Identification of Human Pathways Acting on Nuclear Non-Coding RNAs Using the Mirror Forward Genetic Approach

Rui Che[1, 2 §], Monireh Panah[1, 2 §], Bhoomi Mirani[1, 2], Krista Knowles[1, 2], Anastacia Ostapovich[3], Debarati Majumdar[1, 2], Xiaotong Chen[1], Joseph DeSimone[1], William White[1], Megan Noonan[1], Hong Luo[1], and Andrei Alexandrov[1, 2 *]

[1]Dept. of Genetics and Biochemistry, Clemson University, Clemson, SC 29631, USA

[2]Clemson University Center for Human Genetics, Greenwood, SC 29646, USA

[3]Dept. of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06536, USA

§ authors contributed equally

* correspondence: andreia@clemson.edu

SUMMARY

Despite critical roles in diseases, human pathways acting on strictly nuclear non-coding RNAs have been refractory to forward genetics. To enable their forward genetic discovery, we developed a single-cell approach that “Mirrors” activities of nuclear pathways with cytoplasmic fluorescence. Application of Mirror to two nuclear pathways targeting MALAT1’s 3′ end, the pathway of its maturation and the other, the degradation pathway blocked by the triple-helical Element for Nuclear Expression (ENE), identified nearly all components of three complexes: Ribonuclease P and the RNA Exosome, including nuclear DIS3, EXOSC10, and C1D, as well as the Nuclear Exosome Targeting (NEXT) complex. Additionally, Mirror identified DEAD-box helicase DDX59 associated with the genetic disorder Oral-Facial-Digital syndrome (OFD), yet lacking known substrates or roles in nuclear RNA degradation. Knockout of DDX59 exhibits stabilization of the full-length MALAT1 with a stability-compromised ENE and increases levels of such long non-coding RNAs as NEAT1_1 and NIPBL-DT, as well as 3′-extended forms of small nuclear RNAs. It also exhibits extensive retention of minor introns, including in OFD-associated genes, suggesting a mechanism for DDX59 association with OFD. Mirror efficiently identifies pathways acting on strictly nuclear non-coding RNAs, including essential and indirectly-acting components, and, as a result, uncovers unexpected links to human disease.

INTRODUCTION

Nuclear-localized long non-coding RNAs (lncRNAs) carrying non-canonical 3′-ends with structured triple-helical Elements for Nuclear Expression (ENEs) (**Fig**. **1C**)[1-3] play critical roles in a variety of human diseases, including cancers[4,5] as well as developmental[6] and viral disorders[7,8]. Overexpression of the abundant, stable, vertebrate-specific, 5′-capped nuclear lncRNA Metastasis Associated Lung Adenocarcinoma Transcript 1 (MALAT1)[1,2,9,10] promotes tumor growth by increasing cell proliferation, invasion, and metastasis. It correlates with poor survival in such prevalent cancers as lung, pancreatic, cervical, colorectal, and others[11-16]. The 3′ end of the nuclear lncRNA MALAT1 is not polyadenylated; instead, it is formed by RNase P cleavage (**Fig**. **1A**) upstream of a tRNA-like RNA structure, mascRNA[17,18]. Since downregulation of MALAT1 RNA levels decreases tumor growth and metastasis and promotes cell differentiation[19,20], this lncRNA represents a promising target for cancer treatment[4,10,20-24]. The ENE-containing human multiple endocrine neoplasia (MEN-β) lncRNA (the latter of the NEAT1_1 and NEAT1_2 isoforms), which is a component of nuclear paraspeckles, has also been implicated in a variety of cancers and diseases[5,25-28].

For each of these lncRNAs, a structured triple-helical ENE element (**Fig**. **1C**) has been shown to protect the 3′ end from nuclear RNA degradation machinery[1-3,29]. This 3′ end-acting nuclear degradation machinery is efficient, as the lncRNAs' half-life and steady-state levels critically depend on the protective function of the intact ENE structure[1,2]: a single-nucleotide destabilizing substitution C8351G within the MALAT1 ENE (**Fig**. **1B-C**) results in a 20-fold reduction in MALAT1 levels, whereas the deletion of the entire 80-nucleotide ENE element leads to their dramatic 60-fold decrease[18].

Despite the profound effect of nuclear pathways on steady-state levels of nuclear lncRNAs, and despite the major impact of these steady-state levels on human diseases, it is unknown what human nuclear degradation pathways are blocked from degrading nuclear lncRNAs by the 3′ triple-helical ENE elements. The reason such 3′-end targeting human pathways have not been interrogated using forward genetics is due to the strictly nuclear localization of the respective ENE-containing lncRNAs. Whereas forward genetic interrogation must use lncRNA abundance or processing efficiency as quantifiable screening readouts, their quantification can only be achieved in relatively low throughput. This stems from the fact that, unlike mRNAs, nuclear lncRNAs are not exported into cytoplasm and translated, precluding the use of translation-dependent reporter systems.

Such a technological barrier exists not only for the above-mentioned nuclear lncRNAs, but for nearly 30% of all human non-coding RNAs[30] due to their nuclear localization. Due to the lack of robust genome-wide forward genetic approaches, most of our knowledge about human nuclear RNA pathways and their components is derived from non-genetic approaches and studies in model organisms. This

leads to a gap in knowledge about human-specific nuclear pathways, contributing to incomplete understanding of human diseases.

We overcame these technological barriers by developing the forward genetic Mirror approach, which we applied to identify components of two human nuclear pathways targeting the 3′ end of MALAT1. One of them is the pathway of 3′ end maturation of MALAT1 and the other is the degradation pathway blocked by the triple-helical ENE stability element. Genome-wide Mirror screening identified numerous components of the three major nuclear RNA-processing complexes: Ribonuclease P (RNase P), the nuclear RNA Exosome, and the Nuclear Exosome Targeting (NEXT) complex, as well as three additional genes: DDX59, C1D, and BRF2. We found that the DEAD box helicase DDX59, which lacked a known substrate or role in nuclear RNA degradation, influences nuclear RNA degradation through an extensive retention of U12 introns within components of the RNA Exosome and the NEXT complex, suggesting a previously unknown function of DDX59 in minor intron splicing. Additional retained U12 introns affect genes associated with the rare genetic developmental disorder Oral-Facial-Digital (OFD) syndrome, suggesting a mechanism for DDX59 association with OFD. We further observe effects of DDX59 on endogenous snRNAs and lncRNAs, including NEAT1 isoforms and NIPBL-DT, the latter transcribed divergently from the mRNA of the Cornelia de Lange Syndrome-associated NIPBL. Collectively, we demonstrate that the Mirror forward genetic approach efficiently identifies components of nuclear pathways acting on human nuclear lncRNAs, revealing surprising connections to human disease.

RESULTS

Design and implementation of the Mirror approach for forward genetic discovery of nuclear factors targeting 3′ end of the human MALAT1. Since translation-dependent reporter systems cannot be used directly for forward genetic interrogation of pathways acting on strictly *nuclear* human lncRNAs, we developed the Mirror approach (**Fig**. **2**), which relies on nuclear export of a fragment of non-polyadenylated nuclear RNA of interest (**Fig**. **2C**) to produce fluorescence signal. Here, we employed the previously reported finding that a GFP reporter carrying the downstream 3′ triple-helical ENE sequence of MALAT1 or MEN-β can undergo nuclear export and translation into a functional fluorescent protein[2]. We reasoned that inhibition of nuclear pathways acting on a similarly-designed reporter would affect (i.e., be 'mirrored' by) its steady-state cytoplasmic fluorescence, enabling forward genetic screening for components of human nuclear machinery acting on lncRNAs that would otherwise never leave the nucleus. We further reasoned that whereas the genome-wide forward genetic

screening using such a Mirror reporter would simultaneously identify components of both nuclear (black) and cytoplasmic (grey) pathways acting on it (schematically shown in **Fig**. **2C**), subsequent post-screening analysis of the effects of knockout of individual identified genes on the endogenous nuclear lncRNA (**Fig**. **2C**) will distinguish components of nuclear and cytoplasmic pathways, identifying components of nuclear pathways acting on the strictly nuclear lncRNA of interest. In cases where nuclear export and cytoplasmic degradation would affect the same-cell, orthogonal Mirror reporters similarly, their fluorescence will change synchronously, causing a diagonal fluorescence shift (**Fig**. **2B**), that would effectively exclude non-nuclear factors from screening results.

We employed Mirror for simultaneous forward genetic identification of two human pathways acting on the 3′ end of the strictly nuclear lncRNA MALAT1: one of them is the pathway of the 3′ end maturation of MALAT1 and the other is the pathway of MALAT1 nuclear degradation that is blocked by the triple-helical ENE element. We constructed two orthogonal stable MALAT1 Mirror cell lines (termed “Green” and “Red”), each carrying two genome-integrated Mirror reporters (**Fig**. **2A**). Each Mirror reporter employs multiple sensitivity-enhancing features of the single-cell-based dual-color fluorescence amplification system[31]. Uniquely, each Mirror reporter (**Figs**. **2A** and **S7**) carries two elements of the human lncRNA MALAT1: the 3′ ENE stability element and the tRNA-like mascRNA element[17] (**Figs**. **1** and **2A**). The ENE of one of these reporters carries a single-nucleotide substitution, C8351G (**Figs**. **1B-C** and **2A** and **C**), making it susceptible to nuclear degradation[17,18]. To enhance screening specificity, each individual Mirror cell expresses a second, orthogonal wild-type ENE control reporter that differs from the stability-compromised ENE by just a single nucleotide (**Fig**. **2A**)[31]. Since the ENE(C8351G) Mirror reporters are substrates for nuclear degradation[18], they yield lower fluorescence than those containing ENE(WT). Knockout of a component of nuclear degradation machinery increases the fluorescence of stability-compromised ENE(C8351G) reporters with minimal effect on that with the intrinsically stable ENE(WT) (**Fig**. **2A** and **B**), enabling identification of human nuclear degradation pathways blocked by the MALAT1 ENE(WT). To identify human nuclear factors required for 3′ end maturation of MALAT1, including separation of the tRNA-like mascRNA from the MALAT1 lncRNA precursor (**Figs**. **1A**, **2A**, and **S1**), Mirror leverages the enhanced stability – and, as a result, increased fluorescence – of the unstable ENE(C8351G) Mirror reporter upon inhibition of mascRNA cleavage, which stems from the resulting use of the downstream Bovine Growth Hormone (BGH) cleavage and polyadenylation site (**Fig**. **S1**, compare panels **A** and **B**). Since similar inhibition of the mascRNA separation and the resulting switch to BGH-directed polyadenylation produces little to no change in the signal of the ENE(WT) reporter due to the high intrinsic stability provided by the intact ENE (**Fig**. **S1C-F**), defects in MALAT1 nuclear 3′-end processing are readily identifiable through their differential effects on the same-cell ENE(C8351G) and ENE(WT) Mirror reporters. We chose HeLa cells for Mirror forward

genetic screening of MALAT1-acting human nuclear pathways due to the strong association of steady-state levels of MALAT1 with tumor aggressiveness in cervical cancers[13].

Genome-wide Mirror forward genetic screening for factors targeting the 3′ end of the human lncRNA MALAT1 enriched numerous components of *nuclear* RNA-processing complexes. **Figure 2D-E** shows the outcomes of two iterative rounds of forward genetic Mirror screening (**Fig**. **2F**) to identify components of two human nuclear pathways targeting MALAT1's 3′ end: the pathway of its maturation and the other, the degradation pathway blocked by the ENE element. The screening was conducted in the "Green" and "Red" Mirror cell lines (**Fig**. **2A**) using FACS-based iterative rounds of guide RNA (sgRNA) library enrichment (**Fig**. **2F**) intended to amplify pathway-inhibiting sgRNAs and eliminate stochastic false positives[31]. During the 1st round of screening (**Fig**. **2D** and **E**), the Mirror cell lines were transduced with the original lentiviral GeCKO-lentiCRISPR sgRNA library[32] containing 64,751 sgRNAs targeting 18,080 human genes. During the 2nd round of screening (**Fig**. **2D** and **E**), the original "Green" and "Red" Mirror cell lines (**Fig**. **2A**) were transduced with the enriched lentiviral sgRNA libraries, which were obtained at the end of the 1st round of screening as shown in **Fig**. **2F**. We expected that sgRNAs targeting pathways of MALAT1 3′-end biogenesis and degradation would increase (1) GFP fluorescence of the "Red" MALAT1 Mirror cells (*i.e.*, shift them up, **Fig**. **2B**) and (2) RFP fluorescence of the "Green" MALAT1 Mirror cells (*i.e.*, shift them right, **Fig**. **2B**). Indeed, distinct markedly enriched populations of cells appeared in the blue (**Fig**. **2D**) and green (**Fig**. **2E**) gates after the 2nd round (**Fig**. **2D-F**).

Deep sequencing of sgRNAs in the enriched cell population in the "blue" FACS gate (**Fig**. **2D**) revealed numerous sgRNAs targeting *nuclear* RNA-processing factors (**Fig**. **3A** and **E1**). Indeed, 60 of the top 92 most abundant Mirror-enriched sgRNAs (**Fig**. **3A**) targeted: (i) 9 out of 10 protein components of the human nuclear RNase P[33]: RPP14, RPP21, RPP30, RPP38, RPP40, POP1, POP4 (RPP29), POP5, and POP7 (RPP20); (ii) 9 out of 11 components of the human RNA Exosome[34]: EXOSC2, EXOSC4, EXOSC5, EXOSC6, EXOSC7, EXOSC8, EXOSC9, EXOSC10, including its nuclear DIS3 (but not the cytoplasmic DIS3L) component[35]; (iii) 2 out of 3 components of the human nuclear NEXT (Nuclear Exosome Targeting) complex[36]: RBM7 and SKIV2L2 (hMTR4); and (iv) the nuclear exosome co-localized RNA-binding protein C1D known for its role in 3′-end processing of 5.8S rRNA[37]. For some of these factors, Mirror identified more than one sgRNA (**Fig**. **3A**), and for others, such as RNase P components RPP21 and POP5, nuclear RNA Exosome components EXOSC10, and the NEXT complex component SKIV2L2, it identified every single sgRNA present in the original GeCKO library (**Fig**. **3A** and **Supplemental Data S2**). Additional Mirror-identified and subsequently validated sgRNAs targeted TFIIB-like factor BRF2, known to control initiation of Pol III[38] and the predominantly

nuclear DEAD-box helicase DDX59, which has no known substrate in any organism[39,40]. All Mirror-identified factors are summarized in **Figure 3E1**.

Corroborating the Mirror design, individual knockouts of screening-identified nuclear factors increase Mirror reporter fluorescence. Underscoring Mirror's efficiency in identifying *nuclear* factors acting on *nuclear* RNAs using reporter's *cytoplasmic* fluorescence, transduction of sgRNAs targeting the individual Mirror-enriched factors (**Fig**. **3A** and **3E1**) produces specific cell populations with a 2- to 15-fold increase in GFP fluorescence (**Fig**. **3B-Y** and **D1**). For all 23 genes shown, only minimal effects are observed on the fluorescence of the same-cell control RFP reporter (**Figs**. **2A-B**, **3B-Y**, and **D1**). Notably, 21 out of 23 (91%) of these genes are classified as "Common Essential" in the DepMap database[41], demonstrating the Mirror's effectiveness in identifying genes critical for viability.

We hypothesized that the absence of RPP25 (hPOP6) among the RNase P protein components in Mirror results reflects a limitation of forward genetics in detecting redundant factors. Indeed, whereas individual knockouts of RPP25 and RPP25L[42] produced no populations with increased green fluorescence (**Fig**. **3A1** and **B1**), a simultaneous double knockout of these genes produced such a population (**Fig**. **3C1**), confirming functional redundancy.

Since the GeCKO library lacked sgRNAs targeting the catalytic RNA component of RNase P, RPPH1, identifying it using this library would not have been possible. We demonstrate, however, that not only do sgRNAs targeting protein-encoding genes increase fluorescence in the Mirror system, but also those targeting RNA-encoding genes, as shown by the CRISPR-sgRNA knockout of RPPH1 (**Fig**. **3Z**).

In summary, validating the design, the Mirror's *cytoplasmic* fluorescence efficiently responds to the inactivation of *nuclear* factors acting on *nuclear* RNAs, even when these factors are essential.

DEAD-Box helicase DDX59, identified by Mirror and lacking known substrates, is required for nuclear degradation of the full-length lncRNA MALAT1 with stability-compromised ENE. Whereas the majority of Mirror-identified components associate with established nuclear RNA-acting complexes, such as the RNA Exosome, RNase P, and NEXT complex (**Fig**. **3E1**), the predominantly nuclear DEAD-box helicase DDX59 could not be readily assigned to them. Furthermore, nothing is known about the substrate of DDX59 in any organism, nor has it been determined whether this substrate is RNA or DNA[13]. To ascertain how DDX59 affects MALAT1, we first confirmed that DDX59 knockout (**Fig**. **5K**) specifically affects the fluorescence of Mirror reporters with a stability-compromised 3′ ENE(C8351G) and that additional sgRNAs toward DDX59 can replicate this effect. In addition to reproducing an increase in fluorescence (**Fig**. **3F**) in the "Red" Mirror cell line (**Fig**. **2A**), in which the

first step of validation was performed (**Fig**. **3B-D1**), the individual transductions of four distinct DDX59-targeting sgRNAs into the orthogonal "Green" Mirror cell line (**Fig**. **2A**), in which the ENE(C8351G) and ENE(WT) are swapped, produced an increase in RFP fluorescence without significantly affecting GFP fluorescence (**Fig**. **S2**). This result confirms that DDX59 knockout specifically influences the fluorescence of Mirror reporters with a stability-compromised ENE, and not the wild-type ENE, and ruled out off-target effects of sgRNAs on the fluorescent protein moiety of the reporters.

To confirm that the effects of DDX59 knockout on fluorescence of stability-compromised Mirror reporters represent an accurate reflection of its effects on nuclear degradation of the RNA of MALAT1 with stability-compromised ENE, we analyzed the degradation of the full-length MALAT1(C8351G) RNA expressed from its own native promoter (**Fig**. **3F1**) following the CRISPR sgRNA-based knockout of DDX59 and the cessation of transcription with actinomycin D. Indeed, knockout of DDX59 (quantified in **Fig**. **5K**) results in stabilization of the full-length MALAT1(C8351G) RNA ("red" curve in **Fig**. **3G1**) as compared with non-targeting sgRNA control ("blue" curve in **Fig**. **3G1**).

The stabilizing effect of the DDX59 knockout on full-length MALAT1(C8351G) RNA (**Fig**. **3G1**) parallels the effects observed with knockouts of EXOSC10, SKIV2L2, and RBM7 (quantified in **Fig**. **5K**) for this lncRNA (**Fig**. **3H1-J1**). Since EXOSC10, a catalytic subunit of the nuclear RNA Exosome, as well as RBM7 and SKIV2L2, subunits of the Nuclear Exosome Targeting (NEXT) complex[36], are key constituents of two major established complexes involved in nuclear RNA degradation[43,44], their stabilizing effects corroborate correct identification of these complexes by Mirror. The parallel effects of DDX59 suggest it is also required for the nuclear degradation of stability-compromised MALAT1 through a yet-unidentified mechanism.

DDX59 knockout results in retention of minor (U12) introns, impacting components of the RNA Exosome and NEXT complex. To address the role of DDX59 in the nuclear degradation of stability-compromised lncRNA MALAT1, we analyzed transcriptome-wide effects of DDX59 knockout. Remarkably, knockout of DDX59, rigorously obtained as shown in **Figure 4A**, displays profound retention of minor (U12) spliceosomal introns across numerous genes (**Figs**. **4**, **S3D-I**, **S4**, and **S6**). This retention is more pronounced than that resulting from the knockout of the minor spliceosomal factor ZCRB1[45,46], yet exhibits little to no effect on major (U2) introns (**Figs**. **4**, **S3D-I**, **S4**, and **S6**; also shown for U2 introns of housekeeping genes in **Fig**. **S3A-C**). The U12 intron retention is evident not only for DDX59 knockout cells compared to negative control cells transduced with a non-targeting sgRNA (compare the red "KO" IGV traces with the blue "CTRL-" traces in **Figs**. **4**, **S3D-I**, **S4**, and **S6**) but also for DDX59 knockout cells compared to cells in which the same sgRNA toward DDX59 yielded no FACS-detectable increase in fluorescence (compare the red "KO" IGV traces with the green "Main"

traces in **Figs**. **4**, **S3D-I**, **S4**, and **S6**). These controls mitigate unintended differences in growth conditions, transduction efficiency, and antibiotic selection, as throughout the entire experiment, the cells of the "KO" and "Main" populations grow together on the same plate until sorted by FACS (**Fig**. **4A**). Quantification of U12 intron retention (**Figs**. **4B-D** and **K-M**, and **S6**) demonstrates that DDX59 is required for the splicing of a set of human minor introns.

As shown in **Figures 4B-G** and **S4G**, the genes with pronounced DDX59-dependent retention of minor introns include RNA Exosome components EXOSC1, EXOSC2, and EXOSC5, as well as the scaffold component of the Nuclear Exosome Targeting (NEXT) complex, ZCCHC8. Since Mirror screening and its validation have shown that both the RNA Exosome and the NEXT complex are required for the degradation of stability-compromised lncRNA MALAT1(C8351G), inactivation of integral components of these complexes through U12 intron retention (**Figs**. **4B-G** and **S4G**) and subsequent reduction of their protein levels (**Fig**. **S5**) explains the Mirror-identified role of DDX59 in the stabilization of lncRNA MALAT1 with a compromised 3′-end triple-helical ENE structure.

Retention of numerous minor (U12) introns in cilia-related genes provides an explanation for the role of DDX59 in Oral-Facial-Digital syndrome (OFD). Whereas mutations in the DEAD-box helicase DDX59 are known to associate with the rare genetic developmental disorder Oral-Facial-Digital (OFD) syndrome[39,47], the mechanism linking defects in DDX59 to OFD remains unknown. It has been found that nearly 20 genes encoding proteins that either represent components of cilia or influence ciliogenesis are associated with at least 14 classes of OFD[48-50]. However, DDX59 is a notable exception among OFD-associated genes because it lacks known cilia-associated roles[48].

We found that knockout of DDX59 induces retention of U12 introns in mRNAs of genes with established roles in OFD and OFD-associated ciliopathies such as C2CD3, TCTN3, and TMEM107 (**Figs**. **4H** and **K**, **S4B** and **D**), mutations in which are associated with, and have been proposed to be causative for, OFD subtypes OFDXIV, OFDIV, and OFDVI, respectively[48-50]. Additionally, knockout of DDX59 results in retention of U12 introns in DCTN3, OCRL, C2CD3, ARMC9, KATNIP, SFI1, CUL1, ACTR10, CCDC28B, PPP5C, TCTN3, ACTL6A, TMEM107, KIFAP3, RABL2A (**Figs**. **4H-M**, **S3D-I**, and **S4A-F**), all of which are associated with ciliogenesis or cilia assembly[51].

Whereas knockout of DDX59 produces a major effect on retention of minor (U12) spliceosomal introns, retention of major (U2) spliceosomal introns is virtually unaffected (**Figs**. **4**, **S3**, **S4**, and **S6**). As negative and positive controls, knockout of the Mirror-identified NEXT complex component RBM7 (quantified in **Fig**. **5K**) results in no observable defects in the splicing of either U12 or U2 introns, whereas knockout of ZCRB1 results in retention of U12 introns, as expected for a knockout of a known component of the human minor spliceosome[45,46].

Collectively, the DDX59-induced retention of U12 introns in OFD-associated and cilia-related genes provides an explanation for its role in Oral-Facial-Digital syndrome.

Knockout of DDX59 results in the accumulation of 3′-extended precursors of endogenous small nuclear RNAs (snRNAs). Consistent with the Mirror-identified stabilizing effect of DDX59 knockout on the nuclear MALAT1(C8351G), which undergoes non-canonical 3′-end processing by RNase P, lacks a poly(A) tail, and is partially devoid of the protection of the intact 3′-end triple-helical ENE structure, we observe that CRISPR-Cas9 knockout of DDX59 increases the steady-state levels of 3′-extended precursors of several endogenous Sm-class snRNAs (**Fig**. **5A-C**) known to undergo complex non-canonical 3′-end processing and lack poly(A) tails[52,53]. RT-qPCR quantification and IGV traces of RNA-seq performed without poly(A) selection show a 5.3- and 4.1-fold increase in the levels of 3'-extended forms of snRNAs U4atac and U11, respectively (**Fig**. **5E-F**). Consistent with (i) our finding that knockout of DDX59 impacts the function of the RNA Exosome and the NEXT complex via retention of minor (U12) introns in mRNAs of key components of these two complexes (**Figs**. **4E-G** and **S4G**), and (ii) known roles of both complexes in the 3′ end processing of snRNA precursor transcripts[44,54], knockouts of a known minor spliceosome component ZCRB1, RNA Exosome component EXOSC10, and the NEXT complex component RBM7 similarly increase the levels of 3′-extended forms of U4atac, U11 and U4-2 (**Fig**. **5A-C**). Two lines of evidence suggest that the increase in 3′-extended forms of minor spliceosomal snRNAs is not the primary cause of the minor intron retention observed in DDX59 knockout: First, whereas the effects of DDX59, RBM7, EXOSC10, and ZCRB1 knockouts on the levels of 3′-extended forms of U4atac, U11, and U4-2 snRNAs show significant similarities (**Fig**. **5A-C**), only the knockouts of DDX59 and ZCRB1, not those of EXOSC10 or RBM7, result in U12 intron retention (**Figs**. **4** and **S3-5**). Second, whereas DDX59 knockout increases the levels of 3′-extended forms of both minor (U12) and major (U2) spliceosomal snRNAs (**Fig**. **5A-C**), retention of only U12 introns is observed (**Figs**. **4** and **S3-5**). This suggests that inactivation of DDX59 impacts a subset of endogenous human nuclear non-canonical RNAs lacking poly(A) tails in a manner similar to that resulting from the minor spliceosome inactivation and the ensuring impairment of the RNA Exosome and NEXT complex.

Knockout of DDX59 upregulates certain endogenous lncRNAs, exemplified by NEAT1_1 and NIPBL-DT, the latter transcribed divergently from the mRNA of the Cornelia de Lange Syndrome-associated NIPBL. Currently, only two human-encoded lncRNAs are known to carry non-canonical 3′-end stabilizing triple-helical ENE: one is MALAT1, the second being the long (22,700 nt) isoform of NEAT1, NEAT1_2, also known as MEN-β[1]. The triple-helical ENE structure of NEAT1_2 is similar to that of MALAT1, and the 3′ ends of these lncRNAs are similarly formed *via* RNase P cleavage

of their respective tRNA-like structures[2]. The shorter (3,700 nt) isoform of NEAT1, known as NEAT1_1, is polyadenylated. Its 3′ end is produced through non-canonical cleavage that is reportedly promoted by the components of the Integrator complex[55]. Both NEAT1_1 and NEAT1_2 are transcribed from the same promoter but undergo different steps of 3′-end processing and are potentially subject to distinct 3′ degradation machineries. We found that the CRISPR knockout of DDX59, obtained as shown in **Figure 4A**, results in a substantial, 3.5-fold, increase in the levels of the shorter isoform, NEAT1_1 (**Fig**. **5G**), whereas the ENE-stabilized long isoform NEAT1_2 is only marginally affected (**Fig**. **5G**). This is consistent with the role of the RNA Exosome in regulating NEAT1_1 levels[56], further illustrating effects of DDX59 on non-coding RNAs.

A critical feature of the lncRNA NIPBL-DT is its transcription from the same bidirectional promoter as the mRNA of the Nipped-B-like (NIPBL) protein, also known as delangin, which encodes the major subunit of the cohesin loading complex[57]. Nearly 70% of the multisystem congenital disorder Cornelia de Lange Syndrome (CdLS) patients carry mutations in NIBPL[58]. In mouse studies, CdLS phenotypes are observed with as little as 30% decrease in NIBPL mRNA, implying high sensitivity to transcript levels[59]. Furthermore, not the NIPBL-DT RNA gene product (also called NIPBL-AS1), but its transcription, regulates the expression of the CdLS-associated NIPBL mRNA[60]. We found that knockout of DDX59 results in nearly a 3-fold increase in the steady-state levels of lncRNA NIPBL-DT (**Fig**. **5D** and **H**) with minimal impact on its degradation, as measured by actinomycin D transcription shut off (**Fig**. **5I**), whereas the knockout of the NEXT complex component RBM7 virtually stops NIPBL-DT degradation (**Fig**. **5I**). This suggests that DDX59 disruption upregulates lncRNA NIPBL-DT through a mechanism other than RNA degradation and likely through the transcription of NIPBL-DT, which is known to regulate the transcription of the CdLS-associated NIPBL mRNA itself[60].

The Mirror-identified BRF2 is required for MALAT1-mascRNA processing through expression of the RNA component of RNase P, H1 (RPPH1). To elucidate the effect of BRF2 (**Fig**. **3O**) on the 3′ end of MALAT1, we analyzed cleavage of the endogenous wild-type lncRNA MALAT1 (**Fig**. **1A**) using RT-qPCR across the RNase P cleavage site. The quantification revealed a 12-fold increase in unprocessed MALAT1-mascRNA transcript following BRF2 knockout (**Fig**. **5J** and **K**). This defect parallels a nearly 50-fold increase in RNase P-unprocessed MALAT1 following knockout of RPP21, which is an essential RNase P subunit (**Fig**. **5J** and **K**). Additionally, knockout of BRF2 results in a dramatic 10-fold loss of the steady-state levels of H1 (RPPH1) RNA (**Fig**. **5J**), the catalytic RNA subunit of human RNase P[61], whereas knockout of RPP21 (**Fig**. **5K**) causes less than a 10% reduction (**Fig**. **5J**). Together, these results demonstrate that BRF2, which controls initiation of human Pol III RNA

transcripts[38], is required for expression of Pol III-transcribed H1 RNA, and the loss of the H1 RNA leads to the MALAT1 3′-end processing defect identified by Mirror.

DISCUSSION

We developed the Mirror approach, which enabled forward genetic identification of the human post-transcriptional machinery acting on nuclear non-coding RNAs by employing specialized non-polyadenylated reporters that undergo nuclear export and translation (**Figs**. **2A** and **C**, **S1**, and **S7**). We demonstrate that the fluorescence signal of the Mirror reporters provides a reliable proxy measurement for processing efficiency and/or abundance of a fused fragment of nuclear non-coding RNA of interest (**Figs**. **3** and **5J**), thereby enabling forward genetic screening of strictly nuclear pathways acting on it. Employed to discover two human nuclear pathways acting on the 3′ end of nuclear lncRNA MALAT1 with stability-compromised ENE, Mirror identified nearly all (**Fig**. **3**) components of three major nuclear processing and degradation complexes: RNase P, RNA Exosome, and the NEXT complex. It additionally identified a nuclear exosome-associated factor C1D, TFIIB-like factor BRF2, and the DEAD box helicase DDX59 which previously had no known substrate in any organism or role in nuclear RNA processing and degradation (**Fig**. **3E1**). Since Mirror does not rely on growth metrics and instead detects early changes in cell fluorescence that precede the onset of the knockout-induced lethality, it efficiently identifies essential genes. Underscoring this feature, 21 of the 23 top genes (i.e., 91%) identified by Mirror (**Fig**. **3E1**) are essential[41]. This comprehensive list, alongside with validation of effects on the endogenous lncRNA by at least one component of each of the three identified complexes (**Fig**. **3H1-J1**) and additional factors (**Fig**. **3G1** and **5J**), demonstrates the effectiveness of Mirror in identifying pathways of nuclear non-coding RNAs.

Of the three Mirror-identified genes outside the three major complexes, C1D plays an established role in nuclear RNA degradation as a SKIV2L2- and EXOSC10-interacting partner[37], while BRF2 is involved in the initiation of Pol III transcription[38], a process we confirmed as required for the expression of the RNase P catalytic subunit, RNA H1. Whereas Mirror accurately identified the DEAD box helicase DDX59 as judged by its impact on the degradation of the full-length lncRNA MALAT1(C8351G), the function of DDX59 remained obscure. Examination of the effects of DDX59 on nuclear RNA degradation revealed that they arise from a previously unknown role of DDX59 in minor intron splicing, as knockout of DDX59 produces extensive retention of U12 introns in components of the RNA Exosome EXOSC1, EXOSC2, and EXOSC5 as well as the NEXT complex component ZCCHC8, with the degree of retention being comparable to or greater than that observed with the knockout of the

known minor spliceosome component ZCRB1[45,46]. Whereas retention of U12 introns in key components of the RNA Exosome and NEXT complex could potentially fully account for the observed (i) Mirror-identified stabilization of the full-length lncRNA MALAT1(C8351G), (ii) increased levels of 3′-extended forms of endogenous Sm-class snRNAs, and (iii) increased levels of certain lncRNAs, such as NEAT1_1, resulting from DDX59 knockout, an additional to minor intron splicing role of DDX59 in nuclear RNA degradation cannot be ruled out. Although no obvious physical associations of DDX59 with spliceosomes have been reported, such an interaction could be transient, which has been observed for DEAD-box helicases. Alternatively, DDX59 may be required for minor intron splicing through an indirect mechanism that remains to be elucidated; emerging studies continue to reveal unexpected twists in minor intron recognition[45,62]. Mirror's identification of the roles of DDX59 in nuclear RNA degradation and minor intron splicing underscores the advantages of forward genetics in identifying components that impact pathways of interest, independently on whether such components act directly, indirectly, transiently, or are essential.

A new role of DDX59 in minor intron splicing provides a missing link between defects in DDX59 and the rare developmental genetic disorder Oral-Facial-Digital syndrome, which is often associated with intellectual disability[48]. Our finding that DDX59 knockout induces retention of minor introns in TCTN3, TMEM107, and C2CD3, all associated with distinct subtypes of OFD and linked to cilia and ciliogenesis, has revealed a mechanism by which mutations in DDX59 may affect these and potentially other OFD and cilia genes, contributing to this genetic disorder[39]. This mechanism of DDX59 in OFD is supported by several facts: First, deficiencies in cilia genes can cause OFD. Of the nearly 20 genes linked to at least 14 subtypes of OFD, only two, including DDX59, have not been definitively associated with cilia structure or function[48-50]. Second, a disproportionately high number (6%-7.6%) of cilia-related genes harbor U12 introns, compared to an average of 2% of U12-containing genes in the human genome[51], suggesting a greater impact of minor splicing defects. Third, deficiencies in U12 intron splicing, including those caused by mutations in minor spliceosome snRNAs, are known to result in ciliary defects in humans[51]. Fourth, minor spliceosome deficiencies are known to associate with OFD syndrome via mis-splicing of genes affecting primary cilia[63]. Additionally, consistent with the impaired Sonic Hedgehog (SHH) signaling observed in DDX59-associated OFD[39], DDX59 knockout leads to extensive retention of U12 introns in the mRNA of Derlin 2 (**Fig**. **S6**), a factor that mediates the retro-translocation of the SHH protein at the endoplasmic reticulum[39,63,64]. Altogether, our finding of DDX59's impact on U12 intron splicing suggests that a deficiency in DDX59 may affect as many as 86 known cilia- and SHH-related genes spliced by the minor spliceosome[51,63], contributing to OFD.

As we demonstrate with MALAT1, the Mirror strategy can be applied to more than one nuclear pathway, and may be extended to pathways of degradation, biogenesis, and surveillance of other

nuclear non-coding RNAs. It also can be applied to identify stabilizing factors contributing to increased levels of disease-associated non-coding nuclear RNAs, including the metastasis-promoting lncRNA MALAT1 itself, for which a search for destabilizing small molecules is underway[24,65]. Inhibition or inactivation of such stabilizing factors would result in a decrease in levels of MALAT1, providing additional potential targets for cancers with poor survival linked to high MALAT1 levels[11-16].

In summary, the Mirror approach enables forward genetic discovery of components of human nuclear pathways acting on strictly nuclear non-coding RNAs. It efficiently identifies pathways components essential for cell viability, those that act indirectly, and, as a result, uncovers unexpected connections to human disease.

EXPERIMENTAL PROCEDURES

Construction of plasmids. *Mirror reporter plasmids*: The 171-nucleotide sequences (shown in Supplemental Data S4) of the human wild-type and C8351G mutant MALAT1's ENE-mascRNA were amplified using PCR primers MALAT-Fg and MALAT-Rg, listed in Supplemental Data S4, from constructs C-G(WT) βΔ1,2-MALAT1 ENE+A+mascRNA and G-G(C8351G) βΔ1,2-MALAT1 ENE+A+mascRNA[18], kindly provided by Dr. Jessica Brown. Mirror reporter plasmids were constructed by cloning these 171-mers into the unique PspXI restriction site of the Fireworks reporter plasmids pAVA2598[GFP(PTC-)] and pAVA2515[RFP(PTC-)][31] using Gibson cloning, producing Mirror reporter plasmids pAVA2965[GFP-ENE(WT)-mascRNA], pAVA2987[RFP-ENE(WT)-mascRNA], pAVA3000[GFP-ENE(C8351G)-mascRNA], pAVA2995[RFP-ENE(C8351G)-mascRNA]. *Guide RNA-expressing plasmids*: Plasmids expressing two sgRNAs were constructed by PCR-amplifying a 332-nucleotide fragment comprising the Cas9 scaffold and 7SK promoter from pAVA3129 (Supplemental Data S1), using a forward primer for sgRNA1 and a reverse primer for sgRNA2 (sequences in Supplemental Data S3), and cloned into BsmBI-linearized blasticidin-resistant lentiCRISPR vector[32] to express first sgRNA from the U6 promoter and the second sgRNA from 7SK promoter. Plasmids expressing single guide RNAs (sequences in Supplemental Data S3) were constructed by cloning sgRNA sequences into BsmBI-linearized blasticidin-resistant lentiCRISPR vector[32] using Gibson Assembly (New England Biolabs). The resulting lentiviral sgRNA-expressing constructs for gene knockout were sequences and named as follows: pAVA3250 for RBM7, pAVA3251 for SKIV2L2, pAVA3300 for ZCRB1, pAVA3258 for BRF2, pAVA3259 for RPP21, pAVA3586 for RPPH1, pAVA3318 for EXOSC10, pP18(T)_B9 for C1D, pAVA3791 as sgRNA1 for DDX59, pAVA3796 as sgRNA2 for DDX59, pAVA3317 for as sgRNA3 DDX59, pAVA3794 as sgRNA4 for DDX59, pP18(T)_C3 for

EXOSC2, pP18(T)_C11 for EXOSC4, pP18(T)_B12 for EXOSC5, pP18(T)_B11 for EXOSC6, pP18(T)_C4 for EXOSC7, pP18(T)_B1 for EXOSC8, pP18(T)_A7 for EXOSC9, pP18(T)_A5 for DIS3, pAVA3773 for Control(-), pAVA3497 for POP1, pAVA3523 for RPP38, pAVA3545 for RPP29, pAVA3498 for POP5, pAVA3546 for RPP20, pAVA3507 for RPP14, pAVA3573 for RPP30, pAVA3540 for RPP40, pAVA3522 for RPP25, pAVA3548 for RPP25L. *Plasmids expressing full-length MALAT1*: Genomic DNA fragment consisting of (i) 1.2 kb of MALAT1 5′ flanking sequence, (ii) 8.8 kb sequence corresponding to MALAT1 lncRNA, and (iii) 0.7 kb of MALAT1 3′ flanking sequence was PCR-amplified from genomic DNA of the HeLa Fireworks cell line[31], cloned into the BglII and XhoI restriction sites of the pcDNA5/FRT plasmid (Invitrogen) and verified by sequencing. An 11-nucleotide sequence (5′-cgctcgacgta-3′) was inserted 43 nucleotides upstream of the MALAT1 ENE sequence using the NheI restriction site to distinguish exogenously expressed MALAT1 from its endogenously expressed copies. C8351G mutation was introduced into the resulting 10.8 kb construct using PCR mutagenesis with primers listed in Supplemental Data S4. The resulting plasmids pAVA3169(MALAT1(C8351G)) and pAVA3171(MALAT1(WT)) are schematically shown in **Figure 3F1**. Full sequences of all plasmids are shown in Supplemental Data S1; they will be deposited to the Addgene repository by the time of publication.

Construction of stable cell lines. Stable orthogonal “Red” and “Green” Mirror cell lines shown in **Figure 2A** were obtained by exchanging FRT-integrated Fireworks NMD reporters[31] in the “Green” Fireworks HeLa cell line AVAM526[31] for Mirror reporters using transient co-transfection of pAVA2987 [RFP-ENE(WT)-mascRNA], pAVA3000 [GFP-ENE(C8351G)-mascRNA], pAVA2995 [RFP-ENE(C8351G)-mascRNA], and/or pAVA2965 [GFP-ENE(WT)-mascRNA] with the Flp recombinase-expressing plasmid pOG44 (Invitrogen). Single colonies of cells with stably integrated Mirror reporters were selected using hygromycin (150-300 µg/mL) and puromycin (0.08-0.16 µg/mL). Expanded colonies were FACS-sorted, propagated, and saved as Mirror “Red” AVAM712 (GFP-ENE(C8351G)-mascRNA, RFP-ENE(WT)-mascRNA) and “Green” AVAM742 (GFP-ENE(WT)-mascRNA, RFP-ENE(C8351G)-mascRNA) stable cell lines.

Cell culture and maintenance. The “Red” Mirror cell line was maintained in DMEM (Gibco) media supplemented with 10% fetal bovine serum (FBS, Gibco), 100 U/mL penicillin-streptomycin (Corning), 150 µg/mL hygromycin (InvivoGen), and 0.08 µg/mL puromycin (InvivoGen). The “Green” Mirror cell line was maintained in DMEM media supplemented with 10% FBS, 100 U/mL penicillin-streptomycin, 300 µg/mL hygromycin, and 0.16 µg/mL puromycin. All fluorescent cell lines were routinely FACS-analyzed and sorted for stable GFP and RFP expression. HEK293T cells were maintained in DMEM

supplemented with 10% FBS and 100 U/mL Penicillin-Streptomycin. All cell lines are authenticated using Short Tandem Repeat (STR) analysis and confirmed to be mycoplasma-free using PCR-based tests.

Production of lentiviruses and transduction of human cell lines. To produce lentiviruses, sgRNA-expressing lentiCRISPR plasmids, listed in Supplemental Data S3, were co-transfected with the pCMV-dR8.91 packaging and pMD2.G envelope plasmids into the packaging HEK293T cells (AVAM761) using the TransIT-293 Transfection Reagent (Mirus Bio). Following daily media changes (DMEM, 30% FBS, 100 U/mL penicillin-streptomycin), lentivirus-containing media was collected 48 and 72 hours post-transfection, centrifuged at 900 g for 6 minutes to remove cellular debris, passed through a 0.45 µm filter, supplemented with 15 µg/mL polybrene (Millipore Sigma), and added to the Mirror cell lines growing at about 40% confluency. Selection for blasticidin-expressing infected cells was performed for 3 to 4 days using 3.0 µg/mL blasticidin (InvivoGen), which was added on the 3rd day post-infection.

Forward genetic screening using the Mirror approach. A total of 5x10^{7} cells (2 × 15 cm plates) of the "Red" "Mirror" cell line were transduced with blasticidin-resistant GeCKO-lentiCRISPR[32] viral library and propagated in DMEM media supplemented with 10% FBS, 150 µg/mL hygromycin, 0.08 µg/mL puromycin and 3.0µg/mL blasticidin, with the latter removed 4 days post-infection. Cell populations with increased fluorescence were FACS-isolated 10-11 days post-infection using the Bio-Rad S3e cell sorter. In FACS sorting and analysis, the gates were applied to exclude cell debris [SSC(area) vs FSC(area)], and cell doublets [FSC(width) vs FSC(hight)]. FACS data were collected using 488 nm and 561 nm excitation and 525/30 and 586/25 emission filters, for GFP and RFP fluorescence channels, respectively. The FACS-isolated cells were pelleted by centrifugation for 10 minutes at 1000g and frozen. Genomic DNA was isolated from FACS-isolated cells using phenol extraction. Sequences of guide RNAs were amplified from genomic DNA in two steps. First, a linear[66] sgRNA amplification was performed using Herculase II DNA polymerase using 13 thermal cycles with a single sgRNA promoter-specific primer, RandomF (sequences of all primers are listed in Supplemental Data S4), and the following cycling parameters: 96°C 20s, 63°C 1min, 72°C 90s. Then, the second PCR primer, RandomR, was added to the reaction and a regular PCR was performed for 35 cycles as follows: 96°C 3min; 35 cycles of 96°C 20s, 63°C 1min, 72°C 45s; 72°C 10min; 4°C. The PCR-amplified pools of sgRNAs were Illumina-sequenced and/or cloned into a BsmBI-linearized, blasticidin-resistant lentiCRISPR vector[32] to create an enriched lentiCRISPR sgRNA library for subsequent forward genetic screening rounds, as illustrated in **Figure 2F**. Processing of deep sequencing data was performed as previously described[32].

In vivo analysis of RNA degradation using Actinomycin D. FACS-isolated cell populations of lentivirally-transduced cells expressing Cas9 and gene-specific sgRNAs (pAVA3317 targeting DDX59, pAVA3250 targeting RBM7, pAVA3251 targeting SKIV2L2, pAVA3318 targeting EXOSC10, and pAVA3773 representing negative control; sequences are provided in Supplemental Data S3) were seeded on 10 cm plates at 30% confluency. Eighteen hours later, each plate was transfected using TransIT-293 Transfection Reagent with 18 µg of the plasmid (pAVA3169, full sequence provided in Supplemental Data S1) expressing full-length lncRNA MALAT1 driven by its own promoter and carrying the ENE-destabilizing C8351G mutation. Twenty-four hours post-transfection, the cells were transferred to 6-well plates, recovered for 3 hours, and treated with 5 µg/mL of Actinomycin D for 0, 1, 2, and 3 hours, after which they were lysed using the TRIzol reagent and stored at -80°C for subsequent RNA purification. RNA was purified using TRI reagent (Zymo Research) according to manufacturer's protocol and subjected to two 15-minute rounds of DNase I (Promega) treatment, phenol extraction, and ethanol precipitation. RT-qPCR quantification of RNA levels was performed as described below.

RT-qPCR quantification of RNA levels. For each sample, 1 µg of total RNA was reverse-transcribed using iScript™ Reverse Transcription Supermix according to the manufacturer's protocols. The resulting cDNA was diluted 20 times and qPCR-analysed using Applied Biosystems' SYBR Green PCR Master Mix according to the manufacturer's instructions for the CFX Opus 384 system and the following cycling conditions: 95°C for 3 min; 40 cycles of 95°C for 15 sec, 60°C for 30 sec. Relative quantification of transcript levels was performed using the ΔΔCT method and 18S RNA levels as the reference. Sequences of all primers are listed in Supplemental Data S4.

Analysis of RNase P processing of the endogenous full-length lncRNA MALAT1. The levels of RNase P-unprocessed endogenous full-length lncRNA MALAT1 were measured by RT-qPCR using nested primers across the RNase P cleavage site of MALAT1-mascRNA. RNA extraction and reverse transcription were performed as described above. To quantify the unprocessed MALAT1 transcript, 20 pre-amplification cycles with outside primers (JD08-F and JD08-R; sequences in Supplemental Data S4) were conducted using Applied Biosystems' SYBR Green PCR Master Mix with the PCR cycling parameters: 95°C for 3 min, followed by 20 cycles of 95°C for 15 s, and 60°C for 30 s. Each resulting pre-amplification reaction was diluted 40-fold with water and then used as a template for the subsequent qPCR reaction, which employed inside primers (JD07-F and JD07-R; sequences in Supplemental Data S4) and utilized Applied Biosystems' SYBR Green PCR Master Mix with the PCR cycling parameters: 95°C for 3 min, followed by 40 cycles of 95°C for 15 s, and 60°C for 30 s. The

levels of unprocessed MALAT1-mascRNA RNA were normalized to 18S RNA; ΔΔCt was used to quantify the relative expression levels. Data are presented as means of two biological x four technical replicates; error bars represent standard deviation. Statistically significant differences between knockout and control samples were determined by one-way ANOVA. Posthoc comparisons using Tukey's HSD test were conducted to determine the overall difference between groups, and labeled as "*", P<0.05; "**", P<0.01; "***", P<0.001. Positions of qPCR primers are indicated and their sequences are listed in Supplemental Data S4.

RNA-Seq and genome mapping. Extracted RNA was diluted to 50ng/µl and sent for cDNA library construction (without the use of poly(A) selection) and paired-end sequencing (100-bp paired-end reads) following the manufacturer's protocols (BGI-America, San Jose, CA). 60 millions clean reads received for each sample were mapped to human genome assembly GRCh38.p14 (GCA_000001405.29) using STAR[67] (version 2.7.10b) with BAM output sorted by coordinate and standard output attributes. The resulting alignments were indexed using SAMtools[68] (Version 1.18) and examined using IGV[69] (version 2.16.1).

ACKNOWLEDGMENTS

We thank Suzanne DeGregorio and Elizabeth Greif for help with cloning as well as Drs. Joan A. Steiz, David F. Clayton, Kazimierz Tycowski, and Raul Jobava for critical comments on the manuscript. This work was partially supported by funding to Joan A. Steitz from the Howard Hughes Medical Institute, and by National Institutes of Health grants HG009362 and GM139769 (project 1) to Andrei Alexandrov.

COMPETING INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. Schematics of the MALAT1 3′ end processing and triple-helical ENE stability element that protects the 3′ end of MALAT1 against nuclear degradation pathways. A. Schematics of the MALAT1 3′-end formation by RNase P cleavage of the tRNA-like mascRNA structure and formation of the triple-helical ENE structure[1,17,18] that protects the 3′ end of MALAT1 against nuclear degradation pathways. **B.** The stability-compromised ENE(C8351G) structure provides only partial protection against nuclear degradation pathways. **C.** Crystal structure (PDB ID: 4PLX)[1,18] of MALAT1 triple-helical ENE element with the position of ENE-destabilizing mutation C8351G[18] used in this study indicated by the red arrow. The image is made using Mol* Viewer[70].

Figure 2. The Mirror concept, implementation, and iterative forward genetic Mirror screening for human pathways that are blocked from degrading MALAT1 by its triple-helical 3′-end ENE element. A. The Mirror reporter system. Major features of the Fireworks fluorescence amplification system employed by Mirror are described[31]. The 3′ end of the MALAT1 ENE Mirror reporters is formed by RNase P cleavage upstream of mascRNA, as indicated by the orange arrow. Steady-state levels of the MALAT1 ENE Mirror reporters depend on the (1) integrity of their 3′-end triple-helical ENE stability element, which is inserted after the β-globin 3′ UTR and can be disrupted by a single nucleotide substitution C8351G (shown in red). Two orthogonal reporter cell lines ("Red" and "Green") provide robust controls for cell line- and fluorescent protein-specific effects during genome-wide forward genetic screening and candidate gene validation. **B.** Anticipated effects of inhibiting pathways of nuclear degradation and processing on the fluorescence of "Red" and "Green" Mirror cell lines shown in panel A. **C.** Schematics of the Mirror concept for identifying nuclear pathways acting on nuclear long non-coding lncRNAs by reporters utilizing cytoplasmic fluorescence. **D-E.** Genome-wide iterative FACS-based forward genetic screening for components of human nuclear pathways of MALAT1 3′-end processing and pathways that are blocked from degrading MALAT1 by its triple-helical ENE element. The screening was performed in the orthogonal "Red" (panel D) and "Green" (panel E) Mirror reporter cell lines. **F.** Schematics of iterative forward genetics screening rounds employed in panels D and E.

Figure 3. Mirror-identified components of human nuclear pathways acting on MALAT1's 3′-end and their validation. A. sgRNAs and their gene targets ranked by Mirror enrichment. Components of RNase P, RNA Exosome, and the NEXT complex are highlighted in green, blue, and orange, respectively. Three additional Mirror-identified and validated below genes – DDX59, C1D, and BRF2 – are shown in purple. **B-Y** and **D1.** Validation of Mirror-identified sgRNAs in the "Red" MALAT1 Mirror

cell line, which was transduced with individual sgRNAs and analyzed by FACS as described in Materials and Methods. **Z-D1.** Validation of additional components of Mirror-identified nuclear complexes. **E1.** Schematics of Mirror-identified complexes acting on the 3' end of MALAT1. Mirror-identified complex components are highlighted consistent with their coloring in panel A. **F1.** Schematics of the full-length lncRNA MALAT1(C8351G) driven by its own promoter, followed by its own 3′-flanking sequence, and carrying an insertion of 11 nucleotides in a non-conserved region to distinguish it from endogenously-expressed MALAT1 using RT-qPCR. **G1-J1**. Knockouts of Mirror-identified DDX59, RBM7, SKIV2L2, and EXOSC10 inhibit degradation of the full-length lncRNA MALAT1 with a stability-compromised ENE(C8351G). RT-qPCR-derived MALAT1(C8351G) levels were normalized to 18S RNA and are shown relative to the time of Actinomycin D addition. ΔΔCt was used to quantify the relative expression levels. Data are presented as means of two biological x four technical replicates; error bars represent standard deviation.

Figure 4. DDX59 is required for the proper splicing of U12 introns; its deficiency disrupts the RNA exosome, the NEXT complex, and genes crucial for Oral-Facial-Digital syndrome (OFD) and ciliary function. A. Schematics of generating knockout (KO), Main, and CTRL(-) populations for RNA-seq analysis, in which respective traces are colored in red, green, and blue. These knockouts are rigorously controlled for growth conditions, transduction efficiency, and antibiotic selection, as, throughout the entire experiment, the cell populations grow together on the same plate until sorted by FACS. **B-D.** RT-qPCR quantification of intron retention in mRNAs of the exosome components EXOSC1 and EXOSC2 as well as the scaffold component of the Nuclear Exosome Targeting (NEXT) complex, ZCCHC8 after DDX59 knockout as described in panel A. **E-J.** Integrated Genome Viewer (IGV) traces of RNA-seq reads for minor intron-containing genes, the exosome components EXOSC1 and EXOSC2, as well as the scaffold component of the Nuclear Exosome Targeting (NEXT) complex, ZCCHC8, and genes associated with ciliary function, DCTN3 and OCRL, as well as C2CD3, which is also associated with Oral-Facial-Digital (OFD) syndrome. Cell populations with DDX59, RBM7, ZCRB1, and EXOSC10 knockouts were obtained as shown in panel A and their respective IGV traces are colored accordingly. Major and minor introns are labeled as U2 and U12, respectively. **K-M.** RT-qPCR quantification of intron retention in mRNAs of genes associated with ciliary function, DCTN3 and OCRL, as well as C2CD3, which is also associated with Oral-Facial-Digital (OFD) syndrome. For RT-qPCRs, RNA levels were normalized to 18S RNA; ΔΔCt was used to quantify the relative expression levels. Data are presented as means of two biological x four technical replicates; error bars represent standard deviation. Statistically significant differences between knockout and control samples were determined by one-way ANOVA. Posthoc comparisons using Tukey's HSD test were conducted to determine the

overall difference between groups, and labeled as "*", P<0.05; "**", P<0.01; "***", P<0.001. Positions of qPCR primers are indicated and their sequences are listed in Supplemental Data S4.

Figure 5. Knockout of DDX59 results in the accumulation of 3′-extended precursors of endogenous snRNAs, as well as specific lncRNAs, shown for NEAT1 isoforms and NIPBL-DT, which is transcribed divergently from the mRNA of the Cornelia de Lange Syndrome (CdLS)-associated NIPBL. A-D. Integrated Genome Viewer (IGV) traces of RNA-seq (performed without poly(A) selection) are shown for minor spliceosomal snRNAs, U4atac and U11 (panels A and B), and for a major spliceosomal snRNA, U4 (panel C). Panel D shows similarly obtained traces for lncRNA NIPBL-DT, which is transcribed divergently from the mRNA of the CdLS-associated NIPBL, with its 1st exon shown. Cell populations with knockouts of DDX59, RBM7, ZCRB1, and EXOSC10 were obtained as shown in Figure 4A; their respective IGV traces are colored accordingly. **E-H.** RT-qPCR quantification of the levels of 3′-extended forms of minor spliceosomal snRNAs U4atac (panel E) and U11 (panel F) as well as long non-coding RNAs NEAT1_1 and NEAT1_2 (panel G), and NIPBL-DT (panel H). **I.** RT-qPCR analysis of the time course of degradation of the lncRNA NIBPL-DT in cell populations with knockouts of DDX59 and RBM7 obtained as shown in Figure 4A. RNA levels were normalized to 18S RNA and their remaining fractions are shown relative to those at the time of Actinomycin D addition. **J.** RT-qPCR analysis of the effects of BRF2 knockout on (i) processing of the endogenous full-length lncRNA MALAT1 by RNase P and (ii) RNA levels of its catalytic component RNA H1. The left graph shows analysis of cleavage across the RNase P site of the endogenous full-length lncRNA MALAT1, and the right graph shows RNA levels of RNA H1 in knockout cell populations. Knockout of RPP21, which is unique for RNase P (and absent in the closely-related RNase MRP) and essential for its function[33], is shown as a positive and negative control in the left and right graphs, respectively. **K.** RT-qPCR quantifications of mRNA levels in cell populations with CRISPR-induced knockouts of indicated genes, obtained as shown in Figure 4A. All qPCR quantifications and statistical tests were performed as described in the legend of Figure 4; qPCR primers are listed in Supplemental Data S4.

BIBLIOGRAPHY & REFERENCES CITED

1 Brown, J. A., Valenstein, M. L., Yario, T. A., Tycowski, K. T. & Steitz, J. A. Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc Natl Acad Sci U S A* **109**, 19202-19207, doi:10.1073/pnas.1217338109 (2012).

2 Wilusz, J. E. *et al.* A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev* **26**, 2392-2407, doi:10.1101/gad.204438.112 (2012).

3 Torabi, S. F. *et al.* RNA stabilization by a poly(A) tail 3'-end binding pocket and other modes of poly(A)-RNA interaction. *Science* **371**, doi:10.1126/science.abe6523 (2021).

4 Schmitt, A. M. & Chang, H. Y. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* **29**, 452-463, doi:10.1016/j.ccell.2016.03.010 (2016).

5 Dong, P. *et al.* Long Non-coding RNA NEAT1: A Novel Target for Diagnosis and Therapy in Human Tumors. *Front Genet* **9**, 471, doi:10.3389/fgene.2018.00471 (2018).

6 Schmitz, S. U., Grote, P. & Herrmann, B. G. Mechanisms of long noncoding RNA function in development and disease. *Cell Mol Life Sci* **73**, 2491-2509, doi:10.1007/s00018-016-2174-5 (2016).

7 Tycowski, K. T. *et al.* Viral noncoding RNAs: more surprises. *Genes Dev* **29**, 567-584, doi:10.1101/gad.259077.115 (2015).

8 Conrad, N. K., Mili, S., Marshall, E. L., Shu, M. D. & Steitz, J. A. Identification of a rapid mammalian deadenylation-dependent decay pathway and its inhibition by a viral RNA element. *Mol Cell* **24**, 943-953, doi:10.1016/j.molcel.2006.10.029 (2006).

9 Ji, P. *et al.* MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **22**, 8031-8041, doi:10.1038/sj.onc.1206928 (2003).

10 Gutschner, T., Hammerle, M. & Diederichs, S. MALAT1 -- a paradigm for long noncoding RNA function in cancer. *J Mol Med (Berl)* **91**, 791-801, doi:10.1007/s00109-013-1028-y (2013).

11 Cheng, Y. *et al.* Role of metastasis-associated lung adenocarcinoma transcript-1 (MALAT-1) in pancreatic cancer. *PLoS One* **13**, e0192264, doi:10.1371/journal.pone.0192264 (2018).

12 Gutschner, T. *et al.* The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* **73**, 1180-1189, doi:10.1158/0008-5472.CAN-12-2850 (2013).

13 Yang, L., Bai, H. S., Deng, Y. & Fan, L. High MALAT1 expression predicts a poor prognosis of cervical cancer and promotes cancer cell growth and invasion. *Eur Rev Med Pharmacol Sci* **19**, 3187-3193 (2015).

14 Qiu, M. T., Hu, J. W., Yin, R. & Xu, L. Long noncoding RNA: an emerging paradigm of cancer research. *Tumour Biol* **34**, 613-620, doi:10.1007/s13277-013-0658-6 (2013).

15 Chen, W. *et al.* MALAT1 is a prognostic factor in glioblastoma multiforme and induces chemoresistance to temozolomide through suppressing miR-203 and promoting thymidylate synthase expression. *Oncotarget* **8**, 22783-22799, doi:10.18632/oncotarget.15199 (2017).

16 Malakar, P. *et al.* Long Noncoding RNA MALAT1 Promotes Hepatocellular Carcinoma Development by SRSF1 Upregulation and mTOR Activation. *Cancer Res* **77**, 1155-1167, doi:10.1158/0008-5472.CAN-16-1508 (2017).

17 Wilusz, J. E., Freier, S. M. & Spector, D. L. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**, 919-932, doi:10.1016/j.cell.2008.10.012 (2008).

18 Brown, J. A. *et al.* Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat Struct Mol Biol* **21**, 633-640, doi:10.1038/nsmb.2844 (2014).

19 Xu, S. *et al.* Downregulation of long noncoding RNA MALAT1 induces epithelial-to-mesenchymal transition via the PI3K-AKT pathway in breast cancer. *Int J Clin Exp Pathol* **8**, 4881-4891 (2015).

20 Arun, G. *et al.* Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev* **30**, 34-51, doi:10.1101/gad.270959.115 (2016).

21 Huarte, M. The emerging role of lncRNAs in cancer. *Nat Med* **21**, 1253-1261, doi:10.1038/nm.3981 (2015).

22 Lin, L., Li, H., Zhu, Y., He, S. & Ge, H. Expression of metastasis-associated lung adenocarcinoma transcript 1 long non-coding RNA in vitro and in patients with non-small cell lung cancer. *Oncol Lett* **15**, 9443-9449, doi:10.3892/ol.2018.8531 (2018).

23 Donlic, A. *et al.* Discovery of Small Molecule Ligands for MALAT1 by Tuning an RNA-Binding Scaffold. *Angewandte Chemie (International ed. in English)* **57**, 13242-13247, doi:10.1002/anie.201808823 (2018).

24 Abulwerdi, F. A. *et al.* Selective Small-Molecule Targeting of a Triple Helix Encoded by the Long Noncoding RNA, MALAT1. *ACS chemical biology* **14**, 223-235, doi:10.1021/acschembio.8b00807 (2019).

25 Yu, X., Li, Z., Zheng, H., Chan, M. T. & Wu, W. K. NEAT1: A novel cancer-related long non-coding RNA. *Cell Prolif* **50**, doi:10.1111/cpr.12329 (2017).

26 Nishimoto, Y. *et al.* The long non-coding RNA nuclear-enriched abundant transcript 1_2 induces paraspeckle formation in the motor neuron during the early phase of amyotrophic lateral sclerosis. *Mol Brain* **6**, 31, doi:10.1186/1756-6606-6-31 (2013).

27 Liu, X. *et al.* Long non-coding RNA NEAT1-modulated abnormal lipolysis via ATGL drives hepatocellular carcinoma proliferation. *Mol Cancer* **17**, 90, doi:10.1186/s12943-018-0838-5 (2018).

28 Riva, P., Ratti, A. & Venturin, M. The Long Non-Coding RNAs in Neurodegenerative Diseases: Novel Mechanisms of Pathogenesis. *Curr Alzheimer Res* **13**, 1219-1231 (2016).

29 Mitton-Fry, R. M., DeGregorio, S. J., Wang, J., Steitz, T. A. & Steitz, J. A. Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* **330**, 1244-1247, doi:10.1126/science.1195858 (2010).

30 Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484-1488, doi:10.1126/science.1138341 (2007).

31 Alexandrov, A., Shu, M. D. & Steitz, J. A. Fluorescence Amplification Method for Forward Genetic Discovery of Factors in Human mRNA Degradation. *Mol Cell* **65**, 191-201, doi:10.1016/j.molcel.2016.11.032 (2017).

32 Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87, doi:10.1126/science.1247005 (2014).

33 Wu, J. *et al.* Cryo-EM Structure of the Human Ribonuclease P Holoenzyme. *Cell* **175**, 1393-1404.e1311, doi:10.1016/j.cell.2018.10.003 (2018).

34 Weick, E. M. *et al.* Helicase-Dependent RNA Decay Illuminated by a Cryo-EM Structure of a Human Nuclear RNA Exosome-MTR4 Complex. *Cell* **173**, 1663-1677.e1621, doi:10.1016/j.cell.2018.05.041 (2018).

35 Tomecki, R. *et al.* The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *Embo j* **29**, 2342-2357, doi:10.1038/emboj.2010.121 (2010).

36 Lubas, M. *et al.* Interaction profiling identifies the human nuclear exosome targeting complex. *Mol Cell* **43**, 624-637, doi:10.1016/j.molcel.2011.06.028 (2011).

37 Schilders, G., van Dijk, E. & Pruijn, G. J. C1D and hMtr4p associate with the human exosome subunit PM/Scl-100 and are involved in pre-rRNA processing. *Nucleic Acids Res* **35**, 2564-2572, doi:10.1093/nar/gkm082 (2007).

38 Gouge, J. *et al.* Redox Signaling by the RNA Polymerase III TFIIB-Related Factor Brf2. *Cell* **163**, 1375-1387, doi:10.1016/j.cell.2015.11.005 (2015).

39 Shamseldin, H. E. *et al.* Mutations in DDX59 implicate RNA helicase in the pathogenesis of orofaciodigital syndrome. *Am J Hum Genet* **93**, 555-560, doi:10.1016/j.ajhg.2013.07.012 (2013).

40 Yang, L. *et al.* EGFR and Ras regulate DDX59 during lung cancer development. *Gene* **642**, 95-102, doi:10.1016/j.gene.2017.11.029 (2018).

41 Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* **49**, 1779-1784, doi:10.1038/ng.3984 (2017).

42 Köferle, A. *et al.* Interrogation of cancer gene dependencies reveals paralog interactions of autosome and sex chromosome-encoded genes. *Cell Rep* **39**, 110636, doi:10.1016/j.celrep.2022.110636 (2022).

43 Puno, M. R. & Lima, C. D. Structural basis for MTR4-ZCCHC8 interactions that stimulate the MTR4 helicase in the nuclear exosome-targeting complex. *Proc Natl Acad Sci U S A* **115**, E5506-e5515, doi:10.1073/pnas.1803530115 (2018).

44 Puno, M. R., Weick, E. M., Das, M. & Lima, C. D. SnapShot: The RNA Exosome. *Cell* **179**, 282-282.e281, doi:10.1016/j.cell.2019.09.005 (2019).

45 Li, B. *et al.* RIP-PEN-seq identifies a class of kink-turn RNAs as splicing regulators. *Nat Biotechnol* **42**, 119-131, doi:10.1038/s41587-023-01749-0 (2024).

46 Will, C. L. *et al.* The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *Rna* **10**, 929-941, doi:10.1261/rna.7320604 (2004).

47 Salpietro, V. *et al.* A loss-of-function homozygous mutation in DDX59 implicates a conserved DEAD-box RNA helicase in nervous system development and function. *Hum Mutat* **39**, 187-192, doi:10.1002/humu.23368 (2018).

48 Franco, B. & Thauvin-Robinet, C. Update on oral-facial-digital syndromes (OFDS). *Cilia* **5**, 12, doi:10.1186/s13630-016-0034-4 (2016).

49 Strong, A. *et al.* Expanding the genetic landscape of oral-facial-digital syndrome with two novel genes. *American journal of medical genetics. Part A* **185**, 2409-2416, doi:10.1002/ajmg.a.62337 (2021).

50 Yamada, M. *et al.* IFT172 as the 19th gene causative of oral-facial-digital syndrome. *American journal of medical genetics. Part A* **179**, 2510-2513, doi:10.1002/ajmg.a.61373 (2019).

51 Khatri, D. *et al.* Deficiency of the minor spliceosome component U4atac snRNA secondarily results in ciliary defects in human and zebrafish. *Proc Natl Acad Sci U S A* **120**, e2102569120, doi:10.1073/pnas.2102569120 (2023).

52 Cuello, P., Boyd, D. C., Dye, M. J., Proudfoot, N. J. & Murphy, S. Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *Embo j* **18**, 2867-2877, doi:10.1093/emboj/18.10.2867 (1999).

53 Baillat, D. *et al.* Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265-276, doi:10.1016/j.cell.2005.08.019 (2005).

54 Hrossova, D. *et al.* RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3'-end extended forms of snRNAs. *Nucleic Acids Res* **43**, 4236-4248, doi:10.1093/nar/gkv240 (2015).

55 Barra, J. *et al.* Integrator restrains paraspeckles assembly by promoting isoform switching of the lncRNA NEAT1. *Science advances* **6**, eaaz9072, doi:10.1126/sciadv.aaz9072 (2020).

56 Adriaens, C. *et al.* The long noncoding RNA NEAT1_1 is seemingly dispensable for normal tissue homeostasis and cancer cell growth. *Rna* **25**, 1681-1695, doi:10.1261/rna.071456.119 (2019).

57 Shi, Z., Gao, H., Bai, X. C. & Yu, H. Cryo-EM structure of the human cohesin-NIPBL-DNA complex. *Science* **368**, 1454-1459, doi:10.1126/science.abb0981 (2020).

58 Panarotto, M., Davidson, I. F., Litos, G., Schleiffer, A. & Peters, J. M. Cornelia de Lange syndrome mutations in NIPBL can impair cohesin-mediated DNA loop extrusion. *Proc Natl Acad Sci U S A* **119**, e2201029119, doi:10.1073/pnas.2201029119 (2022).

59 Kawauchi, S. *et al.* Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. *PLoS Genet* **5**, e1000650, doi:10.1371/journal.pgen.1000650 (2009).

60 Zuin, J. *et al.* Regulation of the cohesin-loading factor NIPBL: Role of the lncRNA NIPBL-AS1 and identification of a distal enhancer element. *PLoS Genet* **13**, e1007137, doi:10.1371/journal.pgen.1007137 (2017).

61 Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857, doi:10.1016/0092-8674(83)90117-4 (1983).

62 Bai, R. *et al.* Structure of the activated human minor spliceosome. *Science* **371**, doi:10.1126/science.abg0879 (2021).

63 Iturrate, A. *et al.* Mutations in SCNM1 cause orofaciodigital syndrome due to minor intron splicing defects affecting primary cilia. *Am J Hum Genet* **109**, 1828-1849, doi:10.1016/j.ajhg.2022.08.009 (2022).

64 Huang, C. H., Hsiao, H. T., Chu, Y. R., Ye, Y. & Chen, X. Derlin2 protein facilitates HRD1-mediated retro-translocation of sonic hedgehog at the endoplasmic reticulum. *J Biol Chem* **288**, 25330-25339, doi:10.1074/jbc.M113.455212 (2013).

65 Zablowsky, N. *et al.* High Throughput FISH Screening Identifies Small Molecules That Modulate Oncogenic lncRNA MALAT1 via GSK3B and hnRNPs. *Non-coding RNA* **9**, doi:10.3390/ncrna9010002 (2023).

66 Chen, C. *et al.* Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* **356**, 189-194, doi:10.1126/science.aak9787 (2017).

67 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

68 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, doi:10.1093/gigascience/giab008 (2021).

69 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).

70 Sehnal, D. *et al.* Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res* **49**, W431-w437, doi:10.1093/nar/gkab314 (2021).

Figure 1

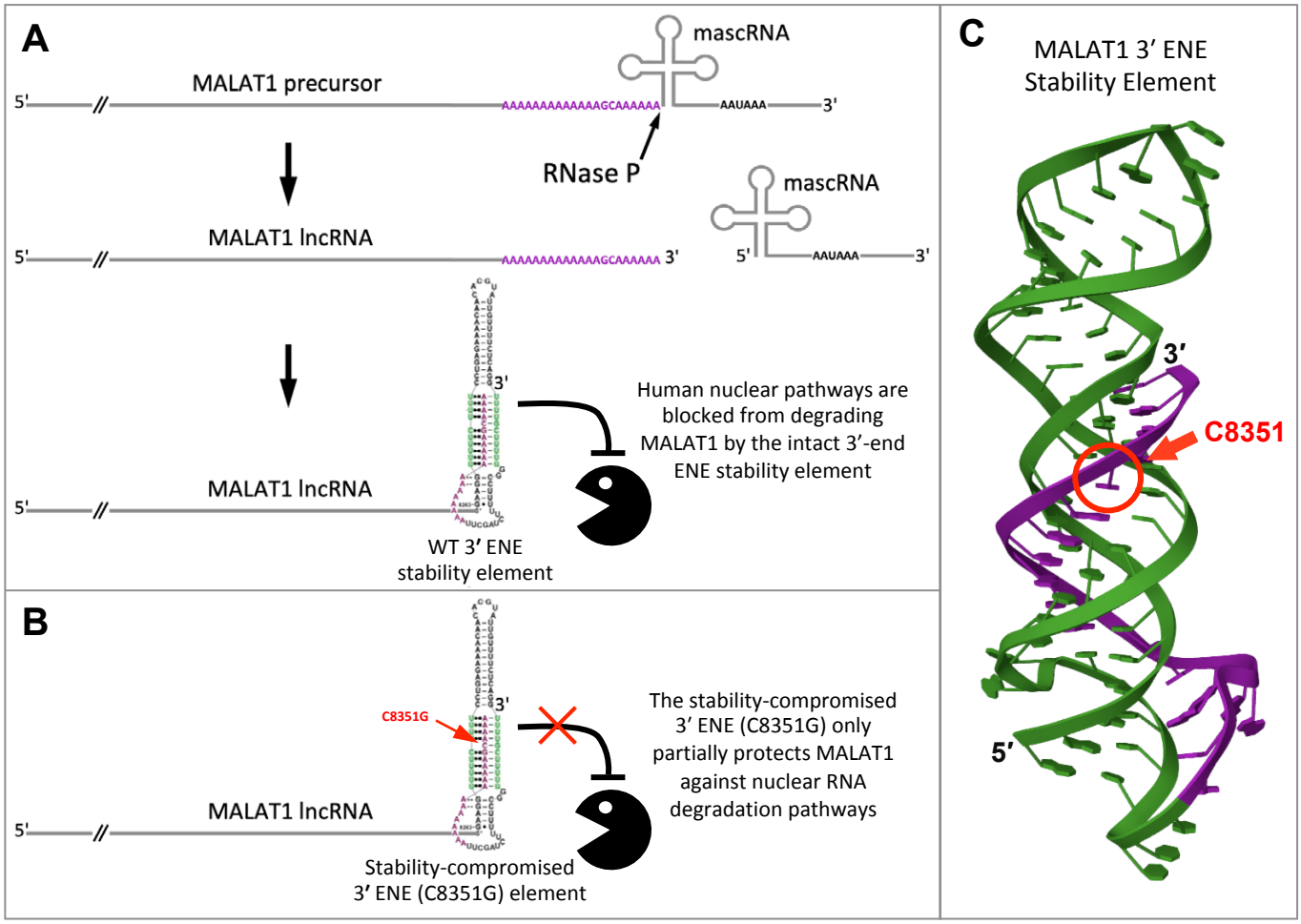

A
mascRNA
MALAT1 precursor
5'
AAAAAAAAAAAAAGCAAAAAA
AAUAAA
3'
RNase P
mascRNA
MALAT1 lncRNA
5'
AAAAAAAAAAAAAGCAAAAAA 3'
5'
AAUAAA
3'
3'
Human nuclear pathways are blocked from degrading MALAT1 by the intact 3'-end ENE stability element
MALAT1 lncRNA
5'
WT 3' ENE stability element
B
C8351G
3'
The stability-compromised 3' ENE (C8351G) only partially protects MALAT1 against nuclear RNA degradation pathways
MALAT1 lncRNA
5'
Stability-compromised 3' ENE (C8351G) element
C
MALAT1 3' ENE Stability Element
3'
C8351
5'

Figure 2

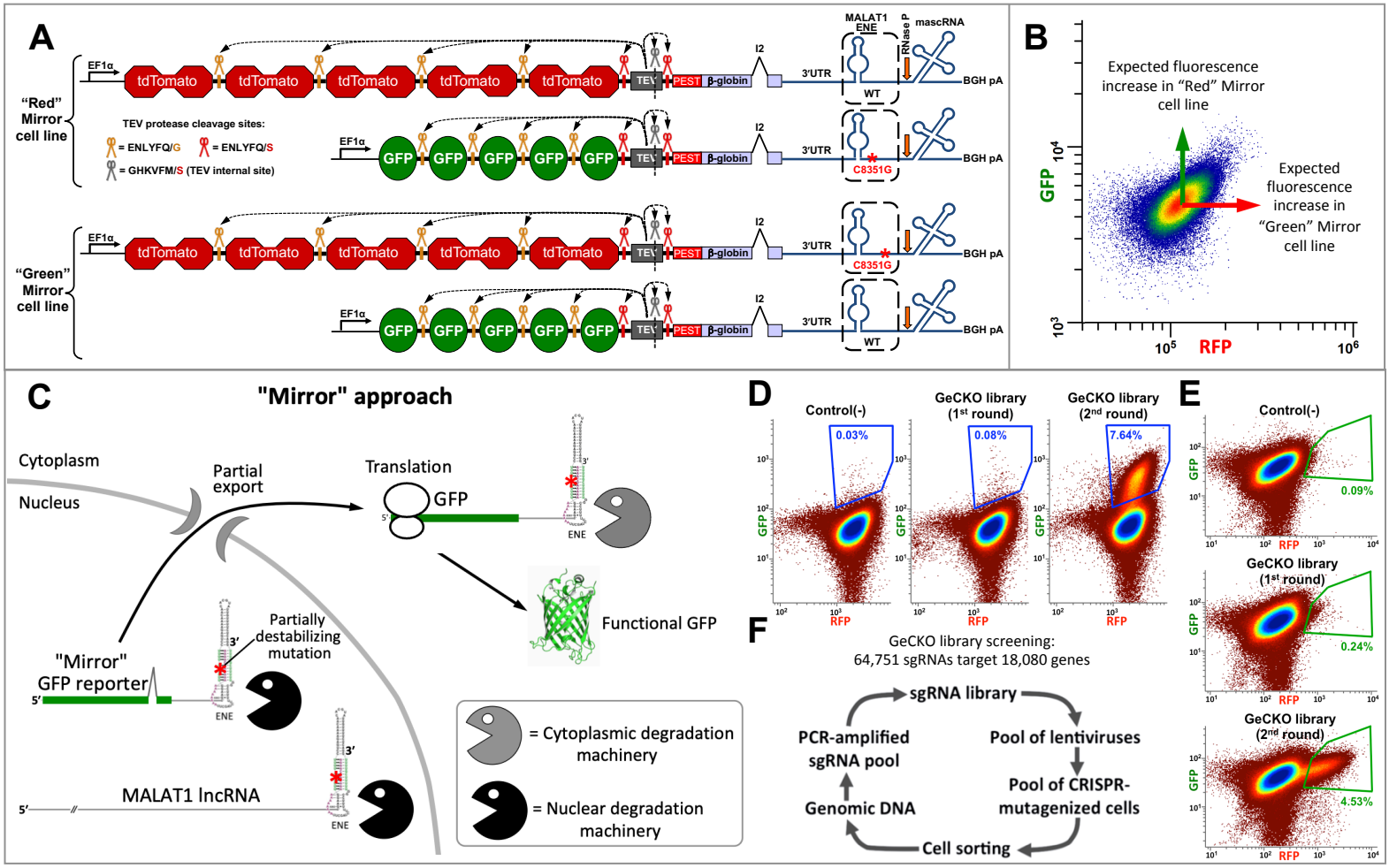

A
"Red" Mirror cell line
"Green" Mirror cell line
EF1α
tdTomato
GFP
TEV
PEST
β-globin
I2
3'UTR
BGH pA
MALAT1 ENE
RNase P
mascRNA
WT
C8351G
TEV protease cleavage sites:
= ENLYFQ/G
= ENLYFQ/S
= GHKVFM/S (TEV internal site)
B
Expected fluorescence increase in "Red" Mirror cell line
Expected fluorescence increase in "Green" Mirror cell line
GFP
RFP
C
"Mirror" approach
Cytoplasm
Nucleus
Partial export
Translation
GFP
ENE
Functional GFP
Partially destabilizing mutation
"Mirror" GFP reporter
MALAT1 lncRNA
= Cytoplasmic degradation machinery
= Nuclear degradation machinery
D
Control(-)
GeCKO library (1st round)
GeCKO library (2nd round)
0.03%
0.08%
7.64%
E
0.09%
0.24%
4.53%
F
GeCKO library screening:
64,751 sgRNAs target 18,080 genes
sgRNA library
Pool of lentiviruses
Pool of CRISPR-mutagenized cells
Cell sorting
Genomic DNA
PCR-amplified sgRNA pool

Figure 3

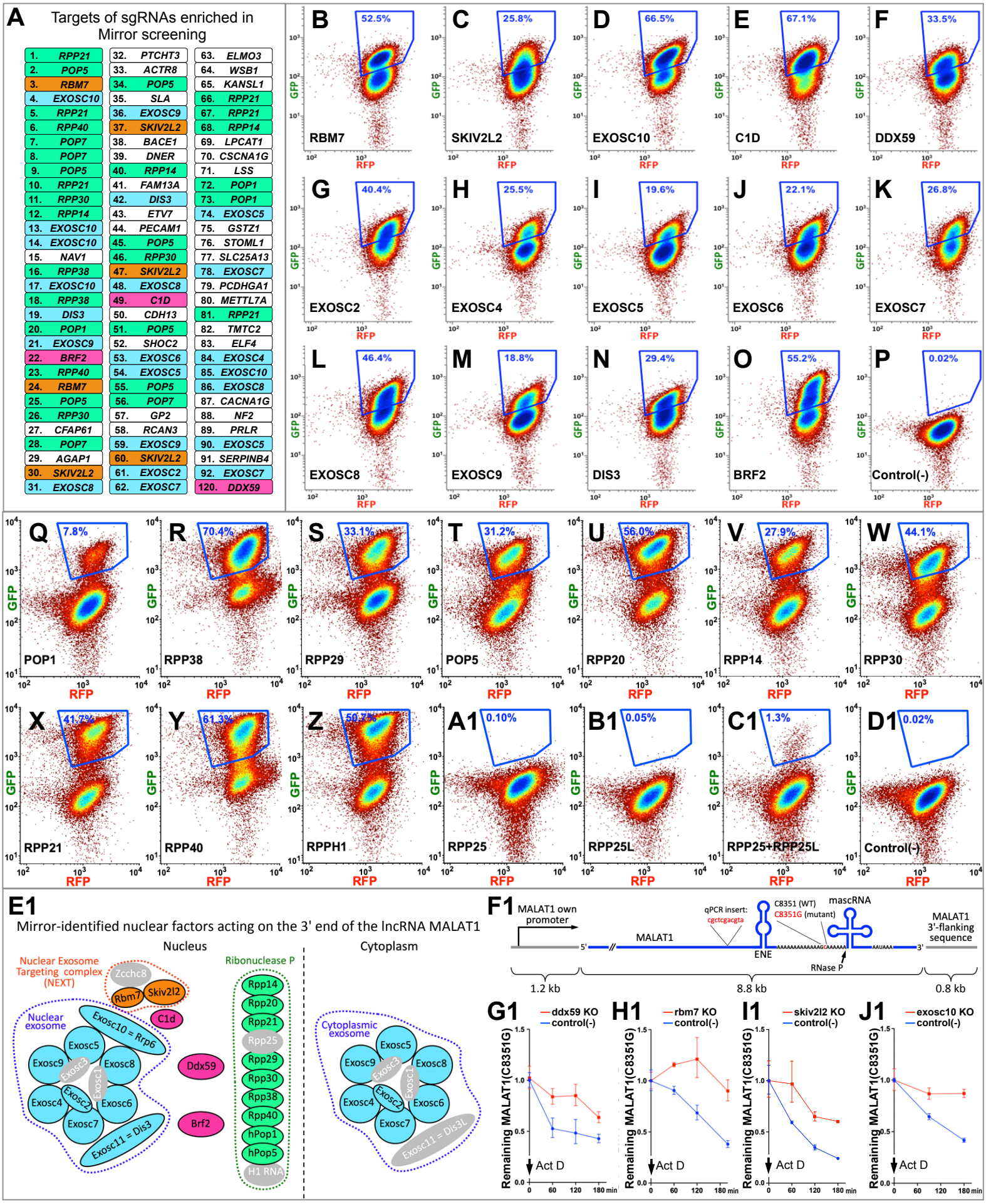

A
Targets of sgRNAs enriched in Mirror screening
1. RPP21
2. POP5
3. RBM7
4. EXOSC10
5. RPP21
6. RPP40
7. POP7
8. POP7
9. POP5
10. RPP21
11. RPP30
12. RPP14
13. EXOSC10
14. EXOSC10
15. NAV1
16. RPP38
17. EXOSC10
18. RPP38
19. DIS3
20. POP1
21. EXOSC9
22. BRF2
23. RPP40
24. RBM7
25. POP5
26. RPP30
27. CFAP61
28. POP7
29. AGAP1
30. SKIV2L2
31. EXOSC8
32. PTCHT3
33. ACTR8
34. POP5
35. SLA
36. EXOSC9
37. SKIV2L2
38. BACE1
39. DNER
40. RPP14
41. FAM13A
42. DIS3
43. ETV7
44. PECAM1
45. POP5
46. RPP30
47. SKIV2L2
48. EXOSC8
49. C1D
50. CDH13
51. POP5
52. SHOC2
53. EXOSC6
54. EXOSC5
55. POP5
56. POP7
57. GP2
58. RCAN3
59. EXOSC9
60. SKIV2L2
61. EXOSC2
62. EXOSC7
63. ELMO3
64. WSB1
65. KANSL1
66. RPP21
67. RPP21
68. RPP14
69. LPCAT1
70. CSCNA1G
71. LSS
72. POP1
73. POP1
74. EXOSC5
75. GSTZ1
76. STOML1
77. SLC25A13
78. EXOSC7
79. PCDHGA1
80. METTL7A
81. RPP21
82. TMTC2
83. ELF4
84. EXOSC4
85. EXOSC10
86. EXOSC8
87. CACNA1G
88. NF2
89. PRLR
90. EXOSC5
91. SERPINB4
92. EXOSC7
120. DDX59
B 52.5% RBM7
C 25.8% SKIV2L2
D 66.5% EXOSC10
E 67.1% C1D
F 33.5% DDX59
G 40.4% EXOSC2
H 25.5% EXOSC4
I 19.6% EXOSC5
J 22.1% EXOSC6
K 26.8% EXOSC7
L 46.4% EXOSC8
M 18.8% EXOSC9
N 29.4% DIS3
O 55.2% BRF2
P 0.02% Control(-)
Q 7.8% POP1
R 70.4% RPP38
S 33.1% RPP29
T 31.2% POP5
U 56.0% RPP20
V 27.9% RPP14
W 44.1% RPP30
X 41.7% RPP21
Y 61.3% RPP40
Z 50.7% RPPH1
A1 0.10% RPP25
B1 0.05% RPP25L
C1 1.3% RPP25+RPP25L
D1 0.02% Control(-)
GFP
RFP
E1
Mirror-identified nuclear factors acting on the 3' end of the lncRNA MALAT1
Nucleus
Cytoplasm
Nuclear Exosome Targeting complex (NEXT)
Zcchc8
Rbm7
Skiv2l2
C1d
Nuclear exosome
Exosc10 = Rrp6
Exosc5
Exosc9
Exosc3
Exosc8
Exosc1
Exosc4
Exosc2
Exosc6
Exosc7
Exosc11 = Dis3
Ddx59
Brf2
Ribonuclease P
Rpp14
Rpp20
Rpp21
Rpp25
Rpp29
Rpp30
Rpp38
Rpp40
hPop1
hPop5
H1 RNA
Cytoplasmic exosome
Exosc11 = Dis3L
F1
MALAT1 own promoter
MALAT1
qPCR insert:
cgctcgacgta
C8351 (WT)
C8351G (mutant)
mascRNA
MALAT1 3'-flanking sequence
ENE
RNase P
1.2 kb
8.8 kb
0.8 kb
G1
ddx59 KO
control(-)
H1
rbm7 KO
control(-)
I1
skiv2l2 KO
control(-)
J1
exosc10 KO
control(-)
Remaining MALAT1(C8351G)
Act D
min

Figure 4

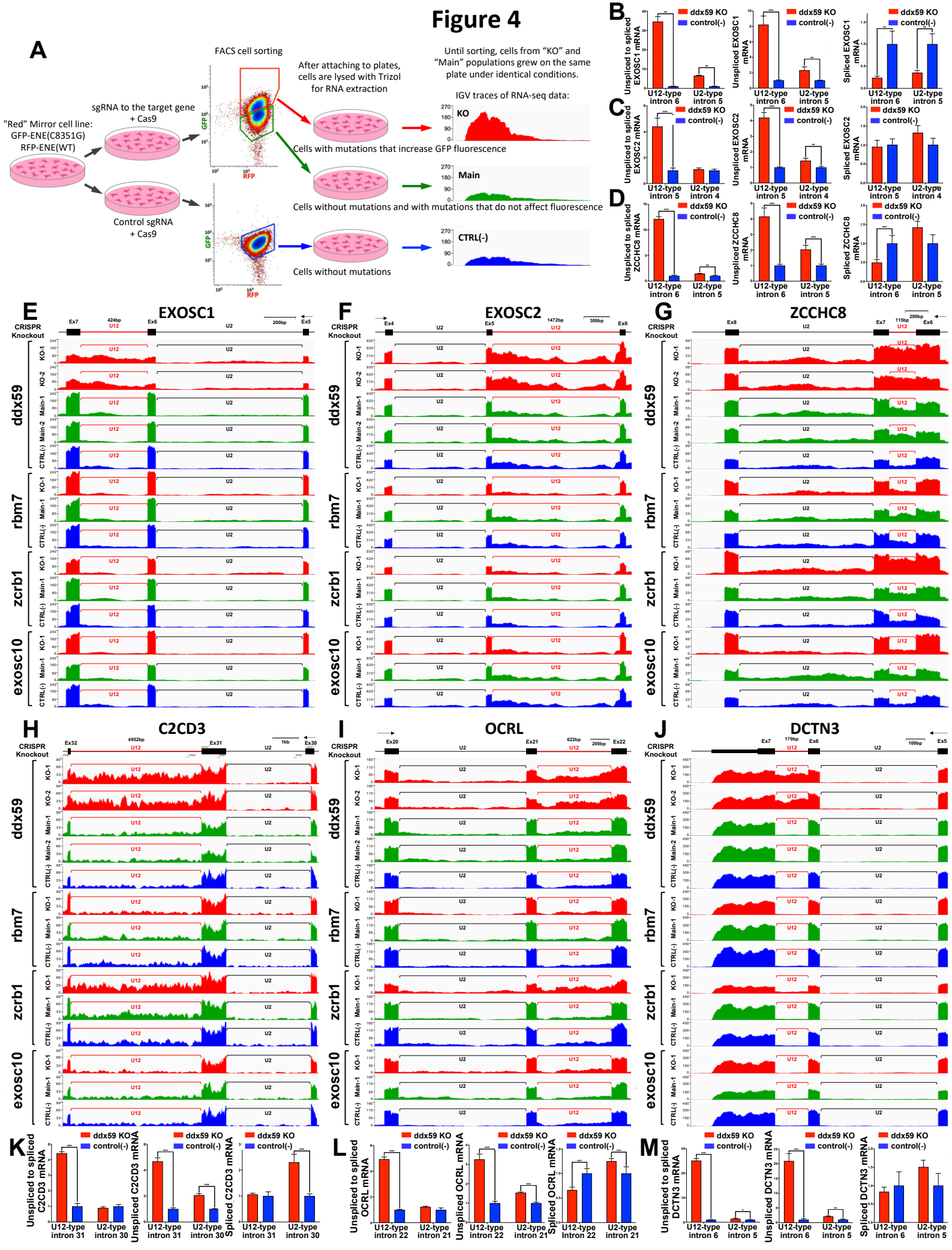
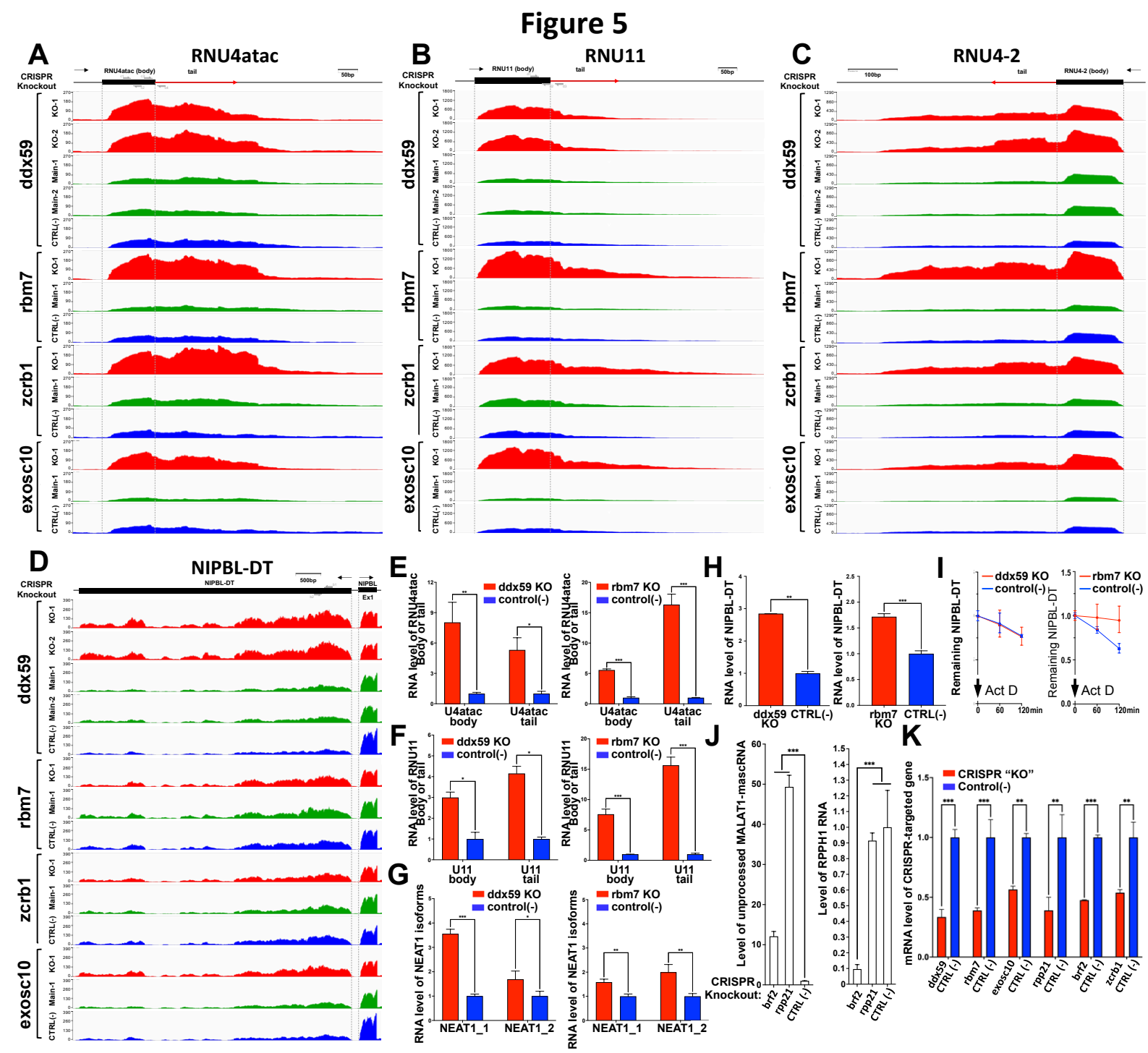

A
FACS cell sorting
"Red" Mirror cell line:
GFP-ENE(C8351G)
RFP-ENE(WT)
sgRNA to the target gene
+ Cas9
Control sgRNA
+ Cas9
After attaching to plates, cells are lysed with Trizol for RNA extraction
Until sorting, cells from "KO" and "Main" populations grew on the same plate under identical conditions.
IGV traces of RNA-seq data:
Cells with mutations that increase GFP fluorescence
Cells without mutations and with mutations that do not affect fluorescence
Cells without mutations
KO
Main
CTRL(-)
GFP
RFP
B
C
D
ddx59 KO
control(-)
Unspliced to spliced EXOSC1 mRNA
Unspliced EXOSC1 mRNA
Spliced EXOSC1 mRNA
Unspliced to spliced EXOSC2 mRNA
Unspliced EXOSC2 mRNA
Spliced EXOSC2 mRNA
Unspliced to spliced ZCCHC8 mRNA
Unspliced ZCCHC8 mRNA
Spliced ZCCHC8 mRNA
U12-type intron 6
U2-type intron 5
U12-type intron 5
U2-type intron 4
E
EXOSC1
F
EXOSC2
G
ZCCHC8
CRISPR Knockout
ddx59
rbm7
zcrb1
exosc10
KO-1
KO-2
Main-1
Main-2
CTRL(-)
U12
U2
H
C2CD3
I
OCRL
J
DCTN3
K
Unspliced to spliced C2CD3 mRNA
Unspliced C2CD3 mRNA
Spliced C2CD3 mRNA
U12-type intron 31
U2-type intron 30
L
Unspliced to spliced OCRL mRNA
Unspliced OCRL mRNA
Spliced OCRL mRNA
U12-type intron 22
U2-type intron 21
M
Unspliced to spliced DCTN3 mRNA
Unspliced DCTN3 mRNA
Spliced DCTN3 mRNA

Figure 5


A
RNU4atac
B
RNU11
C
RNU4-2
D
NIPBL-DT
CRISPR Knockout
ddx59
rbm7
zcrb1
exosc10
KO-1
KO-2
Main-1
Main-2
CTRL(-)
E
RNA level of RNU4atac Body or tail
ddx59 KO
control(-)
rbm7 KO
U4atac body
U4atac tail
F
RNA level of RNU11 Body or tail
U11 body
U11 tail
G
RNA level of NEAT1 isoforms
NEAT1_1
NEAT1_2
H
RNA level of NIPBL-DT
I
Remaining NIPBL-DT
Act D
J
Level of unprocessed MALAT1-mascRNA
Level of RPPH1 RNA
brf2
rpp21
K
mRNA level of CRISPR-targeted gene
CRISPR "KO"
Control(-)