1052-1060.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C.*, et al.* (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol *29*, 731-734.

Hongay, C.F., and Orr-Weaver, T.L. (2011). Drosophila Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. Proc Natl Acad Sci U S A *108*, 14855-14860.

Horiuchi, K., Kawamura, T., Iwanari, H., Ohashi, R., Naito, M., Kodama, T., and Hamakubo, T. (2013). Identification of Wilms' tumor 1-associating protein complex and its role in alternative splicing and the cell cycle. J Biol Chem *288*, 33292-33302.

Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. Journal of bacteriology *190*, 1401-1412.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O.*, et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol *31*, 827-832.

Hu, W., Kaminski, R., Yang, F., Zhang, Y., Cosentino, L., Li, F., Luo, B., Alvarez-Carbonell, D., Garcia-Mesa, Y., Karn, J.*, et al.* (2014). RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. Proc Natl Acad Sci U S A *111*, 11461-11466.

Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S., and Zhang, B. (2011). Heritable gene targeting in zebrafish using customized TALENs. Nat Biotechnol *29*, 699-700.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol *31*, 227-229.

International Human Genome Sequencing, C. (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931-945.

Irimia, M., Weatheritt, R.J., Ellis, J.D., Parikshak, N.N., Gonatopoulos-Pournatzis, T., Babor, M., Quesnel-Vallieres, M., Tapial, J., Raj, B., O'Hanlon, D.*, et al.* (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell *159*, 1511-1523.

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., and Tomari, Y. (2010). Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol Cell *39*, 292-299.

Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Molecular microbiology *43*, 1565-1575.

Jia, G., Yang, C.G., Yang, S., Jian, X., Yi, C., Zhou, Z., and He, C. (2008). Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. FEBS Lett *582*, 3313-3319.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. eLife *2*, e00471.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S.*, et al.* (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science *343*, 1247997.

Johnston, M., Geoffroy, M.C., Sobala, A., Hay, R., and Hutvagner, G. (2010). HSP90 protein stabilizes unloaded argonaute complexes and microscopic P-bodies in human cells. Mol Biol Cell *21*, 1462-1469.

Joung, J.K., and Sander, J.D. (2013). TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol *14*, 49-55.

Kaihatsu, K., Janowski, B.A., and Corey, D.R. (2004). Recognition of chromosomal DNA by PNAs. Chemistry & biology *11*, 749-758.

Karanam, K., Kafri, R., Loewer, A., and Lahav, G. (2012). Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. Mol Cell *47*, 320-329.

Kennedy, E.M., Bassit, L.C., Mueller, H., Kornepati, A.V., Bogerd, H.P., Nie, T., Chatterjee, P., Javanbakht, H., Schinazi, R.F., and Cullen, B.R. (2015). Suppression of hepatitis B virus DNA accumulation in chronically infected cells using a bacterial CRISPR/Cas RNA-guided DNA endonuclease. Virology *476*, 196-205.

Kennedy, E.M., Kornepati, A.V., Goldstein, M., Bogerd, H.P., Poling, B.C., Whisnant, A.W., Kastan, M.B., and Cullen, B.R. (2014). Inactivation of the human papillomavirus E6 or E7 gene in cervical carcinoma cells by using a bacterial CRISPR/Cas RNA-guided endonuclease. Journal of virology *88*, 11965-11972.

Ketting, R.F. (2011). The many faces of RNAi. Dev Cell *20*, 148-161.

Khanna, K.K., and Jackson, S.P. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. Nat Genet *27*, 247-254.

Kim, D.H., and Rossi, J.J. (2007). Strategies for silencing human disease using RNA interference. Nat Rev Genet *8*, 173-184.

Kim, H., and Kim, J.S. (2014). A guide to genome engineering with programmable nucleases. Nat Rev Genet *15*, 321-334.

Kim, H., Um, E., Cho, S.R., Jung, C., Kim, H., and Kim, J.S. (2011). Surrogate reporters for enrichment of cells with nuclease-induced mutations. Nat Methods *8*, 941-943.

Kim, J., Cantor, A.B., Orkin, S.H., and Wang, J. (2009). Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. Nature protocols *4*, 506-517.

Kim, M.S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S.*, et al.* (2014a). A draft map of the human proteome. Nature *509*, 575-581.

Kim, T.K., and Eberwine, J.H. (2010). Mammalian cell transfection: the present and the future. Analytical and bioanalytical chemistry *397*, 3173-3178.

Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E.*, et al.* (2013a). A library of TAL effector nucleases spanning the human genome. Nat Biotechnol *31*, 251-258.

Kim, Y., Phan, D., van Rooij, E., Wang, D.Z., McAnally, J., Qi, X., Richardson, J.A., Hill, J.A., Bassel-Duby, R., and Olson, E.N. (2008). The MEF2D transcription factor mediates stress-dependent cardiac remodeling in mice. The Journal of clinical investigation *118*, 124-132.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A *93*, 1156-1160.

Kim, Y.G., Smith, J., Durgesha, M., and Chandrasegaran, S. (1998). Chimeric restriction enzyme: Gal4 fusion to FokI cleavage domain. Biological chemistry *379*, 489-495.

Kim, Y.H., Ramakrishna, S., Kim, H., and Kim, J.S. (2014b). Enrichment of cells with TALEN-induced mutations using surrogate reporters. Methods *69*, 108-117.

Kim, Y.K., Wee, G., Park, J., Kim, J., Baek, D., Kim, J.S., and Kim, V.N. (2013b). TALEN-based knockout library for human microRNAs. Nat Struct Mol Biol *20*, 1458-1464.

Kok, K.H., Ng, M.H., Ching, Y.P., and Jin, D.Y. (2007). Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. J Biol Chem *282*, 17649-17657.

Konermann, S., Brigham, M.D., Trevino, A.E., Hsu, P.D., Heidenreich, M., Cong, L., Platt, R.J., Scott, D.A., Church, G.M., and Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. Nature *500*, 472-476.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H.*, et al.* (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583-588.

Ku, S., Soragni, E., Campau, E., Thomas, E.A., Altun, G., Laurent, L.C., Loring, J.F., Napierala, M., and Gottesfeld, J.M. (2010). Friedreich's ataxia induced pluripotent stem cells model intergenerational GAATTC triplet repeat instability. Cell stem cell *7*, 631-637.

Kuan, J.Y., and Glazer, P.M. (2004). Targeted gene modification using triplex-forming oligonucleotides. Methods in molecular biology *262*, 173-194.

Kuehn, M.R., Bradley, A., Robertson, E.J., and Evans, M.J. (1987). A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. Nature *326*, 295-298.

Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat Biotechnol *32*, 677-683.

Labuda, M., Labuda, D., Miranda, C., Poirier, J., Soong, B.W., Barucha, N.E., and Pandolfo, M. (2000). Unique origin and specific ethnic distribution of the Friedreich ataxia GAA expansion. Neurology *54*, 2322-2324.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P.Y., Soll, S.J., Dinic, L., Ojo, T., Hafner, M., Zavolan, M., and Tuschl, T. (2008). Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. Rna *14*, 2580-2596.

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. Current biology : CB *14*, 2162-2167.

Ledford, H. (2015). CRISPR, the disruptor. Nature *522*, 20-24.

Lee, M., Kim, B., and Kim, V.N. (2014). Emerging roles of RNA modification: m(6)A and U-tail. Cell *158*, 980-987.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S.*, et al.* (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419.

Lee, Y., Hur, I., Park, S.Y., Kim, Y.K., Suh, M.R., and Kim, V.N. (2006). The role of PACT in the RNA silencing pathway. EMBO J *25*, 522-532.

Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J *21*, 4663-4670.

Li, H., Haurigot, V., Doyon, Y., Li, T., Wong, S.Y., Bhagwat, A.S., Malani, N., Anguela, X.M., Sharma, R., Ivanciu, L.*, et al.* (2011). In vivo genome editing restores haemostasis in a mouse model of haemophilia. Nature *475*, 217-221.

Li, Y., Polak, U., Bhalla, A.D., Rozwadowska, N., Butler, J.S., Lynch, D.R., Dent, S.Y., and Napierala, M. (2015). Excision of Expanded GAA Repeats Alleviates the Molecular Phenotype of Friedreich's Ataxia. Molecular therapy : the journal of the American Society of Gene Therapy *23*, 1055-1065.

Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., Lv, J., Xie, X., Chen, Y., Li, Y.*, et al.* (2015). CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. Protein & cell *6*, 363-372.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annual review of biochemistry *79*, 181-211.

Lin, S.R., Yang, H.C., Kuo, Y.T., Liu, C.J., Yang, T.Y., Sung, K.C., Lin, Y.Y., Wang, H.Y., Wang, C.C., Shen, Y.C.*, et al.* (2014). The CRISPR/Cas9 System Facilitates Clearance of the Intrahepatic HBV Templates In Vivo. Molecular therapy Nucleic acids *3*, e186.

Lindtner, S., Zolotukhin, A.S., Uranishi, H., Bear, J., Kulkarni, V., Smulevitch, S., Samiotaki, M., Panayotou, G., Felber, B.K., and Pavlakis, G.N. (2006). RNA-binding motif protein 15 binds to the RNA transport element RTE and provides a direct link to the NXF1 export pathway. J Biol Chem *281*, 36915-36928.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., and Parker, R. (2005). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. Nat Cell Biol *7*, 719-723.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X.*, et al.* (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol *10*, 93-95.

Lombardo, A., Genovese, P., Beausejour, C.M., Colleoni, S., Lee, Y.L., Kim, K.A., Ando, D., Urnov, F.D., Galli, C., Gregory, P.D.*, et al.* (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nature biotechnology *25*, 1298-1306.

Long, C., McAnally, J.R., Shelton, J.M., Mireault, A.A., Bassel-Duby, R., and Olson, E.N. (2014). Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. Science *345*, 1184-1188.

Lubahn, D.B., Moyer, J.S., Golding, T.S., Couse, J.F., Korach, K.S., and Smithies, O. (1993). Alteration of reproductive function but not prenatal sexual development after insertional disruption of the mouse estrogen receptor gene. Proc Natl Acad Sci U S A *90*, 11162-11166.

Lufino, M.M., Silva, A.M., Nemeth, A.H., Alegre-Abarrategui, J., Russell, A.J., and Wade-Martins, R. (2013). A GAA repeat expansion reporter model of Friedreich's ataxia recapitulates the genomic context and allows rapid screening of therapeutic compounds. Human molecular genetics *22*, 5173-5187.

Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S.A., Zhang, S., and Pederson, T. (2015). Multicolor CRISPR labeling of chromosomal loci in human cells. Proc Natl Acad Sci U S A *112*, 3002-3007.

Maeder, M.L., Angstman, J.F., Richardson, M.E., Linder, S.J., Cascio, V.M., Tsai, S.Q., Ho, Q.H., Sander, J.D., Reyon, D., Bernstein, B.E.*, et al.* (2013). Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat Biotechnol *31*, 1137-1142.

Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A.*, et al.* (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell *31*, 294-301.

Mahiny, A.J., Dewerth, A., Mays, L.E., Alkhaled, M., Mothes, B., Malaeksefat, E., Loretz, B., Rottenberger, J., Brosch, D.M., Reautschnig, P.*, et al.* (2015). In vivo genome editing using nuclease-encoding mRNA corrects SP-B deficiency. Nat Biotechnol *33*, 584-586.

Mak, A.N., Bradley, P., Bogdanove, A.J., and Stoddard, B.L. (2013). TAL effectors: function, structure, engineering and applications. Current opinion in structural biology *23*, 93-99.

Makarova, K.S., Aravind, L., Wolf, Y.I., and Koonin, E.V. (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biology direct *6*, 38.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F.*, et al.* (2011b). Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol *9*, 467-477.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol *31*, 833-838.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., Dicarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-Guided Human Genome Engineering via Cas9. Science.

Maniataki, E., and Mourelatos, Z. (2005). A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. Genes Dev *19*, 2979-2990.

Mansour, S.L., Thomas, K.R., and Capecchi, M.R. (1988). Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature *336*, 348-352.

Marmolino, D., Acquaviva, F., Pinelli, M., Monticelli, A., Castaldo, I., Filla, A., and Cocozza, S. (2009). PPAR-gamma agonist Azelaoyl PAF increases frataxin protein and mRNA expression: new implications for the Friedreich's ataxia therapy. Cerebellum *8*, 98-103.

Marraffini, L.A., and Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature *463*, 568-571.

Martelli, A., Napierala, M., and Puccio, H. (2012a). Understanding the genetic and molecular pathogenesis of Friedreich's ataxia through animal and cellular models. Disease models & mechanisms *5*, 165-176.

Martelli, A., Napierala, M., and Puccio, H. (2012b). Understanding the genetic and molecular pathogenesis of Friedreich's ataxia through animal and cellular models. Dis Model Mech *5*, 165-176.

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563-574.

Martinez, N.J., and Gregory, R.I. (2013). Argonaute2 expression is post-transcriptionally coupled to microRNA abundance. Rna *19*, 605-612.

Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh, H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat Biotechnol *33*, 538-542.

Matranga, C., and Zamore, P.D. (2007). Small silencing RNAs. Current biology : CB *17*, R789-793.

McHugh, C.A., Chen, C.K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A.*, et al.* (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232-236.

McVey, M., and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends in genetics : TIG *24*, 529-538.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. Nat Rev Genet *14*, 447-459.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell *15*, 185-197.

Meister, G., Landthaler, M., Peters, L., Chen, P.Y., Urlaub, H., Luhrmann, R., and Tuschl, T. (2005). Identification of novel argonaute-associated proteins. Current biology : CB *15*, 2149-2155.

Meredith, E.L., Ardayfio, O., Beattie, K., Dobler, M.R., Enyedy, I., Gaul, C., Hosagrahara, V., Jewell, C., Koch, K., Lee, W.*, et al.* (2010a). Identification of orally available naphthyridine protein kinase D inhibitors. Journal of medicinal chemistry *53*, 5400-5421.

Meredith, E.L., Beattie, K., Burgis, R., Capparelli, M., Chapo, J., Dipietro, L., Gamber, G., Enyedy, I., Hood, D.B., Hosagrahara, V.*, et al.* (2010b). Identification of potent and selective amidobipyridyl inhibitors of protein kinase D. Journal of medicinal chemistry *53*, 5422-5438.

Meyer, K.D., and Jaffrey, S.R. (2014). The dynamic epitranscriptome: N6-methyladenosine and gene expression control. Nat Rev Mol Cell Biol *15*, 313-326.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell *149*, 1635-1646.

Mikami, S., Kanaba, T., Takizawa, N., Kobayashi, A., Maesaki, R., Fujiwara, T., Ito, Y., and Mishima, M. (2014). Structural insights into the recruitment of SMRT by the corepressor SHARP under phosphorylative regulation. Structure *22*, 35-46.

Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A.*, et al.* (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol *25*, 778-785.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J.*, et al.* (2011). A TALE nuclease architecture for efficient genome editing. Nat Biotechnol *29*, 143-148.

Miller, J.C., Zhang, L., Xia, D.F., Campo, J.J., Ankoudinova, I.V., Guschin, D.Y., Babiarz, J.E., Meng, X., Hinkley, S.J., Lam, S.C.*, et al.* (2015). Improved specificity of TALE-based genome editing using an expanded RVD repertoire. Nat Methods *12*, 465-471.

Miranda, C.J., Santos, M.M., Ohshima, K., Smith, J., Li, L., Bunting, M., Cossee, M., Koenig, M., Sequeiros, J., Kaplan, J.*, et al.* (2002). Frataxin knockin mouse. FEBS Lett *512*, 291-297.

Misteli, T. (2007). Beyond the sequence: cellular organization of genome function. Cell *128*, 787-800.

Miyanari, Y., Ziegler-Birling, C., and Torres-Padilla, M.E. (2013). Live visualization of chromatin dynamics with fluorescent TALEs. Nat Struct Mol Biol *20*, 1321-1324.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. Journal of molecular evolution *60*, 174-182.

Mojica, F.J., Diez-Villasenor, C., Soria, E., and Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Molecular microbiology *36*, 244-246.

Montermini, L., Andermann, E., Labuda, M., Richter, A., Pandolfo, M., Cavalcanti, F., Pianese, L., Iodice, L., Farina, G., Monticelli, A.*, et al.* (1997). The Friedreich ataxia GAA triplet repeat: premutation and normal alleles. Hum Mol Genet *6*, 1261-1266.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science *326*, 1501.

Moynahan, M.E., and Jasin, M. (2010). Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. Nat Rev Mol Cell Biol *11*, 196-207.

Mukherjee, M., Chow, S.Y., Yusoff, P., Seetharaman, J., Ng, C., Sinniah, S., Koh, X.W., Asgar, N.F., Li, D., Yim, D.*, et al.* (2012). Structure of a novel phosphotyrosine-binding domain in Hakai that targets E-cadherin. EMBO J *31*, 1308-1319.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935-949.

Niu, Y., Shen, B., Cui, Y., Chen, Y., Wang, J., Wang, L., Kang, Y., Zhao, X., Si, W., Li, W.*, et al.* (2014). Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. Cell *156*, 836-843.

Nolte, R.T., Conlin, R.M., Harrison, S.C., and Brown, R.S. (1998). Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. Proc Natl Acad Sci U S A *95*, 2938-2943.

Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. Nat Struct Mol Biol *21*, 528-534.

Oldmeadow, C., Mossman, D., Evans, T.J., Holliday, E.G., Tooney, P.A., Cairns, M.J., Wu, J., Carr, V., Attia, J.R., and Scott, R.J. (2014). Combined analysis of exon splicing and genome wide polymorphism data predict schizophrenia risk loci. Journal of psychiatric research *52*, 44-49.

Orr-Weaver, T.L., Szostak, J.W., and Rothstein, R.J. (1981). Yeast transformation: a model system for the study of recombination. Proc Natl Acad Sci U S A *78*, 6354-6358.

Ouna, B.A., Stewart, M., Helbig, C., and Clayton, C. (2012). The Trypanosoma brucei CCCH zinc finger proteins ZC3H12 and ZC3H13. Molecular and biochemical parasitology *183*, 184-188.

Ousterout, D.G., Perez-Pinera, P., Thakore, P.I., Kabadi, A.M., Brown, M.T., Qin, X., Fedrigo, O., Mouly, V., Tremblay, J.P., and Gersbach, C.A. (2013). Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. Molecular therapy : the journal of the American Society of Gene Therapy *21*, 1718-1726.

Pandolfo, M. (2009). Friedreich ataxia: the clinical picture. J Neurol *256 Suppl 1*, 3-8.

Parker, J.S., Roe, S.M., and Barford, D. (2004). Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J *23*, 4727-4737.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol *31*, 839-843.

Pavletich, N.P., and Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science *252*, 809-817.

Pece, S., and Gutkind, J.S. (2002). E-cadherin and Hakai: signalling, remodeling or destruction? Nat Cell Biol *4*, E72-74.

Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.L.*, et al.* (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol *26*, 808-816.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W.*, et al.* (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods *10*, 973-976.

Ping, X.L., Sun, B.F., Wang, L., Xiao, W., Yang, X., Wang, W.J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y.S.*, et al.* (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell research *24*, 177-189.

Porteus, M.H., and Carroll, D. (2005). Gene targeting using zinc finger nucleases. Nat Biotechnol *23*, 967-973.

Potthoff, M.J., and Olson, E.N. (2007). MEF2: a central regulator of diverse developmental programs. Development *134*, 4131-4140.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology *151*, 653-663.

Puccio, H. (2009). Multicellular models of Friedreich ataxia. J Neurol *256 Suppl 1*, 18-24.

Punga, T., and Buhler, M. (2010). Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. EMBO Mol Med *2*, 120-129.

Punga, T., and Bühler, M. (2010). Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. EMBO Mol Med *2*, 120-129.

Raffel, G.D., Mercher, T., Shigematsu, H., Williams, I.R., Cullen, D.E., Akashi, K., Bernard, O.A., and Gilliland, D.G. (2007). Ott1(Rbm15) has pleiotropic roles in hematopoietic development. Proc Natl Acad Sci U S A *104*, 6001-6006.

Rai, M., Soragni, E., Jenssen, K., Burnett, R., Herman, D., Coppola, G., Geschwind, D.H., Gottesfeld, J.M., and Pandolfo, M. (2008). HDAC inhibitors correct frataxin deficiency in a Friedreich ataxia mouse model. PLoS One *3*, e1958.

Raj, B., and Blencowe, B.J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. Neuron *87*, 14-27.

Ramakrishna, S., Cho, S.W., Kim, S., Song, M., Gopalappa, R., Kim, J.S., and Kim, H. (2014). Surrogate reporter-based enrichment of cells containing RNA-guided Cas9 nuclease-induced mutations. Nature communications *5*, 3378.

Ramirez, C.L., Foley, J.E., Wright, D.A., Muller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., Townsend, J.A.*, et al.* (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods *5*, 374-375.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature protocols *8*, 2281-2308.

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., and Joung, J.K. (2012). FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol *30*, 460-465.

Robb, G.B., and Rana, T.M. (2007). RNA helicase A interacts with RISC in human cells and functions in RISC loading. Mol Cell *26*, 523-537.

Rong, Y.S., and Golic, K.G. (2000). Gene targeting by homologous recombination in Drosophila. Science *288*, 2013-2018.

Rouet, P., Smih, F., and Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol *14*, 8096-8106.

Rudin, N., and Haber, J.E. (1988). Efficient repair of HO-induced chromosomal breaks in Saccharomyces cerevisiae by recombination between flanking homologous sequences. Mol Cell Biol *8*, 3918-3928.

Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. Nat Rev Mol Cell Biol *14*, 630-642.

Sakamoto, N., Chastain, P.D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J.D., and Wells, R.D. (1999). Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. Mol Cell *3*, 465-475.

Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K., and Yeh, J.R. (2011a). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Nat Biotechnol *29*, 697-698.

Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y., *et al.* (2011b). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods *8*, 67-69.

Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G., and Zhang, F. (2012). A transcription activator-like effector toolbox for genome engineering. Nature protocols *7*, 171-192.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res *39*, 9275-9282.

Saveliev, A., Everett, C., Sharpe, T., Webster, Z., and Festenstein, R. (2003). DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. Nature *422*, 909-913.

Scherer, S., and Davis, R.W. (1979). Replacement of chromosome segments with altered DNA sequences constructed in vitro. Proc Natl Acad Sci U S A *76*, 4951-4955.

Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. Science *336*, 1037-1040.

Schirle, N.T., Sheu-Gruttadauria, J., and MacRae, I.J. (2014). Structural basis for microRNA targeting. Science *346*, 608-613.

Schmid-Burgk, J.L., Schmidt, T., Kaiser, V., Honing, K., and Hornung, V. (2013). A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. Nat Biotechnol *31*, 76-81.

Schulz, J.B., Boesch, S., Burk, K., Durr, A., Giunti, P., Mariotti, C., Pousset, F., Schols, L., Vankan, P., and Pandolfo, M. (2009). Diagnosis and treatment of Friedreich ataxia: a European perspective. Nat Rev Neurol *5*, 222-234.

Schwank, G., Koo, B.K., Sasselli, V., Dekkers, J.F., Heo, I., Demircan, T., Sasaki, N., Boymans, S., Cuppen, E., van der Ent, C.K.*, et al.* (2013). Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. Cell stem cell *13*, 653-658.

Schwartz, S., Mumbach, M.R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G.G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D.*, et al.* (2014). Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. Cell Rep *8*, 284-296.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell *48*, 849-862.

Sen, G.L., and Blau, H.M. (2005). Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. Nat Cell Biol *7*, 633-636.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G.*, et al.* (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84-87.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V.*, et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Shi, Y., Downes, M., Xie, W., Kao, H.Y., Ordentlich, P., Tsai, C.C., Hon, M., and Evans, R.M. (2001). Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. Genes Dev *15*, 1140-1151.

Shimizu, Y., Sollu, C., Meckler, J.F., Adriaenssens, A., Zykovich, A., Cathomen, T., and Segal, D.J. (2011). Adding fingers to an engineered zinc finger nuclease can reduce activity. Biochemistry *50*, 5033-5041.

Shiohama, A., Sasaki, T., Noda, S., Minoshima, S., and Shimizu, N. (2007). Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins. Experimental cell research *313*, 4196-4207.

Sirvent, A., Vigy, O., Orsetti, B., Urbach, S., and Roche, S. (2012). Analysis of SRC oncogenic signaling in colorectal cancer by stable isotope labeling with heavy amino acids in mouse xenografts. Mol Cell Proteomics *11*, 1937-1950.

Smallwood, S.A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet *43*, 811-814.

Smih, F., Rouet, P., Romanienko, P.J., and Jasin, M. (1995). Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. Nucleic Acids Res *23*, 5012-5019.

Smith, J., Bibikova, M., Whitby, F.G., Reddy, A.R., Chandrasegaran, S., and Carroll, D. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res *28*, 3361-3369.

Smithies, O. (2001). Forty years with homologous recombination. Nature medicine *7*, 1083-1086.

Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., and Kucherlapati, R.S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature *317*, 230-234.

Soding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res *33*, W244-248.

Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science *305*, 1434-1437.

Southern, E. (2006). Southern blotting. Nature protocols *1*, 518-525.

Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T.*, et al.* (2015). Structural imprints in vivo decode RNA regulatory mechanisms. Nature *519*, 486-490.

Steinberg, S.F. (2012). Regulation of protein kinase D1 activity. Molecular pharmacology *81*, 284-291.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62-67.

Swarts, D.C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R.F., Koonin, E.V., Patel, D.J., and van der Oost, J. (2014). The evolutionary journey of Argonaute proteins. Nat Struct Mol Biol *21*, 743-753.

Taghian, D.G., and Nickoloff, J.A. (1997). Chromosomal double-strand breaks induce gene conversion at high frequency in mammalian cells. Mol Cell Biol *17*, 6386-6393.

Tesson, L., Usal, C., Menoret, S., Leung, E., Niles, B.J., Remy, S., Santiago, Y., Vincent, A.I., Meng, X., Zhang, L.*, et al.* (2011). Knockout rats generated by embryo microinjection of TALENs. Nat Biotechnol *29*, 695-696.

Thomas, C.E., Ehrhardt, A., and Kay, M.A. (2003). Progress and problems with the use of viral vectors for gene therapy. Nat Rev Genet *4*, 346-358.

Thomas, K.R., Folger, K.R., and Capecchi, M.R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. Cell *44*, 419-428.

Tomari, Y., Du, T., Haley, B., Schwarz, D.S., Bennett, R., Cook, H.A., Koppetsch, B.S., Theurkauf, W.E., and Zamore, P.D. (2004a). RISC assembly defects in the Drosophila RNAi mutant armitage. Cell *116*, 831-841.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. (2004b). A protein sensor for siRNA asymmetry. Science *306*, 1377-1380.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A.*, et al.* (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

Uranishi, H., Zolotukhin, A.S., Lindtner, S., Warming, S., Zhang, G.M., Bear, J., Copeland, N.G., Jenkins, N.A., Pavlakis, G.N., and Felber, B.K. (2009). The RNA-binding motif protein 15B (RBM15B/OTT3) acts as cofactor of the nuclear export receptor NXF1. J Biol Chem *284*, 26106-26116.

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature *435*, 646-651.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nat Rev Genet *11*, 636-646.

Valton, J., Dupuy, A., Daboussi, F., Thomas, S., Marechal, A., Macmaster, R., Melliand, K., Juillerat, A., and Duchateau, P. (2012). Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. J Biol Chem *287*, 38427-38432.

Varjosalo, M., Keskitalo, S., Van Drogen, A., Nurkkala, H., Vichalkovski, A., Aebersold, R., and Gstaiger, M. (2013). The protein interaction landscape of the human CMGC kinase group. Cell Rep *3*, 1306-1320.

Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Cowan, C.A., Talkowski, M.E., and Musunuru, K. (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell stem cell *15*, 27-30.

Villasenor, R., Miraglia, L., Romero, A., Tu, B., Punga, T., Knuckles, P., Duss, S., Orth, T., and Buhler, M. (2015). Genome-Engineering Tools to Establish Accurate Reporter Cell Lines That Enable Identification of Therapeutic Strategies to Treat Friedreich's Ataxia. Journal of biomolecular screening *20*, 760-767.

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.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014a). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80-84.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., *et al.* (2014b). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117-120.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell *161*, 1388-1399.

Wang, Y., Li, Y., Toth, J.I., Petroski, M.D., Zhang, Z., and Zhao, J.C. (2014c). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat Cell Biol *16*, 191-198.

Wefers, B., Meyer, M., Ortiz, O., Hrabe de Angelis, M., Hansen, J., Wurst, W., and Kuhn, R. (2013). Direct production of mouse disease models by embryo microinjection of TALENs and oligodeoxynucleotides. Proc Natl Acad Sci U S A *110*, 3782-3787.

Weinthal, D., Tovkach, A., Zeevi, V., and Tzfira, T. (2010). Genome editing in plant cells by zinc finger nucleases. Trends in plant science *15*, 308-321.

Westra, E.R., Swarts, D.C., Staals, R.H., Jore, M.M., Brouns, S.J., and van der Oost, J. (2012). The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. Annu Rev Genet *46*, 311-339.

Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X.*, et al.* (2011). Targeted genome editing across species using ZFNs and TALENs. Science *333*, 307.

Wu, C.L., and Melton, D.W. (1993). Production of a model for Lesch-Nyhan syndrome in hypoxanthine phosphoribosyltransferase-deficient mice. Nat Genet *3*, 235-240.

Wu, Y., Liang, D., Wang, Y., Bai, M., Tang, W., Bao, S., Yan, Z., Li, D., and Li, J. (2013). Correction of a genetic disease in mouse via use of CRISPR-Cas9. Cell stem cell *13*, 659-662.

Wurtz, N.R., and Dervan, P.B. (2000). Sequence specific alkylation of DNA by hairpin pyrrole-imidazole polyamide conjugates. Chemistry & biology *7*, 153-161.

Ye, L., Wang, J., Beyer, A.I., Teque, F., Cradick, T.J., Qi, Z., Chang, J.C., Bao, G., Muench, M.O., Yu, J.*, et al.* (2014). Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Delta32 mutation confers resistance to HIV infection. Proc Natl Acad Sci U S A *111*, 9591-9596.

Yin, H., Xue, W., Chen, S., Bogorad, R.L., Benedetti, E., Grompe, M., Koteliansky, V., Sharp, P.A., Jacks, T., and Anderson, D.G. (2014). Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. Nat Biotechnol *32*, 551-553.

Yuan, Y.R., Pei, Y., Ma, J.B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.Y., Dauter, Z., Tuschl, T., and Patel, D.J. (2005). Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol Cell *19*, 405-419.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol *29*, 149-153.

Zijlstra, M., Li, E., Sajjadi, F., Subramani, S., and Jaenisch, R. (1989). Germ-line transmission of a disrupted beta 2-microglobulin gene produced by homologous recombination in embryonic stem cells. Nature *342*, 435-438.

Zipprich, J.T., Bhattacharyya, S., Mathys, H., and Filipowicz, W. (2009). Importance of the C-terminal domain of the human GW182 protein TNRC6C for translational repression. Rna *15*, 781-793.

List of Figures and Tables

Abbreviations

Ago	Argonaute
AS	Alternative splicing
bp	base-pair
Cas protein	CRISPR associated protein
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
cDNA	complementary DNA
CMV	Cytomegalievirus
DNA	Deoxyribonucleic Acid
D. melanogaster	*Drosophila melanogaster*
dsRNA	double-stranded RNA
DSB	double strand break
ES cell	Embryonic stem cell
FRDA	Friedreich´s Ataxia
FL	Firefly luciferase
GFP	Green fluorescent protein
gRNA	guide RNA
H3K9	Histone H3 Lysine 9
H3K9me	Histone H3 Lysine 9 methylation
HEK293T	Human Embryonic Kidney cell line 293T
HDR	Homologous directed repair
HR	Homologous recombination
HTS	High-throughput screening
HDAC	Histone deacetylase

HDACi	Histone deacetylase inhibitor
hiPSC	human induced Pluripotent Stem Cells
IP	immuno-precipitation
iPSC	induced Pluripotent Stem Cells
lincRNA	long intergenic non-coding RNA
LMW	low molecular weight
lncRNA	long non-coding RNA
MCS	multiple cloning site
mES cell	mouse Embryonic Stem cell
MMEJ	Microhomology-mediated end-joining
miRNA	micro RNA
mRNA	messenger RNA
miRLC	micro RNA loading complex
MID	MID domain
m6A	N6-methyladenosine
MS	Mass spectrometry
NHEJ	Non-homologous end-joining
nt	nucleotide
ORF	Open Reading Frame
PAM	protospacer-adjacent motif
PAZ	Piwi Argonaut Zwille domain
PCR	Polymerase chain reaction
Poi score	probability of interaction score
P-bodies	Processing bodies
PEV	Position Effect Variegation
PIWI	P-element induced wimpy testis
PolII	RNA Polymerase 2

pre-crRNA	precursor CRISPR RNA
pri-miRNA	primary micro RNA
qRT-PCR	quantitative real-time PCR
RING	Really Interesting New Gene
RISC	RNA-induced silencing complex
RGN	RNA-guided nuclease
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing
RRM	RNA recognition motif
rRNA	Ribosomal RNA
RSA	Redundant siRNA Activity (RSA) analysis method
RVD	repeat-variable di-residue
SAINT	Significance Analysis of INTeractome algorithm
SCID	severe combined immunodeficiency
snoRNA	small nucleolar RNA
sRNA	small RNA
ssODN	single-stranded oligodeoxynucleotides
sgRNA	single RNA-hybrid guide
TAP	tandem affinity purification
TALE	transcription activator-like effector
TALEN	transcription activator-like effector nucleases
TRED	trinucleotide repeat expansion disease
tRNA	transfer RNA
tracrRNA	trans-activating crRNA
ZF	zinc finger
ZFN	zinc finfer nuclease

Appendix

Journal of Biomolecular Screening
2015, Vol. 20(6) 760–767

DOI: 10.1177/1087057114568071
jbx.sagepub.com
SAGE

Genome-Engineering Tools to Establish Accurate Reporter Cell Lines That Enable Identification of Therapeutic Strategies to Treat Friedreich´s Ataxia

Rodrigo Villaseñor[1,2*], Loren Miraglia[3*], Angelica Romero[3], Buu Tu[3], Tanel Punga[1,2†], Philip Knuckles[1,2], Stephan Duss[1,2], Tony Orth[3], and Marc Bühler[1,2]

Abstract

Friedreich's ataxia is a neurodegenerative disease caused by deficiency of the mitochondrial protein frataxin. This deficiency results from expansion of a trinucleotide repeat in the first intron of the frataxin gene. Because this repeat expansion resides in an intron and hence does not alter the amino acid sequence of the frataxin protein, gene reactivation could be of therapeutic benefit. High-throughput screening for frataxin activators has so far met with limited success because current cellular models may not accurately assess endogenous frataxin gene regulation. Here we report the design and validation of genome-engineering tools that enable the generation of human cell lines that express the frataxin gene fused to a luciferase reporter gene from its endogenous locus. Performing a pilot high-throughput genomic screen in a newly established reporter cell line, we uncovered novel negative regulators of frataxin expression. Rational design of small-molecule inhibitors of the identified frataxin repressors and/or high-throughput screening of large siRNA or compound libraries with our system may yield treatments for Friedreich's ataxia.

Keywords

Friedreich's ataxia (FRDA), triplet repeat expansion disorder (TRED), frataxin (FXN) gene, high-throughput screening (HTS), zinc-finger nuclease, genome engineering, drug discovery, PRKD1, SBF1

Introduction

Friedreich's ataxia (FRDA) is an autosomal recessive disease resulting from repression of the iron-binding protein frataxin (FXN).[1] Reduced FXN expression results from an expansion of a trinucleotide repeat in the first intron of the frataxin gene[2] and provokes a spectrum of neuropathological defects in patients with FRDA. FRDA is the most common recessively inherited ataxia disorder across Europe, and several epidemiological studies have estimated the prevalence of FRDA as two to three cases per 100,000 people in Caucasian populations.[2–4] Despite years of research, FXN gene regulation remains poorly understood, and treatment options for FRDA are limited.

Therapeutics that reactivate frataxin expression are expected to be beneficial to FRDA patients.[5] The long non-coding GAA triplet repeats found in intron 1 of the FXN gene in FRDA patients impede transcription elongation. Therefore, overcoming this blockade and reactivating FXN expression is an attractive therapeutic strategy. In addition, approaches that enhance transcription rates or increase FXN protein or messenger RNA (mRNA) stability could also correct the FXN protein deficiency in patients with FRDA.[6–8]

[1]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
[2]University of Basel, Basel, Switzerland
[3]Genomics Institute of the Novartis Research Foundation, San Diego, CA, USA
[*]These authors contributed equally to this work.
[†]Present address: Uppsala University, Department of Medical Biochemistry and Microbiology, BMC Uppsala, Sweden

Received Sep 30, 2014, and in revised form Dec 16, 2014. Accepted for publication Dec 18, 2014.

Supplementary material for this article is available on the *Journal of Biomolecular Screening* Web site at http://jbx.sagepub.com/supplemental.

Corresponding Authors:
Marc Bühler, Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, Basel, 4058, Switzerland.
Email: marc.buehler@fmi.ch

Tony Orth, Genomics Institute of the Novartis Research Foundation, 10675 John J Hopkins Dr., San Diego, CA 92121, USA.
Email: aorth@gnf.org

Cell-based assays are essential for the identification of druggable regulators of FXN gene expression and low molecular weight compounds that alleviate FXN deficiency. Previous attempts at developing high-throughput screening (HTS)–friendly assays have used artificial intronic sequences with extended GAA tracts or cell lines containing FXN reporter gene fusions randomly integrated into the genome.[9–14] A potential drawback of these systems is that they do not assess the activity of the endogenous FXN gene in the native chromosomal context. This may limit the identification of modulators of FXN expression, because the endogenous chromatin structure and/or nuclear organization of the human genome are major determinants of FXN gene activity.[15–18] Thus, more advanced reporter systems that enable the monitoring of gene expression from the endogenous FXN locus are desirable.

Here we report the use of genome-engineering tools to establish a novel drug discovery system that combines genetically accurate reporter systems and high-throughput biology to identify novel therapeutics for Friedreich's ataxia. We used zinc-finger nucleases (ZFNs) to generate a cellular model in which a luciferase reporter gene is introduced into the endogenous FXN locus. Using this system in a high-throughput genomic screen, we discovered novel repressors of FXN gene expression. Using our system in future drug discovery efforts may finally yield treatments for FRDA.

Materials and Methods

Cell Culture

Dermal fibroblasts from patients affected by FRDA (GM04078) were obtained from Coriell Cell Repositories (Camden, NJ). GM04078 cells were immortalized by lentiviral delivery of BMI1 and hTERT as described previously,[19] resulting in the cell line FRDA-4078iBT. GM04078 and FRDA-4078iBT were maintained in Eagle's minimal essential medium (Sigma, St. Louis, MO) supplemented with 10% non–heat-inactivated fetal bovine serum (FBS; Sigma) and 1× L-glutamine (Life Technologies, Carlsbad, CA) at 37 °C in 5% CO_2. HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated FBS (Sigma), 100 U/mL^{-1} penicillin, 100 μg/mL^{-1} streptomycin, and 1× L-glutamine (Life Technologies) at 37 °C in 5% CO_2.

Genome Editing

Plasmids encoding FXN-ZFN-L, FXN-ZFN-R, and donor constructs (sequences available as supplementary material) were cotransfected into HEK293T cells with the polyethylenimine (PEI) transfection method. Briefly, exponentially growing cells were trypsinized and seeded in a 10-mm tissue culture dish 1 day before transfection. A solution containing 60 μL of 1 mg/mL PEI and 15 μg DNA (two ZFN plasmids + donor 1) in a 1:5 molar ratio (ZFNs/donor) was prepared in 1 mL of sterile DMEM lacking serum. The solution was briefly vortexed and left for 15 min at room temperature before pipetting dropwise to adherent HEK293 cells. Cells were cultured under standard conditions for 2 to 3 days, followed by culturing the cells in the presence of puromycin-containing growth media for 2 weeks. Single puromycin-resistant clones were picked and expanded for further analysis. Sequences of plasmids used in this study are provided as text files.

Luciferase Assays

Luciferase assays were performed with the Promega (Madison, WI) Dual-Luciferase Reporter Assay system according to the manufacturer's protocol. Briefly, lysis buffer was added directly to confluent cell monolayers. Culture dishes were incubated at 25 °C for 15 min and rocked several times to ensure complete lysis. Cell lysates were centrifuged at maximum speed in an Eppendorf microcentrifuge for 10 min at 4 °C. Luciferase measurements were performed with lysate supernatants in duplicate. All measurements were performed on a Centro LB 960 luminometer (Berthold Technologies, Bad Wildbad, Germany). Bradford assays were carried out to determine total protein concentration, which was used for normalization.

Real-Time RT-PCR

For quantitative reverse transcriptase PCR (qRT-PCR) experiments, total RNA was extracted from GM04078, FRDA-4078iBT, and HEK293T-FF2AP cell lines with the Absolutely RNA Microprep Kit (Stratagene, La Jolla, CA). A 1-μg aliquot of total RNA was reverse transcribed with an AffinityScript enzyme (Stratagene) using the oligo(dT) primer according to the manufacturer's instructions. qRT-PCR was performed on a CFX96 Real-Time PCR System (Bio-Rad, Hercules, CA) using the SsoAdvanced SYBR Green Supermix (#172-5264; Bio-Rad). Relative RNA levels were calculated from C_T values according to the ΔC_T method and normalized to GAPDH mRNA levels. Primer sequences are provided in **Supplemental Table S3**.

RNA Interference Screen

A synthetic small interfering RNA (siRNA) library targeting the druggable genome (4835 genes with 9670 constructs or two constructs per gene) was screened in a duplicate 384-well plate format with two gene-targeting siRNAs per well to identify genes that upregulate frataxin-luciferase fusion expression. The library was reverse transfected into the HEK293T-FF2AP reporter cell line via high-throughput transfection[20,21] using a fully integrated genomic robotic

system (GNF Systems, San Diego, CA). Each plate contained transfection controls targeting cell viability (AllStars Cell Death; Qiagen, Valencia, CA), as well as a nontargeting (AllStars Negative Control; Qiagen) and a frataxin-targeting siRNA (Frataxin-12; Qiagen). Data from each plate were normalized to the plate mean and processed for fold activity. Duplicate readings for each well were geometrically averaged to generate a single activity score. A nonlinear transformation was applied to remove plate variation.[22] Putative hit wells were rescreened as single siRNAs per well for validation. In addition to the first-round hit siRNAs, all siRNA constructs targeting genes of interest from other available libraries (Qiagen; Integrated DNA Technologies, Coralville, IA) were included in the validation screen. An additional 96 nontargeting controls were spotted into each 384-well plate of the validation assay (Qiagen). Data analysis included evaluation of individual and multiple-well RNA interference (RNAi) activities for each gene (Redundant siRNA Activity Analysis).[22]

Short Hairpin RNA Delivery

Short hairpin RNA (shRNA) constructs were delivered to FRDA-4078iBT cells by lentiviral transduction. Lentivirus was produced by calcium phosphate transfection of HEK293T cells.[23] All infections were performed at a multiplicity of infection of approximately 30 viral particles per cell. Cells were harvested 6 days after transduction for total RNA isolation. Lipofection was used to deliver shRNA constructs into HEK293T-FF2AP cells; ~200 cells per mm^2 were seeded 1 day before transfection. Cells were transfected with plasmids using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. shRNA sequences are provided in **Supplemental Table S3**.

Western Blotting

Protein samples were separated on NuPAGE-Novex Bis-Tris 4% to 12% gradient gels (Invitrogen) in MES buffer at 200 V for 40 min. Transfer to Immobilon-P PVDF membrane (Millipore, Billerica, MA) was performed at 25 V (constant) for 45 min in Towbin transfer buffer with 5% methanol and 0.05% sodium dodecyl sulfate (SDS). The membranes were blocked in 5% nonfat dry milk in tris buffered saline with 0.1% Tween-20 (TBST) for at least 30 min at room temperature and subsequently washed once in TBST for 5 to 10 min. The membranes were incubated with the anti-FXN antibody (1:1000, SKU 456300; Invitrogen) in 5% nonfat dry milk in TBST overnight at 4 °C. Membranes were washed in TBST and incubated with secondary antibody (anti–mouse horseradish peroxidase [HRP], 1:10,000) in 5% nonfat dry milk in TBST for 45 to 60 min at room temperature. Blots were developed with Immobilon Western HRP Substrate (Millipore).

Results and Discussion

We aimed at generating genome-editing tools that will enable the FRDA research community to knock in any given sequence tag into the FXN locus in different cell types. To this end, we chose a ZFN-mediated genome-editing approach to tag the 3′ end of the protein-coding region of the endogenous FXN gene with a firefly luciferase (FL) reporter gene.[24,25] To induce a double-strand break approximately 1 kb downstream of the 5′ splice site in the last intron of the FXN gene, we designed a pair of ZFNs that bind to the FXN locus uniquely (**Fig. 1A**). For homologous recombination, we created a transgenic donor DNA fragment encoding the last FXN exon fused to the open reading frame of FL (**Fig. 1A**). To allow clonal selection after targeted insertion, we created donor constructs encoding either puromycin N-acetyl-transferase (Puro) preceded by 2A self-cleaving peptide linked to the FXN-FL fusion (donor 1) or an autonomous neomycin (Neo) selection cassette (donor 2). The left homology arm of the donor constructs was designed to introduce a functional branchpoint and 3′ splice sites to favor efficient splicing between exon 4 and inserted exon 5 (**Fig. 1A**).

To provide proof of concept that the ZFN approach is working and to establish a novel reporter cell line that is suitable for high-throughput genomic screening, we used the ZFN system to introduce the luciferase reporter gene into the FXN locus of human embryonic kidney (HEK) 293T cells. This cell line was selected because it can be expanded to a very large scale at a relatively low cost. As a consequence, our FXN reporter system should be compatible with the technical and financial constraints of most academic and industrial screening platforms. We cotransfected human HEK 293T cells with expression plasmids encoding the two frataxin-ZFNs (ZFN-L and ZFN-R) and a plasmid encoding donor 1 (**Fig. 1A**), which confers puromycin resistance only after correct integration, proper pre-mRNA splicing, and translation of the FXN-FL-2A-PURO (FF2AP) fusion protein (**Fig. 1B**). At 2 to 3 weeks posttransfection, puromycin-resistant colonies were isolated and subjected to Western blot analysis. This revealed specific expression of a 75-kDa FXN-FL fusion protein in puromycin-resistant cells (hereafter referred to as HEK293T-FF2AP; **Fig. 1C**). We also observed a FXN fusion protein of higher molecular weight than the mature FXN-FL fusion protein, demonstrating efficient cotranslational self-cleavage of the 2A peptide, release of the Puro protein, and mitochondrial processing of the FXN-FL fusion protein. Importantly, we detected firefly luciferase activity in lysates from HEK293T-FF2AP but not from parental HEK293T cells (**Fig. 1D**), an activity that could be abrogated upon expression of an shRNA targeting exon 2 of the FXN mRNA (**Fig. 1E**). Thus, we concluded that the firefly luciferase activity in HEK293T-FF2AP cells came solely from the targeted integration at the FXN locus

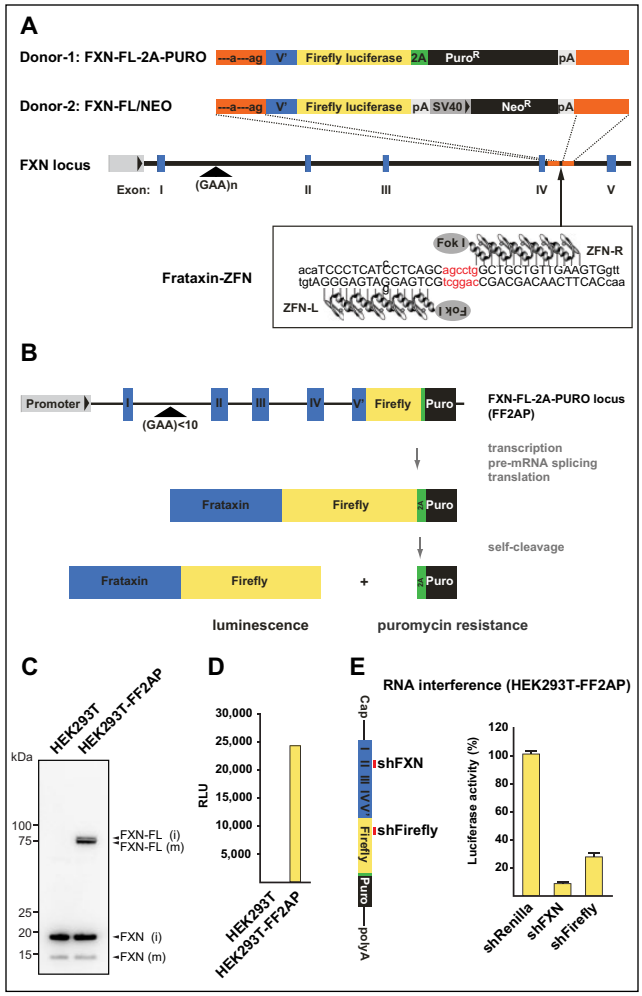


Figure 1. Generation of the HEK293T-FF2AP cell line. (**A**) Scheme of the zinc-finger nuclease (ZFN)–mediated genome-editing strategy to tag the 3′ end of the protein-coding region of the endogenous frataxin (FXN) gene with a firefly luciferase (FL) reporter gene. Each frataxin-ZFN contains the cleavage domain of FokI linked to an array of five zinc fingers that have been designed to specifically recognize sequences (black uppercase letters) flanking the cleavage site (red letters) in intron 4 of the FXN gene. The double-strand break induced by the frataxin-ZFN pair permits site-specific integration of a transgenic donor DNA fragment (donor 1 or 2) encoding the last FXN exon fused to the open reading frame of firefly luciferase (V′-firefly luciferase). Both donor constructs are flanked by regions homologous to the insertion site (orange) to enable homology-directed repair. To ensure correct splicing of the FXN messenger RNA (mRNA) after successful genome editing, the left homology arms were designed to introduce functional branchpoint and 3′ splice sites (—a—ag). Donor 1 encodes puromycin N-acetyl-transferase (Puro) preceded by a 2A self-cleaving peptide linked to the V′-firefly luciferase fusion. Using donor 1, puromycin resistance is expected only after correct integration into the FXN locus and transcription thereof. Donor 2 contains an autonomous neomycin (Neo) selection cassette driven by a SV40 promoter. This makes neomycin resistance independent of the transcriptional status of the FXN locus but might increase the rate of false-positive integration events. The location of the GAA triplet repeat is indicated by a black arrowhead. pA, cleavage and polyadenylation site. (**B**) Schematic representation of the FXN locus after successful genomic insertion of donor 1. Cotranslational self-cleavage of the 2A peptide sequence releases the FXN-FL fusion protein and the puromycin N-acetyl-transferase, which confers resistance against puromycin. (**C**) Western blot showing specific expression of the FXN-FL fusion protein (~75 kDa) in HEK293T-FF2AP cells. The blot was probed with an antibody recognizing FXN. Intermediate (i) and mature (m) forms of FXN-FL and endogenous FXN protein expressed from nontagged alleles are indicated. (**D**) Firefly luciferase activity assayed from nontargeted HEK293T and HEK293T-FF2AP cells. RLU, relative light units. (**E**) RNA interference (RNAi) experiment demonstrating that firefly luciferase activity reliably reflects FXN expression in HEK293T-FF2AP cells. Cells were transfected with a nontargeting short hairpin RNA (shRNA) (shRenilla) and shRNAs targeting either FXN or firefly mRNA (red bars). Firefly luciferase signals were normalized to total protein and shown relative to the nontargeting control. Error bars represent the standard deviation.

and, thus, reliably reports expression of the FXN gene from its endogenous location.

In contrast to previous work,[5] the HEK293T-FF2AP cell line allowed us to screen for modulators of FXN expression in its natural genomic context and in an HTS format. This cell line is easy to transfect and hence well suited for RNAi or complementary DNA (cDNA) overexpression screens. Therefore, to identify potential repressors of FXN expression, we performed an RNAi screen targeting 4835 human genes in the HEK293T-FF2AP cell line, using luciferase signal amplification as the readout (**Fig. 2A**). Two unique siRNAs were assayed in a pooled format (two siRNAs per well) at 72 h posttransfection (**Fig. 2B**). Due to the pooled construct format, we chose a baseline of 1.6-fold luciferase activation over the mean as a cutoff to capture phenotypes generated from wells in which only one siRNA may be active. Maximal luciferase expression was 5.1-fold over baseline (targeting hypothetical gene 1000292277), with an additional five siRNAs scoring between 4- and 5-fold. Minimal expressions were obtained with siRNAs targeting Qiagen AllStars Cell Death control (0.06-fold) and frataxin (0.09-fold), indicating that the siRNA transfection was successful and that the relationship between frataxin and luciferase expression was intact. In subsequent validation screens, positive hit siRNAs were spotted in an arrayed well format of one siRNA per well and screened in duplicate under identical conditions. Using the HEK293T-FF2AP cell line, we performed a reconfirmation screen of the primary screen hits (**Fig. 2C**). In parallel, we screened positive hit siRNAs for non-FXN-specific activators of luciferase activity with an HEK 293T cell line stably expressing firefly luciferase from a cytomegalovirus (CMV) promoter (HEK293-pGF1-CMV) (**Fig. 2D**). This allowed for thorough analysis of independent constructs via redundant siRNA analysis (RSA),[22] which enriched for genes where

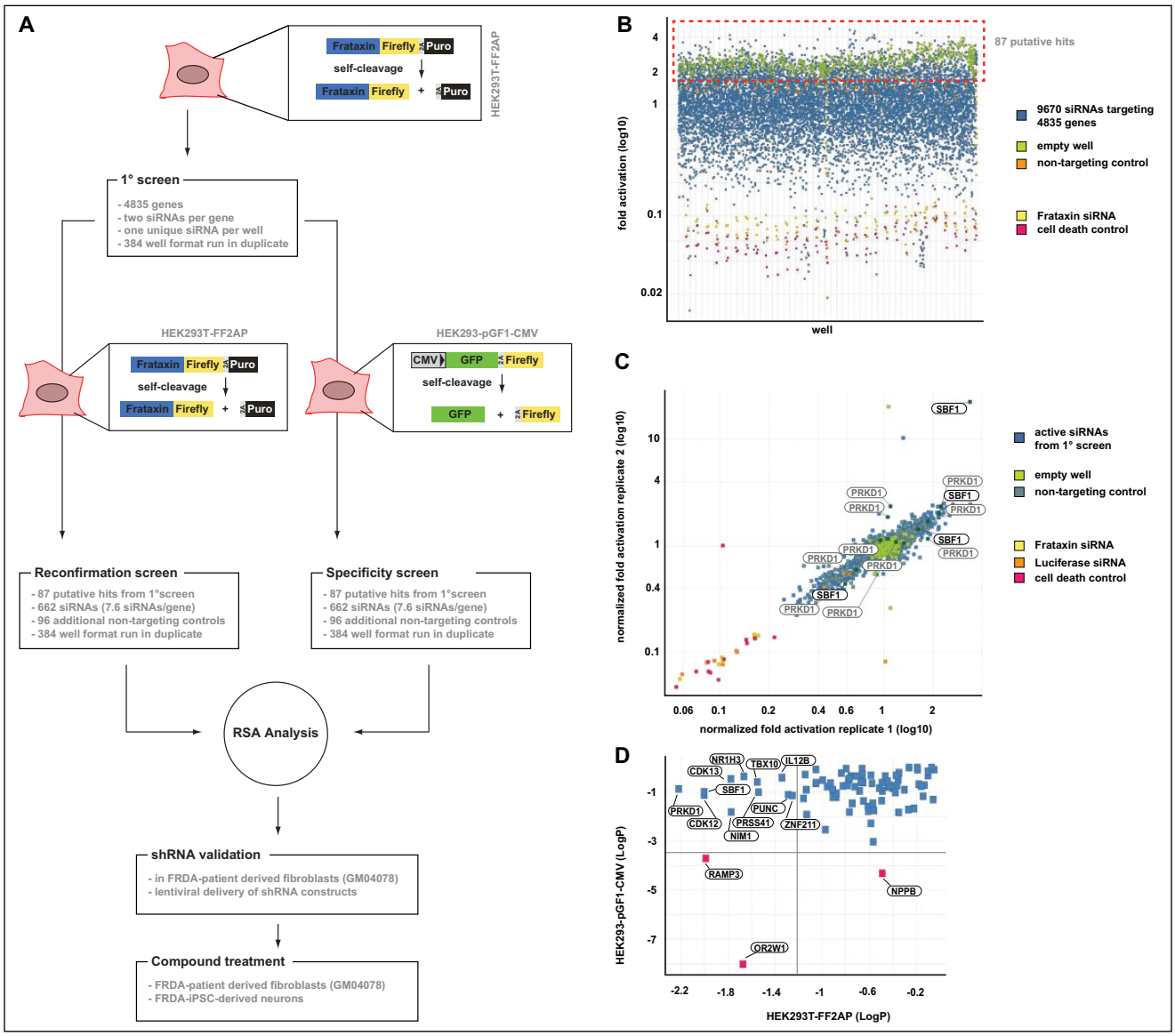


Figure 2. Small interfering RNA (siRNA) screen with the HEK293T-FF2AP cell line uncovers novel negative regulators of frataxin (FXN) expression. (**A**) Scheme of the workflow of the RNA interference (RNAi) screen. For details, see Materials and Methods. (**B**) Primary high-throughput screening (HTS) of the Qiagen (Valencia, CA) V1 Druggome siRNA library was performed in duplicate. Screening data are represented as fold activation in a scatterplot of one set of normalized plates (data, blue squares). Each plate contains the same control wells (cell death control siRNAs, antifrataxin siRNAs, nontargeting control siRNAs, and empty wells). (**C**) A reconfirmation screen of the top 87 hits from the primary screen, including additional targeting as well as nontargeting control siRNAs, was run in duplicate. Results are plotted as normalized replicate values. Activators are at or near the top of the diagonal, while inhibitors such as antifrataxin, antiluciferase, and the cell death siRNA controls are found at the bottom of the diagonal. Reproducibility can be evaluated on the same plot by identifying data points and their position compared with the diagonal. Those points off to the left or bottom from the diagonal indicate disparate values between the replicates. A ranked list of hits after redundant siRNA analysis (RSA) is provided in **Supplemental Table S1**. Individual siRNAs targeting the genes PRKD1 and SBF1, which rank as number 1 and 3 after RSA analysis, respectively, are indicated. (**D**) Filter screen for nonspecific activators of luciferase activity. An HEK293T cell line stably expressing luciferase from a cytomegalovirus (CMV) promoter (HEK293-pGF1-CMV) was transfected in duplicate with siRNAs covering the top 87 hits from the primary screen. Results were normalized to the plate median and then subjected to RSA analysis. A ranked list of hits after RSA is provided in **Supplemental Table S2**. LogP values for genes from the CMV luciferase cell line were plotted against LogP values for the same genes from the HEK293T-FF2AP reporter line to identify outliers (the lower the negative score, the more confidence there is in the gene being associated with the observed phenotype).

two or more independent siRNAs elevated the FXN-luciferase protein levels specifically (**Fig. 3A**). Complete ranked lists of hits after RSA are provided in the supplemental material (**Suppl. Tables S1** and **S2**).

The FXN alleles in the HEK293T-FF2AP cell line used in our primary RNAi screen do not contain pathogenic GAA triplet repeat expansions. The main reason why we designed our screening cell line that way was because high-throughput biology requires large amounts of cells, and FXN deficiency could prevent cell expansion. Moreover, the readily available FRDA patient-derived B lymphoblasts turned out to be difficult to use for genome engineering and are not really well suited for high-throughput RNAi screening approaches. We also encountered poor cutting efficiency/homologous recombination in immortalized FRDA patient-derived fibroblasts that we have generated (FRDA-4078iBT), preventing us from establishing a luciferase reporter cell line that would allow direct monitoring of FXN expression from an expanded GAA allele. Therefore, we performed secondary hit validation experiments in the FRDA-4078iBT fibroblasts by quantitative real-time RT-PCR to determine whether our primary hits could increase FXN gene expression also in cells that contain an expanded repeat tract. Expression analysis revealed that four of the top five candidate genes (**Fig. 3A**) are also expressed in FRDA fibroblasts (PRKD1, CDK12, SBF1, and CDK13) (data not shown). To knock down these genes, we designed lentiviral shRNA constructs and delivered them into fibroblasts by lentiviral transduction. mRNA levels of all expressed candidate genes were reduced significantly upon shRNA expression (**Fig. 3B**,**C** and data not shown). Depletion of the cell cycle–related kinases CDK12 and CDK13 resulted in strong proliferation defects and were omitted from further analysis (data not shown). Even though knockdown efficiencies for PRKD1 or SBF1 were never higher than 60%, we consistently observed a significant increase in FXN mRNA levels (**Fig. 2B**,**C**). These results confirmed that reducing the levels of the serine/threonine-protein kinase PRKD1 or the putative pseudophosphatase SBF1 results in increased FXN expression even in the presence of an expanded repeat tract. Thus, monitoring gene expression from the endogenous FXN locus irrespective of expanded GAA repeats opens up new opportunities for target discovery and the dissection of molecular pathways involved in FRDA. We note that the roughly 1.5-fold increase in FXN mRNA levels after PRKD1 or SBF1 knockdown is quite modest and was hence not reliably detectable by Western blot. This demonstrates the enormous power of using the highly sensitive luciferase approach to monitor changes in FXN expression but also urges us to expand our screens to genome-wide siRNA libraries to identify hits that lead to more drastic changes in FXN expression.

In recent years, a significant body of literature has proposed that genome-engineering technologies and high-throughput biology each have the potential to revolutionize

A

Gene	Accession #	RSA Rank CMV-luc	RSA Rank FXN-luc
PRKD1	NM_002742	38	**1**
CDK12	NM_016507	27	**2**
SBF1	NM_002972	32	**3**
RAMP3	NM_005856	3	**4**
CDK13	NM_003718	65	**5**

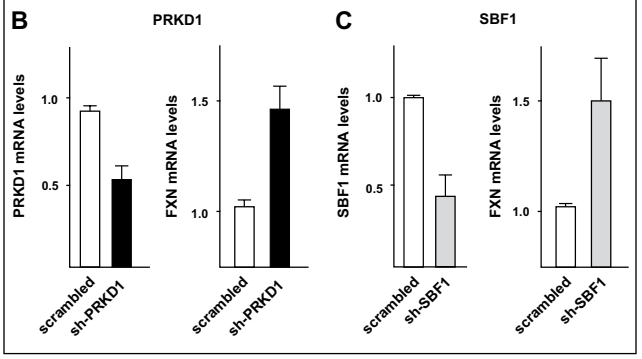


Figure 3. Validation of top five genes in Friedreich's ataxia (FRDA) patient-derived fibroblasts. (**A**) List of top five genes obtained after reconfirmation screens and redundant small interfering RNA analysis (RSA) analysis. RSA ranking in cytomegalovirus (CMV)–luc and frataxin (FXN)–luc screens is shown. Genes marked in bold are expressed in FRDA patient-derived fibroblasts and lymphoblasts (data not shown). The complete lists are available as supplementary tables. (**B**) Short hairpin RNA (shRNA)–mediated knockdown of PRKD1 messenger RNA (mRNA) in FRDA patient-derived fibroblasts (4078iBT cells). Fibroblasts were transduced with a lentiviral vector encoding an shRNA targeting PRKD1 mRNA. PRKD1 (left panel) and FXN (right panel) mRNA levels were determined by quantitative real-time reverse transcriptase PCR (RT-PCR). PRKD1 and FXN mRNA levels are shown relative to the nonsilencing shRNA control. Mean values normalized to GAPDH are shown (n = 4 different transduction experiments). Error bars represent SEM. (**C**) shRNA-mediated knockdown of SBF1 mRNA in FRDA patient-derived fibroblasts (4078iBT cells). Fibroblasts were transduced with a lentiviral vector encoding an shRNA targeting SBF1 mRNA. SBF1 (left panel) and FXN (right panel) mRNA levels were determined by quantitative real-time RT-PCR. SBF1 and FXN mRNA levels are shown relative to the nonsilencing shRNA control. Mean values normalized to GAPDH are shown (n = 5 different transduction experiments). Error bars represent SEM.

biomedicine and pharmaceutical pipelines. This study establishes that these emerging biotechnologies can be effectively combined to perform as proposed, exemplifying a new paradigm for drug discovery. The HEK293T-FF2AP cell line established in this study is the first luciferase reporter–based cellular model, which allows an accurate and effortless assessment of endogenous frataxin gene regulation and is compatible with high-throughput biology. HEK293T-FF2AP cells are easy to transfect and can be expanded to a very large scale at a relatively low cost. Thus, this system should be compatible with the technical and

financial constraints of most academic and industrial screening platforms.

Using the HEK293T-FF2AP cell line in a pilot high-throughput genomic screen, we uncovered novel negative regulators of FXN expression, demonstrating the importance of monitoring gene expression from the endogenous FXN locus. Our results showing that FXN expression can be increased even in the presence of an expanded repeat tract also demonstrate that screening for general regulators of FXN transcription, irrespective of GAA repeat length, is an appropriate strategy. Thus, our approach opens up new opportunities for target discovery, and we anticipate that HTS of genome-wide siRNA libraries will reveal additional potential drug targets in the near future.

Finally, the genome-engineering tools designed in this study will serve as a powerful resource to the FRDA community. The ZNFs and donor constructs developed in this study will enable others to knock in any given sequence tag into the FXN locus in many different cell types, facilitating the dissection of molecular pathways involved in FRDA. Notably, our donor constructs can also be used in combination with the more recently introduced CRISPR/Cas9 technology, which may work more effectively in patient-derived cell lines than the ZNFs that we have used in this study.

In conclusion, we have developed new tools that will serve the FRDA research community as a valuable resource for future mechanistic studies and the identification of new therapeutic strategies for a so far untreatable disease.

Author Contributions

The ZFN targeting approach was designed by T.P. and M.B. in collaboration with Sigma-Aldrich. L.M., A.R., B.T., and T.O. conducted the RNAi screen; R.V., T.P., and P.K. performed the rest of the experiments. S.D. immortalized patient-derived fibroblasts. M.B. designed and oversaw the study, obtained funding, and assisted in data evaluation and interpretation. M.B. and R.V. wrote the manuscript with the help of L.M., T.O., and T.P.

Acknowledgments

We thank Yukiko Shimada and Nathalie Laschet for technical assistance, as well as the members of the Bühler laboratory for fruitful discussions and critical reading of the manuscript. We are grateful to Cécile Blaustein for sharing her expertise and providing helpful advice and to Ira Shulman for providing the HEK293-pGF1-CMV cell line.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: All the contributors are employees of institutes (FMI and GNF) that are financially supported by Novartis. A patent application has been filed based on some of the results described in this contribution.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by funds to M.B. from the Gebert Rüf Stiftung, the Swiss National Science Foundation, and the EMBO Young Investigator Programme. The Genomics Institute of the Novartis Research Foundation (GNF) and the Friedrich Miescher Institute for Biomedical Research are supported by the Novartis Research Foundation. T.P. was supported by a postdoctoral fellowship from the Swedish Society for Medical Research (SSMF).

References

1. Pandolfo, M. Friedreich Ataxia: The Clinical Picture. *J. Neurol*. **2009**, *256*(Suppl 1), 3–8.
2. Campuzano, V.; Montermini, L.; Molto, M. D.; et al. Friedreich's Ataxia: Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion. *Science* **1996**, *271*, 1423–1427.
3. Schulz, J. B.; Boesch, S.; Burk, K.; et al. Diagnosis and Treatment of Friedreich Ataxia: A European Perspective. *Nat. Rev. Neurol*. **2009**, *5*, 222–234.
4. Durr, A.; Cossee, M.; Agid, Y.; et al. Clinical and Genetic Abnormalities in Patients with Friedreich's Ataxia. *N. Engl. J. Med*. **1996**, *335*, 1169–1175.
5. Martelli, A.; Napierala, M.; Puccio, H. Understanding the Genetic and Molecular Pathogenesis of Friedreich's Ataxia through Animal and Cellular Models. *Dis. Model Mech*. **2012**, *5*, 165–176.
6. Kumari, D.; Biacsi, R. E.; Usdin, K. Repeat Expansion Affects Both Transcription Initiation and Elongation in Friedreich Ataxia Cells. *J. Biol. Chem*. **2011**, *286*, 4209–4215.
7. Tsou, A. Y.; Friedman, L. S.; Wilson, R. B.; et al. Pharmacotherapy for Friedreich Ataxia. *CNS Drugs* **2009**, *23*, 213–223.
8. Sturm, B.; Stupphann, D.; Kaun, C.; et al. Recombinant Human Erythropoietin: Effects on Frataxin Expression In Vitro. *Eur. J. Clin. Invest*. **2005**, *35*, 711–717.
9. Sarsero, J. P.; Li, L.; Wardan, H.; et al. Upregulation of Expression from the FRDA Genomic Locus for the Therapy of Friedreich Ataxia. *J. Gene Med*. **2003**, *5*, 72–81.
10. Sarsero, J. P.; Holloway, T. P.; Li, L.; et al. Evaluation of an FRDA-EGFP Genomic Reporter Assay in Transgenic Mice. *Mamm. Genome* **2005**, *16*, 228–241.
11. Grant, L.; Sun, J.; Xu, H.; et al. Rational Selection of Small Molecules That Increase Transcription through the GAA Repeats Found in Friedreich's Ataxia. *FEBS Lett*. **2006**, *580*, 5399–5405.
12. Soragni, E.; Herman, D.; Dent, S. Y.; et al. Long Intronic GAA*TTC Repeats Induce Epigenetic Changes and Reporter Gene Silencing in a Molecular Model of Friedreich Ataxia. *Nucleic Acids Res*. **2008**, *36*, 6056–6065.
13. Li, L.; Voullaire, L.; Sandi, C.; et al. Pharmacological Screening Using an FXN-EGFP Cellular Genomic Reporter Assay for the Therapy of Friedreich Ataxia. *PLoS ONE* **2013**, *8*, e55940.
14. Lufino, M. M.; Silva, A. M.; Nemeth, A. H.; et al. A GAA Repeat Expansion Reporter Model of Friedreich's Ataxia

Recapitulates the Genomic Context and Allows Rapid Screening of Therapeutic Compounds. *Hum. Mol. Genet*. **2013**, *22*, 5173–5187.
15. Herman, D.; Jenssen, K.; Burnett, R.; et al. Histone Deacetylase Inhibitors Reverse Gene Silencing in Friedreich's Ataxia. *Nat. Chem. Biol*. **2006**, *2*, 551–558.
16. Saveliev, A.; Everett, C.; Sharpe, T.; et al. DNA Triplet Repeats Mediate Heterochromatin-Protein-1-Sensitive Variegated Gene Silencing. *Nature* **2003**, *422*, 909–913.
17. Castaldo, I.; Pinelli, M.; Monticelli, A.; et al. DNA Methylation in Intron 1 of the Frataxin Gene Is Related to GAA Repeat Length and Age of Onset in Friedreich Ataxia Patients. *J. Med. Genet*. **2008**, *45*, 808–812.
18. Punga, T.; Buhler, M. Long Intronic GAA Repeats Causing Friedreich Ataxia Impede Transcription Elongation. *EMBO Mol. Med*. **2010**, *2*, 120–129.
19. Duss, S.; Andre, S.; Nicoulaz, A. L.; et al. An Oestrogen-Dependent Model of Breast Cancer Created by Transformation of Normal Human Mammary Epithelial Cells. *Breast Cancer Res*. **2007**, *9*, R38.
20. Aza-Blanc, P.; Cooper, C. L.; Wagner, K.; et al. Identification of Modulators of TRAIL-Induced Apoptosis via RNAi-Based Phenotypic Screening. *Mol. Cell* **2003**, *12*, 627–637.
21. Chanda, S. K.; White, S.; Orth, A. P.; et al. Genome-Scale Functional Profiling of the Mammalian AP-1 Signaling Pathway. *Proc. Natl. Acad. Sci. U. S. A*. **2003**, *100*, 12153–12158.
22. Konig, R.; Chiang, C. Y.; Tu, B. P.; et al. A Probability-Based Approach for the Analysis of Large-Scale RNAi Screens. *Nat. Methods* **2007**, *4*, 847–849.
23. Dull, T.; Zufferey, R.; Kelly, M.; et al. A Third-Generation Lentivirus Vector with a Conditional Packaging System. *J. Virol*. **1998**, *72*, 8463–8471.
24. Lombardo, A.; Genovese, P.; Beausejour, C. M.; et al. Gene Editing in Human Stem Cells Using Zinc Finger Nucleases and Integrase-Defective Lentiviral Vector Delivery. *Nat. Biotechnol*. **2007**, *25*, 1298–1306.
25. Urnov, F. D.; Miller, J. C.; Lee, Y. L.; et al. Highly Efficient Endogenous Human Gene Correction Using Designed Zinc-Finger Nucleases. *Nature* **2005**, *435*, 646–651.

Supplementary Table 1. Results of Confirmation Screen.

siRNA	Gene_ID	accession #	Symbol	siRNA_SCORE	RSA_LogP	Cut-off_Rank	RSA_Rank
siRNA_1	115529462	NM_002742	PRKD1	2.095445003	-2.219118663	20	1
siRNA_2	115529462	NM_002742	PRKD1	1.780323579	-2.219118663	44	1
siRNA_3	115529462	NM_002742	PRKD1	1.609839969	-2.219118663	69	1
siRNA_4	115529462	NM_002742	PRKD1	1.4752528	-2.219118663	108	1
siRNA_5	115529462	NM_002742	PRKD1	1.406779394	-2.219118663	130	1
siRNA_6	115529462	NM_002742	PRKD1	1.188708618	-2.219118663	227	1
siRNA_7	115529462	NM_002742	PRKD1	1.138015488	-2.219118663	260	1
siRNA_8	115529462	NM_002742	PRKD1	1.114720422	-2.219118663	271	1
siRNA_9	115529462	NM_002742	PRKD1	1.044417767	-2.219118663	311	1
siRNA_10	115529462	NM_002742	PRKD1	0.64039326	-2.219118663	520	1
siRNA_1	157817022	NM_016507	CDK12	1.812290686	-2.004268846	36	2
siRNA_2	157817022	NM_016507	CDK12	1.808609379	-2.004268846	37	2
siRNA_3	157817022	NM_016507	CDK12	1.795034143	-2.004268846	41	2
siRNA_4	157817022	NM_016507	CDK12	1.728048453	-2.004268846	49	2
siRNA_5	157817022	NM_016507	CDK12	1.460253976	-2.004268846	113	2
siRNA_6	157817022	NM_016507	CDK12	1.320908772	-2.004268846	159	2
siRNA_7	157817022	NM_016507	CDK12	1.29043562	-2.004268846	177	2
siRNA_8	157817022	NM_016507	CDK12	1.175409293	-2.004268846	236	2
siRNA_9	157817022	NM_016507	CDK12	1.011348085	-2.004268846	330	2
siRNA_10	157817022	NM_016507	CDK12	0.378503601	-2.004268846	642	2
siRNA_1	239735518	NM_002972	SBF1	8.714268322	-2.002520025	1	3
siRNA_2	239735518	NM_002972	SBF1	2.299027371	-2.002520025	10	3
siRNA_3	239735518	NM_002972	SBF1	1.527911534	-2.002520025	94	3
siRNA_4	239735518	NM_002972	SBF1	0.50540777	-2.002520025	586	3
siRNA_1	118572586	NM_005856	RAMP3	2.893583771	-1.989133016	4	4
siRNA_2	118572586	NM_005856	RAMP3	1.997734462	-1.989133016	24	4
siRNA_3	118572586	NM_005856	RAMP3	1.772026491	-1.989133016	45	4
siRNA_4	118572586	NM_005856	RAMP3	1.220584267	-1.989133016	212	4
siRNA_1	281427100	NM_003718	CDK13	1.635956055	-1.775829503	62	5
siRNA_2	281427100	NM_003718	CDK13	1.228109096	-1.775829503	208	5
siRNA_3	281427100	NM_003718	CDK13	1.144972461	-1.775829503	257	5
siRNA_4	281427100	NM_003718	CDK13	1.013877492	-1.775829503	328	5
siRNA_5	281427100	NM_003718	CDK13	0.910704734	-1.775829503	385	5
siRNA_6	281427100	NM_003718	CDK13	0.851012734	-1.775829503	410	5
siRNA_7	281427100	NM_003718	CDK13	0.83204074	-1.775829503	418	5
siRNA_8	281427100	NM_003718	CDK13	0.831155154	-1.775829503	421	5
siRNA_9	281427100	NM_003718	CDK13	0.788663645	-1.775829503	441	5
siRNA_10	281427100	NM_003718	CDK13	0.787857982	-1.775829503	442	5
siRNA_1	37059769	NM_153361	NIM1	2.216073951	-1.773085012	15	6
siRNA_2	37059769	NM_153361	NIM1	1.800919627	-1.773085012	38	6
siRNA_3	37059769	NM_153361	NIM1	1.701318237	-1.773085012	54	6
siRNA_4	37059769	NM_153361	NIM1	1.535428079	-1.773085012	91	6
siRNA_5	37059769	NM_153361	NIM1	1.455615941	-1.773085012	115	6
siRNA_6	37059769	NM_153361	NIM1	1.318261348	-1.773085012	162	6
siRNA_7	37059769	NM_153361	NIM1	0.706872625	-1.773085012	488	6
siRNA_8	37059769	NM_153361	NIM1	0.700411085	-1.773085012	492	6
siRNA_9	37059769	NM_153361	NIM1	0.65874794	-1.773085012	512	6
siRNA_10	37059769	NM_153361	NIM1	0.633037814	-1.773085012	526	6
siRNA_1	169234788	NM_030903	OR2W1	2.290679212	-1.673530837	11	7
siRNA_2	169234788	NM_030903	OR2W1	2.23669789	-1.673530837	14	7
siRNA_3	169234788	NM_030903	OR2W1	1.368323047	-1.673530837	143	7
siRNA_4	169234788	NM_030903	OR2W1	1.167810699	-1.673530837	245	7
siRNA_5	169234788	NM_030903	OR2W1	1.092472964	-1.673530837	286	7
siRNA_6	169234788	NM_030903	OR2W1	0.981698545	-1.673530837	350	7
siRNA_1	194294516	NM_005693	NR1H3	2.118154015	-1.666042816	19	8
siRNA_2	194294516	NM_005693	NR1H3	1.719524785	-1.666042816	51	8
siRNA_3	194294516	NM_005693	NR1H3	1.478086375	-1.666042816	107	8
siRNA_4	194294516	NM_005693	NR1H3	1.314242281	-1.666042816	163	8
siRNA_5	194294516	NM_005693	NR1H3	1.140816122	-1.666042816	259	8
siRNA_6	194294516	NM_005693	NR1H3	0.979786183	-1.666042816	351	8
siRNA_1	312596882	NM_005995	TBX10	1.833016513	-1.553462674	33	9
siRNA_2	312596882	NM_005995	TBX10	1.670682746	-1.553462674	59	9
siRNA_3	312596882	NM_005995	TBX10	1.630511935	-1.553462674	63	9

siRNA_4	312596882	NM_005995	TBX10	1.113951813	-1.553462674	272	9
siRNA_1	34419642	NM_183379	PRSS41	1.579255333	-1.54193238	78	10
siRNA_2	34419642	NM_183379	PRSS41	1.485755726	-1.54193238	106	10
siRNA_3	34419642	NM_183379	PRSS41	1.308762813	-1.54193238	168	10
siRNA_4	34419642	NM_183379	PRSS41	0.982931208	-1.54193238	349	10
siRNA_5	34419642	NM_183379	PRSS41	0.956835191	-1.54193238	366	10
siRNA_6	34419642	NM_183379	PRSS41	0.954874354	-1.54193238	368	10
siRNA_1	24497437	NM_002187	IL12B	1.71995058	-1.344026614	50	11
siRNA_2	24497437	NM_002187	IL12B	1.322787242	-1.344026614	157	11
siRNA_3	24497437	NM_002187	IL12B	1.264218292	-1.344026614	186	11
siRNA_4	24497437	NM_002187	IL12B	1.237491715	-1.344026614	206	11
siRNA_5	24497437	NM_002187	IL12B	1.091965979	-1.344026614	287	11
siRNA_6	24497437	NM_002187	IL12B	1.043507942	-1.344026614	313	11
siRNA_7	24497437	NM_002187	IL12B	0.656108548	-1.344026614	513	11
siRNA_1	134244584	NM_004884	PUNC	1.87579665	-1.289774132	30	12
siRNA_2	134244584	NM_004884	PUNC	1.756272368	-1.289774132	46	12
siRNA_3	134244584	NM_004884	PUNC	1.310815956	-1.289774132	167	12
siRNA_4	134244584	NM_004884	PUNC	0.577522972	-1.289774132	551	12
siRNA_5	221136824	NM_006385	ZNF211	2.541524435	-1.252943148	6	13
siRNA_6	221136824	NM_006385	ZNF211	1.856333999	-1.252943148	31	13
siRNA_7	221136824	NM_006385	ZNF211	1.267050246	-1.252943148	184	13
siRNA_8	221136824	NM_006385	ZNF211	1.175568075	-1.252943148	235	13
siRNA_9	221136824	NM_006385	ZNF211	0.505584925	-1.252943148	585	13
siRNA_1	148839375	NM_138499	PWWP2B	1.569318968	-1.158767218	82	14
siRNA_2	148839375	NM_138499	PWWP2B	1.33161755	-1.158767218	154	14
siRNA_3	148839375	NM_138499	PWWP2B	1.194589127	-1.158767218	222	14
siRNA_4	148839375	NM_138499	PWWP2B	0.998841078	-1.158767218	341	14
siRNA_5	209574322	NM_000614	CNTF	1.620071374	-1.143483865	65	15
siRNA_6	209574322	NM_000614	CNTF	1.44074071	-1.143483865	117	15
siRNA_7	209574322	NM_000614	CNTF	1.18883682	-1.143483865	226	15
siRNA_8	209574322	NM_000614	CNTF	0.992708576	-1.143483865	344	15
siRNA_1	38327639	NM_001334	CTSO	1.320715657	-1.135837832	160	16
siRNA_2	38327639	NM_001334	CTSO	1.299172782	-1.135837832	174	16
siRNA_3	38327639	NM_001334	CTSO	1.174796425	-1.135837832	237	16
siRNA_4	38327639	NM_001334	CTSO	1.168982268	-1.135837832	243	16
siRNA_5	38327639	NM_001334	CTSO	1.06188849	-1.135837832	303	16
siRNA_6	38327639	NM_001334	CTSO	0.532981823	-1.135837832	572	16
siRNA_1	23510288	NM_152386	SGPP2	1.965510882	-1.133369156	26	17
siRNA_2	23510288	NM_152386	SGPP2	1.957082686	-1.133369156	27	17
siRNA_3	23510288	NM_152386	SGPP2	1.080921364	-1.133369156	293	17
siRNA_4	23510288	NM_152386	SGPP2	0.989282069	-1.133369156	346	17
siRNA_5	300193030	NM_003810	TNFSF10	1.673078529	-1.055572329	57	18
siRNA_6	300193030	NM_003810	TNFSF10	1.587328413	-1.055572329	75	18
siRNA_7	300193030	NM_003810	TNFSF10	1.5283987	-1.055572329	93	18
siRNA_8	300193030	NM_003810	TNFSF10	1.012504331	-1.055572329	329	18
siRNA_9	300193030	NM_003810	TNFSF10	0.879229253	-1.055572329	401	18
siRNA_10	300193030	NM_003810	TNFSF10	0.786702313	-1.055572329	443	18
siRNA_1	62420885	NM_001014796	DDR2	1.816550107	-1.054880375	35	19
siRNA_2	62420885	NM_001014796	DDR2	1.259117391	-1.054880375	190	19
siRNA_3	62420885	NM_001014796	DDR2	1.123515274	-1.054880375	267	19
siRNA_4	62420885	NM_001014796	DDR2	1.05630213	-1.054880375	304	19
siRNA_5	62420885	NM_001014796	DDR2	1.017951029	-1.054880375	324	19
siRNA_6	62420885	NM_001014796	DDR2	0.955158089	-1.054880375	367	19
siRNA_7	62420885	NM_001014796	DDR2	0.870127161	-1.054880375	404	19
siRNA_8	62420885	NM_001014796	DDR2	0.849415447	-1.054880375	412	19
siRNA_9	62420885	NM_001014796	DDR2	0.796889446	-1.054880375	436	19
siRNA_10	62420885	NM_001014796	DDR2	0.639740711	-1.054880375	521	19
siRNA_1	154759246	NM_199046	TEPP	1.574553327	-1.050648309	79	20
siRNA_2	154759246	NM_199046	TEPP	1.248335657	-1.050648309	195	20
siRNA_3	154759246	NM_199046	TEPP	1.239260371	-1.050648309	204	20
siRNA_4	154759246	NM_199046	TEPP	0.583633864	-1.050648309	550	20
siRNA_1	157266333	NM_003322	TULP1	1.848192943	-0.995133347	32	21
siRNA_2	157266333	NM_003322	TULP1	1.504971163	-0.995133347	99	21
siRNA_3	157266333	NM_003322	TULP1	1.435853076	-0.995133347	120	21
siRNA_4	157266333	NM_003322	TULP1	0.957844457	-0.995133347	365	21

siRNA_5	157266333	NM_003322	TULP1	0.831465645	-0.995133347	420	21
siRNA_1	289547622	NM_014565	OR1A1	1.338921262	-0.974043292	152	22
siRNA_2	289547622	NM_014565	OR1A1	1.312329438	-0.974043292	166	22
siRNA_3	289547622	NM_014565	OR1A1	1.104014093	-0.974043292	280	22
siRNA_4	289547622	NM_014565	OR1A1	1.022734629	-0.974043292	320	22
siRNA_5	289547622	NM_014565	OR1A1	0.83654968	-0.974043292	417	22
siRNA_6	289547622	NM_014565	OR1A1	0.752837577	-0.974043292	457	22
siRNA_1	55750030	NM_014787	DNAJC6	2.268336084	-0.959092622	12	23
siRNA_2	55750030	NM_014787	DNAJC6	1.974429877	-0.959092622	25	23
siRNA_3	55750030	NM_014787	DNAJC6	1.398967557	-0.959092622	131	23
siRNA_4	55750030	NM_014787	DNAJC6	1.188666884	-0.959092622	228	23
siRNA_5	55750030	NM_014787	DNAJC6	0.629951299	-0.959092622	528	23
siRNA_6	55750030	NM_014787	DNAJC6	0.496776851	-0.959092622	593	23
siRNA_1	51702241	NM_001004023	DYRK3	1.709444804	-0.930547208	53	24
siRNA_2	51702241	NM_001004023	DYRK3	1.500629475	-0.930547208	104	24
siRNA_3	51702241	NM_001004023	DYRK3	1.301495997	-0.930547208	173	24
siRNA_4	51702241	NM_001004023	DYRK3	1.168066589	-0.930547208	244	24
siRNA_5	51702241	NM_001004023	DYRK3	1.017654356	-0.930547208	325	24
siRNA_6	51702241	NM_001004023	DYRK3	0.979576985	-0.930547208	352	24
siRNA_7	51702241	NM_001004023	DYRK3	0.920943603	-0.930547208	378	24
siRNA_8	51702241	NM_001004023	DYRK3	0.91958496	-0.930547208	379	24
siRNA_9	51702241	NM_001004023	DYRK3	0.626547309	-0.930547208	532	24
siRNA_10	51702241	NM_001004023	DYRK3	0.622714815	-0.930547208	536	24
siRNA_1	93004080	NM_006622	PLK2	1.755807497	-0.910835866	47	25
siRNA_2	93004080	NM_006622	PLK2	1.670719959	-0.910835866	58	25
siRNA_3	93004080	NM_006622	PLK2	1.411777339	-0.910835866	127	25
siRNA_4	93004080	NM_006622	PLK2	1.192948173	-0.910835866	225	25
siRNA_5	93004080	NM_006622	PLK2	1.131848335	-0.910835866	261	25
siRNA_6	93004080	NM_006622	PLK2	1.106270478	-0.910835866	279	25
siRNA_7	93004080	NM_006622	PLK2	0.71455747	-0.910835866	483	25
siRNA_8	93004080	NM_006622	PLK2	0.521520275	-0.910835866	578	25
siRNA_9	93004080	NM_006622	PLK2	0.503693272	-0.910835866	590	25
siRNA_1	51464990	XM_291141	MAST4	1.561500149	-0.909846931	85	26
siRNA_2	51464990	XM_291141	MAST4	1.456971588	-0.909846931	114	26
siRNA_3	51464990	XM_291141	MAST4	1.349877085	-0.909846931	148	26
siRNA_4	51464990	XM_291141	MAST4	1.241186726	-0.909846931	202	26
siRNA_5	51464990	XM_291141	MAST4	1.150479985	-0.909846931	254	26
siRNA_6	51464990	XM_291141	MAST4	1.044069785	-0.909846931	312	26
siRNA_7	51464990	XM_291141	MAST4	0.907157278	-0.909846931	388	26
siRNA_8	51464990	XM_291141	MAST4	0.851117902	-0.909846931	409	26
siRNA_9	51464990	XM_291141	MAST4	0.764749279	-0.909846931	453	26
siRNA_10	51464990	XM_291141	MAST4	0.604142565	-0.909846931	540	26
siRNA_1	45439363	NM_004438	EPHA4	2.124304353	-0.885510518	18	27
siRNA_2	45439363	NM_004438	EPHA4	1.568479125	-0.885510518	83	27
siRNA_3	45439363	NM_004438	EPHA4	1.36967125	-0.885510518	142	27
siRNA_4	45439363	NM_004438	EPHA4	1.255145722	-0.885510518	191	27
siRNA_5	45439363	NM_004438	EPHA4	1.180799601	-0.885510518	232	27
siRNA_6	45439363	NM_004438	EPHA4	1.140869617	-0.885510518	258	27
siRNA_7	45439363	NM_004438	EPHA4	1.04198349	-0.885510518	314	27
siRNA_8	45439363	NM_004438	EPHA4	0.570424518	-0.885510518	553	27
siRNA_9	45439363	NM_004438	EPHA4	0.503741924	-0.885510518	589	27
siRNA_10	45439363	NM_004438	EPHA4	0.409138324	-0.885510518	634	27
siRNA_1	87080812	NM_152271	LONRF1	2.984409781	-0.883193309	3	28
siRNA_2	87080812	NM_152271	LONRF1	1.602887912	-0.883193309	72	28
siRNA_3	87080812	NM_152271	LONRF1	0.671970118	-0.883193309	505	28
siRNA_4	87080812	NM_152271	LONRF1	0.660321242	-0.883193309	511	28
siRNA_5	87080812	NM_152271	LONRF1	0.424286906	-0.883193309	626	28
siRNA_6	87080812	NM_152271	LONRF1	0.371208618	-0.883193309	644	28
siRNA_1	62953138	NM_020967	NCOA5	2.417054469	-0.880410447	7	29
siRNA_2	62953138	NM_020967	NCOA5	1.5595136	-0.880410447	86	29
siRNA_3	62953138	NM_020967	NCOA5	1.437218875	-0.880410447	119	29
siRNA_4	62953138	NM_020967	NCOA5	0.882274342	-0.880410447	400	29
siRNA_1	51468634	XM_208545	BMPR1A	1.562604716	-0.839389593	84	30
siRNA_2	51468634	XM_208545	BMPR1A	1.375367085	-0.839389593	137	30
siRNA_3	51468634	XM_208545	BMPR1A	1.312881425	-0.839389593	164	30

siRNA_4	51468634	XM_208545	BMPR1A	1.268819753	-0.839389593	183	30
siRNA_5	51468634	XM_208545	BMPR1A	0.931498868	-0.839389593	372	30
siRNA_6	51468634	XM_208545	BMPR1A	0.872227615	-0.839389593	403	30
siRNA_7	51468634	XM_208545	BMPR1A	0.842005408	-0.839389593	415	30
siRNA_8	51468634	XM_208545	BMPR1A	0.7898328	-0.839389593	440	30
siRNA_9	51468634	XM_208545	BMPR1A	0.744890607	-0.839389593	463	30
siRNA_10	51468634	XM_208545	BMPR1A	0.432239274	-0.839389593	621	30
siRNA_1	345441754	NM_004635	MAPKAPK3	1.623593672	-0.825592174	64	31
siRNA_2	345441754	NM_004635	MAPKAPK3	1.501098544	-0.825592174	102	31
siRNA_3	345441754	NM_004635	MAPKAPK3	1.415771962	-0.825592174	124	31
siRNA_4	345441754	NM_004635	MAPKAPK3	1.179200152	-0.825592174	234	31
siRNA_5	345441754	NM_004635	MAPKAPK3	1.020339151	-0.825592174	321	31
siRNA_6	345441754	NM_004635	MAPKAPK3	0.811886502	-0.825592174	428	31
siRNA_7	345441754	NM_004635	MAPKAPK3	0.79742863	-0.825592174	435	31
siRNA_8	345441754	NM_004635	MAPKAPK3	0.767393621	-0.825592174	449	31
siRNA_9	345441754	NM_004635	MAPKAPK3	0.740255275	-0.825592174	465	31
siRNA_10	345441754	NM_004635	MAPKAPK3	0.421372343	-0.825592174	628	31
siRNA_1	114520614	NM_001954	DDR1	1.460642928	-0.799621967	112	32
siRNA_2	114520614	NM_001954	DDR1	1.40830533	-0.799621967	129	32
siRNA_3	114520614	NM_001954	DDR1	1.040144296	-0.799621967	315	32
siRNA_4	114520614	NM_001954	DDR1	1.035090313	-0.799621967	317	32
siRNA_5	114520614	NM_001954	DDR1	0.98599546	-0.799621967	347	32
siRNA_6	114520614	NM_001954	DDR1	0.970045787	-0.799621967	358	32
siRNA_7	114520614	NM_001954	DDR1	0.918428522	-0.799621967	380	32
siRNA_8	114520614	NM_001954	DDR1	0.893059861	-0.799621967	395	32
siRNA_9	114520614	NM_001954	DDR1	0.590615024	-0.799621967	546	32
siRNA_10	114520614	NM_001954	DDR1	0.28376567	-0.799621967	659	32
siRNA_1	100913195	NM_152999	STEAP2	1.1588314	-0.787223431	249	33
siRNA_2	100913195	NM_152999	STEAP2	1.090506454	-0.787223431	288	33
siRNA_3	100913195	NM_152999	STEAP2	1.075106082	-0.787223431	298	33
siRNA_4	100913195	NM_152999	STEAP2	0.811512588	-0.787223431	429	33
siRNA_5	100913195	NM_152999	STEAP2	0.746685395	-0.787223431	462	33
siRNA_1	110349798	NM_007170	TESK2	1.514928323	-0.77806555	97	34
siRNA_2	110349798	NM_007170	TESK2	1.373085148	-0.77806555	138	34
siRNA_3	110349798	NM_007170	TESK2	1.335044826	-0.77806555	153	34
siRNA_4	110349798	NM_007170	TESK2	1.200248788	-0.77806555	220	34
siRNA_5	110349798	NM_007170	TESK2	1.075244565	-0.77806555	297	34
siRNA_6	110349798	NM_007170	TESK2	0.991351975	-0.77806555	345	34
siRNA_7	110349798	NM_007170	TESK2	0.861382158	-0.77806555	407	34
siRNA_8	110349798	NM_007170	TESK2	0.819153216	-0.77806555	425	34
siRNA_9	110349798	NM_007170	TESK2	0.732197511	-0.77806555	472	34
siRNA_10	110349798	NM_007170	TESK2	0.502414065	-0.77806555	591	34
siRNA_1	301336149	NM_001065	TNFRSF1A	1.397347854	-0.770102819	132	35
siRNA_2	301336149	NM_001065	TNFRSF1A	1.187129468	-0.770102819	229	35
siRNA_3	301336149	NM_001065	TNFRSF1A	1.153100878	-0.770102819	253	35
siRNA_4	301336149	NM_001065	TNFRSF1A	0.886013477	-0.770102819	399	35
siRNA_5	301336149	NM_001065	TNFRSF1A	0.809345194	-0.770102819	430	35
siRNA_6	301336149	NM_001065	TNFRSF1A	0.698050095	-0.770102819	494	35
siRNA_1	47078234	NM_001495	GFRA2	1.547577669	-0.767942399	89	36
siRNA_2	47078234	NM_001495	GFRA2	1.372326111	-0.767942399	139	36
siRNA_3	47078234	NM_001495	GFRA2	1.146939789	-0.767942399	256	36
siRNA_4	47078234	NM_001495	GFRA2	1.117558085	-0.767942399	268	36
siRNA_5	47078234	NM_001495	GFRA2	0.950751949	-0.767942399	369	36
siRNA_6	47078234	NM_001495	GFRA2	0.505824793	-0.767942399	584	36
siRNA_1	98986449	NM_022048	CSNK1G1	0.985563626	-0.754165912	348	37
siRNA_2	98986449	NM_022048	CSNK1G1	0.851340008	-0.754165912	408	37
siRNA_3	98986449	NM_022048	CSNK1G1	0.836617116	-0.754165912	416	37
siRNA_4	98986449	NM_022048	CSNK1G1	0.79554152	-0.754165912	438	37
siRNA_5	98986449	NM_022048	CSNK1G1	0.738272274	-0.754165912	466	37
siRNA_6	98986449	NM_022048	CSNK1G1	0.667339859	-0.754165912	507	37
siRNA_7	98986449	NM_022048	CSNK1G1	0.644016449	-0.754165912	518	37
siRNA_1	61744443	NM_052841	TSSK3	1.586520907	-0.74655924	76	38
siRNA_2	61744443	NM_052841	TSSK3	1.272206944	-0.74655924	182	38
siRNA_3	61744443	NM_052841	TSSK3	1.250226118	-0.74655924	194	38
siRNA_4	61744443	NM_052841	TSSK3	1.164055075	-0.74655924	247	38

siRNA_5	61744443	NM_052841	TSSK3	1.045740256	-0.74655924	310	38
siRNA_6	61744443	NM_052841	TSSK3	0.848919775	-0.74655924	413	38
siRNA_7	61744443	NM_052841	TSSK3	0.752187042	-0.74655924	460	38
siRNA_8	61744443	NM_052841	TSSK3	0.695927696	-0.74655924	496	38
siRNA_9	61744443	NM_052841	TSSK3	0.644259767	-0.74655924	516	38
siRNA_10	61744443	NM_052841	TSSK3	0.562552916	-0.74655924	559	38
siRNA_1	262118281	NM_032242	PLXNA1	1.931797138	-0.725107733	28	39
siRNA_2	262118281	NM_032242	PLXNA1	1.302009027	-0.725107733	172	39
siRNA_3	262118281	NM_032242	PLXNA1	1.187114993	-0.725107733	230	39
siRNA_4	262118281	NM_032242	PLXNA1	1.115463613	-0.725107733	270	39
siRNA_5	262118281	NM_032242	PLXNA1	0.636715744	-0.725107733	523	39
siRNA_6	262118281	NM_032242	PLXNA1	0.490990403	-0.725107733	596	39
siRNA_1	48717322	NM_001001667	OR6V1	2.362915186	-0.710355313	8	40
siRNA_2	48717322	NM_001001667	OR6V1	1.26184623	-0.710355313	188	40
siRNA_3	48717322	NM_001001667	OR6V1	1.18120999	-0.710355313	231	40
siRNA_4	48717322	NM_001001667	OR6V1	1.112028133	-0.710355313	273	40
siRNA_5	48717322	NM_001001667	OR6V1	0.708696658	-0.710355313	486	40
siRNA_6	48717322	NM_001001667	OR6V1	0.624481938	-0.710355313	534	40
siRNA_1	153946420	NM_001433	ERN1	1.245228934	-0.684347831	197	41
siRNA_2	153946420	NM_001433	ERN1	1.164333089	-0.684347831	246	41
siRNA_3	153946420	NM_001433	ERN1	1.111785099	-0.684347831	274	41
siRNA_4	153946420	NM_001433	ERN1	1.015840424	-0.684347831	326	41
siRNA_5	153946420	NM_001433	ERN1	0.908848946	-0.684347831	386	41
siRNA_6	153946420	NM_001433	ERN1	0.673974878	-0.684347831	502	41
siRNA_7	153946420	NM_001433	ERN1	0.602230784	-0.684347831	541	41
siRNA_8	153946420	NM_001433	ERN1	0.562594756	-0.684347831	558	41
siRNA_9	153946420	NM_001433	ERN1	0.55756138	-0.684347831	562	41
siRNA_10	153946420	NM_001433	ERN1	0.547896306	-0.684347831	567	41
siRNA_1	33239441	NM_006754	SYPL1	1.446796447	-0.673317459	116	42
siRNA_2	33239441	NM_006754	SYPL1	1.359713683	-0.673317459	146	42
siRNA_3	33239441	NM_006754	SYPL1	0.555113225	-0.673317459	564	42
siRNA_4	33239441	NM_006754	SYPL1	0.30062609	-0.673317459	657	42
siRNA_1	189095272	NM_000215	JAK3	1.616085804	-0.661246407	67	43
siRNA_2	189095272	NM_000215	JAK3	1.471251511	-0.661246407	110	43
siRNA_3	189095272	NM_000215	JAK3	1.414819794	-0.661246407	126	43
siRNA_4	189095272	NM_000215	JAK3	1.32050119	-0.661246407	161	43
siRNA_5	189095272	NM_000215	JAK3	1.172460087	-0.661246407	239	43
siRNA_6	189095272	NM_000215	JAK3	0.978146101	-0.661246407	353	43
siRNA_7	189095272	NM_000215	JAK3	0.777023697	-0.661246407	447	43
siRNA_8	189095272	NM_000215	JAK3	0.766121578	-0.661246407	451	43
siRNA_9	189095272	NM_000215	JAK3	0.665065449	-0.661246407	508	43
siRNA_10	189095272	NM_000215	JAK3	0.537850164	-0.661246407	570	43
siRNA_1	168480147	NM_005164	ABCD2	1.371731776	-0.65937004	140	44
siRNA_2	168480147	NM_005164	ABCD2	1.238350338	-0.65937004	205	44
siRNA_3	168480147	NM_005164	ABCD2	1.077706797	-0.65937004	295	44
siRNA_4	168480147	NM_005164	ABCD2	0.752801208	-0.65937004	458	44
siRNA_5	168480147	NM_005164	ABCD2	0.729751367	-0.65937004	473	44
siRNA_6	168480147	NM_005164	ABCD2	0.716825763	-0.65937004	480	44
siRNA_7	168480147	NM_005164	ABCD2	0.704779961	-0.65937004	490	44
siRNA_8	168480147	NM_005164	ABCD2	0.584673187	-0.65937004	549	44
siRNA_1	51464023	XM_497921	LOC391533	2.183492334	-0.657768808	17	45
siRNA_2	51464023	XM_497921	LOC391533	1.665250433	-0.657768808	60	45
siRNA_3	51464023	XM_497921	LOC391533	1.382612533	-0.657768808	134	45
siRNA_4	51464023	XM_497921	LOC391533	1.376653876	-0.657768808	136	45
siRNA_5	51464023	XM_497921	LOC391533	1.358928651	-0.657768808	147	45
siRNA_6	51464023	XM_497921	LOC391533	1.326227321	-0.657768808	155	45
siRNA_7	51464023	XM_497921	LOC391533	1.229043892	-0.657768808	207	45
siRNA_8	51464023	XM_497921	LOC391533	1.13020613	-0.657768808	262	45
siRNA_9	51464023	XM_497921	LOC391533	0.902210228	-0.657768808	390	45
siRNA_10	51464023	XM_497921	LOC391533	0.754400455	-0.657768808	456	45
siRNA_11	51464023	XM_497921	LOC391533	0.716516249	-0.657768808	481	45
siRNA_12	51464023	XM_497921	LOC391533	0.707973325	-0.657768808	487	45
siRNA_13	51464023	XM_497921	LOC391533	0.60212568	-0.657768808	542	45
siRNA_14	51464023	XM_497921	LOC391533	0.500487004	-0.657768808	592	45
siRNA_15	51464023	XM_497921	LOC391533	0.483258617	-0.657768808	600	45

siRNA_1	194306627	NM_022350	LRAP	1.534842686	-0.651131068	92	46
siRNA_2	194306627	NM_022350	LRAP	1.245073866	-0.651131068	198	46
siRNA_3	194306627	NM_022350	LRAP	1.217931656	-0.651131068	214	46
siRNA_4	194306627	NM_022350	LRAP	1.069791311	-0.651131068	300	46
siRNA_5	194306627	NM_022350	LRAP	0.898688004	-0.651131068	394	46
siRNA_6	194306627	NM_022350	LRAP	0.496488687	-0.651131068	594	46
siRNA_1	295148042	NM_138814	PNPLA5	2.009824369	-0.648982291	23	47
siRNA_2	295148042	NM_138814	PNPLA5	1.340401196	-0.648982291	151	47
siRNA_3	295148042	NM_138814	PNPLA5	0.722069423	-0.648982291	477	47
siRNA_4	295148042	NM_138814	PNPLA5	0.402655472	-0.648982291	636	47
siRNA_1	203097723	NM_001007540	CDH4	2.353563819	-0.610724372	9	48
siRNA_2	203097723	NM_001007540	CDH4	1.303705586	-0.610724372	171	48
siRNA_3	203097723	NM_001007540	CDH4	1.049799523	-0.610724372	309	48
siRNA_4	203097723	NM_001007540	CDH4	0.742236343	-0.610724372	464	48
siRNA_5	203097723	NM_001007540	CDH4	0.678305016	-0.610724372	500	48
siRNA_6	203097723	NM_001007540	CDH4	0.633842918	-0.610724372	525	48
siRNA_1	153267406	NM_001004105	GRK6	1.381471872	-0.608467754	135	49
siRNA_2	153267406	NM_001004105	GRK6	1.295986475	-0.608467754	175	49
siRNA_3	153267406	NM_001004105	GRK6	1.052877719	-0.608467754	307	49
siRNA_4	153267406	NM_001004105	GRK6	0.995013325	-0.608467754	343	49
siRNA_5	153267406	NM_001004105	GRK6	0.94231957	-0.608467754	370	49
siRNA_6	153267406	NM_001004105	GRK6	0.73444076	-0.608467754	470	49
siRNA_7	153267406	NM_001004105	GRK6	0.692837605	-0.608467754	497	49
siRNA_8	153267406	NM_001004105	GRK6	0.609317225	-0.608467754	539	49
siRNA_9	153267406	NM_001004105	GRK6	0.538002469	-0.608467754	568	49
siRNA_1	111038119	NM_139246	TSTD2	1.572940605	-0.586638063	80	50
siRNA_2	111038119	NM_139246	TSTD2	1.344693327	-0.586638063	149	50
siRNA_3	111038119	NM_139246	TSTD2	0.825565184	-0.586638063	423	50
siRNA_1	150378531	NM_017572	MKNK2	1.558729073	-0.585226047	87	51
siRNA_2	150378531	NM_017572	MKNK2	1.240678434	-0.585226047	203	51
siRNA_3	150378531	NM_017572	MKNK2	1.226629055	-0.585226047	209	51
siRNA_4	150378531	NM_017572	MKNK2	1.226305744	-0.585226047	210	51
siRNA_5	150378531	NM_017572	MKNK2	1.193426508	-0.585226047	223	51
siRNA_6	150378531	NM_017572	MKNK2	1.081890151	-0.585226047	292	51
siRNA_7	150378531	NM_017572	MKNK2	1.005230309	-0.585226047	335	51
siRNA_8	150378531	NM_017572	MKNK2	0.808215485	-0.585226047	431	51
siRNA_9	150378531	NM_017572	MKNK2	0.549689433	-0.585226047	565	51
siRNA_10	150378531	NM_017572	MKNK2	0.429271623	-0.585226047	624	51
siRNA_1	83627720	NM_199289	NEK5	1.647202876	-0.576148315	61	52
siRNA_2	83627720	NM_199289	NEK5	1.598695344	-0.576148315	73	52
siRNA_3	83627720	NM_199289	NEK5	1.569363514	-0.576148315	81	52
siRNA_4	83627720	NM_199289	NEK5	1.293139417	-0.576148315	176	52
siRNA_5	83627720	NM_199289	NEK5	1.260775656	-0.576148315	189	52
siRNA_6	83627720	NM_199289	NEK5	1.159760465	-0.576148315	248	52
siRNA_7	83627720	NM_199289	NEK5	1.018987726	-0.576148315	322	52
siRNA_8	83627720	NM_199289	NEK5	1.014739958	-0.576148315	327	52
siRNA_9	83627720	NM_199289	NEK5	0.923239421	-0.576148315	376	52
siRNA_10	83627720	NM_199289	NEK5	0.7523574	-0.576148315	459	52
siRNA_11	83627720	NM_199289	NEK5	0.67353876	-0.576148315	503	52
siRNA_12	83627720	NM_199289	NEK5	0.431370607	-0.576148315	623	52
siRNA_13	83627720	NM_199289	NEK5	0.343114123	-0.576148315	650	52
siRNA_1	8922129	NM_017555	EGLN2	1.795146199	-0.569870075	40	53
siRNA_2	8922129	NM_017555	EGLN2	1.326168058	-0.569870075	156	53
siRNA_3	8922129	NM_017555	EGLN2	0.960600657	-0.569870075	364	53
siRNA_4	8922129	NM_017555	EGLN2	0.721240351	-0.569870075	478	53
siRNA_1	46395495	NM_207172	NPSR1	2.553735396	-0.568965338	5	54
siRNA_2	46395495	NM_207172	NPSR1	1.253759325	-0.568965338	192	54
siRNA_3	46395495	NM_207172	NPSR1	1.218872423	-0.568965338	213	54
siRNA_4	46395495	NM_207172	NPSR1	1.06320531	-0.568965338	301	54
siRNA_5	46395495	NM_207172	NPSR1	1.062023698	-0.568965338	302	54
siRNA_6	46395495	NM_207172	NPSR1	0.716339113	-0.568965338	482	54
siRNA_7	46395495	NM_207172	NPSR1	0.617830157	-0.568965338	537	54
siRNA_8	46395495	NM_207172	NPSR1	0.43865581	-0.568965338	619	54
siRNA_1	14589930	NM_002589	PCDH7	1.156998464	-0.562536377	250	55
siRNA_2	14589930	NM_002589	PCDH7	1.005360572	-0.562536377	334	55

siRNA_3	14589930	NM_002589	PCDH7	0.917135682	-0.562536377	383	55
siRNA_4	14589930	NM_002589	PCDH7	0.824674796	-0.562536377	424	55
siRNA_5	14589930	NM_002589	PCDH7	0.732746655	-0.562536377	471	55
siRNA_6	14589930	NM_002589	PCDH7	0.72683151	-0.562536377	475	55
siRNA_7	14589930	NM_002589	PCDH7	0.646586346	-0.562536377	515	55
siRNA_8	14589930	NM_002589	PCDH7	0.563927795	-0.562536377	556	55
siRNA_9	14589930	NM_002589	PCDH7	0.562335364	-0.562536377	560	55
siRNA_10	14589930	NM_002589	PCDH7	0.506417518	-0.562536377	583	55
siRNA_1	71143111	NM_018124	RFWD3	1.711857807	-0.516885682	52	56
siRNA_2	71143111	NM_018124	RFWD3	1.688811313	-0.516885682	55	56
siRNA_3	71143111	NM_018124	RFWD3	1.549272409	-0.516885682	88	56
siRNA_4	71143111	NM_018124	RFWD3	1.098074499	-0.516885682	283	56
siRNA_5	71143111	NM_018124	RFWD3	0.81686954	-0.516885682	426	56
siRNA_6	71143111	NM_018124	RFWD3	0.50410566	-0.516885682	587	56
siRNA_1	164663809	NM_173642	RIMKLA	1.272879381	-0.495870198	181	57
siRNA_2	164663809	NM_173642	RIMKLA	1.264264006	-0.495870198	185	57
siRNA_3	164663809	NM_173642	RIMKLA	1.125864239	-0.495870198	266	57
siRNA_4	164663809	NM_173642	RIMKLA	1.018287051	-0.495870198	323	57
siRNA_5	164663809	NM_173642	RIMKLA	0.528394155	-0.495870198	573	57
siRNA_6	164663809	NM_173642	RIMKLA	0.446007815	-0.495870198	616	57
siRNA_1	148613885	NM_022841	RXF7	2.037748129	-0.491877781	22	58
siRNA_2	148613885	NM_022841	RXF7	1.581646479	-0.491877781	77	58
siRNA_3	148613885	NM_022841	RXF7	1.004777342	-0.491877781	336	58
siRNA_4	148613885	NM_022841	RXF7	0.674626524	-0.491877781	501	58
siRNA_1	119637838	NM_002498	NEK3	1.826154222	-0.490808551	34	59
siRNA_2	119637838	NM_002498	NEK3	1.683289655	-0.490808551	56	59
siRNA_3	119637838	NM_002498	NEK3	1.209444561	-0.490808551	216	59
siRNA_4	119637838	NM_002498	NEK3	1.087804464	-0.490808551	289	59
siRNA_5	119637838	NM_002498	NEK3	1.085506467	-0.490808551	290	59
siRNA_6	119637838	NM_002498	NEK3	0.926097237	-0.490808551	375	59
siRNA_7	119637838	NM_002498	NEK3	0.918148543	-0.490808551	382	59
siRNA_8	119637838	NM_002498	NEK3	0.660379298	-0.490808551	510	59
siRNA_9	119637838	NM_002498	NEK3	0.520696056	-0.490808551	579	59
siRNA_10	119637838	NM_002498	NEK3	0.495568748	-0.490808551	595	59
siRNA_1	83700236	NM_002521	NPPB	1.431154926	-0.48878903	122	60
siRNA_2	83700236	NM_002521	NPPB	1.385059586	-0.48878903	133	60
siRNA_3	83700236	NM_002521	NPPB	1.004650515	-0.48878903	337	60
siRNA_4	83700236	NM_002521	NPPB	0.568452998	-0.48878903	554	60
siRNA_1	169790812	NM_003239	TGFB3	1.430273075	-0.466900651	123	60
siRNA_2	169790812	NM_003239	TGFB3	1.246960212	-0.466900651	196	60
siRNA_3	169790812	NM_003239	TGFB3	1.127503061	-0.466900651	265	60
siRNA_4	169790812	NM_003239	TGFB3	0.8871071	-0.466900651	397	60
siRNA_5	169790812	NM_003239	TGFB3	0.792483445	-0.466900651	439	60
siRNA_6	169790812	NM_003239	TGFB3	0.445422303	-0.466900651	617	60
siRNA_1	301897246	NM_006296	VRK2	1.414978442	-0.44705262	125	61
siRNA_2	301897246	NM_006296	VRK2	1.1293585	-0.44705262	264	61
siRNA_3	301897246	NM_006296	VRK2	1.093030745	-0.44705262	285	61
siRNA_4	301897246	NM_006296	VRK2	1.072529576	-0.44705262	299	61
siRNA_5	301897246	NM_006296	VRK2	0.967902616	-0.44705262	360	61
siRNA_6	301897246	NM_006296	VRK2	0.849970077	-0.44705262	411	61
siRNA_7	301897246	NM_006296	VRK2	0.59332044	-0.44705262	545	61
siRNA_8	301897246	NM_006296	VRK2	0.487484865	-0.44705262	598	61
siRNA_1	29729561	XM_291105	TADA2B	1.797073701	-0.43824466	39	62
siRNA_2	29729561	XM_291105	TADA2B	1.242947087	-0.43824466	199	62
siRNA_3	29729561	XM_291105	TADA2B	1.129955664	-0.43824466	263	62
siRNA_4	29729561	XM_291105	TADA2B	1.001125383	-0.43824466	339	62
siRNA_5	29729561	XM_291105	TADA2B	0.738257332	-0.43824466	467	62
siRNA_6	29729561	XM_291105	TADA2B	0.466191556	-0.43824466	611	62
siRNA_1	296179393	NM_175866	UHMK1	1.790143354	-0.429462348	43	63
siRNA_2	296179393	NM_175866	UHMK1	1.541156078	-0.429462348	90	63
siRNA_3	296179393	NM_175866	UHMK1	1.149411213	-0.429462348	255	63
siRNA_4	296179393	NM_175866	UHMK1	1.084959901	-0.429462348	291	63
siRNA_5	296179393	NM_175866	UHMK1	1.055465659	-0.429462348	305	63
siRNA_6	296179393	NM_175866	UHMK1	0.888260858	-0.429462348	396	63
siRNA_7	296179393	NM_175866	UHMK1	0.784081619	-0.429462348	444	63

siRNA_8	296179393	NM_175866	UHMK1	0.720569399	-0.429462348	479	63
siRNA_9	296179393	NM_175866	UHMK1	0.622784811	-0.429462348	535	63
siRNA_10	296179393	NM_175866	UHMK1	0.478478481	-0.429462348	601	63
siRNA_1	219521905	NM_145024	CES5A	1.465928115	-0.400890958	111	64
siRNA_2	219521905	NM_145024	CES5A	1.4397527	-0.400890958	118	64
siRNA_3	219521905	NM_145024	CES5A	1.312825347	-0.400890958	165	64
siRNA_4	219521905	NM_145024	CES5A	1.251899589	-0.400890958	193	64
siRNA_5	219521905	NM_145024	CES5A	1.10769192	-0.400890958	278	64
siRNA_6	219521905	NM_145024	CES5A	0.927811752	-0.400890958	374	64
siRNA_7	219521905	NM_145024	CES5A	0.848111365	-0.400890958	414	64
siRNA_8	219521905	NM_145024	CES5A	0.734751839	-0.400890958	468	64
siRNA_9	219521905	NM_145024	CES5A	0.469706338	-0.400890958	610	64
siRNA_10	219521905	NM_145024	CES5A	0.463668465	-0.400890958	612	64
siRNA_1	157779133	NM_001699	AXL	2.088914055	-0.388394121	21	65
siRNA_2	157779133	NM_001699	AXL	1.474431884	-0.388394121	109	65
siRNA_3	157779133	NM_001699	AXL	1.242659362	-0.388394121	201	65
siRNA_4	157779133	NM_001699	AXL	1.226173997	-0.388394121	211	65
siRNA_5	157779133	NM_001699	AXL	1.029267731	-0.388394121	318	65
siRNA_6	157779133	NM_001699	AXL	1.027475191	-0.388394121	319	65
siRNA_7	157779133	NM_001699	AXL	0.867035579	-0.388394121	405	65
siRNA_8	157779133	NM_001699	AXL	0.470560054	-0.388394121	608	65
siRNA_9	157779133	NM_001699	AXL	0.453245889	-0.388394121	615	65
siRNA_10	157779133	NM_001699	AXL	0.396160897	-0.388394121	638	65
siRNA_1	260166660	NM_001003788	STRADA	1.518693118	-0.377588642	96	66
siRNA_2	260166660	NM_001003788	STRADA	1.321007019	-0.377588642	158	66
siRNA_3	260166660	NM_001003788	STRADA	1.107908051	-0.377588642	276	66
siRNA_4	260166660	NM_001003788	STRADA	1.096662324	-0.377588642	284	66
siRNA_5	260166660	NM_001003788	STRADA	1.053219116	-0.377588642	306	66
siRNA_6	260166660	NM_001003788	STRADA	0.918164227	-0.377588642	381	66
siRNA_7	260166660	NM_001003788	STRADA	0.517599833	-0.377588642	581	66
siRNA_8	260166660	NM_001003788	STRADA	0.511125476	-0.377588642	582	66
siRNA_9	260166660	NM_001003788	STRADA	0.422297395	-0.377588642	627	66
siRNA_1	51477706	NM_032037	TSSK6	1.754944419	-0.371602452	48	67
siRNA_2	51477706	NM_032037	TSSK6	1.617500136	-0.371602452	66	67
siRNA_3	51477706	NM_032037	TSSK6	1.609331301	-0.371602452	70	67
siRNA_4	51477706	NM_032037	TSSK6	1.504434072	-0.371602452	101	67
siRNA_5	51477706	NM_032037	TSSK6	0.977440935	-0.371602452	355	67
siRNA_6	51477706	NM_032037	TSSK6	0.863867048	-0.371602452	406	67
siRNA_7	51477706	NM_032037	TSSK6	0.690238203	-0.371602452	498	67
siRNA_8	51477706	NM_032037	TSSK6	0.64306403	-0.371602452	519	67
siRNA_9	51477706	NM_032037	TSSK6	0.594665459	-0.371602452	544	67
siRNA_10	51477706	NM_032037	TSSK6	0.470493175	-0.371602452	609	67
siRNA_1	148922837	NM_021966	TCL1A	1.614594293	-0.363469621	68	68
siRNA_2	148922837	NM_021966	TCL1A	1.196836101	-0.363469621	221	68
siRNA_3	148922837	NM_021966	TCL1A	1.078027829	-0.363469621	294	68
siRNA_4	148922837	NM_021966	TCL1A	0.964472355	-0.363469621	362	68
siRNA_5	148922837	NM_021966	TCL1A	0.562771481	-0.363469621	557	68
siRNA_6	148922837	NM_021966	TCL1A	0.488895762	-0.363469621	597	68
siRNA_1	32130539	NM_012395	PFTK1	3.679457131	-0.355102121	2	69
siRNA_2	32130539	NM_012395	PFTK1	1.605848237	-0.355102121	71	69
siRNA_3	32130539	NM_012395	PFTK1	1.280198373	-0.355102121	180	69
siRNA_4	32130539	NM_012395	PFTK1	1.156425744	-0.355102121	251	69
siRNA_5	32130539	NM_012395	PFTK1	0.928252727	-0.355102121	373	69
siRNA_6	32130539	NM_012395	PFTK1	0.90741366	-0.355102121	387	69
siRNA_7	32130539	NM_012395	PFTK1	0.50385149	-0.355102121	588	69
siRNA_8	32130539	NM_012395	PFTK1	0.397143467	-0.355102121	637	69
siRNA_9	32130539	NM_012395	PFTK1	0.35429567	-0.355102121	647	70
siRNA_1	38569459	NM_015191	SIK2	1.589105472	-0.35411698	74	70
siRNA_2	38569459	NM_015191	SIK2	1.501025346	-0.35411698	103	70
siRNA_3	38569459	NM_015191	SIK2	1.371227875	-0.35411698	141	70
siRNA_4	38569459	NM_015191	SIK2	1.216852924	-0.35411698	215	70
siRNA_5	38569459	NM_015191	SIK2	0.967598904	-0.35411698	361	70
siRNA_6	38569459	NM_015191	SIK2	0.806986283	-0.35411698	432	70
siRNA_7	38569459	NM_015191	SIK2	0.801532323	-0.35411698	433	70
siRNA_8	38569459	NM_015191	SIK2	0.795734006	-0.35411698	437	70

siRNA_9	38569459	NM_015191	SIK2	0.662804405	-0.35411698	509	70
siRNA_10	38569459	NM_015191	SIK2	0.566684641	-0.35411698	555	70
siRNA_11	38569459	NM_015191	SIK2	0.369000524	-0.35411698	646	70
siRNA_1	112363079	NM_015112	MAST2	1.902074056	-0.349873842	29	71
siRNA_2	112363079	NM_015112	MAST2	1.523056537	-0.349873842	95	71
siRNA_3	112363079	NM_015112	MAST2	1.504534927	-0.349873842	100	71
siRNA_4	112363079	NM_015112	MAST2	1.102217671	-0.349873842	281	71
siRNA_5	112363079	NM_015112	MAST2	0.900759309	-0.349873842	392	71
siRNA_6	112363079	NM_015112	MAST2	0.898728246	-0.349873842	393	71
siRNA_7	112363079	NM_015112	MAST2	0.701302595	-0.349873842	491	71
siRNA_8	112363079	NM_015112	MAST2	0.62465674	-0.349873842	533	71
siRNA_9	112363079	NM_015112	MAST2	0.379812373	-0.349873842	641	71
siRNA_1	51466088	XR_000292	CLK2	2.264122037	-0.30200523	13	72
siRNA_2	51466088	XR_000292	CLK2	0.977669393	-0.30200523	354	72
siRNA_3	51466088	XR_000292	CLK2	0.938147199	-0.30200523	371	72
siRNA_4	51466088	XR_000292	CLK2	0.901917188	-0.30200523	391	72
siRNA_5	51466088	XR_000292	CLK2	0.713501806	-0.30200523	484	72
siRNA_6	51466088	XR_000292	CLK2	0.629117267	-0.30200523	530	72
siRNA_7	51466088	XR_000292	CLK2	0.537765636	-0.30200523	571	72
siRNA_8	51466088	XR_000292	CLK2	0.473163843	-0.30200523	606	72
siRNA_9	51466088	XR_000292	CLK2	0.453321873	-0.30200523	614	72
siRNA_1	191252790	NM_020341	PAK7	1.173185436	-0.267400888	238	73
siRNA_2	191252790	NM_020341	PAK7	1.170745476	-0.267400888	241	73
siRNA_3	191252790	NM_020341	PAK7	1.154630424	-0.267400888	252	73
siRNA_4	191252790	NM_020341	PAK7	1.107776163	-0.267400888	277	73
siRNA_5	191252790	NM_020341	PAK7	0.962958956	-0.267400888	363	73
siRNA_6	191252790	NM_020341	PAK7	0.830056562	-0.267400888	422	73
siRNA_7	191252790	NM_020341	PAK7	0.765038684	-0.267400888	452	73
siRNA_8	191252790	NM_020341	PAK7	0.635958384	-0.267400888	524	73
siRNA_9	191252790	NM_020341	PAK7	0.555556408	-0.267400888	563	73
siRNA_10	191252790	NM_020341	PAK7	0.394092133	-0.267400888	639	73
siRNA_1	133908632	NM_001024847	TGFBR2	1.286741761	-0.263746657	178	74
siRNA_2	133908632	NM_001024847	TGFBR2	1.180481097	-0.263746657	233	74
siRNA_3	133908632	NM_001024847	TGFBR2	1.075801744	-0.263746657	296	74
siRNA_4	133908632	NM_001024847	TGFBR2	0.777686083	-0.263746657	445	74
siRNA_5	133908632	NM_001024847	TGFBR2	0.766644208	-0.263746657	450	74
siRNA_6	133908632	NM_001024847	TGFBR2	0.672239255	-0.263746657	504	74
siRNA_7	133908632	NM_001024847	TGFBR2	0.484127263	-0.263746657	599	74
siRNA_8	133908632	NM_001024847	TGFBR2	0.471778091	-0.263746657	607	74
siRNA_9	133908632	NM_001024847	TGFBR2	0.434499025	-0.263746657	620	74
siRNA_1	62952497	NM_020397	CAMK1D	1.511697521	-0.222934146	98	75
siRNA_2	62952497	NM_020397	CAMK1D	1.209033384	-0.222934146	217	75
siRNA_3	62952497	NM_020397	CAMK1D	1.193051522	-0.222934146	224	75
siRNA_4	62952497	NM_020397	CAMK1D	1.117502052	-0.222934146	269	75
siRNA_5	62952497	NM_020397	CAMK1D	1.006825915	-0.222934146	333	75
siRNA_6	62952497	NM_020397	CAMK1D	0.902293089	-0.222934146	389	75
siRNA_7	62952497	NM_020397	CAMK1D	0.696466618	-0.222934146	495	75
siRNA_8	62952497	NM_020397	CAMK1D	0.687568272	-0.222934146	499	75
siRNA_9	62952497	NM_020397	CAMK1D	0.431676755	-0.222934146	622	75
siRNA_10	62952497	NM_020397	CAMK1D	0.420750345	-0.222934146	629	75
siRNA_11	62952497	NM_020397	CAMK1D	0.413023585	-0.222934146	633	75
siRNA_1	126513135	NM_002773	PRSS8	2.190266502	-0.193268778	16	76
siRNA_2	126513135	NM_002773	PRSS8	1.207313113	-0.193268778	218	76
siRNA_3	126513135	NM_002773	PRSS8	0.521937302	-0.193268778	577	76
siRNA_4	126513135	NM_002773	PRSS8	0.318784605	-0.193268778	653	76
siRNA_5	126513135	NM_002773	PRSS8	0.291712399	-0.193268778	658	76
siRNA_6	126513135	NM_002773	PRSS8	0.253848445	-0.193268778	662	76
siRNA_1	207029159	NM_002732	PRKACG	1.171996402	-0.190556998	240	77
siRNA_2	207029159	NM_002732	PRKACG	1.004112507	-0.190556998	338	77
siRNA_3	207029159	NM_002732	PRKACG	0.996041724	-0.190556998	342	77
siRNA_4	207029159	NM_002732	PRKACG	0.974632251	-0.190556998	356	77
siRNA_5	207029159	NM_002732	PRKACG	0.70597353	-0.190556998	489	77
siRNA_6	207029159	NM_002732	PRKACG	0.699161822	-0.190556998	493	77
siRNA_7	207029159	NM_002732	PRKACG	0.562031361	-0.190556998	561	77
siRNA_8	207029159	NM_002732	PRKACG	0.373069918	-0.190556998	643	77

siRNA_1	156151440	NM_025144	ALPK1	1.305512706	-0.180276614	169	78
siRNA_2	156151440	NM_025144	ALPK1	1.009464565	-0.180276614	332	78
siRNA_3	156151440	NM_025144	ALPK1	0.708760689	-0.180276614	485	78
siRNA_4	156151440	NM_025144	ALPK1	0.667756802	-0.180276614	506	78
siRNA_5	156151440	NM_025144	ALPK1	0.537876835	-0.180276614	569	78
siRNA_6	156151440	NM_025144	ALPK1	0.440759841	-0.180276614	618	78
siRNA_7	156151440	NM_025144	ALPK1	0.428388214	-0.180276614	625	78
siRNA_1	254675141	NM_173649	C2orf61	1.304801997	-0.159410796	170	79
siRNA_2	254675141	NM_173649	C2orf61	0.728011152	-0.159410796	474	79
siRNA_3	254675141	NM_173649	C2orf61	0.644233348	-0.159410796	517	79
siRNA_4	254675141	NM_173649	C2orf61	0.474137913	-0.159410796	605	79
siRNA_1	38016933	NM_003733	OASL	1.283313016	-0.143200994	179	80
siRNA_2	38016933	NM_003733	OASL	1.169204795	-0.143200994	242	80
siRNA_3	38016933	NM_003733	OASL	0.96905856	-0.143200994	359	80
siRNA_4	38016933	NM_003733	OASL	0.755092049	-0.143200994	455	80
siRNA_5	38016933	NM_003733	OASL	0.458193197	-0.143200994	613	80
siRNA_6	38016933	NM_003733	OASL	0.256488001	-0.143200994	661	80
siRNA_1	187937178	NM_001080395	AATK	1.365568547	-0.133862349	145	81
siRNA_2	187937178	NM_001080395	AATK	1.341122148	-0.133862349	150	81
siRNA_3	187937178	NM_001080395	AATK	1.206349735	-0.133862349	219	81
siRNA_4	187937178	NM_001080395	AATK	1.101410694	-0.133862349	282	81
siRNA_5	187937178	NM_001080395	AATK	0.877859732	-0.133862349	402	81
siRNA_6	187937178	NM_001080395	AATK	0.831815373	-0.133862349	419	81
siRNA_7	187937178	NM_001080395	AATK	0.775360813	-0.133862349	448	81
siRNA_8	187937178	NM_001080395	AATK	0.630078873	-0.133862349	527	81
siRNA_9	187937178	NM_001080395	AATK	0.5495295	-0.133862349	566	81
siRNA_10	187937178	NM_001080395	AATK	0.474440054	-0.133862349	604	81
siRNA_11	187937178	NM_001080395	AATK	0.30282682	-0.133862349	656	81
siRNA_1	21536426	NM_030753	WNT3	1.493373194	-0.102966609	105	82
siRNA_2	21536426	NM_030753	WNT3	1.435433224	-0.102966609	121	82
siRNA_3	21536426	NM_030753	WNT3	1.366591619	-0.102966609	144	82
siRNA_4	21536426	NM_030753	WNT3	1.038968586	-0.102966609	316	82
siRNA_5	21536426	NM_030753	WNT3	0.617419181	-0.102966609	538	82
siRNA_6	21536426	NM_030753	WNT3	0.589339616	-0.102966609	547	82
siRNA_7	21536426	NM_030753	WNT3	0.52572397	-0.102966609	576	82
siRNA_8	21536426	NM_030753	WNT3	0.518753642	-0.102966609	580	82
siRNA_9	21536426	NM_030753	WNT3	0.337931009	-0.102966609	651	82
siRNA_10	21536426	NM_030753	WNT3	0.310866103	-0.102966609	655	82
siRNA_1	187171274	NM_030662	MAP2K2	1.242780364	-0.102742714	200	83
siRNA_2	187171274	NM_030662	MAP2K2	1.108309193	-0.102742714	275	83
siRNA_3	187171274	NM_030662	MAP2K2	1.049859399	-0.102742714	308	83
siRNA_4	187171274	NM_030662	MAP2K2	0.973694581	-0.102742714	357	83
siRNA_5	187171274	NM_030662	MAP2K2	0.813816654	-0.102742714	427	83
siRNA_6	187171274	NM_030662	MAP2K2	0.762298251	-0.102742714	454	83
siRNA_7	187171274	NM_030662	MAP2K2	0.59962819	-0.102742714	543	83
siRNA_8	187171274	NM_030662	MAP2K2	0.415203013	-0.102742714	632	83
siRNA_9	187171274	NM_030662	MAP2K2	0.405806657	-0.102742714	635	83
siRNA_10	187171274	NM_030662	MAP2K2	0.369279447	-0.102742714	645	83
siRNA_11	187171274	NM_030662	MAP2K2	0.344328345	-0.102742714	649	83
siRNA_1	103471998	NM_006281	STK3	0.922339599	-0.073161481	377	84
siRNA_2	103471998	NM_006281	STK3	0.799592168	-0.073161481	434	84
siRNA_3	103471998	NM_006281	STK3	0.777493426	-0.073161481	446	84
siRNA_4	103471998	NM_006281	STK3	0.734468405	-0.073161481	469	84
siRNA_5	103471998	NM_006281	STK3	0.627302234	-0.073161481	531	84
siRNA_6	103471998	NM_006281	STK3	0.528102536	-0.073161481	574	84
siRNA_7	103471998	NM_006281	STK3	0.477356191	-0.073161481	603	84
siRNA_8	103471998	NM_006281	STK3	0.417787047	-0.073161481	630	84
siRNA_9	103471998	NM_006281	STK3	0.416993665	-0.073161481	631	84
siRNA_10	103471998	NM_006281	STK3	0.337423476	-0.073161481	652	84
siRNA_1	51467060	XM_498294	CDK5	1.792060293	-0.058115447	42	85
siRNA_2	51467060	XM_498294	CDK5	1.409267192	-0.058115447	128	85
siRNA_3	51467060	XM_498294	CDK5	1.263717272	-0.058115447	187	85
siRNA_4	51467060	XM_498294	CDK5	1.010101208	-0.058115447	331	85
siRNA_5	51467060	XM_498294	CDK5	0.917023396	-0.058115447	384	85
siRNA_6	51467060	XM_498294	CDK5	0.886833927	-0.058115447	398	85

siRNA_7	51467060	XM_498294	CDK5	0.749160966	-0.058115447	461	85
siRNA_8	51467060	XM_498294	CDK5	0.65581623	-0.058115447	514	85
siRNA_9	51467060	XM_498294	CDK5	0.637040027	-0.058115447	522	85
siRNA_10	51467060	XM_498294	CDK5	0.629201067	-0.058115447	529	85
siRNA_11	51467060	XM_498294	CDK5	0.57420218	-0.058115447	552	85
siRNA_12	51467060	XM_498294	CDK5	0.526692983	-0.058115447	575	85
siRNA_13	51467060	XM_498294	CDK5	0.384178158	-0.058115447	640	85
siRNA_14	51467060	XM_498294	CDK5	0.311310612	-0.058115447	654	85
siRNA_15	51467060	XM_498294	CDK5	0.277413616	-0.058115447	660	85
siRNA_1	211938419	NM_002898	RBMS2	1.000782772	-0.049844941	340	86
siRNA_2	211938419	NM_002898	RBMS2	0.722962985	-0.049844941	476	86
siRNA_3	211938419	NM_002898	RBMS2	0.589266685	-0.049844941	548	86
siRNA_4	211938419	NM_002898	RBMS2	0.478108127	-0.049844941	602	86
siRNA_5	211938419	NM_002898	RBMS2	0.345652981	-0.049844941	648	86

siRNA: indicates the number of siRNAs targeting each gene
Gene_ID: Entrez GeneID
Accession #: RefSeq database identification
siRNA_Score: Normalized fold activation of individual siRNAs
RSA_LogP: Evidence based score based on RSA analysis
Cut-off_Rank: Ranking of individual siRNA based on activity
RSA_Rank: Rank of genes based on activities of all targeting siRNAs

Supplementary Table 2. RSA rank comparison.

accession #	Symbol	RSA Rank_CMV-luc_screen	RSA Rank_FXN-luc_screen
NM_002742	PRKD1	38	1
NM_016507	CDK12	27	2
NM_002972	SBF1	32	3
NM_005856	RAMP3	3	4
NM_003718	CDK13	65	5
NM_153361	NIM1	9	6
NM_030903	OR2W1	1	7
NM_005693	NR1H3	68	8
NM_005995	TBX10	60	9
NM_183379	PRSS41	31	10
NM_002187	IL12B	67	11
NM_004884	PUNC	28	12
NM_006385	ZNF211	25	13
NM_138499	PWWP2B	21	14
NM_000614	CNTF	76	15
NM_001334	CTSO	36	16
NM_152386	SGPP2	7	17
NM_003810	TNFSF10	83	18
NM_001014796	DDR2	54	19
NM_199046	TEPP	58	20
NM_003322	TULP1	63	21
NM_014565	OR1A1	5	22
NM_014787	DNAJC6	44	23
NM_001004023	DYRK3	55	24
NM_006622	PLK2	37	25
XM_291141	MAST4	23	26
NM_004438	EPHA4	26	27
NM_152271	LONRF1	22	28
NM_020967	NCOA5	29	29
XM_208545	BMPR1A	78	30
NM_004635	MAPKAPK3	47	31
NM_001954	DDR1	85	32
NM_152999	STEAP2	49	33
NM_007170	TESK2	64	34
NM_001065	TNFRSF1A	66	35
NM_001495	GFRA2	10	36
NM_022048	CSNK1G1	69	37
NM_052841	TSSK3	16	38
NM_032242	PLXNA1	43	39
NM_001001667	OR6V1	70	40
NM_001433	ERN1	34	41
NM_006754	SYPL1	59	42
NM_000215	JAK3	84	43
NM_005164	ABCD2	71	44

XM_497921	LOC391533	77	45
NM_022350	LRAP	56	46
NM_138814	PNPLA5	18	47
NM_001007540	CDH4	15	48
NM_001004105	GRK6	11	49
NM_139246	TSTD2	74	50
NM_017572	MKNK2	6	51
NM_199289	NEK5	53	52
NM_017555	EGLN2	17	53
NM_207172	NPSR1	4	54
NM_002589	PCDH7	12	55
NM_018124	RFWD3	40	56
NM_173642	RIMKLA	19	57
NM_022841	RXF7	57	58
NM_002498	NEK3	33	59
NM_002521	NPPB	2	60
NM_003239	TGFB3	72	61
NM_006296	VRK2	61	62
XM_291105	TADA2B	42	63
NM_175866	UHMK1	50	64
NM_145024	CES5A	39	65
NM_001699	AXL	46	66
NM_001003788	STRADA	30	67
NM_032037	TSSK6	75	68
NM_021966	TCL1A	48	69
NM_012395	PFTK1	62	70
NM_015191	SIK2	8	71
NM_015112	MAST2	79	72
XR_000292	CLK2	35	73
NM_020341	PAK7	13	74
NM_001024847	TGFBR2	87	75
NM_020397	CAMK1D	14	76
NM_002773	PRSS8	86	77
NM_002732	PRKACG	82	78
NM_025144	ALPK1	52	79
NM_173649	C2orf61	45	80
NM_003733	OASL	24	81
NM_001080395	AATK	51	82
NM_030753	WNT3	80	83
NM_030662	MAP2K2	73	84
NM_006281	STK3	88	85
XM_498294	CDK5	20	86
XM_498294	CDK5	41	86
NM_002898	RBMS2	81	87

Supplementary Table 3. Primer and shRNA target sequences used in this study

qRT-PCR primers

Transcript	Forward	Reverse	Internal #
GAPDH	ACTCCTCCACCTTTGACGC	GTTGCTGTAGCCAAATTCGTT	mb1402 / 1403
FXN	ACAAGCAGACGCCAAACAAGCA	ACCCAGTTTTTCCCAGTCCAGTCA	mb3675 / 3676
PRKD1	ACGTCCGCGAGATGGCTTGC	TCACTGGCCGCTTTCACCAGC	mb3663 / 3664
CDK12	AGAGCCAGCAGGCAGTCTGGAG	CAGGGGGCCTCTTCTCTGGTGG	mb3665 / 3666
SBF1	CAGCCCACATGGGCAGTCACG	GGTCCGGACACTGCCCCACT	mb3667 / 3668
RAMP3	GATGGGCAAGGTGGACGTCTGG	CAGGGGGTTGGGCCAGTAGCA	mb3878 / 3879
CDK13	GCTGTGGCTGTATCCTTGGCG	GGCCACACTGCAGGACATGGAC	mb3888 / 3889

RNAi

Transcript	shRNA	target sequence		Reference
FXN	shFXN	AACGUGGCCUCAACCAGAUUU	pMB762	Sequence used in Lu et al., Biochimica et Biophysica Acta 2009
Firefly Luciferase	shFirefly	CUUACGCUGAGUACUUCGA	pMB763	Sequence used in Elbashir et al., Nature 2001
Renilla Luciferase	shRenilla	AAACAUGCAGAAAAUGCUGTT		Sequence used in Elbashir et al., Nature 2001
PRKD1	shPRKD1	GTTCCCTGAATGTGGTTTC	pMB896	
SBF1	shSBF1		pMB957	Thermo scientific RHS4430_200262153 (cloneId V3LHS_385794)
Negative control	non-silencing shRNA (scramble)		pMB049	Thermo scientific RHS4346

TAP – MS bench protocol (Rodrigo Villaseñor, FMI 2015)

General points

- Work during the isolation of protein complexes in cold room at 4°C.
- Replace buffers (stock solutions) every 6 months to obtain optimal results.

Where should be everything?

	Where?	Master stock
TrypLE	Tissue culture room, shelf	Tissue culture room
DMEM	Tissue culture room, fridge	Cold room
PBS	Tissue culture room, fridge	Tissue culture room
bMe	Bottom drawer; dilution in fridge	
LIF	-20 °C in a box (LIF enriched medium)	
mES FBS	50 ml aliquots (already heat inactivated) in 3rd drawer of -20 °C => should be thawed @ 4°C o/n	Heat 56°C for 30 min then aliquot (45 ml)
Na pyruvate	Fridge	Cold room
Trypsin	Fridge	Freezer
PBS	Fridge	Tissue culture room
Pen strep	Unlabelled; 3rd drawer of -20 °C	Ground floor -20 °C
Trypsin	(V5111, Promega). Request at PAF, store at -20°C	10-ul aliquots [0.1 ug/ul]
TCEP	Request at PAF, store at -20°C	

Buffers

TAP lysis buffer

Component	Final concentration
NaCl	150 mM
Tris-HCl (pH 7.5)	20 mM
NP-40	0.5% (vol/vol)
EDTA	1 mM
Glycerol	10% (vol/vol)
DTT	1 mM (add fresh)

Digest buffer

Component	Final concentration
Tris-HCl (pH 8.0)	50 mM
CaCl	1 mM
TCEP	1 mM
Trypsin	0.2 ug (per sample)

Maintenance Of Cells

Everything you need to know about maintenance of mouse ES cells can be found in the “cell cuture bench protocol” from Alex Tuck (Bühler lab protocols). Read it!

Passaging And Expanding Cells

1. With mES cells, aim for ~70% confluency. Passage the cells before medium gets orange (**do not** let medium get yellowish as it will affect the viability of your cells). Usually split 1:3-1:5. But to expand them, take all from a 10-cm dish and put into two 15-cm dishes. Then, split the cells with 1:2 ratio into 15-cm dishes. Now you have 4 × 15-cm dishes for each cell line and ready for TAP-purification.
   Note: The following numbers assume the passaging of cells in 15-cm culture dishes.
2. Aspirate medium with Pasteur (flamed).
3. Add 15 ml PBS, tilt plate a couple of times, then aspirate.
4. Washing step is not needed when using TrypLE reagent but recommended. The crucial advantage of TrypLE over conventional trypsin is that it is very gentle on cells and dilution alone inactivates TrypLE, avoiding the need for trypsin inhibitors.
5. Add 2 ml TrypLE, then put at 37 °C for 1-2 min.
6. Add 10 ml pre-warmed PBS. Resuspend cells gently on plate to break clumps.
7. Repeat step to collect all cells.
8. Transfer to 50-ml Falcon tube.
9. Spin at 1,000 rpm for 3-5 min.
10. Aspirate supernatant, and resuspend in 10-ml of new medium.
11. Optional – count cells (See Alex´s cell culture bench protocol).
12. Take appropriate amount of cells and add to new 15-cm plate (gelatin-coated) containing pre-warmed ES medium.

Note: mES cells require fresh media <u>every</u> day.

Protein Extraction

13. ES cells should be ~80% confluent on the day of protein extraction. One 15cm dish per sample usually yields sufficient material for one TAP-purification.
14. Follow the steps 2-8 to harvest cells by trypsinization.
15. Keep cells on ice during the next steps!
16. Centrifuge the cells in 50-ml conical tubes for 15 min using an Eppendorf centrifuge at 4 °C and 2,400g.
17. Carefully decant the supernatant, resuspend the cell pellet with 10 ml of ice-cold PBS (from media kitchen). Optional: count cells.
18. Centrifuge the cells in 50-ml conical tubes for 10 min using an Eppendorf centrifuge at 4 °C and 2,400g.
19. Carefully decant the supernatant and aspirate the remaining supernatant with a disposable 0.2-ml tip. Change tip for every sample.
20. Resuspend with 1 ml of ice-cold TAP lysis buffer with freshly added 1 mM DTT (final conc.) and protease inhibitor cocktail (for 50-ml buffer, dissolve 1 complete EDTA-free tablet). I usually prepared the supplemented TAP lysis buffer during the washing steps. See first page to prepare TAP lysis buffers.
21. Transfer lysate to a cold 1.5-ml tube. Shake tube at 1,000 rpm in cold room for 30 min.
22. Clear lysate by centrifugation at 4 °C and max. speed.
23. Carefully take supernatant (try not to take pieces from pellet) and transfer to a new cold 1.5-ml tube.
24. Determine the protein concentration of each sample using the Bradford assay (Bio-Rad Dye).

Bradford Assay

- Make BSA standards:

2 mg/ml BSA	H_2O	Final conc (ug/ml)
5 ul	495 ul	20
5 ul	245 ul	40
6 ul	194 ul	60
5 ul	120 ul	80
5 ul	95 ul	100
7.5 ul	92.5 ul	150

- Add 3x 10 ul of each standard, and 3x 10 ul of each of your samples (diluted 1:50 in water) to a 96-well plate, then add 200 ul of working conc Bradford reagent (stock diluted 1 in 5 in water)
- Incubate 5 min and read absorbance (595 nm) in plate reader.

25. Use equal amounts of lysate (as determined by Bradford assay) from the control sample (parental cells, RosaB) and the bait-protein sample (gene tagged cells). Transfer required input material to new tube.
    Critical step: Optimal results are obtained with 5 mg of input material for some proteins. However, input amount might need adjustment. Equal amount of input material for ALL samples is important for MS-data analysis. A good starting point is 5 mg of input material, but can be adjusted depending on the expression level of bait-protein.

26. Add an appropriate amount of cold TAP lysis buffer to each sample to adjust the final sample concentration to be ~5 mg/ml.
27. This should be carried out just before the beads are required. Equilibrate FLAG M2 beads (10 µl of packed beads ~ 20 ul of slurry per 5 mg protein / sample) in tubes containing 1 ml of cold TAP lysis buffer. Rotate in cold room for 5-10 min, place tube on magnet and decant buffer. Repeat once. Resuspend packed beads in 50 µl of cold TAP lysis.
28. Transfer 50 ul of pre-equilibrated FLAG M2 beads to pre-cleared lysate from step 26. Place tube on end-over-end rotating wheel at 4°C overnight.

Tandem Affinity Purification Of Protein Complexes (TAP)

29. Place tubes from overnight incubation on magnet and remove supernatant (unbound material). Optional: aliquot from unbound material can be loaded on gel to check purification efficiency.
30. Wash FLAG-beads four times with 1ml of ice-cold TAP lysis buffer (add protease inhibitors). Rotate tubes at 4°C for 15 min per wash (first wash can be done for 5 min).
31. After the final wash, remove most of supernatant with 200-ul tip.
32. Elute bound material from beads with 250 ul of 0.3 mg/ml FLAG peptide. Elute beads four times, each time rotating tubes for 1-1.5 h at 4°C. Combine total 1 ml of each sample. Discard beads. Optional: aliquot of discarded beads can be boiled and loaded on gel to check how much material remains bound.
33. Proceed with streptavidin purification immediately.
34. Prepare streptavidin beads: equilibrate streptavidin dynabeads (10 µl of packed beads ~ 20 ul of

slurry per 5 mg protein / sample) in tubes containing 1 ml of cold TAP lysis buffer. Rotate in cold room for 5-10 min, place tube on magnet and decant buffer. Repeat once. Resuspend packed beads in 50 µl of cold TAP lysis.

35. Add 50 ul of pre-equilibrated streptavidin dynabeads to 1 ml of FLAG-eluate from step 32 and place on end-over-end rotating wheel at 4°C overnight.
36. Place tubes from overnight incubation on magnet and remove supernatant (unbound material).
37. Wash FLAG-beads two times with 1ml of ice-cold TAP lysis buffer (add protease inhibitors). Rotate tubes at 4°C for 15 min per wash (first wash can be done for 5 min).
38. After the final wash, remove all supernatant with 200-ul tip.
39. **Pause Point:** quick-freeze samples in liquid N_2 at this point and store at – 80°C (fine for weeks)

Trypsin-Digest Of Protein Complexes On Beads

Note: I recommend doing this step one day before MS analysis. Plan accordingly with PAF facility.

40. Wash beads two times with ice-cold TAP lysis buffer (without NP-40 detergent!!!, add protease inhibitors). Rotate tubes at 4°C for 15 min per wash.
41. Resuspend beads in 20-ul of Digest buffer with added Trypsin
42. Digest protein complexes over-night at 37°C
43. Next day: samples are ready for further processing.