Fakultät für Medizin
Technische Universität München

Technische Universität München

Development of non-invasive tools for interrogating alternative splicing of coding genes and monitoring the expression of non-coding RNA

Vollständiger Abdruck der von der Fakultät für Medizin
der der Technische Universität München zur Erlangung des
akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten
Dissertation.

Dong-Jiunn Jeffery TRUONG

Vorsitzender: Prof. Dr. Dr. Stefan ENGELHARDT

Prüfende der Dissertation:

1. Prof. Dr. Gil Gregor WESTMEYER
2. Prof. Dr. Wolfgang WURST

Die Dissertation wurde am 21.08.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 11.02.2020 angenommen.

Development of non-invasive tools for interrogating alternative splicing of coding genes and monitoring the expression of non-coding RNA

ABSTRACT

In this thesis, two non-invasive methods have been developed that allow longitudinal gene expression analysis of alternatively spliced proteins and non-coding RNAs.

The first part describes the development of a method to track exon-specific isoform expression, a result of RNA alternative splicing of coding genes, through protein splicing. Alternative splicing of pre-mRNA is a central post-transcriptional process in eukaryotes that enables the dynamic expression of a variety of protein isoforms from a fixed number of genes. Current methodologies to investigate alternative splicing are destructive, laborious, and not suitable for processing large quantities of samples. Therefore, an exon-specific isoform expression reporter system (EXSISERS) was developed that is non-invasively and longitudinally indicating the translation of a defined exon-containing isoform of an endogenous protein-coding gene by excising different reporter proteins from the nascent polypeptide chain scarlessly via fast, intein-mediated protein splicing. After initial optimization, EXSISERS was used to quantify the inclusion of the exon 10 in microtubule-associated protein tau (MAPT), involved in many diseases, and to optimize programmable RNA-targeting effector systems, including classical shRNA approaches and the latest Cas13 systems. With a modified EXSISERS, which couples cell survival to the inclusion of exon 18b of FOXP1 that plays a prominent role in pluripotency maintenance of embryonic stem cells (ESCs), it could be confirmed that MBNL1 is a primary regulator for exon 18b exclusion. Hence, EXSISERS facilitates the sensitive non-disruptive monitoring of exon-specific expression with cellular resolution with a reporter modality of choice and empowers high-throughput screening applications of exon-specific therapeutic perturbations.

The second part of the thesis deals with the non-coding part of the genome, which currently cannot be tracked longitudinally and non-invasively. For this purpose, a second method has been developed based on RNA-introns to track ncRNA non-invasively. Non-coding transcripts represent approximately 80% of the transcriptome and are involved in the regulation of various cellular processes, including many diseases. Current methods to quantify non-coding transcripts are consuming, not longitudinal, and do not allow for high-throughput screening with cellular resolution. Therefore, an intron-specific exon-independent coding transcript system (INSPECT) was developed that introduces a reporter gene hidden in an engineered intron, allowing scarless transcriptional coupling of the non-coding host gene to the reporter modality. After transcription, the synthetic or native intron, which is equipped with elements that enable nuclear export and cap-independent translation of the reporter gene hidden in the intron, is spliced out of the naive transcript, leaving the non-coding RNA scarlessly religated. By introducing the INSPECT system into the long non-coding RNA (lncRNA) *NEAT1*, it could be shown that one can couple the expression of a reporter gene to that of the host lncRNA without a trace. In addition, *NEAT1* was knocked out completely using an INSPECT-integrated Cre recombinase-mediated off-switch. Thus, INSPECT is a versatile tool, compatible with high-throughput screenings, which allows longitudinal monitoring of lncRNAs with cellular resolution without consuming the cells as necessary for all currently available methods.

In summary, EXSISERS and INSPECT represent two new, non-invasive monitoring methods that will enable new findings regarding two fundamental biological phenomena, exon-specific isoform expression and non-coding RNAs.

Entwicklung nicht-invasiver Werkzeuge zur Untersuchung alternativen Spleißens kodierender Gene und zur Verfolgung der Expression nicht-kodierender RNA

ZUSAMMENFASSUNG

In dieser Arbeit wurden zwei nicht-invasive Methoden entwickelt, die eine longitudinale Genexpressionsanalyse von alternativ gespleißten Proteinen und nicht-kodierenden RNAs ermöglichen.

Der erste Teil beschreibt die Entwicklung einer Methode zur Verfolgung des exonspezifischen Isoform-Expression, ein Ergebnis des alternativen RNA-Spleißens kodierender Gene durch Proteinspleißen. Alternatives Spleißen von prä-mRNA ist ein zentraler posttranskriptioneller Prozess in Eukaryoten, der die dynamische Expression einer Vielzahl von Protein-Isoformen aus einer festen Anzahl von Genen ermöglicht. Die derzeitigen Methoden zur Untersuchung des alternativen Spleißens sind verbrauchend, arbeitsintensiv und für die Verabeitung großer Probenmengen nicht geeignet. Deshalb wurde ein exonspezifisches Isoform-Expressions-Reportersystem (EXSISERS) entwickelt, das nicht-invasiv und longitudinal die Translation einer definierten exonhaltigen Isoform eines endogenen proteinkodierenden Gens anzeigt, indem verschiedene Reporterproteine aus der entstehenden Polypeptidkette durch schnelles, inteinvermitteltes Proteinspleißen narbenlos herausgeschnitten werden. Nach einer ersten Optimierung wurde EXSISERS zur Quantifizierung der Inklusion des Exons 10 im Mikrotubuli-assoziiertes Protein Tau (MAPT), das an vielen Krankheiten beteiligt ist, und zur Optimierung programmierbarer RNA-Targeting Effektorsysteme, einschließlich klassischer shRNA Ansätze und der neuesten Cas13 Systeme, eingesetzt. Mit einem modifizierten EXSISERS, der das Überleben der Zellen an die Inklusion von Exons 18b von FOXP1 koppelt, das eine entscheidende Rolle bei der Erhaltung der Pluripotenz embryonaler Stammzellen (ESCs) spielt, konnte bestätigt werden, dass MBNL1 ein primärer Regulator für die Exklusion von Exon 18b ist. Folglich erleichtert EXSISERS die sensitive, nicht-zerstörende Verfolgung der exonspezifischen Expression mit zellulärer Auflösung und ermöglicht Hochdurchsatz-Screening-Anwendungen exonspezifischer therapeutischer Maßnahmen.

Der zweite Teil der Thesis beschäftigt sich mit dem nicht kodierenden Teil des Genoms, der derzeit nicht longitudinal und nicht-invasiv verfolgt werden kann. Zu diesem Zweck wurde eine zweite Methode entwickelt, die auf der Basis von RNA-Intronen entwickelt wurde, um ncRNA nicht-invasiv zu verfolgen. Nicht-kodierende Transkripte machen etwa 80 % des Transkriptoms aus und sind an der Regulation verschiedener zellulärer Prozesse, darunter viele Krankheiten, beteiligt. Aktuelle Methoden zur Quantifizierung nicht-kodierender Transkripte sind verbrauchend, nicht longitudinal und erlauben kein Hochdurchsatz-Screening mit zellulärer Auflösung. Daher wurde ein System für ein Intron-spezifisches Exon-unabhängiges kodierendes Transkript (INSPECT) entwickelt, das ein Reportergen, versteckt in einem synthetischen Intron, einführt, was eine narbenlose transkriptionelle Kopplung des nicht-kodierenden Wirtsgens an die Reporter-Modalität ermöglicht. Nach der Transkription wird das synthetische oder native Intron, das mit Elementen ausgestattet ist, die einen Kernexport und eine cap-unabhängige Translation des im Intron versteckten Reportergens ermöglichen, aus dem naiven Transkript herausgespleißt, wodurch die nicht-kodierende RNA narbenlos religiert wird. Durch die Einführung des INSPECT-Systems in die lange nicht-kodierende RNA (lncRNA) *NEAT1* konnte gezeigt werden, dass man die Expression eines Reportergens spurlos an die der Wirts-lncRNA koppeln kann. Darüber hinaus konnte *NEAT1* mittels eines INSPECT-integriertem Cre-Rekombinase-vermittelten Ausschalters komplett inaktiviert werden. Damit ist INSPECT ein vielseitiges Werkzeug, das mit Hochdurchsatz-Screenings kompatibel ist und die longitudinale Untersuchung von lncRNAs mit zellulärer Auflösung ermöglicht, ohne die Zellen zu verbrauchen, wie es für alle derzeit verfügbaren Methoden erforderlich ist.

Zusammenfassend stellen EXSISERS und INSPECT zwei neue, nicht-invasive Untersuchungsmethoden dar, die neue Erkenntnisse über zwei grundlegende biologische Phänomene ermöglichen werden, exonspezifische Isoform-Expression und nicht-kodierende RNAs.

BACKGROUND & MOTIVATION

From a molecular biologist's view, life can be described as a self-sustaining multi-layered network comprising different complex substrates, *e.g.,* DNA, RNA, proteins/peptides, (an)organic metabolites, and ions. One of the most critical layers is the DNA→RNA→protein-pathway, also called the "central dogma," which describes the canonical gene expression of protein-coding genes (**Figure 1**), whereas non-coding gene expression lacks the protein part. Also, via alternative splicing, a variety of different isoforms can be generated by a single gene (**Figure 1**). The dysregulation of this pathway might lead to pathological conditions and might be genetically inherited, epigenetically imprinted, or caused by (a)biotic environmental factors. Currently established methods to assess this pathway, basically, the canonical gene expression, include end-point measurement/quantifications, *e.g.,* RT-qPCR (DNA/RNA), (sm)FISH (DNA/RNA)[1], DNA/RNA-sequencing[2], immunoblots and immunofluorescence labeling (proteins/peptides) and protein/peptide mass-spectrometry[3]. Those methods have their own unique capability, multiplexability, and limitations regarding (sub)cellular resolution, sensitivity, and technical requirements. All those methods lack longitudinal capabilities due to the consumptive nature but do not require prior genetic manipulation of the sample. For longitudinal non-consumptive measurements, one has to couple the expression (transcription or translation) of the endogenous gene of interest to an introduced reporter protein or tag via nuclease-assisted (CRISPR/Cas9, TALENs or ZFNs) homologous recombination[4]. However, the existing methods to tag coding genes are invasive since they are either not scarless (introducing scars on protein level, *e.g.,* 2A ribosome skipping peptides[5] and fusion proteins) and/or change the untranslated region (*e.g.,* using internal ribosome entry sites IRES[6]) and thereby might change the efficiency of translation initiation (5'-UTR) or the half-life of the mRNA (3'-UTR) resulting in a change of the protein level and thus a potential change in the physiology. Moreover, those methods have only pan-gene resolution and cannot distinguish between isoforms resulting from alternative splicing, which occurs in more than 90% of the genes and thus are associated with many diseases[7–9]. Also, the presence of mRNA cannot be equally interpreted as the presence of the encoded protein, since mRNA can be translationally arrested, *e.g.,* in during the transport along the axons or in stress granules, thus statements about protein abundance via RNA quantification can be misleading[10]. Besides this, the methods mentioned above cannot be applied for non-coding RNAs ncRNAs since they do not encode for proteins or are nuclear-localized, where translation does not occur. The only method for longitudinal monitoring of is a highly invasive two-component system, composed of multiple repeats of aptamer motifs introduced into an ncRNA and a constitutively expressed aptamer-biding protein (ABP) fused to a fluorescent protein (FP)[11]. To overcome those limitations, a non-invasive high-throughput screening-compatible exon-specific isoform expression reporter system (EXSISERS) was developed first in this work, based on fast-splicing split-protein introns (split-inteins) and coiled-coil domains, to assess and monitor RNA alternative splicing on protein-level with cellular resolution. Secondly, an intron-specific exon-independent coding transcript (INSPECT) was developed to monitor non-invasively the expression of lncRNAs longitudinally with cellular resolution, by embedding the CDS of a reporter gene on an intronic sequence of natural or synthetic origin and additionally equip it with RNA elements enabling intronic nuclear export and cytosolic protein translation. EXSISERS and INSPECT will complement existing methods to assess gene expression of coding and non-coding genes and combine their advantages with non-invasiveness, single-cell resolution, isoform-specificity, and high-throughput screening-compatibility.

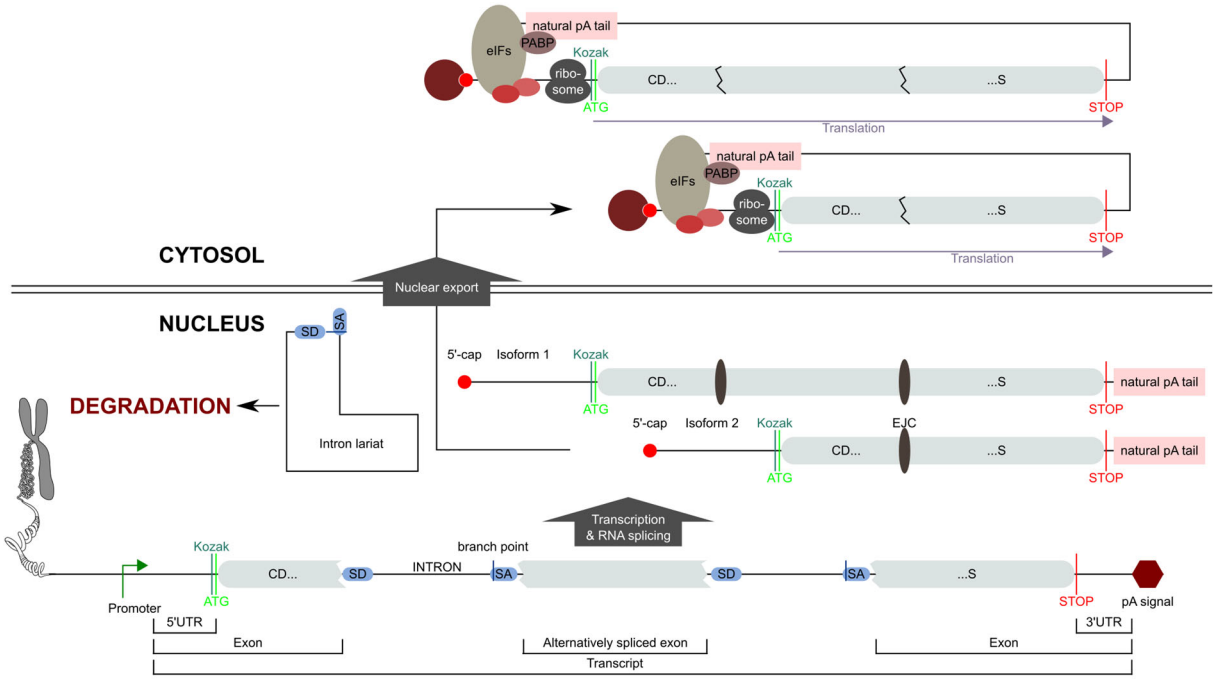

Figure 1 | Schematic of gene transcription and transcript modification and export. Canonical gene expression of most protein-coding genes is driven by an RNA-polymerase II promoter, and 95% of them contain introns that are excised co-/post-transcriptionally, leaving the remaining exons ligated scarlessly. This mechanism is called RNA-splicing and is one of the major steps beside 5'-capping (addition of a 7-methylguanylate cap to the 5'-end of the de-novo transcribed RNA) and 3'-polyadenylation (addition of poly(A) tail to the RNA) resulting in a mature mRNA. Some exons are alternatively spliced, resulting in isoforms with and without this exon. A complex called exon-junction-complex (EJC) will mark the position ~50 nt upstream of an exon-exon-junction after splicing. Afterward, a variety of proteins binds to the 5'-cap and the poly(A) tail stimulating the nuclear export of the mature mRNA. The excised intron is degraded after the 2'-5' phosphodiester bonds of the circular intron is debranched by DBR1. Afterward, the exported mRNA, the 5'-cap-binding and poly(A) binding proteins initiate translation of the CDS by recruiting the ribosomal subunits. The 5'-and 3'-untranslated region (upstream of the start codon ATG and downstream of the stop codon TAA/TGA/TAG) are called 5'-UTR and 3'-UTR.

STATE-OF-THE-ART & CURRENT LIMITATIONS

Genetic manipulation strategies

Transient delivery of DNA

Current methods to track gene expression can be divided into two classes in which foreign DNA is transiently or stably introduced into the cell. In contrast to *in cellulo* (in distinction from *in vitro*, which is used here for cell-free extracts), *in vivo,* only stable methods can be used with a few exceptions. DNA is introduced into cells *in cellulo* either chemically (polyethyleneimine, calcium phosphate, lipids, and polymers of proprietary formulation)[12] or by the application of a short pulse of a defined voltage (electroporation or nucleofection)[13]. Among other things, integrase-defective lentiviruses[14], recombinant adeno-associated viruses (rAAVs)[15], and other replication and integrase-defective viruses can also be used. The disadvantage of the method is that the DNA coding for the reporter is usually present in large excess in the cell and therefore represents a non-physiological condition (a gene is usually present as a 2 or 4-fold copy in the cell in diploid organisms), *i.e.,* the ratio of the factors to the target gene/mRNA/protein is disproportionate and might lead to artifacts/misinterpretations.

Stable integration of DNA

The same methods are used to generate stable *in cellulo* lines, but the following steps are taken to ensure that the genetic material is not lost. The simplest method is to rely on the random integration of the genetic material (mostly circular or linearized plasmid) in the genome (classical transgenesis); since this method is very inefficient, transgenic cells are selected using an antibiotic selection cassette, which, if desired, is removed from the genome afterward via recombinases or transposases[16]. In addition to classical transgenesis, directed integration/modification in a target sequence in the genome is becoming increasingly common. For this purpose, the transgene is flanked with homology arms, which correspond to the 5'- and 3'-flanks of the position to be inserted in the genome. The homology arms are recognized by the endogenous cell system (in the S- and G2-phase of the cell cycle) and mediated by homologous recombination (HR) the transgenic information is copied or transferred to the corresponding position in the genome[17]. Since HR (also known as homology-directed repair: HDR) is normally activated during DNA damage (*e.g.,* double-strand break (DSB) or single-strand break/nick (SSB)), the targeted integration frequency without DSB/SSB is low. The integration frequency can be dramatically increased by a targeted DSB (or SSB), for which great progress has been made in recent years with regard to the ease of use. The first programmable site-directed DNA nucleases (SDNs) began with meganucleases (MNs), followed by zinc-finger nucleases (ZFNs), transcription factor-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9 (CRISPR/Cas9)[18]. The latter represents a breakthrough in feasibility since instead of repeatedly adapting the respective nuclease to the new target DNA by protein engineering for a new target, only a consecutive 20 nucleotide stretch (also called spacer) is changed, which pairs with the respective target DNA via Watson-Crick base pairing, mediated by the generic nuclease Cas9[19–21]. CRISPR/Cas9 was originally derived from bacterial adaptive defense systems and is adapted and simplified for eukaryotes. The best known and characterized Cas9 system is from *Streptococcus pyogenes* (SpCas9 or SpyCas9). The RNA component which guides Cas9 to its target DNA is a chimeric fusion of the tracrRNA and the crRNA and is called single-guide RNA (sgRNA or gRNA), which represents a simplification of the natural system because instead of two RNA molecules only one is expressed[19–21]. The 5'-end of the sgRNA, which is 19-21 nucleotides (nt) long (for SpCas9), is the region responsible for the recognition of the target sequence. The 3'-end of the sgRNA, on the other hand, is the constant stem and is responsible for the recognition and incorporation of the sgRNA into the generic Cas9 nuclease. In addition to the 20 nts, a protospacer-adjacent motif (PAM; 2-6 base pair DNA sequence immediately following the target DNA sequence (protospacer); NGG for SpCas9)) is important for binding and recognition of the target DNA, which ensures that Cas9 loaded with sgRNA

does not target the locus in the bacterial system that encodes for the spacer-containing crRNA direct repeats (avoidance of self-targeting), since the PAM is not coded on the crRNA/sgRNA. Therefore, it is important for the SpCas9 that the target protospacer in the genome (which corresponds exactly to the sequence of the spacer) is followed by NGG, whereby site-specific hybrid (R-loop) with its complementary DNA sequence is formed while displacing the non-complementary strand (contains the protospacer followed by the PAM). If everything matches, then the two nuclease domains of Cas9 (RuvC and HNH) introduces a DSB 3-4 nts upstream of the PAM[22,23]. The DSB introduced by Cas9 stimulates the endogenous repair pathways, which are canonical and alternative non-homologous end-joining (c- and alt-NHEJ), homologous recombination (HR), single-strand annealing (SSA) and microhomology-mediated end-joining (MMEJ, mediated by NHEJ)[24]. For targeted insertion of a transgene, HR is the desired outcome. Thus, methods have been developed to stimulate HR or suppressing other repair pathways. Beside the Cas9 system, which has many orthologues from different organisms and different PAMs, there are a vast variety of RNA-guided dsDNA nuclease systems, *e.g.,* Cas12a (formerly Cpf1), Cas12b (formerly C2c1), Cas12c (formerly C2c3), Cas12d (formerly CasY), Cas12e (formerly CasX), Cas12h, Cas12i (prefers dsDNA nicking instead of a DSB), all with different PAMs (different length, sequence and 5'- or 3'-located from the protospacer) and cutting pattern[25]. Since the eukaryotic DNA in the nucleus is separated from the rest of the cell by a double membrane, the nuclease must be actively transported into the nucleus after translation. This is achieved by adding a nuclear localization sequence (NLS), which is of a bipartite or monopartite nature[26].

Reporter systems

Transgenic lines with surrogate reporters

Due to a lack of alternatives, the first tools to understand canonical gene expression of coding genes (**Figure 2a**) were so-called substitute reporters, *i.e.,* gene expression was not directly measured but was observed with the help of a stable or transient transgene. The simplest way was to couple the promoter of the gene to be investigated with a reporter protein, and a polyadenylation signal, which was then specifically integrated as a transgene into a safe harbor locus (*e.g., AAVS1* in human or *Rosa26* in murine systems) or randomly integrated into the genome (*e.g.,* with lentiviral vectors or transposases) or introduced as a plasmid (transient) into cells. The advantage is the simplicity of the system, but the statement of such systems should only be taken with caution, as the influence of the direct regions on the gene (*e.g.,* epigenetics, distal gene enhancers/silencers) and co-/post-transcriptional regulations on the mRNA are not considered. Moreover, 90% of genes are alternatively spliced and some alternatively spliced mRNAs result in completely unrelated products. A promoter-driven reporter does not have the capability to resolve those events. One prominent example is calcitonin and CRGP, both are encoded by the same gene, but as a result of alternative splicing (AS), the gene products are completely different[27]: calcitonin is regulating calcium homeostasis[28,29] and CRGP is triggering vasodilatation in the nervous system[30]. Therefore, transgenic promoter-driven gene reporters should always be used with caution.

Fusion proteins

Reporter proteins, such as fluorescent proteins, can also be fused to a protein of interest (POI) (**Figure 2**). Fusion can take place either at the N- or C-terminus. The advantage of the system is that the localization of the POI can be tracked. The disadvantage, however, is that not every protein tolerates fusions, either for steric reasons, or because the N- or C-terminus contains signal sequences for the correct localization of the protein that is destroyed by a fusion. Fusions can be created using CRISPR/Cas9-mediated insertion of a reporter moiety to the N- or C-terminus of the POI. However, no alternative spliced isoforms can be distinguished, as the fluorescent protein would have to be inserted into an exon, whereby the protein isoform is interrupted by the reporter protein and is thus destroyed. Even, if the POI tolerates N- or C-terminal fusions, the 5'- and 3'-proximal untranslated region

might be negatively affected since the 5'-region dictates the translation initiation efficiency[31] and the 3'-region the half-life of the mRNA[32] both resulting in an altered protein level.

Internal ribosome entry sites and 2A ribosomal skipping sequences

Since fusion proteins are too invasive in many cases and subcellular localization is less critical, processes adapted from viruses are used to add a reporter gene to a protein-coding gene on the same mRNA without obtaining a fusion protein. Two elements are commonly used, 2A ribosome skipping peptides and internal ribosome entry sites (IRES) (**Figure 2c,d**). A 2A sequence, also called CHYSEL for cis-acting hydrolase element, which is usually about 20 amino acids long, contains a non-conserved sequence of amino acids with a strong α-helical tendency, followed by the consensus sequence -D(V/I)ExNPG|P, where x is any amino acid[5]. The co-translational separation of the polypeptide occurs between G and P. The 2A sequences of the porcine teschovirus-1 and *Thosea asigna* virus appear to be the most efficient in generating two separate proteins from a continuous coding sequence (CDS), the efficiency being determined by the cleavage/uncleaved rate[33]. This method is not entirely scarless because, depending on the direction of fusion, ~20 amino acids remain at the POI at the C-terminus or a proline remains at the N-terminus. Since the termini also have a tremendous influence on the half-life of a protein (N- and C-degrons)[34], the half-life of the protein can be changed dramatically.

IRES, on the other hand, is not a co-translational separation of polypeptides, but a 5'-cap-independent translation in which the factors for the translation initiation and the ribosome are recruited directly or indirectly to the IRES via its secondary and tertiary structure[6]. Together with a reporter protein inserted into the 3'-UTR after the CDS, the POI CDS is typically translated cap-dependently and the downstream IRES-reporter is translated cap-independently. However, since the 3'-UTR is altered, the half-life of the mRNA is potentially altered[31]; furthermore, there are indications that an IRES has a negative impact on the first cistron of the mRNA, either due to mRNA destabilization or due to the interference with the cap-dependent translation by the secondary structure-rich IRES[35]. Compared to the 2A method, however, it has the advantage that the CDS of the POI is not changed. Also, here, no isoforms generated by AS can be distinguished with 2A or IRES elements.

Reporter-protein fusions to intrabodies

Intrabodies are a class of binding proteins, in most cases, single-chain nanobodies derived from camelid antibodies, which, unlike human antibodies, can be expressed intracellularly**.** By the fusion to fluorescent proteins, the localization of a POI can be indirectly determined (**Figure 2e**)[36]. However, this only works for abundant proteins, such as histones and cytoskeletal elements, and only for very high local concentrations. The POI must have a non-even localization in the cytosol or the nucleus, as else no contrast can be generated. The reason for this is that the fusion constructs between intrabody and fluorescent protein are always fluorescing, regardless of whether the intrabody part is bound to a target or not. Therefore, the detection is not done by measuring the fluorescence but by detecting non-diffuse structures and thereby reaching a threshold-fluorescence, such as cytoskeleton filaments, membrane proteins, DNA interacting proteins. Therefore, the method is not suitable for the quantification of proteins, although it could theoretically distinguish splice isoforms from each other.

Aptamer-binding proteins

Non-coding RNA (ncRNA) cannot be analyzed with any of the above methods, as they are often located in the nucleus where no translation takes place (**Figure 2f**). The methods are also not suitable for cytosolically localized ncRNA since 2A and fusions depend on an existing CDS. An IRES-mediated translation would be possible for cytosolically localized ncRNA. Since ncRNA carries out its function on the RNA level, the insertion of foreign RNA is risky, as this alters the secondary structure of the RNA and may, therefore, no longer be able to carry out its actual function. Moreover, the IRES would recruit translation initiation factors to the RNA, which would change the

actual function of the ncRNA. The only currently available method, which is also highly invasive, is based on RNA aptamer and aptamer-binding proteins, which are fused with fluorescent proteins (**Figure 2g**)[11]. Tandem repeats of aptamer motifs are specifically introduced into an ncRNA, and attention should be paid not to alter physiologically relevant regions. The second component of the system is an aptamer-binding protein (ABP), which additionally must be expressed constitutively. Both components are typically introduced via CRISPR/Cas9, whereas the second component can also be introduced via lentiviral vectors or transposases (PiggyBac or Sleeping Beauty)[37]. As a result, RNA can be indirectly observed, similar to the nanobody-approach mentioned above for POIs.

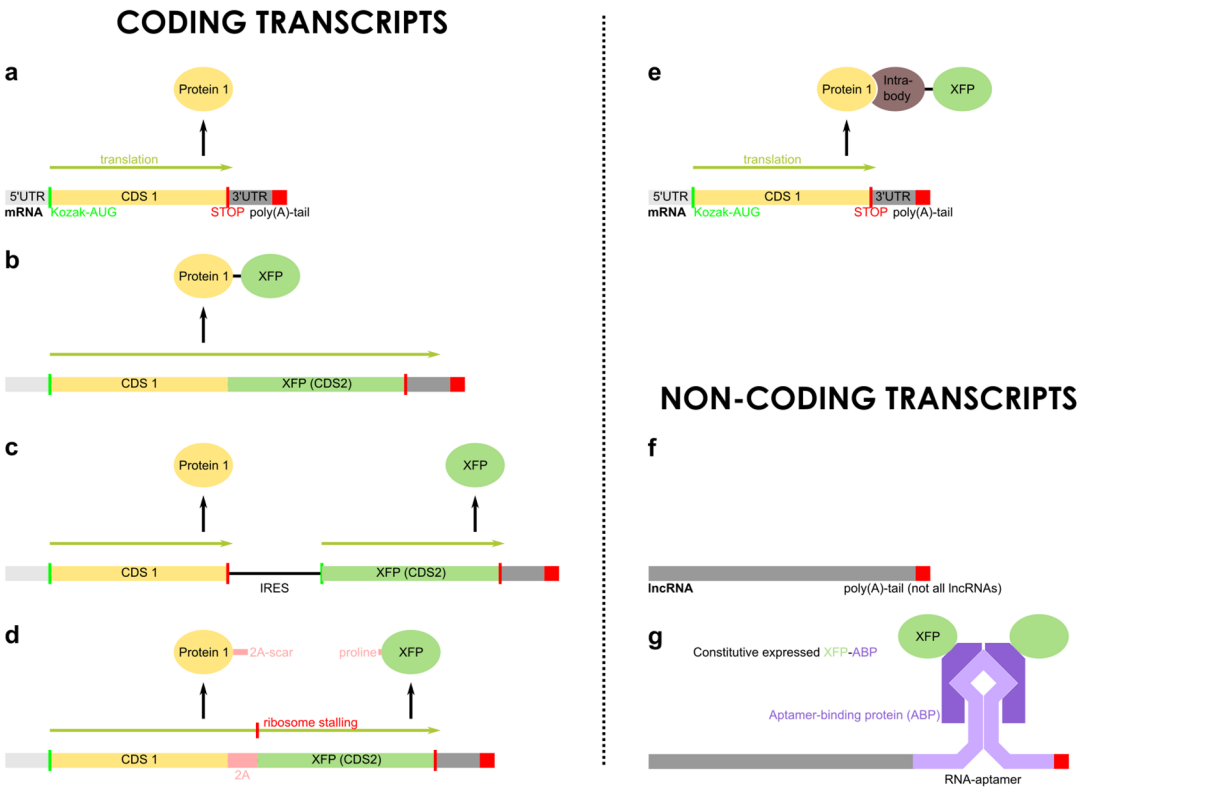

Figure 2 | Current methods to monitor gene expression of coding and non-coding transcripts. a, Protein-coding genes usually are expressed from an RNA polymerase II promoter carrying a 5'-cap (m7G) and are polyadenylated. **b,** Classical N- or C-terminal fusion proteins can be used to determine subcellular localization. **c,** Using a viral internal ribosome entry site (IRES), multi-cistronic mRNAs can be created such that an endogenous gene can be tagged by the insertion of an IRES-reporter downstream of the stop codon of the coding sequence (CDS) in the 3'-UTR. **d,** Also derived from virus elements are 2A peptides that enable the co-translational formation of independent proteins in one translation round via a ribosome skipping mechanism. **e,** Intrabody fusions to fluorescent proteins allow the indirect subcellular tracking of a POI. **f,** The methods from b-c for coding genes are not applicable for non-coding RNAs since many of them are located in the nucleus where translation does not occur. Moreover, these methods are invasive as they heavily modify the RNA sequence and structure. **g,** The only established method to track RNA longitudinally and obtain subcellular resolution are aptamer-based two-component systems, where the first is a multidentate RNA-aptamer motif introduced into the DNA encoding the RNA of interest and a second part is an aptamer-binding-protein to fluorescent protein fusion. The latter is constitutively expressed from a safe-harbor locus (*AAVS1* locus in human cells, *Rosa26* in human and murine systems). This method necessitates modifications of the lncRNA with possibly adverse consequences regarding the stability and lifetime of the sequence.

Development of an exon-specific isoform expression reporter system (EXSISERS)

INTRODUCTION I

Alternative splicing (AS) is a major mechanism to generate genetic diversity within a limited number of genes and occurs in >90% of genes; disruptions of this complex regulatory system (**Figure 3a**) are associated with a variety of neuromuscular and neurodegenerative diseases such as spinal muscular atrophy and Parkinson's disease[7–9]. Existing methods for splicing analysis quantifies mRNAs by endpoint-measurements (RT-qPCR, (sm)FISH[1], RNA-sequencing[2]), proteins by immunochemistry (immunoblot analysis, immunofluorescence staining), or mimic the complex genetic regulations via a simplified mini-gene analysis[38–40]. Nevertheless, studies at the mRNA level can lead to wrong conclusions, since post-transcriptional and co-translational control does not necessarily change the mRNA levels, *e.g.,* in cases of translation-arrested[41], ribosomal-frameshift-regulated[42], or locally translated mRNA[43,44]. Moreover, current analytical methods at the protein-level are limited by the number of available exon-specific antibodies. Most importantly, all of the existing techniques are destructive and thus prevent longitudinal analyses, a common challenge when the sample number cannot be scaled up due to limited resources.

Thus, an exon-specific isoform expression reporter system (EXSISERS) was developed, which non-invasively reports the exon usage at the protein level via a scarless post-translational excision of an exon-resident effector domain (**Figure 3b**). After the inclusion of the exon of interest (EOI) during mRNA biogenesis and its cytosolic trafficking, the reporter or effector moiety is co-translated and rapidly released by intein-mediated[45] protein splicing resulting in an unmodified isoform on protein-level and preserving the original isoform ratios (**Figure 3b**). EXSISERS' modularity could be demonstrated using different self-excising reporter and effector domains that can be easily integrated into the EOI by targeted nuclease-technologies, *e.g.*, CRISPR/Cas9[20], enabling quantification and cellular imaging over time of disease-associated exons (**Figure 3c-e**), as well as enrichment of EOI expressing cells for the rapid identification of splicing modulators and high-throughput analysis of exon-specific perturbations.

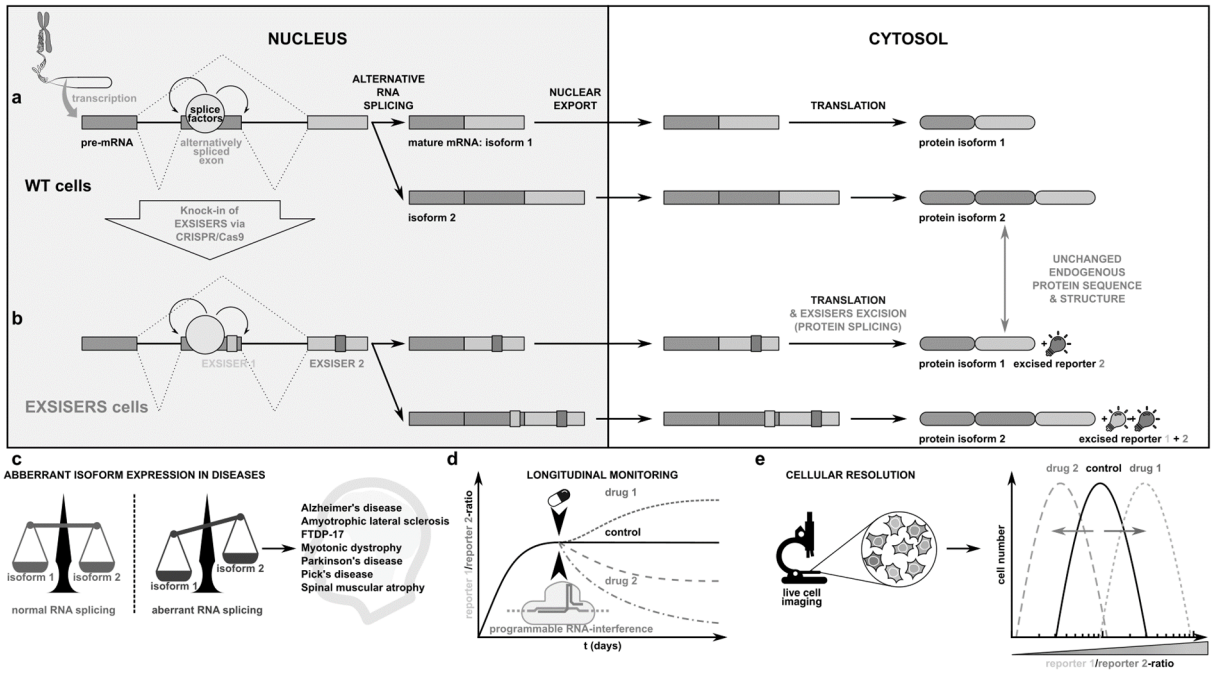

Figure 3 | Principle of the exon-specific isoform expression reporter system (EXSISERS). a, Schematic depiction of biogenesis, maturation, and subcellular trafficking of protein-coding mRNAs in mammalian systems. **b,** Genomic integration of EXSISERS into an exon of interest (EOI) results in a post-translational intein-mediated, scarless excision of an exon-specific reporter domain. **C,** Dysregulation of alternative splicing (AS) results in anomalous isoform ratios involved in several neurodegenerative diseases. EXSISERS facilitates **d,** longitudinal protein-level monitoring of isoforms with **e,** cellular resolution such that the distribution of exon-specific isoform expression can be quantified. Partial results and subfigures have been submitted as a part for a publication with the title "Non-invasive and high-throughput interrogation of exon-specific isoform expression".

RESULTS I

Scarless excision of exon inclusion reporters

Inteins are proteinogenic domains which are found as parasitic elements in all kingdoms of life. They are typically inserted within genes encoding for proteins without disrupting the native expression product (extein) on the protein level, due to their post-translational scarless excision[46]. So far, no systematic attempts have been made to exploit this mechanism for the analysis of proteinogenic isoform expression.

Ratiometric monitoring of exon-specific isoform expression of tau protein

To show that non-invasive monitoring of an exon of interest (EOI) within the natural genomic context is possible without changing the natural gene expression products, cell lines harboring the dual-luciferase EXSISERS for ratiometric monitoring of exon 10 of the gene encoding microtubule-associated protein tau (*MAPT)* were generated (**Figure 4a**).

The microtubule-associated protein tau is mainly expressed in cells of neuronal lineage and mediates microtubule polymerization and hence its stabilization[47]. Alternative splicing (AS) of *MAPT* exon 10 can be misregulated, causing an imbalance of the isoforms contributing to numerous neurodegenerative diseases[48,49]. In the mature human central nervous system, tau is expressed in six "adult" isoforms produced by alternative splicing of the *MAPT* exons 2, 3, and 10 (**Figure 4a**). Depending on the presence of exon 10, tau isoforms contain three (3R-tau) or four (4R-tau) tandem repeats of a microtubule-binding motif[50].

NanoLuc luciferase (NLuc)[51] was inserted into exon 10 flanked by the recently discovered ultra-fast splicing split-intein pair gp41-1[45], and homozygosity was validated using PCR (data not shown). Crucially, to further enhance intein splicing of the exon usage reporters, synthetic anti-parallel coiled-coil (CC) domains were added between the inteins, which actively improve the co-folding for the split-intein binary complex[52] (**Figure 4b** and **Figure 5**).

Ratiometric readout of exon 10 containing isoforms referenced against total tau (pan-tau) was realized by homozygously inserting firefly luciferase (FLuc) flanked by a second bioorthogonal set of CCs and fast splicing inteins (NrdJ-1[45]) into the non-alternatively spliced exon 11 (**Figure 4a,b**). Consequently, the NLuc signal represents the usage of exon 10-specific 4R-isoforms, while FLuc, which can be detected via orthogonal luciferase substrates, reports the cumulative expression of all tau isoforms (**Figure 4a,b**).

Stable *MAPT* 4R reporter cell lines carrying the auto-excising dual-luciferase reporters (EXSISERS$_{MAPT:10NLuc/11FLuc}$) were generated homozygously via CRISPR/Cas9 for wild type as well as for the well-known disease-associated IVS10+16 c>t mutation[53–55]. *MAPT* expression was enforced using CRISPR/dCas9-VPR transactivators[56] targeted upstream of its transcription start site since *MAPT* expression is only weak in HEK293T cells. EXSISERS$_{MAPT:10NLuc/11FLuc}$ HEK293T cells (**Figure 4c**) showed the typical cytosolic tau-staining in after CRISPR/dCas9-VPR-mediated induction. Also, immunoblot analysis (**Figure 4d**, first two lanes) confirmed that the splice pattern was not changed by EXSISERS' insertion compared to unmodified HEK293T cells not harboring the EXSISERS constructs (first lane), indicating effective post-translational excision of the reporters. Additional introduction of the disease-associated IVS10+16 c>t hairpin-destabilizing transition mutation led to a visible increase of the 4R isoform (**Figure 4d**, last lane).

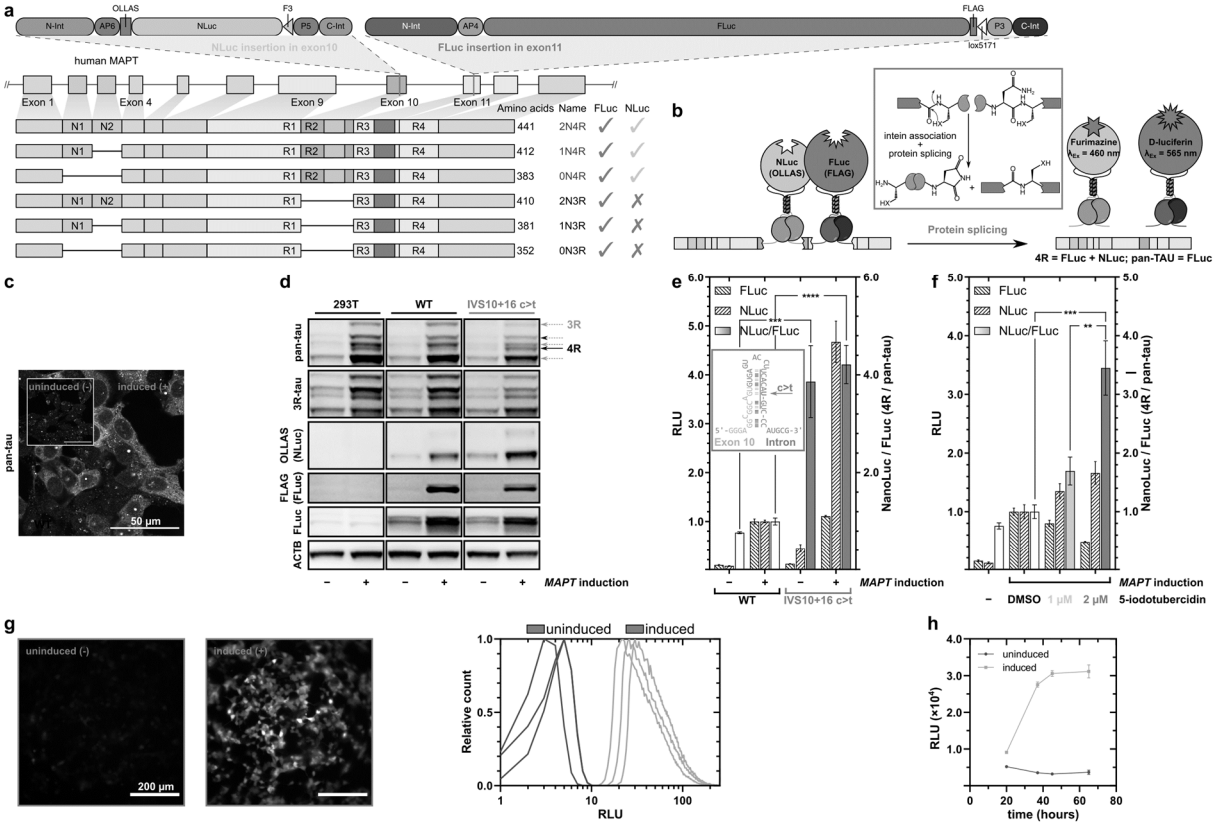

Figure 4 | Non-invasive ratiometric monitoring of exon 10-specific isoforms of human *MAPT*. a, Structure of the genomic organization of human *MAPT* locus with NLuc inserted into the alternatively spliced exon 10 and FLuc into the constitutive exon 11. Both luciferases are flanked with coiled-coils and inteins, resulting in a ratiometric reporter system that detects the fractional expression of 4R referenced against pan-tau. **b,** Schematic representation of post-translational excision of the reporter modalities via intein splicing in 4R-tau. NLuc specifically indicates the presence of 4R-tau and FLuc the presence of pan-tau (total amount of tau). The intein-flanked luciferases are excised scarlessly from the initially tagged translation product and can be read out independently via orthogonal substrates. **c,** Anti-tau immunostaining of the dual-luciferase EXSISERS cell line showed the typical tau pattern of tau 72 h post-transfection with CRISPR/dCas9-VPR transactivators targeting *MAPT* promoter. **d,** Immunoblot analysis of dephosphorylated lysate from unmodified HEK293T cells and EXSISERS cell lines harboring WT *MAPT* or the disease-associated IVS10+16 c>t mutation, confirmed that both FLuc and NLuc were clearly increased after *MAPT*-induction indicating tau – luciferase coupling; anti-pan-tau revealed the presence of all tau isoforms in EXSISERS$_{MAPT}$ cells that are also present in wild type cells. The 4R-tau band was identified by its negative stain for 3R-tau. **e,** Relative luminescence units (RLU) of FLuc and NLuc luciferase (striped bars, left y-axis) and the ratio (NLuc/FLuc-ratio, filled bars, right y-axis) of the denoted exon 10 and 11 EXSISERS-tagged HEK293T lines with wild type and the pathological IVS10+16 c>t mutation. **f,** RLU of FLuc and NLuc, and their ratio as a function of the DYRK1A/GSK3A inhibitor 5-iodotubercidin. **g,** Bioluminescence imaging enables longitudinal monitoring of 4R-tau with cellular resolution in the *MAPT* exon 10-tagged HEK293T line; images were taken 72 hours post-induction. **h,** 4R-tau expression was tracked longitudinally via Nluc and its substrate furimazine over a time course of 63 hours post-induction in the double-tagged HEK293T cell line. Error bars represent standard deviation (n = 3). Only selected results of ANOVA post-hoc tests are shown with *,**, ***, **** denoting p-values smaller than 0.05, 0.01, 0.001 and 0.0001 (Bonferroni correction). Results shown were performed together with Teeradon Phlairaharn as a part of his research internship. Partial results and subfigures have been submitted as a part for a publication with the title “Non-invasive and high-throughput interrogation of exon-specific isoform expression”.

Furthermore, the amount of excised NLuc immunoblot (OLLAS tagged) was markedly increased in the IVS10+16 c>t reporter line in both conditions, before and after induction, corresponding to a relative increase of the 4R-tau isoform. This change in isoform-ratio could also be conveniently read out via a dual-luciferase assay that revealed an approximately four-fold increase of the NLuc/FLuc ratio (p<0.001 for uninduced and p<0.0001 for induced, post-hoc tests of two-way ANOVA), when comparing the c>t hairpin modification to WT, in line with the literature (2–6-fold increase)[57,58] (**Figure 4e**).

Next, the effects of small molecules on exon 10 inclusion were examined and whether the changes can be reliably followed by ratiometric EXSISERS$_{MAPT:10NLuc/11FLuc}$. Hence, the inhibitor 5-iodotubercidin[59–61] (ITU) was chosen, which inhibits the kinases DYRK1A and GSK3A, leading to a more active SC35 (SRSF2) that in turn binds to exon 10 of *MAPT* pre-mRNA[62–64] promoting its inclusion. After the application of ITU, a concentration-dependent increase of the NLuc/FLuc-ratio to up to 3-fold (p<0.001, post-hoc tests of one-way ANOVA) could be observed (**Figure 4f**), and dose-dependent expression of 4R-tau was also confirmed by immunoblot in wild type HEK293T cells (**Figure 6**). Also, 4R-tau expression could be measured non-invasively via bioluminescence microscopy, and thus, the NLuc signal from single cells could be registered to determine the distribution of *MAPT* exon 10 usages in the cell population (**Figure 4g**). Furthermore, longitudinal monitoring of the NLuc signal showed that maximal *MAPT* exon 10 expression was reached 50 hours post-induction (**Figure 4h**).

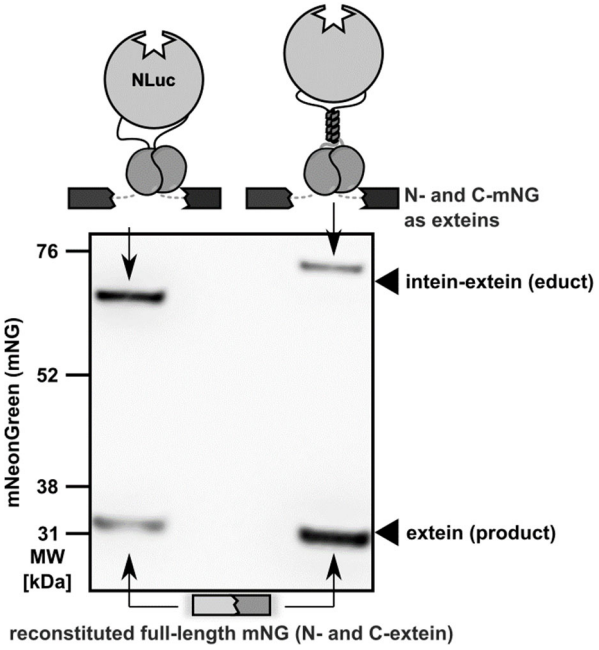

Figure 5 | Enhancement of the cis-splicing split-inteins by synthetic coiled-coils (CCs) for rapid excision of the reporter module. The schematic depiction in the top panel shows the construct containing split-inteins without (left) and with (right) the introduction of the CCs (yellow) flanked by N- and C-mNeonGreen[65] (mNG) which models exteins with fast folding rates of less than 10 minutes[65]. The immunoblot analysis (anti-mNG) analysis shows a substantial increase of the extein product for the variant containing CCs 24 hours post-transfection (experiment performed together with Enikő Baligács as a part of her bachelor's thesis). Partial results and subfigures have been submitted as a part for a publication with the title "Non-invasive and high-throughput interrogation of exon-specific isoform expression".

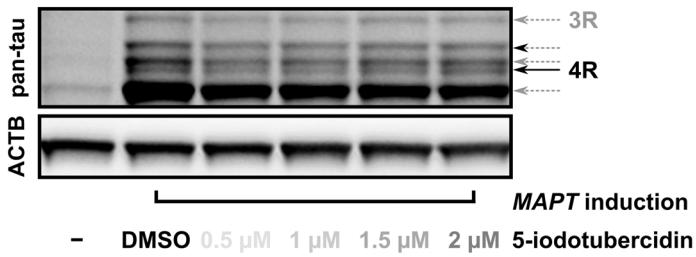

Figure 6 | Effects of 5-iodotubercidin on tau RNA-splicing in unmodified wild-type HEK293T cells. Immunoblot analysis shows a concentration-dependent inclusion of 4R-tau as a function of the kinase inhibitor 5-iodotubercidin directed against the kinases DYRK1A or GSK3A. Results shown were performed together with Teeradon Phlairaharn as a part of his research internship. Partial results and subfigures have been submitted as a part for a publication with the title "Non-invasive and high-throughput interrogation of exon-specific isoform expression".

Optimization of RNA-targeting systems for the isoform-specific knock-down of MAPT

After confirming that the ratiometric EXSISERS$_{MAPT:10\text{-}NLuc/11\text{-}FLuc}$ cell line worked as expected, the next goal was to use EXSISERS to systematically optimize programmable CRISPR/Cas13 effectors for tau perturbation with respect to Cas13 type and subcellular localization, spacer length and targeting site, and its impact on different splicing isoforms in direct comparison to the latest generation of short hairpin RNAs (shRNAs). These effectors influence isoform-specific expression at different subcellular sites of action during the mRNA maturation process (**Figure 7**).

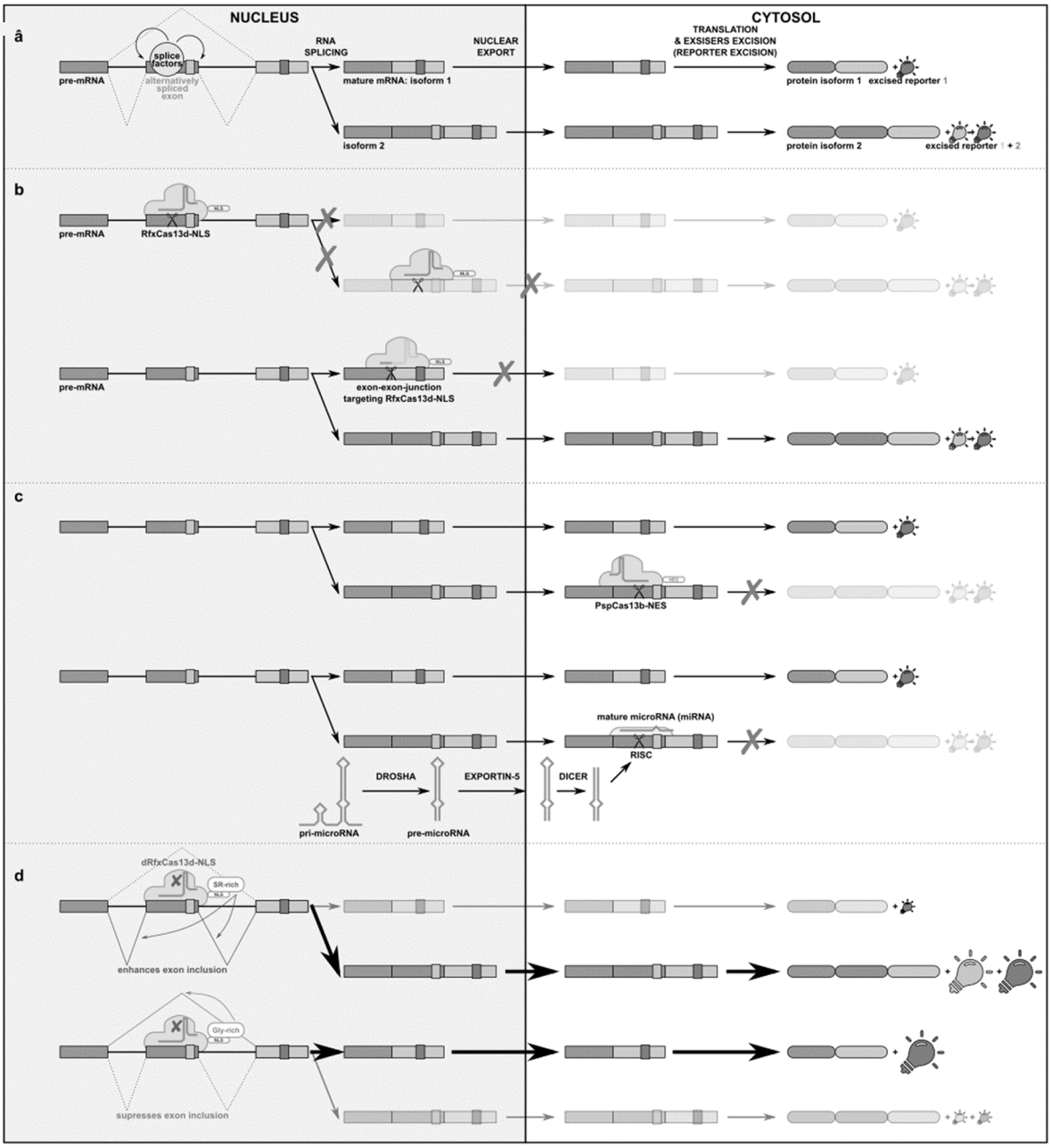

Figure 7 | Schematic of the subcellular sites of action for the different programmable RNA-targeting systems. a, Alternative splicing (AS) generates different isoforms from a single pre-mRNA, regulated by splice enhancers and suppressors. **b,** RNA-guided perturbation using RfxCas13d-NLS in the nucleus is expected to be target any isoform but can be made isoform-specific by targeting the isoform-specific exon-exon-junctions. **c**, In contrast, cytosolic PspCas13b-NES acts in the cytosolic environment and can thereby target isoforms without the requirement for targeting the exon-spanning junctions. Artificially reprogrammed microRNAs also act in the cytosolic environment. In comparison to PspCas13b-NES, however, a few restrictions on the choice of the target RNA sequence may apply. **d,** By using the HEPN nuclease domain inactivated variant of dRfxCas13d-NLS, splice enhancers/suppressors can be recruited to an exon of interest to alter exon in- or exclusion resulting in an isoform shift. Partial results and subfigures have been submitted as a part for a publication with the title “Non-invasive and high-throughput interrogation of exon-specific isoform expression”.

Cas13d localization and spacer length

First of all, EXSISERS$_{MAPT:10\text{-}NLuc/11\text{-}FLuc}$ cells were used to verify recent reports, according to which RfxCas13d[66] (from *Ruminococcus flavefaciens XPD3002*) with a 22 nt spacer displayed a more potent RNA perturbation in the nucleus than in the cytosol[66] when 4R-tau is targeted. A small decrease in the 4R/pan-tau ratio was observed when RfxCas13d was equipped with a nuclear export signal (NES) to localize it in the cytosol in comparison to when it was targeted to the nucleus via a nuclear localization signal (NLS) (**Figure 8a**).

Still, the overall knockdown (KD) was unexpectedly weak in the experimental conditions adjusted for low levels of plasmids used for transfection. Thus, it was investigated whether the activity could be improved by extending the length of the spacer in mammalian cells. Intriguingly, RfxCas13d suppressed 4R-tau (NLuc) more effectively with a prolonged 30 nt spacer compared to the published 22 nt spacer[66] ($p<0.0001$ for the effect of spacer length, post-hoc tests of two-way ANOVA, **Figure 8a,b**). When the extended 30 nt spacer was combined with a nuclear-localized RfxCas13d-NLS, a robust reduction in 4R-tau (NLuc) was observed ($p<0.0001$, post-hoc tests of two-way ANOVA). However, a substantial reduction of pan-tau (FLuc) was also detected ($p<0.0001$, post-hoc tests of one-way ANOVA), indicating RfxCas13d-NLS activity on the *MAPT* pre-mRNA (**Figure 7b**).

Targeting exon-spanning regions with nuclear-localized Cas13d

One approach to make RfxCas13d more specific towards different isoforms is to target the mature spliced mRNA (**Figure 7b**) by targeting the exon-exon junctions that are not present in the pre-mRNA. Compared to a crRNA against exon 10, targeting of the exon 9-10 junction with RfxCas13d-NLS showed a decreased activity in NLuc ($p<0.01$, post-hoc tests of one-way ANOVA) but also in FLuc ($p<0.0001$) resulting in an unchanged NLuc/FLuc-ratio ($p>0.05$) **(Figure 8c)**. Conversely, targeting the 10-11 junction showed a better isoform specificity with NLuc unchanged ($p>0.05$) but FLuc significantly less depleted (**Figure 8c,** $p<0.0001$).

As expected, RfxCas13d targeting of the constitutive exon 9 and the alternatively spliced exon-junction 1-4 junction showed the lowest isoform-specificity regarding 4R-tau of all tested RfxCas13d crRNAs (**Figure 8c,d**) because exon 9 is not alternatively spliced (constitutive exon) and the inclusion/exclusion of exons 2 and 3 are exon 10 independent. On the contrary, when 3R-tau was targeted via the exon 9-11 junction, the NLuc (4R) signal was unchanged ($p>0.05$) while the FLuc (pan) signal was markedly reduced ($p<0.0001$), causing an increase of the NLuc/FLuc-ratio ($p<0.0001$) suggesting a 3R-specific reduction. An alternative RfxCas13d crRNA targeting the exon 9-10 (^{2}9-10) junction with an asymmetrical design (24 nt on exon 9, 6 nt on exon 10, **Figure 8d**) showed an elevated isoform-promiscuity (NLuc/FLuc increased compared to ^{1}9-10, $p<0.05$). This result was not unexpected since the spacer also matches the 9-11 junction with only a single nucleotide terminal mismatch (**Figure 9**), which has been reported to be well tolerated by all known Cas13 systems[67,68].

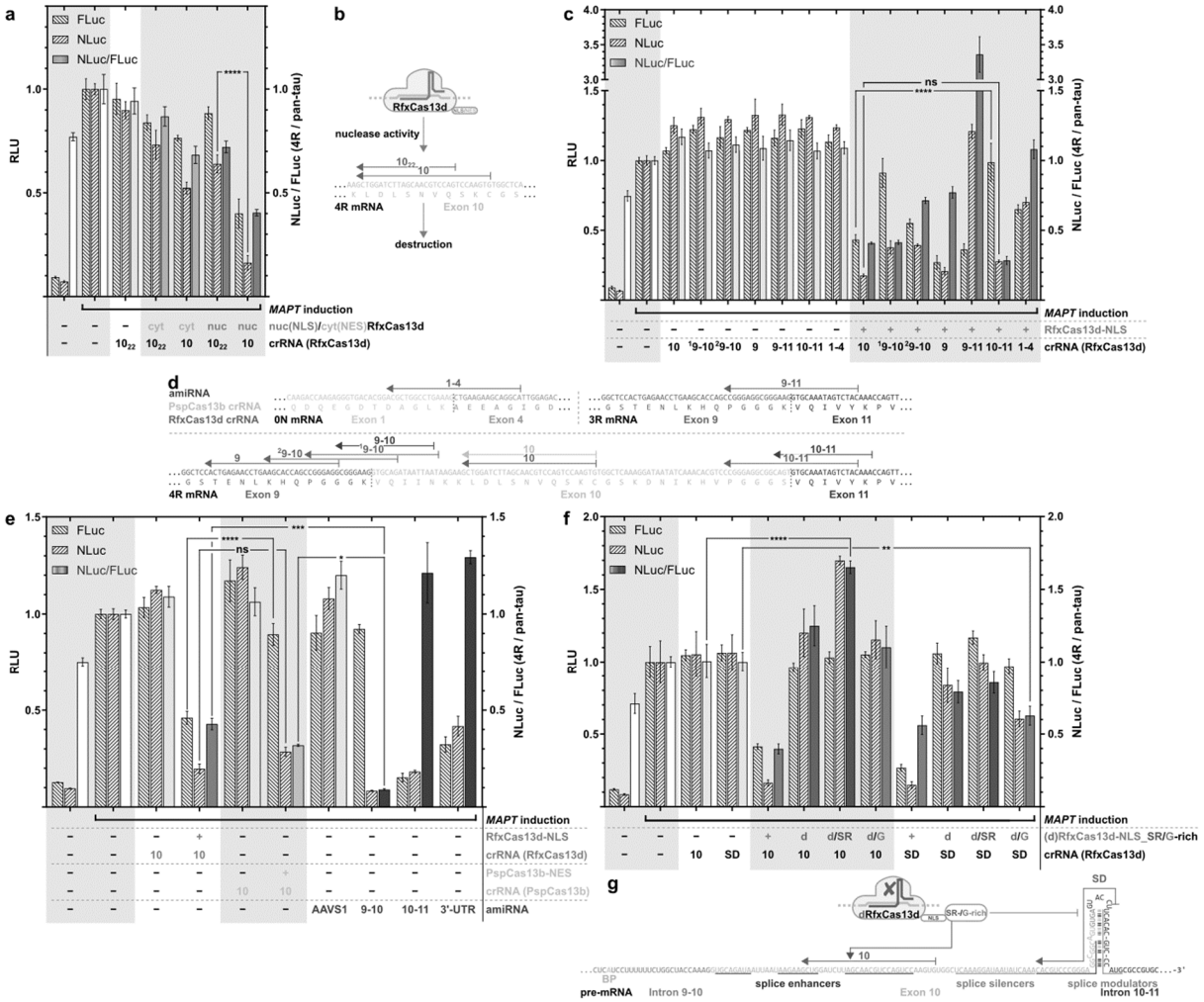

Figure 8 | Optimization of programmable RNA-targeting systems for isoform-specific perturbation of *MAPT*. **a,** Optimization of spacer length for RfxCas13d assessed via depletion of induced tau, tracked via bioluminescence signal from NLuc (4R-tau) and FLuc (pan-tau) 72 h after transfection. RLU of the bioluminescence signals are given for FLuc and NLuc separately and as a ratio (NLuc/FLuc). nuc: nuclear localization by a C-terminal NLS; cyt: cytoplasmic localization by a C-terminal NES. **b,** Binding sites of RfxCas13d-NLS/NES with a canonical (22 nt) and an extended spacer targeting *MAPT* exon 10 (30 nt, not indicated as subscript because all subsequent experiments uses 30 nt spacers). **c,** Directing RfxCas13d to exon-exon junctions to increase isoform specificity of RfxCas13d-NLS measured via NLuc and FLuc. **d,** Depiction of binding sites of RfxCas13d, PspCas13b, and amiRNA on the mature *MAPT* 0N/3R/4R mRNA. **e,** Specificity of cytosolic targeting of mature mRNA by Cas13b compared to amiRNA on different tau isoforms measured via dual luciferase assay. **f,** Isoform specificity of RNA-guided splice modulation via dRfxCas13d-NLS fusions to splice suppressor/enhancer domains; d: dRfxCas13d-NLS; d/SR: dRfxCas13d-NLS_SR-rich-domain; d/G: dRfxCas13d-NLS_G-rich-domain. **g,** Binding sites of RfxCas13d and its chimeras on *MAPT* pre-mRNA; dRfxCas13d nuclease-inactivated version of RfxCas13d; SR-rich: Serine/arginine-rich domain of SC35; G-rich: Glycine-rich domain from HNRNPA1 isoform A1-B**.** Error bars represent the standard deviation for all graphs (n = 3). Only the most relevant results of ANOVA post-hoc tests are shown, where *,**, *** denote p-values smaller than 0.05, 0.01, and 0.001 (Bonferroni correction). Results shown in a, c, e, and f were performed together with Teeradon Phlairaharn as a part of his research internship. Partial results and subfigures have been submitted as a part for a publication with the title “Non-invasive and high-throughput interrogation of exon-specific isoform expression”.

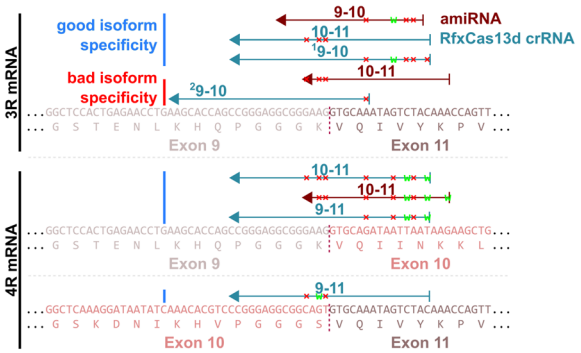

Figure 9 | Depiction of potential off-target binding sites of RfxCas13d/PspCas13b crRNAs. The number over the binding sites shows the intended *MAPT* targeting region (*e.g.,* 10-11 is intended to target the exon-junction 10-11). 'x' indicates mismatched nucleotides on the isoform off-target sites, and 'w' denotes wobble base pairs (G-U pairings). amiRNA directed against 10-11 has 3 terminal mismatches, which are well tolerated in shRNAs, resulting in an isoform-unspecific depletion of both 4R and 3R tau isoforms. ^{2}9-10 crRNA (RfxCas13d-NLS) is also matching on the 9-11 junction with a near-perfect matching resulting in an unspecific depletion of both 4R and 3R tau. Partial results and subfigures have been submitted as a part for a publication with the title "Non-invasive and high-throughput interrogation of exon-specific isoform expression".

Targeted mRNA knockdown in the cytosol

Another approach to enhance isoform-specificity is to target mature *MAPT* mRNA in the cytosol, where unspliced nuclear-localized pre-mRNA cannot be unintentionally altered (**Figure 7c)**. Since RfxCas13d showed its superior performance in the nuclear compartment (**Figure 8a**), PspCas13b was chosen instead (from *Prevotella sp. P5–125*), which has been previously shown to mediate robust RNA-targeting activity in the cytosol[66].

Cytosolic PspCas13b-NES directed against the same region of exon 10 (**Figure 8e, orange bar**) resulted in less efficient depletion of total tau (FLuc) as compared with the corresponding RfxCas13d-NLS (p<0.0001, post-hoc tests of one-way ANOVA) but with a comparable NLuc signal (p>0.05) (**Figure 8e, blue bar**).

The performance of RfxCas13d-NLS was then compared with established short hairpin RNAs (shRNAs), which also act preferentially in the cytosol (**Figure 7c)**. The latest generation of shRNA scaffolds mimics the native microRNA biogenesis pathway. Combined with the latest design rules, increased overall activity and fidelity were achieved[69,70]. A strong isoform-specific KD was observed using the latest generation of artificial microRNAs (amiRNA) targeting the exon 9-10 junction leading to the greater isoform-specificity compared to the Cas13b/d systems (p<0.05 vs. Cas13b-NES and p<0.001 vs. Cas13d-NLS, **Figure 8e**). Using an asymmetrical 10-11 junction (similar to the exon ^{2}9-10 asymmetrical targeting RfxCas13d-NLS (**Figure 8c,d and Figure 9**)) and a 3'-UTR targeting amiRNA, the KD was isoform-unspecific as expected (**Figure 8e**). In contrast, the control AAVS1-targeting amiRNA did not result in any change in FLuc and NLuc depletion (**Figure 8e**).

Repurposing dRfxCas13d-NLS as an RNA-guided splicing enhancer or suppressor

Additionally to the nuclease-active variants, different nuclease-**d**ead versions of RfxCas13d-NLS (dRfxCas13d-NLS) were created to harness RfxCas13d as a programmable splice enhancer or suppressor[66] (**Figure 7d**). The fusion of dRfxCas13d-NLS to the glycine-rich (G-rich) domain of hnRNP A1 (aa 187-320 from the major isoform A1)[66,71] created an RNA-guided splice-suppressor (**Figure 8f,g**). Vice versa, the fusion to the serine/arginine-rich (SR-rich) domain of SC35 (alias SRSF2, aa 90-271)[71,72] resulted in an RNA-guided splice enhancer. Two different approaches were tested, the first one is targeting the exonic splice-enhancers in exon 10 directly and the second one the 3'-exon-intron-junction including the splice donor and the critical regulatory hairpin (**Figure 8g**). A combination of the exon 10 targeting crRNA with both dRfxCas13d-NLS chimeras showed that only the fusion to the SR-rich domain yielded a significant rise in exon 10 inclusion (p<0.0001, post-hoc tests of one-way ANOVA, **Figure 8f** green bar middle panel). In contrast, combinations with the SD/hairpin binding crRNA showed that only

the fusion to the G-rich domain resulted in a significant decline in exon 10 inclusion compared to the crRNA only control ($p<0.01$, **Figure 8f**, brown bar right panel).

Modular reporters and effectors

Besides the high-sensitivity bioluminescence detection mode, fluorescent imaging of exon 10 usage was also facilitated by replacing the luciferase with a membrane-presented HaloTag[73] (EXSISERS$_{MAPT:10Halo}$) that allow cell-enrichment via fluorescence-activated cell sorting (FACS) or capturing the cell surface handle (**Figure 10a-f**). As an alternative membrane functionalization, pseudo-tetrameric single-chain avidin (scAvidin) was inserted instead, which shows an exceptional high affinity towards biotin-functionalized ligands with each pseudo-domain with picomolar affinity; this can be harnessed for magnetic cell separation system (MACS) enrichments or also labeling with fluorescent dyes and subsequent enrichment by FACS (**Figure 10g,h**).

For non-consumptive, supernatant-based, and time-resolved monitoring of exon usage, the TMDs were equipped with NLuc. Furin endoprotease cleavage sites lead to the reporter's shedding into the extracellular space for convenient longitudinal sampling from the supernatant (**Figure 10i**). To check whether the membrane-shedding mechanism of NLuc worked as expected, NLuc-TMD-intein, with or without furin cleavage sites, was inserted into mNG as surrogate exteins and transfected into HEK293T cells. Supernatant samples showed a significant increase in the NLuc signal over time when the furin cleavage sites were present ($p<0.001$, post-hoc tests of 2-way ANOVA with repeated measures at 24 hours, **Figure 10j**).

Established fluorescent proteins can also serve as a reporter for EXSISERS. As an example, mNG flanked by inteins was knocked-in into the mouse *Tubb3*'s second coding exon resulting in visible fluorescence without the disruption of the typical Tubb3 filaments. Its non-invasiveness of the scarless protein splicing was additionally shown by immunoblot analysis (**Figure 10k-m**). Summarized, this set of effectors enables a variety of usages, including convenient sampling from the supernatant and non-invasive live-staining, and high-throughput-screening applications.

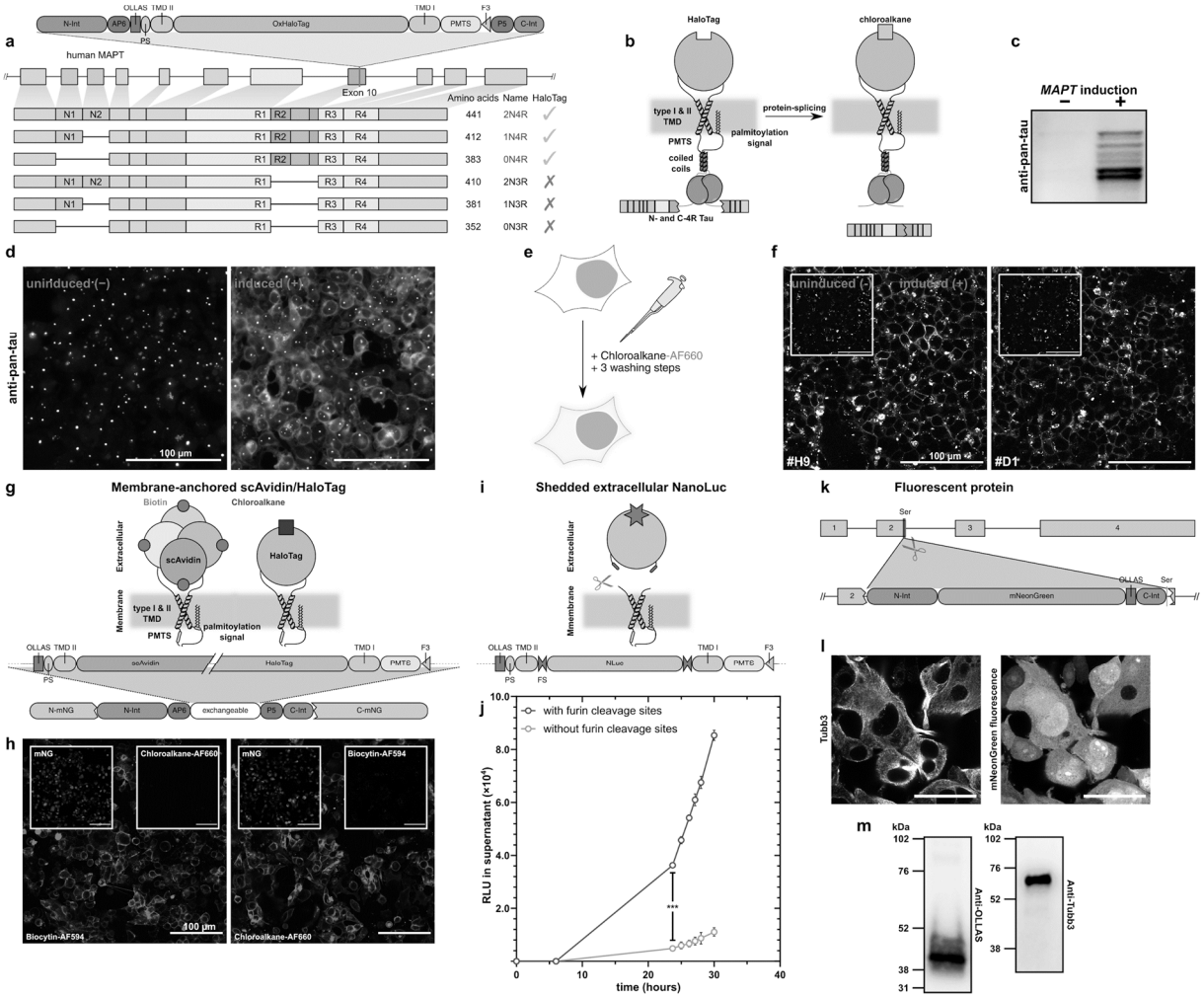

Figure 10 | EXSISERS can be equipped with a variety of effectors. a,b HaloTag was flanked by type I and type II transmembrane domains (TMDs) and the corresponding coiled-coil split-intein moieties, resulting in a HaloTag-functionalization of the cell membrane when *MAPT* exon 10 is expressed. **c,** Immunoblot analysis of pan-TAU showed the classic TAU ladder demonstrating the traceless nature of exon tagging after TAU-induction. **d,** TAU was again induced by CRISPR/dCas9-NLS-VPR, resulting in a visible filamentous TAU staining (epifluorescence microscopy). **e,** Cells with and without induction were incubated with chloroalkane-functionalized Alexa Fluor 660 (AF660) for 10 min at 37 °C and washed twice with HBSS before live-imaging. **f,** *MAPT*-induction (or without induction shown in the inset) for two different clones resulted in a distinct membrane stain with chloralkane-functionalized AF660. **g,** Alternatively, scAvidin was used, a tetravalent, high-affinity binding domain for biotinylated compounds. Constructs for both, membrane-bound scAvidin and HaloTag, and mNG as surrogate exteins were tested which are expected to be spatially separated after the post-translational excision of the intein-domains. **h,** Membrane localization of scAvidin or HaloTag in transfected HEK293T was verified by biocytin-AF594 and chloroalkane-AF660, while mNG-NLS was expectedly localized in the nucleus. **i,** Another variant for non-consumptive quantification was created by replacing the binding moiety of the dual-TMD with furin sites (FS) flanking NLuc. After the transition to the Golgi-network, NLuc is released by the Golgi-resident furin protease into the supernatant, enabling a non-classical signal-peptide-independent secretion. Constructs were transfected in HEK293T and compared to a control construct without furin-sites. **j,** NLuc activity were detected over time in the cells' supernatant expressing the cleavable NLuc without and with furin sites. **k,** mNeonGreen (mNG) flanked by an N- and C-intein was knocked-in into the second coding exon of the *Tubb3* gene in murine N2a cells. **l,** Cells carrying the insertion of the intein-mNG reporter into *Tubb3* showed typical Tubb3 filaments (left) indicating intact Tubb3. The uniform fluorescent signal of mNG (right) in the whole cell indicated effective post-translational excision of the intein-flanked mNG (scale bar: 50 µm) **m,** Immunoblot analysis against Tubb3 and the OLLAS-tagged inteins confirmed successful excision of intein-mNG and ligation of the remaining exteins. Error bars represent standard deviation (n = 3). *** denotes $p<0.001$ of post-hoc tests of two-way ANOVA with repeated measures (Bonferroni correction). Results shown in c, d, and h were performed together with Deniz Tümen as a part of his master's thesis. Results shown in f were performed together with Eva Lederer as a part of her research internship. Results shown in l and m were performed together with Eva Magdalena Beck as a part of her bachelor's thesis. Partial results and subfigures have been submitted as a part for a publication with the title "Non-invasive and high-throughput interrogation of exon-specific isoform expression".

Identification of regulators for isoform-specific expression

To demonstrate the modularity of EXSISERS, a modification was made that facilitates library-based discovery of regulators of alternative splicing by coupling cellular viability to exon inclusion. The Blasticidin-S deaminase[74] (BSD), an enzyme mediating resistance towards Blasticidin-S, flanked by CC-enhanced inteins, was inserted into the exon 18b of the forkhead family transcription factor (*FOXP1*) which encodes the domain for the DNA-binding domain of an embryonic stem cell-specific FOXP1 isoform (**Figure 11a**). In differentiated and somatic cells, exon 18b inclusion is repressed by the splice regulators muscleblind-like protein 1 and 2 (MBNL1/2)[75,76] (**Figure 11a**). At the titrated minimal lethal concentration of blasticidin-S (~5 µg/ml), the homozygously exon 18b-tagged cells did not display altered sensitivity compared to the parental HEK293T cells (data not shown). CRISPR/Cas9-induced *MBNL1/2* knockout (KO) was compared to a control 'KO' of the safe-harbor *AAVS1* locus encoding for a dispensable phosphatase (*PPP1R12C*) in the exon 18b-tagged cell line as a proof-of-concept for exon-inclusion-mediated survival. Treatment with blasticidin-S resulted in the formation of colonies only in the *MBNL1* and *MBNL2* KO condition (**Figure 11b**), indicating a functional coupling of the presence of MBNL1/2 and cell survival dependent on *FOXP1* exon 18b inclusion. Blasticidin-S dose-dependent enrichment of cells with exon 18b inclusion was validated in parallel by semi-quantitative RT-PCR (**Figure 11c**) and by immunoblot detection of OLLAS-tagged BSD after KO of MBNL1/2 (**Figure 11d**). The integrity of FOXP1 in the genetically exon 18b-tagged reporter line was authenticated for all conditions by immunoblot analysis, also proving the scarless removal of BSD from the FOXP1 (**Figure 11d**). Sanger sequencing of the *MBNL1/2* loci in the blasticidin-S resistant cell populations revealed an overall dose-dependent accumulation of *MBNL1/2* insertions and deletions (InDels). Analysis of the sequencing results by sequence decomposition[77] revealed a remaining 10.2% of residual *MBNL1* WT-allele in the most stringent condition (12.5 µg/ml blasticidin-S) and 57.6% of *MBNL1* WT-allele in the minimal selection condition. Meanwhile, the *MBNL2* InDels showed no dose-dependency (**Figure 11e**). The correlation between *MBNL1* KO and *FOXP1* exon 18b inclusion mediated by EXSISERS$_{FOXP1:18b\text{-}BSD}$ successfully confirmed MBNL1 as the dominant repressor of exon 18b inclusion[76].

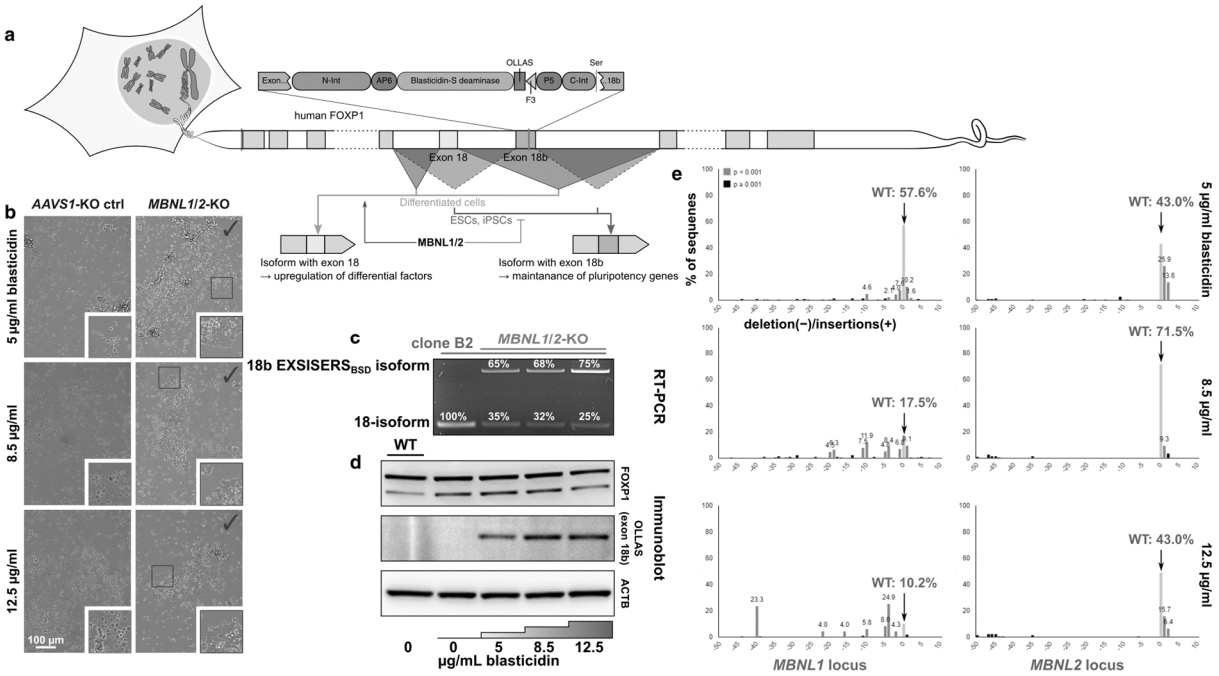

Figure 11 | Traceless exon-inclusion-dependent selection marker for the discovery of regulators of exon in-/exclusion. a, Split-intein-CC-flanked blasticidin-S deaminase (BSD) was inserted into *FOXP1* exon 18b in HEK293T cells assisted by CRISPR/Cas9. **b,** Tagged cells (EXSISERS$_{FOXP1:18b\text{-}BSD}$) were treated with an increasing concentration of blasticidin-S (top to bottom) after CRISPR/Cas9-mediated KO of *MBNL1* and *MBNL2* which are solely expressed in differentiated cells and is reported to mediate 18b inclusion. The safe-harbor *AAVS1* locus was targeted as control (AAVS1-KO ctrl). **c**, RT-PCR of cells after MBNL1/2-KO treated with increasing blasticidin-S concentration revealed a dose-dependent enrichment of exon 18b. **d,** While FOXP1 were detected in all conditions, excised OLLAS-tagged BSD was only detected in *MBNL1/2*-KO cells with blasticidin selection. **e**, PCR analysis via sequence decomposition of the *MBNL1/2* locus from cells selected with blasticidin at three different concentrations (top row to bottom row). The frequency of insertions and deletions were approximated for the *MBNL1* (left column) or *MBNL2* locus (right column) from Sanger sequencing chromatograms. Cells resistant to an increasing concentration of blasticidin-S showed an increased appearance of mutated *MBNL1* validating the prominent role of MBNL1 in repressing *FOXP1* exon 18b inclusion. Results shown in b were performed together with Deniz Tümen as a part of his master's thesis. Results shown in c, d, and e were performed together with Eva Lederer as a part of her research internship. Partial results and subfigures have been submitted as a part for a publication with the title "Non-invasive and high-throughput interrogation of exon-specific isoform expression".

DISCUSSION & OUTLOOK I

In this part of the thesis, a flexible exon-specific isoform expression reporter system (EXSISERS) was established that converts a specific exon expression event into reporter signals or genetically controlled handles for cell enrichment and selection. These features are facilitated by the combination of fast-splicing inteins with coiled-coil domains to guarantee immediate removal of effector proteins and traceless formation of the unchanged, endogenous protein product. EXSISERS was utilized in this work to improve RNA-targeting strategies for isoform-specific RNA degradation or modulation of alternative splicing, as well as for identifying their regulators.

Particularly, EXSISERS verified MBNL1 as the major splice suppressor of *FOXP1* exon 18b inclusion[75,76].

Also, EXSISERS was used to measure tau isoforms non-invasively with cellular resolution and disclosed that physiological and pathological effects (increased 4R-tau level) could be read out conveniently via high-throughput-compatible ratiometric bioluminescence assays and longitudinal quantification. EXSISERS also showcased that the impact of small molecule inhibitors on the inclusion of exon 10 of MAPT , demonstrating that pharmacological agents can be screened for 3R/4R up- or down-regulation conveniently using EXSISERS.

Likewise, the ratiometric readout of EXSISERS, which indicates the fractional 4R-tau expression, was used to characterize programmable RNA perturbation systems regarding their efficiency and isoform-specificity.

The significance of optimizing crRNA design was emphasized in this work, *e.g.*, by targeting exon-spanning regions to avoid targeting of the pre-mRNA nuclear Cas13d activity, which is particularly relevant for genes encoding repetitive domains, common in structural proteins and receptors. It was also demonstrated that isoform-specificity can be enhanced by selectively targeting mature mRNA using cytosolic Cas13b, which is superior to Cas13d in the cytosol but still was exceeded in performance by amiRNAs of the latest generation with regard to efficiency and specificity.

Since EXSISERS reports the translation of mature mRNAs, it is robust against alterations in mRNA levels that do not result in differences in isoform expression[78]. EXSISERS could thus also become particularly informative to non-invasively report local exon-specific protein translation as it occurs in neurons[44]. In addition, EXSISERS could be expanded to create exon-dependent logic circuits in eukaryotic systems, by using intein-flanked recombinases, CRISPR RNA/DNA-targeting effectors, transcription factors, or proteinogenic effector domains without disruption of the host gene.

EXSISERS will complement existing tools due to its non-disruptive nature and longitudinal monitoring capability combined with cellular resolution and will empower *in vivo* imaging studies in genetic model organisms to decode the spatiotemporal dynamics of exon-specific expression.

Development of an intron-specific exon-independent coding transcript (INSPECT) to monitor non-coding RNA expression

INTRODUCTION II

Only less than 2% of the human genome encodes proteins. However, it has been shown that more than three-quarters of all nucleotides are transcribed detectably under certain conditions[79,80]. Those "junk DNA/RNA" has attracted increasing attention in the last decade due to high-throughput technologies such as next-generation sequencing (NGS), which allowed extensive examination of the non-coding genome. The non-coding genome consists of functional and regulatory short non-coding RNAs such as tRNAs, rRNAs, snRNAs, miRNAs, siRNAs, and piRNAs, respectively. Non-protein-coding transcripts longer than 200 nucleotides belong to a class called long non-coding RNAs (lncRNAs). Those can further be classified by their genomic location, mechanism of action, or effector location. Many of those transcripts are not merely "transcriptional noise" but can carry out a variety of functions, such as transcriptional regulation in cis and trans, (re)-organization of the nuclear architecture, scaffolding, or modulation of enzymatic activities[81–84].

Dysregulation of lncRNAs is often involved in a variety of diseases, as in cancer and neurodegenerative diseases[85–88]. Prominent examples of lncRNAs include the X-inactive specific transcript (*XIST*), the nuclear enriched abundant transcript 1 (*NEAT1*) or the homeobox antisense intergenic RNA (*HOTAIR*)[89]. While NGS approaches can be applied for the discovery of lncRNAs, monitoring levels of known lncRNAs are classically detected by RT-qPCR or (sm)FISH[1]. These methods are tedious, consumptive, lack cellular resolution, and are not suited for processing a large number of samples.

Moreover, even for coding genes, current classical gene expression reporters are normally invasive., *e.g.*, the direct fusion of a reporter moiety to the protein of interest's N- or C-terminus or methods using ribosome-skipping peptides (*e.g.,* P2A, T2A) are not scarless since they leave a ~20 aa scar when placed downstream of a protein-coding gene (upstream of the native stop codon) and an N-terminal proline, when placed upstream of a protein-coding gene, which might influence the proteins function or half-life due to its modified termini (N-end rule)[34]. Also, many proteins have an N-terminal or C-terminal signal sequence, which cannot be modified (signal peptide, C-prenylation motifs, N-myristoylation, C-terminal GPI-anchors). For those proteins, placing an IRES-reporter construct into the genes 3'-UTR is the method of choice. Since the 3'-UTR might contain the zip code[90] of a transcript and determines the half-life of the mRNA[31], its modification might be detrimental. The only method to monitor the activity of a gene of interest without modifying its protein sequence or modifying its UTR is the insertion of the promoter region of the GOI (the proximal DNA region upstream of TSS) with a downstream reporter gene into a transgenic locus, *e.g.*, the *AAVS1* (human) or the *Rosa26* (mouse and human) locus.

Therefore, a non-consumptive, modular intron-specific exon-independent coding transcript (INSPECT) system was developed, which allows tagging virtually any transcribed sequence of the genome. Briefly, the INSPECT system mediates nuclear export of an otherwise degraded native or synthetic intronic RNA into the cytoplasmic compartment, where a 5'-cap-independent mechanism enables the translation of a reporter modality of interest. An intronless transcript, independent of its coding potential, can be tagged with INSPECT using a synthetic intron embedded in a suitable region. Especially for ncRNAs where the function of the effector is on the RNA level, it is of utmost importance to not change its primary sequence, *e.g.*, by introducing RNA aptamer loops using the MS2/MCP or PP7/PCP system[11] which has been shown to even unintentionally extend the half-life of the tagged RNA leading to misinterpretation of RNA dynamics[91].

To overcome these shortcomings, INSPECT is based on a reporter gene embedded in a native or artificial intron, together with RNA elements ensuring nuclear export and efficient cap-independent translation. Utilizing the endogenous splicing machinery, the intron-resident construct is excised in an entirely traceless manner.

RESULTS II

There are several canonical methods to monitor the expression of coding genes from an endogenous promoter by tagging the 3'-UTR with an IRES-dependent reporter gene or classical fusion proteins in-frame with the protein of interest or by using a 2A peptide sequence which produces two separate peptides via a ribosome skipping mechanism (see above). However, modification of mRNA may lead to altered half-life times and scars from 2A. More importantly, all those methods only work for coding transcripts and cannot be applied for the majority of the transcriptome that is non-coding.

Over 90% of the coding genes and greater than 60% of the non-coding genes are already containing introns[92] and for those, which do not have introns, a synthetic intron can be introduced artificially without gene disruption. Here, an approach is developed to couple the natural promoter activity to a gene of interest, independent of its coding potential, via a traceless intron tagging system that is inserted using CRISPR/Cas9. The system consists of a reprogrammed intron-resident reporter system that escapes the natural degradation pathways of introns and instead is exported into the cytosol, where the reporter gene is expressed by a cap-independent translation.

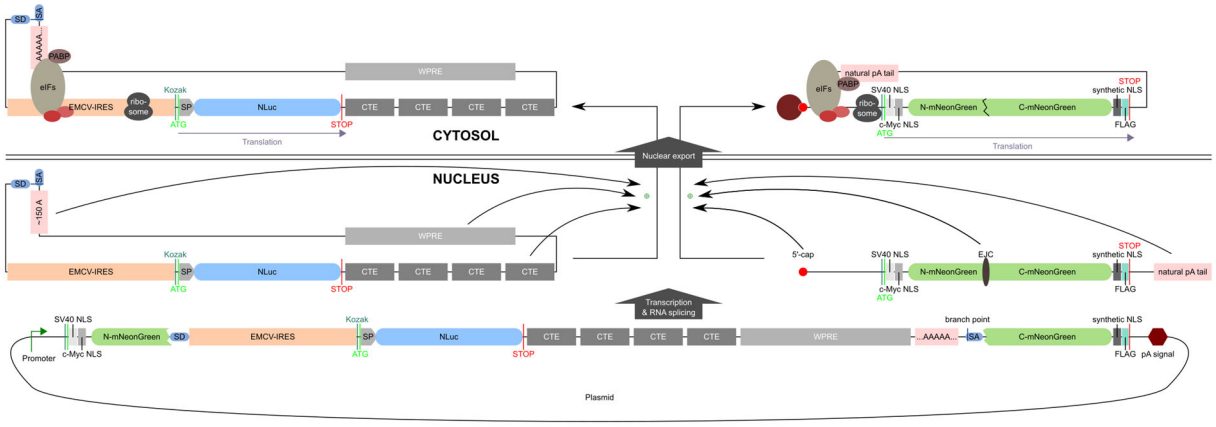

Figure 12 | Introduction of elements of endogenous or synthetic introns into exonic sequences. This schematic describes how intronic sequences can be embedded into exonic sequences such that the transcriptional activity of a gene of interest can be read out without changing its mature mRNA or lncRNA sequence. To test the feasibility of this approach, an mRNA encoding the CDS for mNeonGreen was expressed transiently from a plasmid. Additionally, within the CDS, a synthetic intron was embedded, including an intron-encoded CDS for a secretory NanoLuc luciferase (NLuc). Different elements from RNA viruses known to mediate nuclear export of the viral genome were combined to create an intron-encoded cap-independent translation in a non-canonical way that generates a functional eukaryotic intron-encoded protein which is independent of the co-transcribed mRNA but still reports the transcriptional activity of its host promoter. Elements stimulating nuclear export: CTE: constitutive transport element from Mason-Pfizer monkey virus (MPMV), WPRE: Woodchuck Hepatitis virus post-transcriptional regulatory element (WPRE), poly(A): homopolymeric tracts of adenine bases. Elements enabling cap-independent translation: internal ribosome entry sites (IRES) from a hepatitis C virus (HCV) or encephalomyocarditis virus (EMCV).

Design

To establish INSPECT, a surrogate reporter comprised of a constitutive promoter-driven nuclear-localized fluorescent protein was developed (**Figure 12**). A synthetic intron, derived from a modified rabbit β-globin intron 1, was inserted into the CDS of mNeonGreen[65] **(Figure 12**). To test the efficiency of equipping introns with coding sequences, elements for cap- and poly(A)-independent nuclear export and translation were inserted.

Typically, nuclear export of mature mRNA transcripts to the cytoplasm is mediated by binding of several proteins and protein complexes to the mRNA, *e.g.*, the cap-binding complex (CBC, composed CBP20 and CBP80), TAP (NXF1), p15 (NXT1) and the poly(A)-binding protein PABP2 (PAPBN1)[93]. Those components stimulate the nuclear export of the mRNA[93]. The splicing machinery removes introns of the pre-mRNA, and usually, the 5'-2' linked intron lariat is debranched by DBR1, followed by exonuclease-mediated degradation[94].

Nuclear export of an mRNA is followed by translation, where the initiation is described by a scanning model, in which the 40S subunit of the ribosome is recruited initially to the 5'-cap multimeric complex of the mRNA, forming the 43S preinitiation complex (PIC) and migrates until finding the first AUG codon within an optimal consensus (Kozak) sequence[95,96].

Since many viral transcripts are neither 5'-capped nor polyadenylated, they exploit an alternative strategy for exporting their transcripts from the nucleus and translation initiation. One prominent example for RNA export is the retroviral REV-RRE system from HIV that mediates its RNA-genome export via a REV-mediated binding and nuclear export in its late life-cycle[97,98].

To create an export system which only relies on host factors, a one-component system from another retrovirus was used, the Mason-Pfizer monkey virus (MPMV) constitutive transport element (CTE), a region on the RNA that recruits TAP and p15 from the host export machinery to ensure the export of the viral transcript to the cytoplasm[99–101].

A better-known system improving nuclear export of an RNA is the Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element (WPRE), which has widely been used in transgenic expression systems to enhance mRNA stability and protein yield[102]. WPRE stimulates the nuclear export via karyopherin (CRM1), which explains its positive effect on gene expression on non-polyadenylated transcripts of lentiviral vectors[103]. CRM1 acts as a protein export receptor and exports a subset of endogenous RNAs as well as viral RNAs via adaptor proteins[104].

Translation initiation is mediated in many RNA viruses by an internal ribosome entry site (IRES) located in the 5'-UTR. In contrast to CITE, which is cap-independent but still requires a 5'-3'-direction scanning[105], an IRES does not require scanning of the ribosome but serves as a ribosome landing pad[6] and promotes cap-independent, internal initiation of RNA translation. In the following experiments, the IRES efficiencies of the hepatitis C virus (HCV) and encephalomyocarditis virus (EMCV) were compared. Capped mRNAs recruit the eIF4F complex (consisting of eIF4E, eIF4A, and eIF4G) to the 5'-cap, which allows binding of the 43S pre-initiation complex (40S ribosomal subunit–eIF3–Met-tRNA$_i$–eIF2–GTP–eIF1–eIF1A) and initiation of the scanning process (**Figure 1**)[106]. The EMCV-IRES recruits the 43S particle through direct interaction, whereas the HCV-IRES specifically recognizes the 40S subunit and eIF3 (**Figure 13)**[6].

For classic mRNA translation in eukaryotes, the "closed-loop" model describes the circularization of the mRNA via the mRNA binding proteins on its 5'-cap and on its 3'-end (**Figure 1**). This process enhances mRNA stability and the probability of translation reinitiation[107,108]. The model proposes that the PABP and the eukaryotic translation initiation factor 4E (eIF4E) bind to the 3'-poly(A)-tail and the 5'-cap, respectively, while eIF4G acts as an adaptor protein in-between.

In the reporter system, the closed-loop model is mimicked by the IRES on the one site (5'-end), which recruits the 40S subunit of the ribosome indirectly via a cap-independent binding of translation initiation factors (*e.g.,* EMCV IRES) or directly (*e.g.,* HCV IRES), on the other site (3'-end) by encoding a polyadenylic acid polymer (poly(A)) on

the 3'-end of the intron, which recruits PABP and circularizes to the 5'-end. The poly(A) tail is directly encoded since a poly(A)-signal instead would lead to transcription termination and thus the KO of the host-gene. This is crucial since the intronic reporter should not have an impact on the transcription of the tagged gene of interest. Also, the circular covalently linked intron lariat mimics the closed-loop state of a translation-competent mRNA and should, therefore, be beneficial for translation.

Optimization of INSPECT

To optimize the INSPECT system, a minimal set of different constructs with mNeonGreen (mNG) as a constitutively expressed exon-encoded protein was carefully designed. A synthetic intron derived from the first intron of the rabbit β-globin gene was inserted into mNG between Gln-849 and Val-850 (CAG|GTG) since those nucleotides follow the consensus sequence flanking an intronic sequence and thus ensure optimal splicing efficiency. As an intron-encoded protein, NanoLuc luciferase (NLuc) with an N-terminal secretion peptide (SP) from *Gaussia princeps* luciferase[109,110] was used. Different elements enabling cap-independent translation and cap- and poly(A) independent nuclear export elements were permuted and tested transiently in HEK293T cells (**Figure 13a**). First, the IRES from the hepatitis C virus was combined with different nuclear export elements. When transiently transfected in cells, a time-dependent increase of NLuc signal in the supernatant could be observed with different slopes. As expected for SP-NLuc with solely HCV-IRES, only a marginal increase could be detected. Most likely, the intron escaped the nuclear compartment during cell division and was then translated cap-independently via the HCV-IRES (**Figure 13b**). By contrast, by inserting WPRE downstream of HCV-IRES_SP-NLuc, the "intron-encoding capacity (IEC)" could be dramatically increased. CTE elements also have an equal or even better effect since a tandem CTE pair alone downstream of the HCV-IRES_SP-NLuc raised the signal even more than WPRE alone (**Figure 13b**). Also, the addition of directly encoded poly(A)s downstream of the elements increased the IEC. The highest signal was measured with all components combined (**Figure 13b**). Next, the IRES from the encephalomyocarditis virus was tested for its potency to drive cap-independent translation since it has recently been shown that EMCV IRES not only recruits the ribosomal subunits indirectly by the translation initiation factors but also recruits the 40S ribosomal subunit in the absence of eIF4G/4A[111]. Additionally, the commonly used EMCV-IRES (*e.g.,* pCITE-1 and pIRES) contained non-optimal mutant variants of the IRES, such as an adenine insertion in the bifurcation loop, and thus are attenuated[112]. The HCV-IRES element for some key constructs was replaced by a mutation-free EMCV-IRES. Seemingly, the EMCV-IRES drove the SP-NLuc translation much more efficiently compared to HCV-IRES since the EMCV-counterpart with only a single WPRE-element already was comparable as the best candidate with HCV-IRES (**Figure 13c**). Additionally equipped with CTE elements and poly(A)s, the signal almost tripled (**Figure 13c**). All constructs tested showed a similar expression of the exonic mNeonGreen, indicating the non-invasiveness of those reprogrammed introns (**Figure 13d**).

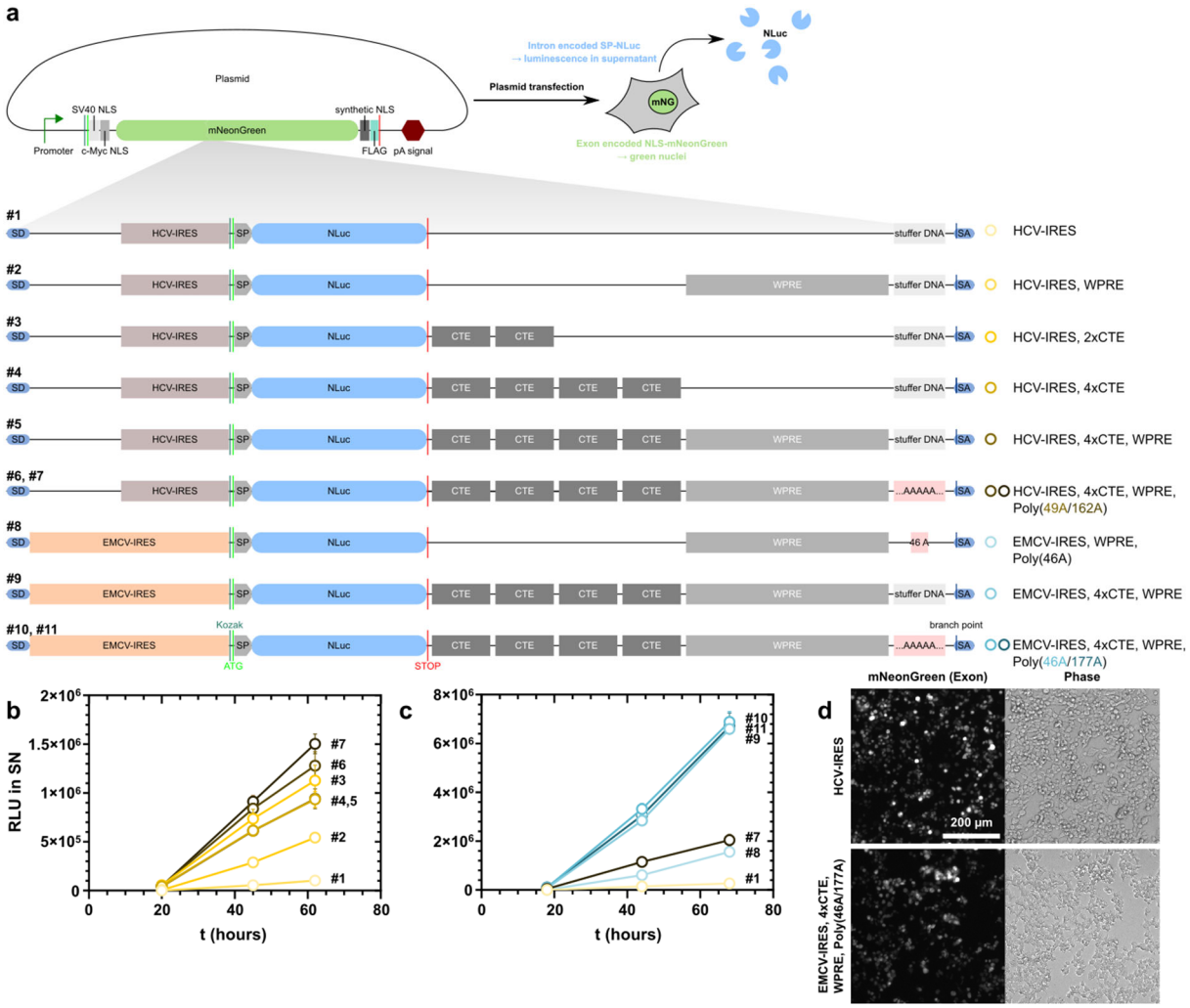

Figure 13 | Engineering of an eukaryotic intron-encoded extranuclear cap-independent protein-coding transcript. a, To assess the ability to encode proteins within an intronic sequence, a secreted Nanoluc luciferase (NLuc) as intron-encoded protein was used and inserted into the intronic sequence within an exonic mRNA encoding for a nuclear-localized mNeonGreen driven by a constitutive hybrid mammalian CAG promoter. To enable the translation of an intron-encoded CDS, the intron first has to be exported out of the nucleus after its excision while escaping the native degradation pathway and secondly, a cap-independent translation has to be initiated. Different elements from RNA viruses known to mediate nuclear export of the viral genome and cap-independent translation were combined to generate a functional eukaryotic intron-encoded protein. Elements stimulating nuclear export: CTE: constitutive transport element from Mason-Pfizer monkey virus (MPMV), WPRE: Woodchuck Hepatitis virus post-transcriptional regulatory element, poly(A): homopolymeric tracts of adenine bases. Elements enabling cap-independent translation: internal ribosome entry sites (IRES) from Hepatitis C virus (HCV) or encephalomyocarditis virus (EMCV) **b,** Different elements were combined or put in tandem to optimize the nuclear export and translation efficiency of the intronic RNA containing HCV-IRES; read-out via the intron-encoded secreted NLuc. The supernatant of the samples was collected at the indicated time points post-transfection. **c,** Different elements were combined or put in tandem to optimize the nuclear export and translation efficiency of the intronic RNA containing EMCV-IRES. Intron-encoded secreted NLuc was read-out in the supernatant. The supernatant of the samples was collected at the indicated time points post-transfection. **d,** Representative epifluorescence images cells expressing the exon-encoded mNeonGreen-NLS transfected with the indicated constructs. Results shown in b, c, and d were performed together with Gerald Raffl as a part of his research internship.

Modularity of the intron-encoded protein

After optimization of the intron-encoding capability of INSPECT, more complex proteins were tested if they can be intronically expressed. For this, the multipass transmembrane sodium-iodide symporter (NIS alias SLC5A5)[113] was selected, which is initially inserted into the membrane at the endoplasmic reticulum. The expression of NIS can be monitored by measuring the accumulation of radioactive iodine ($^{131}I^-$)[114,115], which usually is not absorbed by non-thyroid cells (**Figure 14a**). SP-NLuc was used as an intron-encoded protein for control. After 48 h post-transfection, the cells were incubated at the specified times, and the accumulated iodine was read out via a γ-scintillator (**Figure 14a**). Cells transfected with the intron-encoded NIS showed a dramatic incubation-time-dependent increase in accumulated radioactivity (**Figure 14b**), which shows that complex multipass transmembrane proteins can also be encoded in the intron. Surprisingly, the 3-fold larger size of NIS compared to SP-NLuc did not change the splicing efficiency, as shown by the comparable fluorescence of the exon-encoded nuclear mNG (**Figure 14c**, measurements were performed by Volker Morath at the Klinikum rechts der Isar) indicating the general usability of introns to encode complex proteins. The intron-encoded NIS may already prove to be a valuable tool for tracking genes with non-invasive imaging. Besides the $^{131}I^-$, there are also isotopes such as $^{124}I^-$ (β^- and β^+ emitter), which are excellent isotopes for positron emission tomography imaging. For example, engineered (CAR)-T-cells could be tracked non-invasively in pre-clinical or clinical settings, where $INSPECT_{NIS}$ is inserted into *IL2*, an early response marker for activated T-cells[116]. Those activated (CAR)-T-cells express the NIS without the gene for *IL2* being modified at the mRNA level since INSPECT is excised at the pre-mRNA level and is translated independently (**Figure 14d**). Also, NIS is not immunogenic because it is a human protein unchanged in its sequence, which eases its usage under clinical settings.

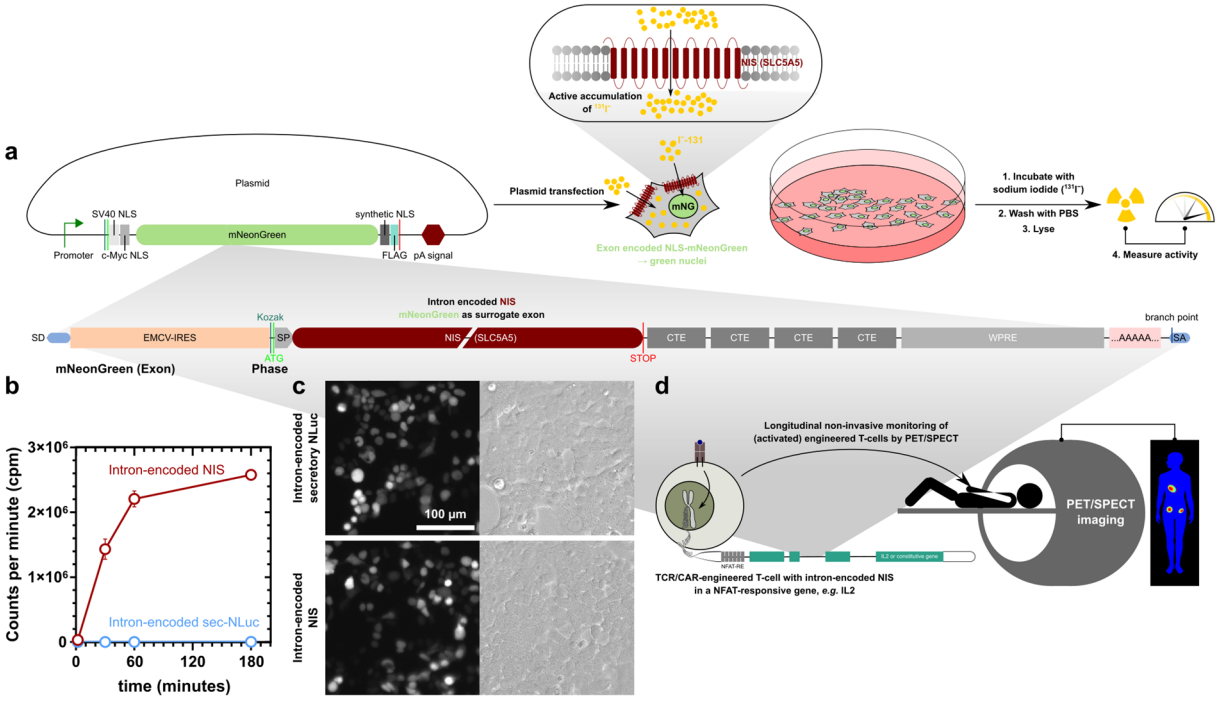

Figure 14 | Application of the intron-encoded extranuclear transcript for non-invasive expression of a translocon-dependent multipass-transmembrane protein. a, As a prototype intron-encoded multipass transmembrane protein, sodium-iodide symporter (NIS alias SLC5A5) was used, which was transfected into HEK293T cells. Its expression was quantified via the accumulation of the -emitter $^{131}I^-$. **b,** After the indicated incubation time with sodium iodide (^{131}I isotope), the accumulated $^{131}I^-$ in the lysed samples was measured via a γ-scintillator. (Experiment performed in collaboration with Volker Morath) **c,** Epifluorescence microscopy images of exonic mNeonGreen-NLS expressing the indicated intron-encoded NIS or secretory NLuc. **d,** The intron-encoded NIS could be integrated within the *IL2* gene, which is transcriptionally induced in activated (CAR)-T-cells enabling longitudinal non-invasive monitoring of activated (CAR)-T-cells using positron emission tomography (PET) and single-photon emission computed tomography (SPECT) via the accumulation of radioactive I^- isotopes. Results shown in b and c were performed together with Gerald Raffl as a part of his research internship.

Design and integration of a non-leaky and efficient KO-switch

Many biological questions regarding the physiological function of a gene are still solved by classic (conditional) knock-outs (KOs). Thus, a KO-switch, which is integrated into INSPECT in a way that does not disturb the host gene in its non-activated basal state, would be ideal. For this purpose, the off-switch was placed upstream of the IRES, consisting of the following elements: three inverted poly(A) signals, composed of those of the SV40 late poly(A) signal[117], the rabbit β-globin poly(A) signal[118] and a synthetic poly(A) signal[119] (**Figure 15a)**. As the SV40 late poly(A) signal also encodes a poly(A) signal in the reverse complementary direction (early poly(A) signal), two mutations were introduced, which destroyed the two AAUAAA motifs in the early poly(A) direction[120]. Additionally, an inverted splice acceptor from the second rabbit β-globin intron was placed downstream of the inverted triple poly(A) signal (**Figure 15a**). Two semi-orthogonal loxP sites (loxP-WT and lox2272, both are recognized by the Cre recombinase but can only recombine with its sequence-identical site but not with each other and thus, are called semi-orthogonal[121]) are positioned upstream and downstream of the inverted SA_3×poly(A) in a way, that upon Cre recombinase expression, the inverted SA_3×poly(A) is re-inverted into its active functional state[122], resulting in a KO of the host-gene. The splice acceptor (SA) ensures the usage of the poly(A) signals for transcript termination upon Cre-mediated activation of the KO-switch. Without the proximal SA, the poly(A) site could potentially be skipped without being cleaved, since splicing of the intron splice donor (SD) and acceptor of INSPECT is highly efficient and might be faster than the poly(A)-signal-mediated cleavage resulting in a functional host mRNA/ncRNA[123,124]. The SA of the SA_3×poly(A) ensures the usage of the poly(A) by preventing the usage of the downstream SA of the original intron-encoded construct. The off-switch is placed upstream of the IRES to not only couple the on/off-state to the host gene but also the intron-encoded protein to this switch. To facilitate an easy selection of cells containing the INSPECT system in the gene of interest, an inverted EF1α promoter-driven puromycin N-acetyltransferase (PuroR) was coupled to a Herpex simplex thymidine kinase (HSV-Tk) ORF downstream of the inverted poly(A) signal enabling puromycin-mediated selection. Afterward, the cassette is removed upon Flp recombinase expression[125], and the cells are counter-selected with ganciclovir. Ganciclovir kills cells that still contain the cassette because HSV-TK converts ganciclovir into a DNA-damaging phosphorylated product[126].

This KO-switch was tested again transiently in the exonic mNeonGreen-NLS system and co-expressed Cre or Flp recombinases to benchmark the KO-efficiency (**Figure 15a**). Upon Flp recombinase expression, both the mNeonGreen and the NLuc activity in the supernatant increased, which can be explained that by the excision of the inverted EF1α-driven cassette, the transcriptional interference of the CAG-driven mNeonGreen by the EF1α-promoter does not occur anymore (**Figure 15b,d,e**). Upon Cre recombinase expression, the exonic mNeonGreen signal and the intronic NLuc signal is dramatically decreased, indicating an efficient Cre-mediated off-switch (**Figure 15c,d,e**).

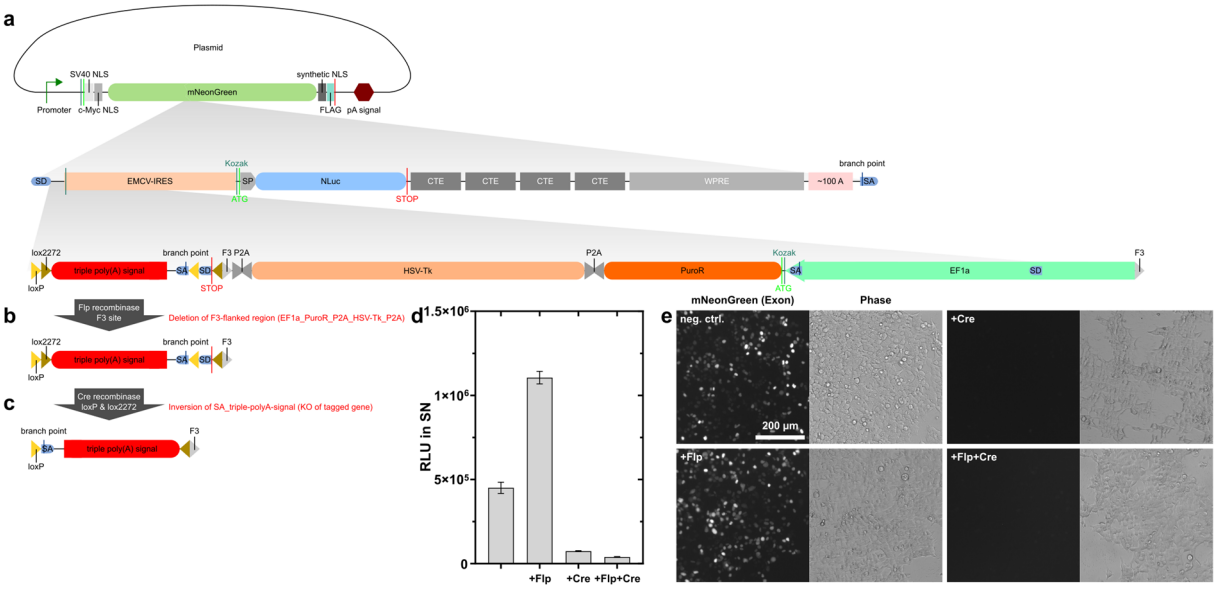

Figure 15 | Design of the Cre-inducible KO-switch based on the intron-encoded extranuclear transcript system. a, A plasmid-driven mNeonGreen expression cassette was used as a surrogate gene to test the KO-switch. Additionally to the intron-encoded reporter system, an inverted EF1a promoter-driven selection cassette was integrated encoding for the puromycin N-acetyltransferase (PuroR) and the viral thymidine kinase (HSV-Tk), separated via a P2A ribosome skipping peptide. The selection cassette enables positive selection after nuclease-mediated KI of the intron-encoded transcript into the gene of interest. **b,** Afterward, the cassette is removed by Flp recombinases. Only the promoter-CDS moiety is flanked by mutant variant F3 of FRT-sites and thus is excised via transfection of a plasmid encoding for Flp recombinases. The inverted composite part comprising the splice donor (SD), splice acceptor (SA), and the triple poly(A) (pA) signal, is thus not removed. **c,** The SA-pA part is "FLExed", meaning two different semi-orthogonal loxP sites (lox2272 and loxP WT sites are both not compatible but are both recognized by the same Cre recombinase) are flanking the SA-pA part in a way, that upon Cre recombinase expression, this part will be irreversible flipped in its non-inverted direction. The SD part is positioned in a way that it will be removed after Cre-mediated SA-pA inversion. Since Cre recombinase leads to the restoration of the SA-pA in the sense direction of any tagged gene, it will inevitably lead to the KO of the gene by premature polyadenylation by the restored poly(A) signal. The SA ensures that the poly(A) signal is not accidentally skipped since some introns splice within seconds, which might lead to an ineffective premature transcript termination. The SA from the switch prevents the usage of the downstream SA. The SA_poly(A) transcript is redefined as an exonic sequence after Cre-mediated inversion into the gene's sense direction and thus ensures the premature transcript termination. The effect of Flp or Cre recombinases on the plasmid-based test-constructs expressing exonic mNeonGreen and intron secretory NLuc with the Cre-inducible KO-switch are readout via **d,** bioluminescence signal of NLuc in the supernatant and **e,** epifluorescence microscopy of the nuclear-localized mNeonGreen. Results shown in d and e were performed together with Gerald Raffl as a part of his research internship.

Non-invasive transcriptional coupling of the lncRNA *NEAT1* using INSPECT

Ultimately, the aim was to show that one can transcriptionally couple a non-coding RNA non-invasively via INSPECT to a secretory luciferase and knock it out afterward via Cre recombinase. As a target, the long non-coding RNA (lncRNA) *NEAT1* was selected, which plays a role in pluripotency maintenance in human iPSC and ESC by controlling the phase separation of TDP-43 (TARDBP)[127]. *NEAT1* is expressed in two isoforms, the short (v1, 3.7 kbp) and the extended version (v2, 22.7 kbp). TDP-43, which usually shows an increased expression in stem cells, stimulates the premature polyadenylation of *NEAT1_v1* and thus, mediates the exclusive expression of *NEAT1_v1* in stem cells. If the level of TDP-43 decreases during cell differentiation, *NEAT1_v2* is also expressed more frequently because the alternative poly(A) site (APA) of *NEAT1_v1* is used less. Since *NEAT1_v2* is an essential part of so-called nuclear bodies called paraspeckles (an agglomeration of *NEAT1* RNA and sequestered proteins), differentiation also will induce paraspeckle formation. *NEAT1_v2* also contains elements that bind TDP-43, and consequently, induction of *NEAT1_v2* leads to the phase separation of TDP-43. Hence, the expression of *NEAT1_v2* triggers a positive feedback loop where more and more TDP-43 is taken from the solution and is sequestered into paraspeckles[127,128]. *NEAT1* is also induced in a variety of cellular stress, such as viral infections, DNA damage, cancer, hypoxia, and heat shock[129–135].

INSPECT$_{\text{SP-NLuc}}$ was introduced using CRISPR/Cas9 into the shared region of *NEAT1_v1* and *NEAT1_v2* (**Figure 16a**). After successful knock-in and selection (puromycin), and Flp-mediated cassette excision (**Figure 16b**) and counter-selection (Ganciclovir), only homozygous clones were used for further analysis. A subclone with homozygous NEAT-KO was also created by transfecting a homozygous INSPECT$_{\textit{NEAT1}\text{:SP-NLuc}}$ clone with a plasmid expressing Cre recombinase (**Figure 16c**). Using smFISH analysis, it could be shown that both, the INSPECT$_{\textit{NEAT1}\text{:SP-NLuc}}$ and the unmodified HEK293T cells have paraspeckles, but not the subclone INSPECT$_{\textit{NEAT1}\text{:SP-NLuc+Cre}}$, where the inverted SA_3xpoly(A) signal is flipped into its sense direction (**Figure 16d**). Consequently, the NLuc signal is also barely detectable in the KO clone, clearly demonstrating a transcriptional coupling between the gene and that of the intron-encoded reporter (**Figure 16e**). At the same time, it is also shown that the protein encoded in the intron has no relevant upstream promoter-like sequences that generate false-positive background luciferase activity. Otherwise, a residual signal would be evident despite Cre recombinase. The quantification of the images of INSPECT$_{\textit{NEAT1}\text{:SP-NLuc}}$ and unmodified HEK293T cells (representative examples shown in **Figure 16d**) also show that the number of paraspeckles-containing cells remained unchanged between both (**Figure 16f**).

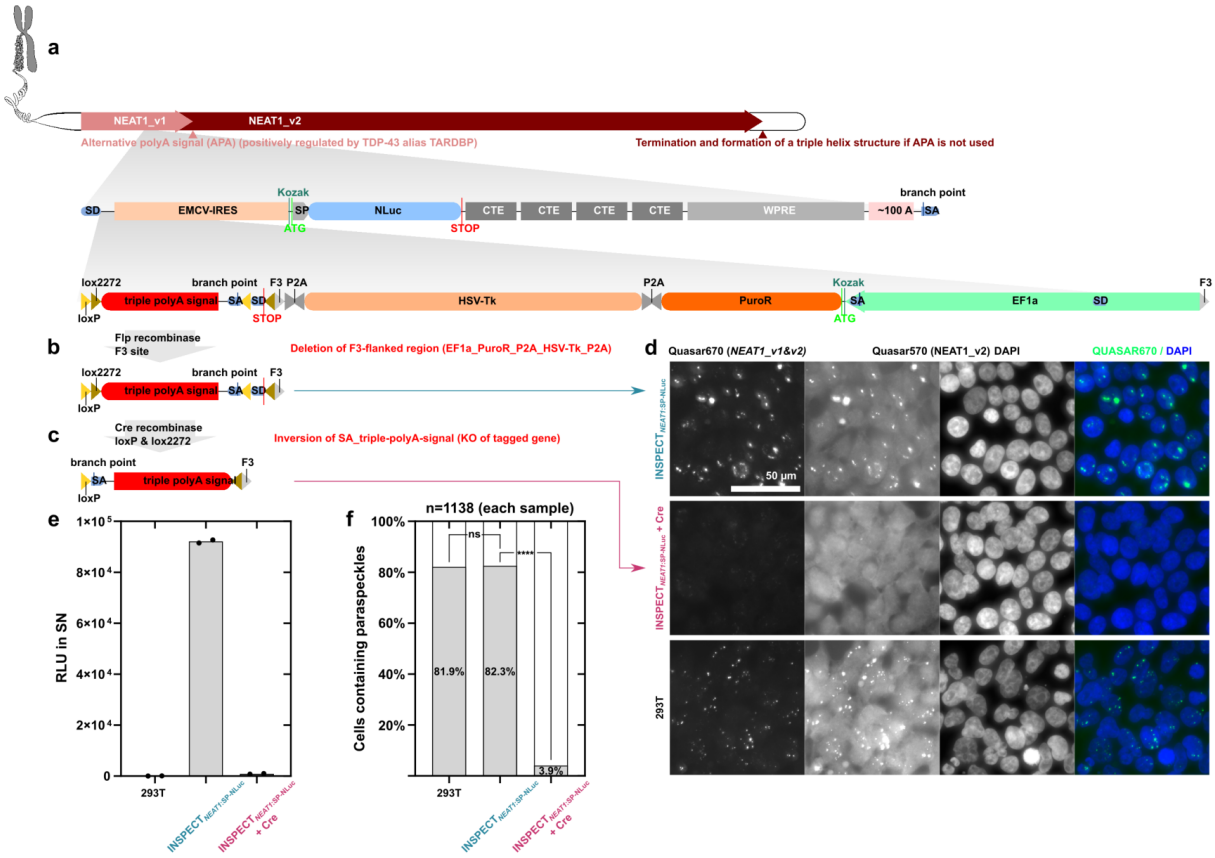

Figure 16 | The intron-encoded extranuclear transcript system enables non-invasive and longitudinal monitoring of long non-coding RNAs (lncRNAs) with an integrated Cre-inducible KO-system. a, The reporter construct was knocked in into the lncRNA *NEAT1_v1,* which is also a part of the long isoform *NEAT1_v2*. **b,** Flp recombinase-mediated the excision of the EF1a-PuroR-P2A-HSV-Tk and **c,** the Cre recombinase-mediated KO of *NEAT1*. **d,** Representative smFISH images of probes binding to the region of *NEAT1_v1/v2* and *NEAT1_v2* of unmodified 293T cells, the reporter without (INSPECT$_{NEAT1}$:SP-NLuc) and with Cre-activated off-switch. **e,** Relative luminescence of the supernatant 48 h post-seeding of indicated cells (unmodified HEK293T, INSPECT$_{NEAT1}$:SP-NLuc, INSPECT$_{NEAT1}$:SP-NLuc +Cre, technical duplicates shown as data points). **f,** Quantification of paraspeckle containing cells (using Quasar670 signal of *NEAT1_v1/v2*). **** denoting *p*-values smaller than 0.0001 (binomial test, two-tailed). Results shown in d, e, and f were performed together with Tobias Santl and Maren Beyer as a part of their bachelor's thesis.

DISCUSSION & OUTLOOK II

The results showed that several sequences could be tagged by INSPECT, which is exported from the nucleus to the cytoplasm and could be used to express a reporter gene of choice. INSPECT is traceless since it did not interfere with *NEAT1* lncRNA paraspeckle formation due to INSPECT's co- or post-transcriptional excision from the tagged lncRNA. This reporter system enables longitudinal tracking of transcriptional states of the host gene regardless of any influence by posttranscriptional regulatory mechanisms. Screening of changes in expression levels is non-consumptive *in cellulo* and potentially also *in vivo,* and might enable high-throughput screening applications.

It remains to be determined, to which extent the intron exists as a lariat structure. The nuclear export of the transcript competes kinetically with intron degradation, whose first step consists of debranching the lariat structure by DBR1[94].

Moreover, embedding the tagging system in existing, endogenous introns could be challenging since endogenous splice sites are not always perfectly conserved, and therefore splicing at those sites could be slower than the nuclear export of the transcript stimulated by CTE and WPRE, resulting in unspliced RNA in the cytosol. Therefore, it is recommended to embed the construct within an artificial, highly efficient intron to ensure fast and traceless excision of the embedded intron-reporter moiety. Analogous to the anti-parallel coiled-coil domains of the EXSISERS constructs, one might consider adding similar elements, such as randomized sense and anti-sense motifs immediately downstream and upstream of the splice donor and acceptor to improve RNA-splicing efficiency further.

Also, a second optimization round can be applied in INSPECT to remove all potential cryptic splice sites, enhancer, suppressor and modulator motifs.

One should note that when inserting INSPECT into an RNA without introns, the exon-exon-junction most probably will have an exon junction complex (EJC) ~20-24 nt upstream of the exon-junction[93,136]. EJCs are known to stimulate the nuclear export of mRNAs[93,136] and might increase the chance for an undesired export of a transcript, which is naturally located in the nucleus.

The Cre-inducible KO tool allows highly efficient spatiotemporally controlled depletion of a GOI, as shown for *NEAT1*. Due to the excellent availability of many constitutive or inducible Cre-driver mice, such a system would enable a tissue-specific KO by breeding such a mouse with a Cre-driver line. When introducing this reporter system with NIS (SLC5A5) as a non-invasive reporter in an *in vivo* mouse model, it might enable observation of lncRNA expression throughout the whole body via PET. NIS is nearly orthogonal because NIS can be down-regulated with high doses of iodide[137]. This effect is mediated by the diminution of thyroid function by depletion of TH biosynthesis (Wolff-Chaikoff effect)[137]. It allows near-orthogonal imaging by autoradiography, γ-camera scintigraphy, or positron emission tomography (PET) using NIS as a non-immunogenic, endogenous reporter gene[114,115,138,139].

Herein, a novel approach was presented to tag and quantify transcribed sequences regardless of their coding potential. This method allows the interrogation of transcriptional changes of any transcribed sequence in a non-invasive and high-throughput-compatible manner. Since the construct resides in an intron, tagging is practically traceless and does not affect the mature transcript of interest. If further improved, INSPECT might be a novel and highly promising tool, allowing precise investigation of RNA transcription independent of its coding potential.

METHODS

The following sections contain paragraphs that were part of the method section that were submitted for publication with the title “Non-invasive and high-throughput interrogation of exon-specific isoform expression”.

Molecular cloning

PCR for molecular cloning

Single-stranded primer deoxyribonucleotides were diluted to 100 µM in nuclease-free water (Integrated DNA Technology (IDT)). PCR reaction with plasmid and genomic DNA template was performed with Q5 Hot Start High-Fidelity 2x Master Mix or with 5x High-Fidelity DNA Polymerase and 5x GC-enhancer (New England Biolabs (NEB)) according to manufacturer‘s protocol. Samples were purified by DNA agarose gel electrophoresis and subsequent purification using Monarch® DNA Gel Extraction Kit (NEB).

DNA digestion with restriction endonucleases

Samples were digested with NEB restriction enzymes according to the manufacturer‘s protocol in a total volume of 40 µl with 2-3 µg of plasmid DNA. Afterward, fragments were gel-purified by DNA agarose gel electrophoresis and subsequent purification using Monarch® DNA Gel Extraction Kit (NEB).

Molecular cloning using DNA ligases and Gibson assembly

Agarose-gel purified DNA fragment concentrations were determined by a spectrophotometer (NanoDrop 1000, Thermo Fisher Scientific). Ligations were carried out with 50-100 ng backbone-DNA (DNA fragment containing the ori) in 20 µl volume, with molar 1:1-3 backbone:insert ratios, using T4 DNA ligase (Quick Ligation™ Kit, NEB) at room temperature for 5-10 min. Gibson assemblies were performed with 75 ng backbone DNA in a 15 µl reaction volume and a molar 1:1-5 backbone:insert ratios, using NEBuilder® HiFi DNA Assembly Master Mix (2x) (NEB) for 20-60 min at 50 °C.

DNA agarose gel electrophoresis

Gels were prepared with 1% agarose (Agarose Standard, Carl Roth) in 1x TAE-buffer and 1:10.000 SYBR Safe stain (Thermo Fisher Scientific), running for 20-40 min at 120 V. For analysis 1 kb Plus DNA Ladder (NEB) was used. Samples were mixed with Gel Loading Dye (Purple, 6x) (NEB).

Bacterial strains (*E. coli*) for molecular cloning

Chemically- and electrocompetent Turbo/Stable cells (NEB) were used for the transformation of circular plasmid DNA. For plasmid amplification, carbenicillin (Carl Roth) was used as a selection agent at a final concentration of 100 µg/ml. All bacterial cells were incubated in Lysogeny Broth-Medium (LB) and on LB agar plates, including the respective antibiotics.

Bacterial transformation with plasmid DNA

For electroporation, either 1-5 µl Ligation or Gibson reaction was dialyzed against MilliQ water for 10-20 min on an MF-Millipore membrane filter (Merck). Afterward, 1-5 µl dialysate was mixed with 50 µl of thawed, electrocompetent cells, transferred to a pre-cooled 2 mm electroporation cuvette (Bio-Rad), shocked at 2.5 kV (Gene Pulser Xcell™ Electroporation Systems, Bio-Rad), and immediately mixed with 950 µl SOC-medium (NEB). The chemical transformation was performed by mixing 1-5 µl of Ligation or Gibson reaction with 50 µl thawed, chemically competent cells and incubated on ice for 30 min. Cells were then heat shocked at 42 °C for 30 s, further incubated on ice for 5 min, and finally mixed with 950 µl SOC-medium (NEB). Transformed cells were then plated

on agar plates containing an appropriate type of antibiotic and concentrations according to the supplier's information. Plates were incubated overnight at 37 °C or over 48 hours at room temperature.

Plasmid DNA purification and Sanger-sequencing

Plasmid DNA transformed clones were picked and inoculated from agar plates in 2 ml LB medium with appropriate antibiotics and incubated for about 6 h (NEB Turbo) or overnight (NEB Stable). Plasmid DNA intended for sequencing or molecular cloning was purified with QIAprep Plasmid MiniSpin (QIAGEN) according to the manufacturer‘s protocol. Clones that were intended to be used in cell culture experiments were inoculated in 100 ml antibiotic-medium and grown overnight at 37 °C containing the appropriate antibiotic. Plasmid DNA was purified with the Plasmid Maxi Kit (QIAGEN). Plasmids were sent for Sanger sequencing (GATC-Biotech) and analyzed by Geneious Prime (Biomatters) sequence alignments.

Mammalian cell culture

Cell lines and cultivation

HEK293T cells (ECACC: 12022001, Sigma-Aldrich) were maintained at 37 °C, in 5% CO_2, H_2O saturated atmosphere in advanced Gibco™ Advanced DMEM (Gibco™, Thermo Fisher Scientific) supplemented with 10% FBS (Gibco™, Thermo Fisher Scientific), GlutaMAX (Gibco™, Thermo Fisher Scientific) and penicillin-streptomycin (Gibco™, Thermo Fisher Scientific) at 100 µg/ml. Cells were passaged at 90% confluency by removing the medium, washing with DPBS (Gibco™, Thermo Fisher Scientific), and separating the cell with 2.5 ml of an Accutase® solution (Gibco™, Thermo Fisher Scientific). Cells were then incubated for 5-10 min at room temperature until a visible detachment of the cells was observed. Accutase™ was subsequently inactivated by adding 7.5 ml pre-warmed DMEM including 10% FBS and all supplements. Cells were then transferred into a new flask at an appropriate density or counted and plated on 96-well, 48-well, or 6-well format for plasmid transfection.

Plasmid transfection

Cells were transfected with X-tremeGENE HP (Roche) according to the protocol of the manufacturer. DNA amounts were kept constant in all transient experiments to yield reproducible complex formation and comparable results. In 96-well plate experiments, a total amount of 100 ng of plasmid DNA was used, in 48-well plates, a total amount of 300 ng of plasmid DNA was used, and in 6-well plates, a total amount of 2.4 µg of plasmid DNA was used per well. Cells were plated one day before transfection (25,000 cells/well in 100 µl for 96-well plates, 75,000 cells/well in 500 µl for 48-well plates, 600,000 cells/well in 3 ml for 6-well plate). 24 h post-transfection, 100 µl fresh medium was added on 96-well transfection per well, 48 h post-transfection 100 µL medium was removed and replaced with fresh medium on 96-well transfections per well. For modulation of alternative splicing with 5-iodotubercidin (5-ITU) (Sigma-Aldrich), 24 hours post-transfection 5-ITU (in DMSO) were applied on the cells. Control cells received the same volume of DMSO.

Generation of stable cell lines via CRISPR/Cas9

To generate a stable cell line (HEK293T, N2a), plasmids expressing a mammalian codon-optimized Cas9 from *S. pyogenes* (SpyCas9) with a tandem C-terminal SV40 nuclear localization signal (SV40 NLS) (CBh hybrid RNA-polymerase II promoter-driven) and a single-guide-RNA (sgRNA/gRNA, human U6 RNA-polymerase III promoter-driven) with a 19-21 bp (SpyCas9 | *Tubb3* exon 2 (between Gly-81 and Ser-82); spacer: G+TGTGGAGCGGTACTCACAGG | *MAPT* exon 10; spacer: G+CCAGTCCAAGTGTGGCTCAA | *MAPT* exon 11; spacer: GTTGCCTAATGAGCCACACT | *FOXP1* exon 18b; spacer: AGGATGAGTTTGGGTCCTTT) cloned spacer targeting the exon-of-interest were used. The insertions site for EXSISERS was always upstream of serine, cysteine, or threonine due to the downstream extein requirements of inteins. The efficiency of CRISPR/Cas9 for a

target site was performed by T7 endonuclease I assay (NEB) according to the manufacturer's protocol after 48-72 h post-transfection of cells with plasmids encoding Cas9 and the targeting sgRNA on a 48-well plate. Optionally, an i53 expression plasmid (a genetically encoded 53bp1 inhibitor140) was co-transfected to enhance homologous recombination (HR) after the Cas9-mediated double-strand break at the spacer-guided genomic site. Donor DNA plasmid contains the intein-flanked moiety, including the selection-cassette to select for cells undergoing successful Cas9-mediated HR; moreover, the donor DNA plasmid contains homology arms of at least 800 bps flanking the intein-reporter construct. 48 hours post-transfection (48-well or 6-well format), the medium was replaced with medium containing 50 µg/ml puromycin, if not otherwise indicated. Cells were observed daily and were detached with Accutase™ and replated with puromycin when surviving colonies reach the colony size of about 50 cells. This step was repeated until no significant puromycin-mediated cell death could be observed. Those cells were plated without puromycin on a 48-well plate and were transfected with a CAG-hybrid promoter-driven nuclear-localized Cre or Flp recombinase with and a low amount of a green fluorescent protein (Xpa-H62Q141) in a 10:1 ratio to excise the selection cassette. The green fluorescent protein was co-transfected in order to enrich cells successfully co-transfected with the recombinase expressing plasmid. Green cells were enriched with the BD FACSaria II (controlled with the BD FACSDiva Software (Version 6.1.3, BD Biosciences)) and replated on a suitable dish/plate. After one week, enriched cells were single-cell-sorted in 96-well plates and grown mono-clonally until colony size was big enough to be duplicated onto a second 96-well plate containing 2 µg/ml puromycin. Cells that underwent successful cassette excision should not survive puromycin treatment indicating that the original clone from which it was duplicated did not anymore contain the puromycin-N-acetyltransferase and was a potential candidate for genotyping for zygosity. Those clones were detached and expanded on 48-well plates until confluency and half of the cell mass were then used subsequently for isolation of genomic DNA using Wizard® Genomic DNA Purification Kit (Promega). Genotyping of the genomic DNA was performed using LongAmp® Hot Start Taq 2X Master Mix (NEB) according to manufacturer's protocol with primer deoxynucleotides pairs (IDT) with at least one primer binding outside of the homology arms. The PCR product from clones, where the genotyping indicates homozygosity, were sent for Sanger-sequencing to verify its sequence integrity. MAPT exon 10 was genotyped with following primers: CCTGCTTATGGAAGAGCTGAGAAAAAG & CCTCACAGCAGACAAACTGG. MAPT exon 11 was genotyped with the following primers: GACTGAGATCAGCTGGCAGC & CTGCCGATGGTGAAGTGTCTC. FOXP1 exon 18b was genotyped with following primers: GAACCTATTTTGGGCTTGTATGC & AGAGGATGGAAATTATGATACTGCTG. Tubb3 exon 2 was genotyped with following primers: CTGTGTCTTGAGTCCTAGTCCATTC & GGAACTTACCAAAGATAAAGTTGTCGG.

For INSPECT constructs, a similar approach was performed, with the only difference that after transfection with a plasmid encoding Flp recombinase, the cells were counter-selected one week after transfection with ganciclovir (2 and 10 µM) for another two weeks, before they were monoclonalized via the FACS in 96-well plates. The spacer used for SpyCas9 was G+CATCTGAAAACCTTTACCCC for *NEAT1*. *NEAT1* was genotyped with following primers: CAAAAGTTGTGGCAAGTCCAGCC & ACGTGATCCTGCTCACACCTTTG. The INSPECT-integrated KO-switch status was genotyped with: CAGAGTTACGTAGCGATCGCTAG & CAAATAGAATAGGAACTTCCGAAGG.

Gene expression manipulation with CRISPR/Cas9

Gene expression of *MAPT* was induced in HEK293T cells by co-transfecting CAG-driven mammalian-codon optimized nuclease-defect *S. pyogenes* Cas9 (D10A, H840A) fused to a tripartite trans-activation domain and SV40 NLS[56] and a plasmid mix expressing (human U6 promoter) three spacer-truncated sgRNAs (14-15 nt truncated spacer instead of 19-21)[142,143] targeting the 5'-upstream region of the *MAPT* transcription start site (TSS): G+GAGGGCAGCGCCGAG, GGAGAAGGCTCCCG, and GCGAGCCTCCCCAG. Uninduced control was co-transfected instead with an empty sgRNA cloning plasmid.

mRNA manipulation with CRISPR/Cas13.

CAG-driven mammalian codon-optimized RfxCas13d (Cas13d from *Ruminococcus flavefaciens* XPD3002)[66] with a C-terminal triple NLS (SV40 NLS + MYC NLS + synthetic NLS) or PspCas13b (Cas13b from *Prevotella sp. P5-125*)[67] with a C-terminal nuclear export signal from HIV Rev protein were co-transfected with a plasmid encoding for the crRNA of the Cas13 system (human U6 RNA polymerase III driven) targeting the RNA of interest indicated in the figures. All constructs were generated using oligodeoxynucleotides (IDT) and gene fragments (gBlocks®, IDT).

mRNA manipulation with artificial microRNAs

CAG-driven mammalian codon-optimized iRFP720 was interrupted with a modified intron derived from rabbit β-globin. Within the synthetic intron, the artificial mir-30-based synthetic microRNA backbone containing the critical region for efficient microRNA biogenesis and a type IIS restriction enzyme cloning site were embedded[69]. Guide sequences were designed with the help of splashRNA[70] and cloned into the intron-embedded microRNA backbone with type IIS restriction enzymes. All constructs were generated using oligodeoxynucleotides (IDT) and gene fragments (gBlocks®, IDT).

KO of *MBNL1* and *MBNL2* with CRISPR/Cas9

To knock-out *MBNL1/2* in HEK293T cells, which carry a blasticidin resistance gene flanked by inteins within the *FOXP1* exon 18b, two plasmids expressing a mammalian codon-optimized SpyCas9-NLS and sgRNAs targeting *MBNL1* (G+TGGTGCCCCATTACAACCCG) and *MBNL2* (G+ TCAACCTGACAACTTTTGGG) were co-transfected. 72 h later, cells were replated in a proper format and medium were supplemented with indicated blasticidin concentration. The control sample was transfected with the same conditions, but with a sgRNA that is targeting the control locus *AAVS1* (*PPP1R12C*: GGGACCACCTTATATTCCCA). Genomic DNA was isolated from blasticidin-treated surviving colonies with Wizard® Genomic DNA Purification Kit (Promega) according to the manufacturer's protocol.

Proteinbiochemical analysis

Immunoblot analysis

Cells were lysed with a proper volume of M-PER (Thermo Fisher Scientific), including a protease inhibitor cocktail (Halt Protease Inhibitor Cocktail, Thermo Fisher Scientific), according to manufacturer's protocol. Cleared lysates were then equalized against the relative protein concentration determined using NanoDrop 1000 (Thermo Fisher Scientific) and diluted with M-PER. If indicated, samples were additionally dephosphorylated at 30 °C for 1 hour using λ (Lambda) protein phosphatase (NEB) according to the manufacturer's protocol. Equalized lysates were prepared for SDS-gel-electrophoresis using XT Sample Buffer (Bio-Rad) and XT Reducing Agent (Bio-Rad) and denatured at 70 °C for 10 min or 95 °C for 5 min. Samples were loaded in 18-well 4-12% Criterion™ XT Bis-Tris Protein Gel and electrophoresis were run at 150 V for 1.5 hours in XT MOPS Running Buffer (Bio-Rad). Subsequently, an immunoblot was performed onto an Immobilon®-P PVDF membrane (Merck) with a wet blotting

system (Criterion™ Blotter, Bio-Rad) in ice-cold Towbin buffer (Bio-Rad) with 20% Methanol (Carl Roth) overnight (15 V, 4 °C). Afterward, the free valences on the PVDF membrane were blocked in blocking buffer containing 5% skimmed milk (Carl Roth) in TBS-T (pH 7.6) with 0.1% Tween-20 (Sigma-Aldrich) at room temperature for 1 h. Antibodies were diluted at 1:1000 (only anti-pan-tau, PC1C6, Merck, was diluted 1:200) in blocking buffer and either incubated at room temperature for 2 hours or overnight at 4 °C, followed by at least 3 washing steps (room temperature, 5 min, 60 rpm). The HRP-conjugated secondary antibody (Abcam) was also diluted in blocking buffer (1:10 000 – 1:20 000) and subsequently again washed with TBS-T at least four times. HRP-detection was performed with the SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) on a Fusion FX7/SL advanced imaging system (Vilber Lourmat). The primary and secondary antibodies used were: mouse M2 anti-FLAG (Sigma-Aldrich), rat L6 anti-OLLAS (Thermo Fisher Scientific), mouse PC1C6 anti-pan-tau (Merck), mouse 8E6/C11 anti-3R-tau (Merck), rat EPR4114 anti-FOXP1 (abcam), mouse 32F6 anti-mNeonGreen (ChromoTek), rabbit anti-firefly luciferase (ab21176, abcam), rabbit D71G9 anti-TUBB3 (Cell Signaling Technology (CST)), mouse AC-15 anti- β-Actin (HRP) (abcam), goat anti-mouse IgG H&L (HRP) (ab97023, abcam), goat anti-rat IgG H&L (HRP) (ab97057, abcam) and goat anti-rabbit IgG H&L (HRP) (ab6721, abcam).

DNA analysis

To analyze the genome editing outcome of *MBNL1*/*MBNL2*, the intended genomic DSB region was PCR-amplified using LongAmp® Hot Start Taq 2x Master Mix (NEB) according to the manufacturer's protocol with primer deoxynucleotides pairs (IDT) flanking the expected mutation region. After DNA agarose gel electrophoresis, the PCR fragments were cut out and purified with Monarch® DNA Gel Extraction Kit (NEB). Indel analysis was performed using TIDE[77].

RNA-analysis

Semiquantitative RT-PCR

Cells were harvested for 5 min at 200×g, and the RNA was isolated with RNeasy Mini Kit (QIAGEN) according to the manufacturer's protocol. Reverse transcription was performed with SuperScript IV VILO Master Mix (Invitrogen™, Thermo Fisher Scientific) as described by the manufacturer's protocol, followed by a semi-quantitative PCR under non-saturating conditions using Q5 Polymerase (NEB).

Single-molecule fluorescence in-situ hybridization smFISH of *NEAT1*

HEK293T or its derived INSPECT clones were plated on 2-well µ-slides (Ibidi) 24 hours before fixation (300,000 in 1.2 ml medium). Before fixation, cells were washed with DPBS (Gibco™, Thermo Fisher Scientific) and fixed for 10 min in 10% neutral buffered formalin (Sigma-Aldrich). After further three DPBS washing steps à 5 min, the cells were permeabilized for either overnight hours at 4°C with ice-cold 70% ethanol or at RT for 1 hour. After three DPBS washing steps à 5 min, the samples were then incubated for 15 minutes with the hybridization buffer prepared with 2x saline sodium citrate (SSC) solution + 10% deionized formamide (Calbiochem®, Merck). Hybridization with Stellaris FISH probes was carried out in a total volume of 50 µl hybridization buffer containing 50 µg competitor tRNA from *E. coli* (Roche), 10% dextran sulfate (9011-18-1, VWR), 2 mg/ml UltraPure BSA (Thermo Fisher Scientific) and 10 mM ribonucleoside vanadyl complex (NEB) with probes in a final concentration of 1 ng/µl. The preparations were covered with parafilm and incubated at 37 °C for at least 5 hours or overnight, and then washed twice with 37 °C preheated 2xSCC + 10% deionized formamide at 37°C for 30 minutes. Finally, the preparations were washed twice with DPBS at RT and then mounted with 10 µl ProLong Glass Antifade Mountant with NucBlue Stain (Thermo Fisher Scientific). The probes were pre-designed by Biosearch Technologies and supplied by the same. The probes included were human *NEAT1* middle segment conjugated to

Quasar570® (SMF-2037-1, Biosearch Technologies) and human *NEAT1* 5'-segment conjugated to Quasar670® (VSMF-2247-5). The automated quantification of the hybridization signal was performed with the Fiji is just ImageJ (Fiji) software that includes the BioVoxxel toolbox plug-in.

Fluorescence, chemo/bioluminescence, and γ-scintillator detection

Immunofluorescence labeling

Medium from cells for immunofluorescence was removed and washed with DPBS (Gibco™, Thermo Fisher Scientific) and fixed for 15 min in 10% neutral buffered formalin (Sigma-Aldrich) at room temperature. Primary antibodies with indicated concentration were diluted 1:1000 in BSA blocking buffer (only anti-pan-tau, PC1C6, Merck, was diluted 1:200). Blocking buffer was prepared using DPBS (Gibco™, Thermo Fisher Scientific) with 1% BSA (Sigma-Aldrich) containing 0.5% Triton X-100 (Sigma-Aldrich). Cells were washed 3x after fixation with DPBS (Gibco™, Thermo Fisher Scientific) for 5 min at room temperature and blocking buffer containing the suitable fluorescent dye coupled secondary antibodies (1:1000, Thermo Fisher Scientific) were applied for 2 hours at room temperature or overnight at 4 °C. The primary and secondary antibodies used were: mouse PC1C6 anti-pan-tau (Merck), rabbit D71G9 anti-TUBB3 (Cell Signaling Technology (CST)), Cy3-conjugated cross-adsorbed goat anti-mouse IgG (H+L) (Thermo Fisher Scientific), Cy5-conjugated cross-adsorbed goat anti-mouse IgG (H+L) (Thermo Fisher Scientific) and Alexa Fluor 633-conjugated cross-adsorbed goat anti-rabbit IgG (H+L) (Thermo Fisher Scientific).

Epifluorescence microscopy

Epifluorescence microscopy was performed on an EVOS™ FL Auto Cell Imaging System (Invitrogen™, Thermo Fisher Scientific) under identical conditions for all samples across conditions.

Confocal microscopy

Confocal microscopy was conducted on a Leica SP5 system (Leica Microsystems). For life-imaging of cells, warm phenol-red-free DMEM/F12 supplemented with HEPES (Gibco™, Thermo Fisher Scientific) and 10% FBS was used as medium, and imaging was performed at 37 °C with 5% CO_2.

Bioluminescence microscopy

Bioluminescence life-imaging was performed on an LV200 bioluminescence imaging system (Olympus) in 8-well µ-slides (Ibidi). Transfection conditions of cells on 8-well µ-slides (Ibidi) were identical to 48-well plates. As a substrate solution for NLuc, the Nano-Glo® Live Cell Assay System (Promega) was used according to the manufacturer's protocol. Images were analyzed with Fiji ImageJ.

Bioluminescence quantification

For bioluminescence bulk quantifications, cells were plated and transfected in 96-well format. For bioluminescence detection of secreted/shedded NLuc, the supernatant was sampled (10 µl) at the indicated time points and detected using the Nano-Glo® Luciferase Assay System (Promega) on the Centro LB 960 (Berthold Technologies) plate reader with 0.1 s acquisition time. For dual-luciferase readout using Nano-Glo® Dual-Luciferase® Reporter Assay System (Promega), NLuc and FLuc signals were read out on-plate 72 hours post-transfection on a Centro LB 960 (Berthold Technologies) plate reader with 0.5 s acquisition time after 10 min of reagent 1 (ONE-Glo™ EX Luciferase) addition for FLuc and after 20 min of reagent 2 (NanoDLR™ Stop & Glo®) addition for NLuc which includes a FLuc inhibitor. The relative luminescence units (RLU) of FLuc and NLuc signals were expressed relative to the signals obtained after *MAPT* induction, defined as 1 to make the y-axis range consistent across experiments.

For live-cell NLuc quantification, Nano-Glo® Endurazine™ (Promega) Life Cell Substrates were used according to the manufacturer's protocol for a maximum duration of 70 hours.

γ-scintillator measurements

48 hours post-transfection with 600 ng plasmid DNA on 24-well plate format with 150,000 cells per well (in 1 ml medium), each well was incubated with 1 ml of medium containing 584 kBq/ml sodium iodide ($^{131}I^-$, day of usage was 48 d after calibration day). The cells were incubated for 2 min, 30 min, 60 min, and 180 min. After rinsing the wells three times with 1 ml PBS before, the cells were lysed in 1 ml NaOH, and quantification of the intracellularly accumulated $^{131}I^-$ was measured with a γ-scintillator (Wizard^2 1-Detector Gamma Counter, PerkinElmer) for 1 min.

Design considerations of EXSISERS constructs

All constructs were generated using gene fragments (gBlocks®) and oligodeoxynucleotides (IDT) and were codon-optimized for mammalian expression. If not otherwise noted, protein sequences are always mammalian codon-optimized and correspond to the WT proteins. The N-terminal starting methionine was not included in cases where it is not necessary for the protein's structural integrity or function.

EXSISERS$_{TMD\text{-}HaloTag}$

The type II TMD acts as a start-transfer sequence and translocates the subsequent binding moiety to the luminal/extracellular space, whereas the type I TMD stops the translocation process; the sequence before and after the type II and type I is therefore located in the cytosolic compartment, the sequence between the TMDs consequently in the extra-cytosolic compartments (**Figure 10b**). Two mutations (C61V and C262A) were additionally introduced to prevent the accidental formation of disulfide bonds in the oxidative extra-cytosolic compartments and thus might induce misfolding as previously shown for many fluorescent proteins[144]. The mouse Fcer2 membrane-spanning region was chosen as the type II TMD, and the flanking amino acids were also adopted since the N-terminally positively charged amino acids on the N-terminal (cytoplasmic) site ensure proper domain topology ('positive-inside rule') and two TMD-preceding palmitoylatable cysteines might also improve membrane association and topology. The human GYPA was used as the prototypical type-I TMD because of its positively charged amino acids C-terminally (cytosolic site) adjacent to the TMD. Additionally, G102L was introduced to disrupt the GxxxG TMD-dimerization motif, and a plasma membrane trafficking signal (PMTS) was additionally added[145].

EXSISERS$_{NLuc/FLuc}$

Here, the NanoLuc luciferase (NLuc) was flanked with P5/AP6 CCs and a gp41-1 split-intein pair, whereas EXSISERS$_{FLuc}$ was created by flanking a quintuple thermostable mutant of firefly luciferase (F14R, L35Q, V182K, I232K, F465R)[146,147] with another set of CCs (P3/AP4) and an NrdJ-1 split-intein pair.

EXSISERS$_{BSD}$

In this construct, flanking blasticidin-S deaminase with P5/AP6 CCs and a gp41-1 split-intein pair was used in analogy to the aforementioned constructs.

STATISTICS

Statistics were calculated with Prism 8 (Graphpad) as indicated with at least three biological triplicates.

REFERENCES

1. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).

2. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).

3. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).

4. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20–36 (2017).

5. Donnelly, M. L. *et al.* Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J. Gen. Virol.* **82**, 1013–1025 (2001).

6. Lee, K.-M., Chen, C.-J. & Shih, S.-R. Regulation Mechanisms of Viral IRES-Driven Translation. *Trends Microbiol.* **25**, 546–561 (2017).

7. Daguenet, E., Dujardin, G. & Valcárcel, J. The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. *EMBO Rep.* **16**, 1640–1655 (2015).

8. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics* **40**, 1413–1415 (2008).

9. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).

10. Van Driesche, S. J. & Martin, K. C. New frontiers in RNA transport and local translation in neurons. *Dev. Neurobiol.* **78**, 331–339 (2018).

11. Pichon, X., Lagha, M., Mueller, F. & Bertrand, E. A Growing Toolbox to Image Gene Expression in Single Cells: Sensitive Approaches for Demanding Challenges. *Mol. Cell* **71**, 468–480 (2018).

12. Jin, L., Zeng, X., Liu, M., Deng, Y. & He, N. Current progress in gene delivery technology based on chemical methods and nano-carriers. *Theranostics* **4**, 240–255 (2014).

13. Shi, J. *et al.* A Review on Electroporation-Based Intracellular Delivery. *Molecules* **23**, (2018).

14. Banasik, M. B. & McCray, P. B., Jr. Integrase-defective lentiviral vectors: progress and applications. *Gene Ther.* **17**, 150–157 (2010).

15. Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **18**, 358–378 (2019).

16. Cheng, J. K. & Alper, H. S. The genome editing toolbox: a spectrum of approaches for targeted modification. *Curr. Opin. Biotechnol.* **30**, 87–94 (2014).

17. Sung, P. & Klein, H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat. Rev. Mol. Cell Biol.* **7**, 739–750 (2006).

18. Khan, S. H. Genome-Editing Technologies: Concept, Pros, and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application. *Mol. Ther. Nucleic Acids* **16**, 326–334 (2019).

19. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).

20. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).

21. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).

22. Li, Y. *et al.* A versatile reporter system for CRISPR-mediated chromosomal rearrangements. *Genome Biol.* **16**, 111 (2015).

23. Zuo, Z. & Liu, J. Cas9-catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations. *Scientific Reports* **6**, (2016).

24. Ceccaldi, R., Rondinelli, B. & D'Andrea, A. D. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol.* **26**, 52–64 (2016).

25. Yan, W. X. *et al.* Functionally diverse type V CRISPR-Cas systems. *Science* **363**, 88–91 (2019).

26. Lange, A. *et al.* Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J. Biol. Chem.* **282**, 5101–5105 (2007).

27. Amara, S. G., Jonas, V., Rosenfeld, M. G., Ong, E. S. & Evans, R. M. Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. *Nature* **298**, 240–244 (1982).

28. Copp, D. H. & Cameron, E. C. Demonstration of a hypocalcemic factor (calcitonin) in commercial parathyroid extract. *Science* **134**, 2038 (1961).

29. Copp, D. H. & Cheney, B. Calcitonin—a Hormone from the Parathyroid which Lowers the Calcium-level of the Blood. *Nature* **193**, 381–382 (1962).

30. Brain, S. D., Williams, T. J., Tippins, J. R., Morris, H. R. & MacIntyre, I. Calcitonin gene-related peptide is a potent vasodilator. *Nature* **313**, 54–56 (1985).

31. Sample, P. J., Wang, B., Reid, D. W. & Presnyak, V. Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. *Nature* (2019).

32. Sharova, L. V. *et al.* Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res.* **16**, 45–58 (2009).

33. Kim, J. H. *et al.* High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One* **6**, e18556 (2011).

34. Varshavsky, A. N-degron and C-degron pathways of protein degradation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 358–366 (2019).

35. Abe, N. *et al.* Rolling Circle Translation of Circular RNA in Living Human Cells. *Sci. Rep.* **5**, 16435 (2015).

36. Keller, B.-M. *et al.* Chromobodies to Quantify Changes of Endogenous Protein Concentration in Living Cells. *Mol. Cell. Proteomics* **17**, 2518–2533 (2018).

37. Wu, S. C.-Y. *et al.* piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15008–15013 (2006).

38. Zhang, M. L., Lorson, C. L., Androphy, E. J. & Zhou, J. An in vivo reporter system for measuring increased inclusion of exon 7 in SMN2 mRNA: potential therapy of SMA. *Gene Ther.* **8**, 1532–1538 (2001).

39. Deshpande, A., Win, K. M. & Busciglio, J. Tau isoform expression and regulation in human cortical neurons. *FASEB J.* **22**, 2357–2367 (2008).

40. Stoilov, P., Lin, C.-H., Damoiseaux, R., Nikolic, J. & Black, D. L. A high-throughput screening strategy identifies cardiotonic steroids as alternative splicing modulators. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 11218–11223 (2008).

41. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front. Endocrinol.* **9**, 402 (2018).

42. Matsufuji, S. *et al.* Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell* **80**, 51–60 (1995).

43. Anderson, P. & Kedersha, N. RNA granules. *J. Cell Biol.* **172**, 803–808 (2006).

44. Baser, A. *et al.* Onset of differentiation is post-transcriptionally controlled in adult neural stem cells. *Nature* **566**, 100–104 (2019).

45. Carvajal-Vallejos, P., Pallissé, R., Mootz, H. D. & Schmidt, S. R. Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. *J. Biol. Chem.* **287**, 28686–28696 (2012).

46. Aranko, A. S., Wlodawer, A. & Iwaï, H. Nature's recipe for splitting inteins. *Protein Eng. Des. Sel.* **27**, 263–271 (2014).

47. Mietelska-Porowska, A., Wasik, U., Goras, M., Filipek, A. & Niewiadomska, G. Tau protein modifications and interactions: their role in function and dysfunction. *Int. J. Mol. Sci.* **15**, 4671–4713 (2014).

48. Adams, S. J., DeTure, M. A., McBride, M., Dickson, D. W. & Petrucelli, L. Three repeat isoforms of tau inhibit assembly of four repeat tau filaments. *PLoS One* **5**, e10810 (2010).

49. Panda, D., Samuel, J. C., Massie, M., Feinstein, S. C. & Wilson, L. Differential regulation of microtubule dynamics by three- and four-repeat tau: implications for the onset of neurodegenerative disease. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9548–9553 (2003).

50. Wang, Y. & Mandelkow, E. Tau in physiology and pathology. *Nat. Rev. Neurosci.* **17**, 5–21 (2016).

51. Hall, M. P. *et al.* Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol.* **7**, 1848–1857 (2012).

52. Gradišar, H. *et al.* Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat. Chem. Biol.* **9**, 362–366 (2013).

53. Baker, M. *et al.* Localization of frontotemporal dementia with parkinsonism in an Australian kindred to chromosome 17q21-22. *Ann. Neurol.* **42**, 794–798 (1997).

54. Hutton, M. *et al.* Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* **393**, 702–705 (1998).

55. Sposito, T. *et al.* Developmental regulation of tau splicing is disrupted in stem cell-derived neurons from frontotemporal dementia patients with the 10 + 16 splice-site mutation in MAPT. *Hum. Mol. Genet.* **24**, 5260–5269 (2015).

56. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).

57. Grover, A. *et al.* 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J. Biol. Chem.* **274**, 15134–15143 (1999).

58. Connell, J. W. *et al.* Quantitative analysis of tau isoform transcripts in sporadic tauopathies. *Brain Res. Mol. Brain Res.* **137**, 104–109 (2005).

59. Walte, A. *et al.* Mechanism of dual specificity kinase activity of DYRK1A. *FEBS J.* **280**, 4495–4511 (2013).

60. Dirice, E. *et al.* Inhibition of DYRK1A Stimulates Human β-Cell Proliferation. *Diabetes* **65**, 1660–1671 (2016).

61. Massillon, D., Stalmans, W., van de Werve, G. & Bollen, M. Identification of the glycogenic compound 5-iodotubercidin as a general protein kinase inhibitor. *Biochem. J* **299 (Pt 1)**, 123–128 (1994).

62. Qian, W. *et al.* Regulation of the alternative splicing of tau exon 10 by SC35 and Dyrk1A. *Nucleic Acids Res.* **39**, 6161–6171 (2011).

63. Yin, X. *et al.* Dyrk1A overexpression leads to increase of 3R-tau expression and cognitive deficits in Ts65Dn Down syndrome mice. *Sci. Rep.* **7**, 619 (2017).

64. Hernández, F. *et al.* Glycogen synthase kinase-3 plays a crucial role in tau exon 10 splicing and intranuclear distribution of SC35. Implications for Alzheimer's disease. *J. Biol. Chem.* **279**, 3801–3806 (2004).

65. Shaner, N. C. *et al.* A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat. Methods* **10**, 407–409 (2013).

66. Konermann, S. *et al.* Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665–676.e14 (2018).

67. Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* **358**, 1019–1027 (2017).

68. Zhang, C. *et al.* Structural Basis for the RNA-Guided Ribonuclease Activity of CRISPR-Cas13d. *Cell* **175**, 212–223.e17 (2018).

69. Fellmann, C. *et al.* An optimized microRNA backbone for effective single-copy RNAi. *Cell Rep.* **5**, 1704–1713 (2013).

70. Pelossof, R. *et al.* Prediction of potent shRNAs with a sequential classification algorithm. *Nat. Biotechnol.* **35**, 350–353 (2017).

71. Wang, Y., Cheong, C.-G., Hall, T. M. T. & Wang, Z. Engineering splicing factors with designed specificities. *Nat. Methods* **6**, 825–830 (2009).

72. Philipps, D., Celotto, A. M., Wang, Q.-Q., Tarng, R. S. & Graveley, B. R. Arginine/serine repeats are sufficient to constitute a splicing activation domain. *Nucleic Acids Res.* **31**, 6502–6508 (2003).

73. Los, G. V. *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).

74. Kimura, M., Takatsuki, A. & Yamaguchi, I. Blasticidin S deaminase gene from Aspergillus terreus (BSD): a new drug resistance gene for transfection of mammalian cells. *Biochim. Biophys. Acta* **1219**, 653–659 (1994).

75. Gabut, M. *et al.* An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* **147**, 132–146 (2011).

76. Han, H. *et al.* MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* **498**, 241–245 (2013).

77. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).

78. Trabzuni, D. *et al.* MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies. *Hum. Mol. Genet.* **21**, 4094–4103 (2012).

79. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).

80. Dhanoa, J. K., Sethi, R. S., Verma, R., Arora, J. S. & Mukhopadhyay, C. S. Long non-coding RNA: its evolutionary relics and biological implications in mammals: a review. *Hanguk Tongmul Chawon Kwahakhoe Chi* **60**, 25 (2018).

81. Hüttenhofer, A., Schattner, P. & Polacek, N. Non-coding RNAs: hope or hype? *Trends Genet.* **21**, 289–297 (2005).

82. Kung, J. T. Y., Colognori, D. & Lee, J. T. Long noncoding RNAs: past, present, and future. *Genetics* **193**, 651–669 (2013).

83. Kopp, F. & Mendell, J. T. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* **172**, 393–407 (2018).

84. Salviano-Silva, A., Lobo-Alves, S. C., Almeida, R. C. de, Malheiros, D. & Petzl-Erler, M. L. Besides Pathology: Long Non-Coding RNA in Cell and Tissue Homeostasis. *Noncoding RNA* **4**, (2018).

85. 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).

86. Ji, Y., Wang, M., Li, X. & Cui, F. The Long Noncoding RNA NEAT1 Targets miR-34a-5p and Drives Nasopharyngeal Carcinoma Progression via Wnt/β-Catenin Signaling. *Yonsei Med. J.* **60**, 336–345 (2019).

87. Bao, Z. *et al.* LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.* **47**, D1034–D1037 (2019).

88. Schmitt, A. M. & Chang, H. Y. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* **29**, 452–463 (2016).

89. Yu, B. & Shan, G. Functions of long noncoding RNAs in the nucleus. *Nucleus* **7**, 155–166 (2016).

90. Jansen, R. P. mRNA localization: message on the move. *Nat. Rev. Mol. Cell Biol.* **2**, 247–256 (2001).

91. Tutucci, E. *et al.* An improved MS2 system for accurate reporting of the mRNA life cycle. *Nat. Methods* **15**, 81–89 (2018).

92. Hubé, F. & Francastel, C. Mammalian introns: when the junk generates molecular diversity. *Int. J. Mol. Sci.* **16**, 4429–4452 (2015).

93. Carmody, S. R. & Wente, S. R. mRNA nuclear export at a glance. *J. Cell Sci.* **122**, 1933–1937 (2009).

94. Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763–776 (2009).

95. Kozak, M. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**, 1109–1123 (1978).

96. Kozak, M. The scanning model for translation: an update. *J. Cell Biol.* **108**, 229–241 (1989).

97. Pollard, V. W. & Malim, M. H. The HIV-1 Rev protein. *Annu. Rev. Microbiol.* **52**, 491–532 (1998).

98. Cullen, B. R. Nuclear mRNA export: insights from virology. *Trends Biochem. Sci.* **28**, 419–424 (2003).

99. Pasquinelli, A. E. *et al.* The constitutive transport element (CTE) of Mason-Pfizer monkey virus (MPMV) accesses a cellular mRNA export pathway. *EMBO J.* **16**, 7500–7510 (1997).

100. Braun, I. C., Rohrbach, E., Schmitt, C. & Izaurralde, E. TAP binds to the constitutive transport element (CTE) through a novel RNA-binding motif that is sufficient to promote CTE-dependent RNA export from the nucleus. *EMBO J.* **18**, 1953–1965 (1999).

101. Teplova, M., Wohlbold, L., Khin, N. W., Izaurralde, E. & Patel, D. J. Structure-function studies of nucleocytoplasmic transport of retroviral genomic RNA by mRNA export factor TAP. *Nat. Struct. Mol. Biol.* **18**, 990–998 (2011).

102. Donello, J. E., Loeb, J. E. & Hope, T. J. Woodchuck hepatitis virus contains a tripartite posttranscriptional regulatory element. *J. Virol.* **72**, 5085–5092 (1998).

103. Oh, T., Bajwa, A., Jia, G. & Park, F. Lentiviral vector design using alternative RNA export elements. *Retrovirology* **4**, 38 (2007).

104. Popa, I., Harris, M. E., Donello, J. E. & Hope, T. J. CRM1-dependent function of a cis-acting RNA export element. *Mol. Cell. Biol.* **22**, 2057–2067 (2002).

105. Shatsky, I. N., Dmitriev, S. E., Terenin, I. M. & Andreev, D. E. Cap- and IRES-independent scanning mechanism of translation initiation as an alternative to the concept of cellular IRESs. *Mol. Cells* **30**, 285–293 (2010).

106. Leppek, K., Das, R. & Barna, M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.* **19**, 158–174 (2018).

107. Tomek, W. & Wollenhaupt, K. The 'closed loop model' in controlling mRNA translation during development. *Anim. Reprod. Sci.* **134**, 2–8 (2012).

108. Vicens, Q., Kieft, J. S. & Rissland, O. S. Revisiting the Closed-Loop Model and the Nature of mRNA 5′–3′ Communication. *Mol. Cell* **72**, 805–812 (2018).

109. Stern, B., Olsen, L. C., Trösße, C., Ravneberg, H. & Pryme, I. F. Improving mammalian cell factories: The selection of signal peptide has a major impact on recombinant protein synthesis and secretion in mammalian cells. *Trends Cell Mol. Biol.* **2**, 1–17 (2007).

110. Trösse, C., Ravneberg, H., Stern, B. & Pryme, I. F. Vectors encoding seven oikosin signal peptides transfected into CHO cells differ greatly in mediating Gaussia luciferase and human endostatin production although mRNA levels are largely unaffected. *Gene Regul. Syst. Bio.* **1**, 303–312 (2007).

111. Chamond, N., Deforges, J., Ulryck, N. & Sargueil, B. 40S recruitment in the absence of eIF4G/4A by EMCV IRES refines the model for translation initiation on the archetype of Type II IRESs. *Nucleic Acids Res.* **42**, 10373–10384 (2014).

112. Bochkov, Y. A. & Palmenberg, A. C. Translational efficiency of EMCV IRES in bicistronic vectors is dependent upon IRES sequence and gene location. *Biotechniques* **41**, 283–4, 286, 288 passim (2006).

113. Darrouzet, E., Lindenthal, S., Marcellin, D., Pellequer, J.-L. & Pourcher, T. The sodium/iodide symporter: state of the art of its molecular characterization. *Biochim. Biophys. Acta* **1838**, 244–253 (2014).

114. Schmohl, K. A. *et al.* Imaging and targeted therapy of pancreatic ductal adenocarcinoma using the theranostic sodium iodide symporter (NIS) gene. *Oncotarget* **8**, 33393–33404 (2017).

115. Penheiter, A. R., Russell, S. J. & Carlson, S. K. The Sodium Iodide Symporter (NIS) as an Imaging Reporter for Gene, Viral, and Cell-based Therapies. *Current Gene Therapy* **12**, 33–47 (2012).

116. Sojka, D. K., Bruniquel, D., Schwartz, R. H. & Singh, N. J. IL-2 secretion by CD4+ T cells in vivo is rapid, transient, and influenced by TCR-specific competition. *J. Immunol.* **172**, 6136–6143 (2004).

117. Carswell, S. & Alwine, J. C. Efficiency of utilization of the simian virus 40 late polyadenylation site: effects of upstream sequences. *Mol. Cell. Biol.* **9**, 4248–4258 (1989).

118. Lanoix, J. & Acheson, N. H. A rabbit beta-globin polyadenylation signal directs efficient termination of transcription of polyomavirus DNA. *EMBO J.* **7**, 2515–2522 (1988).

119. Levitt, N., Briggs, D., Gil, A. & Proudfoot, N. J. Definition of an efficient synthetic poly(A) site. *Genes Dev.* **3**, 1019–1025 (1989).

120. Kessler, M. M., Beckendorf, R. C., Westhafer, M. A. & Nordstrom, J. L. Requirement of A-A-U-A-A-A and adjacent downstream sequences for SV40 early polyadenylation. *Nucleic Acids Res.* **14**, 4939–4952 (1986).

121. Araki, K., Araki, M. & Yamamura, K.-I. Site-directed integration of the cre gene mediated by Cre recombinase using a combination of mutant lox sites. *Nucleic Acids Res.* **30**, e103 (2002).

122. Schnütgen, F. *et al.* A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. *Nat. Biotechnol.* **21**, 562–565 (2003).

123. Carmo-Fonseca, M. & Kirchhausen, T. The timing of pre-mRNA splicing visualized in real-time. *Nucleus* **5**, 11–14 (2014).

124. Beyer, A. L. & Osheim, Y. N. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* **2**, 754–765 (1988).

125. Takata, Y., Kondo, S., Goda, N., Kanegae, Y. & Saito, I. Comparison of efficiency between FLPe and Cre for recombinase-mediated cassette exchange in vitro and in adenovirus vector production: RMCE efficiency of FLPe and Cre. *Genes Cells* **16**, 765–777 (2011).

126. Balzarini, J. *et al.* Engineering of a single conserved amino acid residue of herpes simplex virus type 1 thymidine kinase allows a predominant shift from pyrimidine to purine nucleoside phosphorylation. *J. Biol. Chem.* **281**, 19273–19279 (2006).

127. Modic, M. *et al.* Cross-Regulation between TDP-43 and Paraspeckles Promotes Pluripotency-Differentiation Transition. *Mol. Cell* **74**, 951–965.e13 (2019).

128. Yamazaki, T. *et al.* Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Mol. Cell* **70**, 1038–1053.e7 (2018).

129. Lellahi, S. M. *et al.* The long noncoding RNA NEAT1 and nuclear paraspeckles are up-regulated by the transcription factor HSF1 in the heat shock response. *J. Biol. Chem.* **293**, 18965–18976 (2018).

130. Imamura, K. *et al.* Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol. Cell* **53**, 393–406 (2014).

131. Adriaens, C. *et al.* p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat. Med.* **22**, 861–868 (2016).

132. Beeharry, Y., Goodrum, G., Imperiale, C. J. & Pelchat, M. The Hepatitis Delta Virus accumulation requires paraspeckle components and affects NEAT1 level and PSP1 localization. *Sci. Rep.* **8**, 6031 (2018).

133. Choudhry, H. *et al.* Tumor hypoxia induces nuclear paraspeckle formation through HIF-2α dependent transcriptional activation of NEAT1 leading to cancer cell survival. *Oncogene* **34**, 4546 (2015).

134. Ma, H. *et al.* The long noncoding RNA NEAT1 exerts antihantaviral effects by acting as positive feedback for RIG-I signaling. *Journal of* (2017).

135. Zhang, Q., Chen, C.-Y., Yedavalli, V. S. R. K. & Jeang, K.-T. NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *MBio* **4**, e00596–12 (2013).

136. Le Hir, H., Gatfield, D., Izaurralde, E. & Moore, M. J. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20**, 4987–4997 (2001).

137. Wolff, J., Chaikoff, I. L., Goldberg, R. C. & Meier, J. R. THE TEMPORARY NATURE OF THE INHIBITORY ACTION OF EXCESS IODIDE ON ORGANIC IODINE SYNTHESIS IN THE NORMAL THYROID. *Endocrinology* **45**, 504–513 (1949).

138. Lv, J. *et al.* A Novel Ideal Radionuclide Imaging System for Non-invasively Cell Monitoring built on Baculovirus Backbone by Introducing Sleeping Beauty Transposon. *Sci. Rep.* **7**, 43879 (2017).

139. Zhou, X., Li, B., Wang, J., Yin, H. & Zhang, Y. The feasibility of using a baculovirus vector to deliver the sodium-iodide symporter gene as a reporter. *Nucl. Med. Biol.* **37**, 299–308 (2010).

140. Canny, M. D. *et al.* Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat. Biotechnol.* **36**, 95–102 (2018).

141. Tsutsui, H. *et al.* A diffraction-quality protein crystal processed as an autophagic cargo. *Mol. Cell* **58**, 186–193 (2015).

142. Kiani, S. *et al.* Cas9 gRNA engineering for genome editing, activation and repression. *Nat. Methods* **12**, 1051–1054 (2015).

143. Dahlman, J. E. *et al.* Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat. Biotechnol.* **33**, 1159–1161 (2015).

144. Costantini, L. M. *et al.* A palette of fluorescent proteins optimized for diverse cellular environments. *Nat. Commun.* **6**, 7670 (2015).

145. Gradinaru, V. *et al.* Molecular and cellular approaches for diversifying and extending optogenetics. *Cell* **141**, 154–165 (2010).

146. Law, G. H., Gandelman, O. A., Tisi, L. C., Lowe, C. R. & Murray, J. ALTERING THE SURFACE HYDROPHOBICITY OF FIREFLY LUCIFERASE. in *Bioluminescence and Chemiluminescence* 37–40 (WORLD SCIENTIFIC, 2002).

147. Law, G. H. E., Gandelman, O. A., Tisi, L. C., Lowe, C. R. & Murray, J. A. H. Mutagenesis of solvent-exposed amino acids in Photinus pyralis luciferase improves thermostability and pH-tolerance. *Biochem. J* **397**, 305–312 (2006).

ACKNOWLEDGMENTS

I want to thank Prof. Gil Gregor Westmeyer for the supervision of my thesis, especially for the discussions, the freedom and support to realize my ideas. I would also like to thank Prof. Wolfgang Wurst and Dr. Florian Giesert for the exciting discussions and helpful remarks during my thesis. Many thanks to Prof. Stefan Engelhardt for his participation as the chairman of my PhD-thesis defense. I would also like to thank Anja Stelzl and Hannes Rolbieski for keeping the laboratory intact and ensuring that everything runs smoothly. Many thanks also to the remaining members of the AG Westmeyer for the pleasant working atmosphere. Many thanks also go to the members of the Institute for Developmental Genetics, in particular, Christoph Gruber and Jessica Schwab for numerous helpful discussions. Special thanks are also to my hard-working students, who have actively helped with all projects, without whom, I would be screwed, and special thanks to the ones, who contributed with their own ideas: Teeradon Phlairaharn, Christoph Gruber, Deniz Tümen, Enikő Baligács, Niklas Armbrust, Maren Beyer, Tobias Santl, Eva-Maria Lederer, Julian Geilenkeuser, Simone Göppert, Luisa Krumwiede, Gerald Raffl, Milica Živanić, Eva Magdalena Beck, Christian Grätz, Dominic Schwarz, Martin Zirngibl, Johann Dietmar Körner, Franziska Winzig and Francesco Vaccaro. Many thanks also to Markus Grosch, who helped me out with the smFISH protocols for *NEAT1*. I also want to thank Volker Morath, who helped out with the NIS experiments. Thanks for the many cool and nerdy discussions with Artem Romanov and Valentin Evsyukov. I also want to thank my parents Yi-Xuan Zhang and So-Trung Truong, and my sister Dong-Shiow Jenny Truong for supporting me throughout my life. Finally, I want to thank Leonie Anna Reichart for the many discussions, the readings and suggestions to my manuscripts, and especially for supporting me in all matters of life in her very own charming way.