Inaugural dissertation
for
obtaining the doctoral degree
of the
Combined Faculty of Mathematics, Engineering and Natural Sciences
of the
Ruprecht - Karls - University
Heidelberg

Presented by M.Sc. Julia (Iuliia) Kornienko
born in: Saint Petersburg, Russia
Oral examination: 26th of September 2023

RBM20 mislocalization in dilated cardiomyopathy

Referees:
Prof. Dr. Michael Boutros
Dr. Kyung-Min Noh

To twelve-year old me, who once said:
“What is the point of all this, I will never become a scientist anyway”.

Acknowledgments

First, I would like to say thanks to my mentor – Lars Steinmetz. You are a lot of inspiration and I learned so much from you. I thank you for having me in the lab and teaching me to think outside the box. Thank you also for all the fun lab activities and keeping the lab atmosphere so great.

Then, many thanks go to my amazing post and present lab mates. Thank you, Marta Rodríguez-Martínez for the wonderful times working together – my PhD project would not have been the same without you, and I will miss working with you. Thank you, Markus Grosch for helping me with early stages of my project, and you, dear Kai Fenzl, for helping me with the late stages of my PhD and for forming the lunchy lunch club. Thanks to my first master student – Brigit Tunaj, for the great work. Thanks to everyone who helped me with the project and/or just made the lab days so enjoyable, including but not limited to Melanie Krausse, Anastasiia Korosteleva, Linda Müller, John Hawkins, Daniel Schraivogel, Annique Claringbould, Sandra Clauder-Münster, Shengdi Li, Balint Csorgo, Benjamin Story, Katie Zeier, Benedikt Rauscher, and many others. And thank you, Amanda Hughes, for all your delicious cakes. It has been a pleasure working with you all and getting to know you, also outside the lab. You all are the best colleagues one could think of!

Thanks to all the EMBL Core Facilities for the excellent scientific services, support and helpful discussions. Especially I would like to thank Diana Ordonez, Beata Ramasz, Daniel Gimenes, and Michael Bonadonna from the Flow Cytometry Core Facility; Frank Stein and Per Haberkant from the Proteomics Core Facility; Ferris Jung, Laura Villacorta, Jessica Lynn Sutter, and Vladimir Benes from the Genomics Core Facility; Ernesto de la Cueva and Frank Diego Montoya Castillo from the Laboratory Animal Resources; Aliaksandr Halavatyi, and Arif Khan from the Advanced Light Microscopy Core Facility.

Thanks to my TAC members Kyung-Min Noh, Judith Zaugg, and Michael Boutros, for their valuable input and feedback on my project, as well as for helpful discussions. Thanks to our collaborators – Maarten van den Hoogenhof, Michael Gotthardt, Victoria Parikh, and others who contributed into this project.

Next, I would like to say special thanks to my beloved friends. I have met so many incredible people before and during my PhD with whom I share so many happy memories. I thank Nastya Zaikina – my oldest friend with whom I was lucky to spend high school years which were really fun. I thank my university friends – Natasha Rodina, Ira Utkina, and many others, who shared with me the tough days of studying higher maths and quantum physics and made them (almost) enjoyable. I thank my friends that I met during the time in Heidelberg: Dewi Moonen, Dominik Lindenhofer, and Laura Schraft who I am also lucky to have as my lab mates; my former roommate Karel Mocaer; Habibi Ahmad Al Alwash, Joschy Hey, Anita Ray, Antonia Hauth, Luca Marconato, Vera Janssen, Jess Bourn, Ale Dulja, Ana Kojic, Karen Grimes, Sergio Triana, Clement Potel, Linda Schatz, Agnes Walk, Donnacha Fitzgerald, Luoyan Yong, Martin Larralde, and the rest of the EMBL PhD cohort 2019, as well as many more lovely people.

Next, I would like to thank my sweetest partner Björn Arnold. Thank you for always being there for me during the last, and the most stressful part of my PhD. Thanks for making me happy every day. You are my joy and support, and I am so lucky to having met you, love!

Last but not least, I would like to thank my loving family. I cannot describe how grateful I am for having you supporting me throughout my life. Thank you, mum, for teaching me how to believe in myself. Having you as an inspiring example, I was always enjoying studying. Thanks to you, I am where I am now, and I am endlessly thankful that you invested so much in me. I thank my grandmother for teaching me that hard work and discipline always pay off, and my grandad for his joyful and humorous attitude to life. I thank my dad for inspiring and motivating me to pursue academic career, and for always believing in me. I thank my step dad for always being there for me and my mum, and for all of his adorable grumpiness. I thank my great grandmother who took care of babysitting me when both my mum and my grandparents had to work hard to ensure I can live happily and get good education. I am so lucky to have such strong women as my mum, grandma, and great grandma in my family as role models.

Thank you all for the great journey!

Summary

Protein mislocalization and aggregation is at the core of several pathologies, including neurological diseases and cardiomyopathies. Up to the date, it has been challenging to systematically study these phenotypes due to the lack of a technology allowing to distinguish healthy from pathological phenotypes in a high-throughput manner. The recent development of the high-throughput image-enabled cell sorting technology (ICS, Schraivogel et al., *Science,* 2022) has opened the door to such studies. In my PhD project, I applied the ICS technology for the first time to decipher the pathobiology of a human disease.

Dilated cardiomyopathy (DCM) is amongst the most common causes of heart failure. However, targeted therapy approaches to treat DCM patients are still largely absent in clinics. Mutations in the heart-specific alternative splicing (AS) regulator RBM20 are usually associated with a particularly severe form of the disease in patients. All identified pathogenic RBM20 mutations result in loss of RBM20's splicing function. In addition, a subset of mutations located in the protein's RS-rich region is linked to cytoplasmic mislocalization. This aberrant cytoplasmic localization has been shown to worsen the disease phenotype, leading to a more aggressive form of DCM in vivo. To date, it has been unclear why these mutations lead to cytoplasmic mislocalization, and whether relocalizing these variants to the nucleus may be beneficial. Here, I combined the ICS technology with RNA-sequencing, CRISPR screens and other functional studies to decipher the mechanism of pathological RS-domain RBM20 variants. This led to the identification of a potential therapeutic target and strategy which I then and tested in vivo.

I found that, unlike other known mutations in the RS-domain, a newly identified P633L substitution leads to a mixed mislocalization pattern, with a significant fraction of the protein still localizing to the nucleus. Using the ICS technology together with RNA-sequencing, I showed that nuclear-localised RBM20-P633L functions similar to the wild type (WT) in iPSC-derived cardiomyocytes (iPSC-CMs), displaying functional AS activity. Forcing nuclear localization of RBM20 harbouring another, more severely mislocalizing pathological mutation in the RS-domain (R634Q) by fusing it with the SV40 NLS sequence, rescued splicing of RBM20 targets. In addition, protein

interactors were mainly unchanged for nuclear-localised RBM20-WT, –P633L, or -R634Q in HeLa cells. These results indicate that RS-domain mutations do not affect RBM20 splicing activity but rather affect its ability to be efficiently imported to the nucleus.

To identify factors responsible for RBM20 nuclear import in an unbiased way, I performed two genome-wide CRISPR-based screens with ICS-measured RBM20-WT or -R634Q localization as readouts in HeLa cells. I identified TNPO3 to be the main nuclear importer of RBM20. I showed that, loss of TNPO3 resulted in cytoplasmic mislocalization of RBM20-WT protein in iPSC-CMs, similar to what is observed in the presence of RS-domain mutations. This once more indicates that the pathogenic phenotype of these variants depends mainly on cellular localization. I, together with Dr. Marta Rodríguez-Martínez, then used biochemistry and mass spectrometry to show that RBM20 interacts with TNPO3 and that their interaction is disrupted by RS-domain mutations. More severely mislocalizing mutant proteins lost the interaction the most. Notably, since TNPO3 is still able to bind RS-domain mutated RBM20 to some extent, I hypothesised that it may serve as an anchor for developing therapeutic strategies targeted to stabilize this interaction and restore the nuclear import.

To test this hypothesis, I first delivered a full-length *TNPO3* to iPSC-CMs with RS-domain mutant RBM20. I showed that it restored nuclear localization and proportionally restored splicing of RBM20 targets. Next, I tested AAv9-*Tnpo3* delivery in mouse models mimicking patients' mutations P633L and R634Q. I found that *Tnpo3* overexpression in vivo restored alternative splicing of RBM20 targets, proportionally restoring cardiac ejection fraction, without causing systemic changes in gene expression or alternative splicing. These results demonstrate that, enhancing nuclear import of RBM20 variants either by up-regulating TNPO3, or by other means, can serve as a promising therapeutic strategy.

Altogether, in this PhD project, I found the molecular mechanism responsible of RBM20 nuclear transport and deciphered the cause of RBM20 mislocalisation in DCM. In addition, I showed that restoring nuclear localization of RBM20 variants presents a viable strategy to rescue both disease phenotypes: dissolve detrimental cytoplasmic granules and restore splicing deficiency.

Zusamennfassung

Die falsche Lokalisierung und Aggregation von Proteinen steht im Mittelpunkt verschiedener Krankheiten, einschließlich neurologischer Erkrankungen und Kardiomyopathien. Bislang war es schwierig, diese Phänotypen systematisch zu untersuchen, da es keine Technologie gab, mit der sich gesunde von pathologischen Phänotypen im Hochdurchsatzverfahren unterscheiden ließen. Die jüngste Entwicklung der bildgestützten Hochdurchsatz-Zellsortierungstechnologie (image-enabled cell sorting (ICS), Schraivogel et al., Science, 2022) hat die Tür zu solchen Studien geöffnet. In meinem Promotionsprojekt habe ich die ICS-Technologie zum ersten Mal eingesetzt, um die Pathobiologie einer menschlichen Krankheit zu entschlüsseln.

Die dilatative Kardiomyopathie (DCM) gehört zu den häufigsten Ursachen von Herzversagen, doch fehlt es in den Kliniken noch weitgehend an gezielten Therapieansätzen zur Behandlung von DCM-Patienten. Mutationen im herzspezifischen alternativen Spleißregulator (AS) RBM20 sind in der Regel mit einer besonders schweren Form der Erkrankung bei Patienten verbunden. Alle identifizierten pathogenen RBM20-Mutationen führen zu einem Verlust der Spleißfunktion von RBM20. Darüber hinaus ist eine Untergruppe von Mutationen in der RS-reichen Region des Proteins mit einer zytoplasmatischen Fehllokalisierung verbunden, die zu einem ungewollten Funktionsgewinn von RBM20 in Zytoplasma führt. Diese abnormale zytoplasmatische Lokalisierung verschlimmert nachweislich den Krankheitsphänotyp und führt zu einer aggressiveren Form von DCM in vivo. Bislang war jedoch unklar, warum diese Mutationen zu einer zytoplasmatischen Fehllokalisierung führen und ob eine Relokalisierung dieser Varianten von Vorteil sein könnte. Ich habe hier die ICS-Technologie mit RNA-Sequenzierung, CRISPR-Screens und anderen funktionellen Studien kombiniert, um den Mechanismus dieser pathologischen RBM20 Varianten zu entschlüsseln und ein dadurch identifiziertes therapeutisches Ziel in vivo zu testen.

Ich habe herausgefunden, dass im Gegensatz zu anderen bekannten Mutationen in der RS-Domäne eine neu identifizierte P633L-Substitution zu einem gemischten Fehllokalisierung führt, wobei ein erheblicher Teil des Proteins immer noch im

Zellkern lokalisiert ist. Mit Hilfe der ICS-Technologie und der RNA-Sequenzierung konnte ich zeigen, dass das im Zellkern lokalisierte RBM20-P633L in iPSC-Kardiomyozyten (iPSC-CMs) ähnlich funktioniert wie der Wildtyp (WT) und eine funktionelle AS-Aktivität zeigt. Die Erzwingung der nukleären Lokalisierung durch Fusion mit der SV40 NLS-Sequenz von RBM20 mit einer anderen, stärker fehllokalisierenden pathologischen Mutation in der RS-Domäne (R634Q) führte zu einer deutlichen Rückgewinnung der Spleiß-Aktivität von RBM20. Darüber hinaus waren die Protein-Protein-Interaktoren für kernlokalisiertes RBM20-WT, -P633L oder -R634Q in HeLa-Zellen weitgehend identisch. Diese Ergebnisse deuten darauf hin, dass Mutationen der RS-Domäne nicht die Spleißaktivität von RBM20 beeinträchtigen, sondern eher seine Fähigkeit, effizient in den Zellkern importiert zu werden.

Um Faktoren, die für den RBM20-Kernimport verantwortlich sind, unvoreingenommen zu identifizieren, habe ich zwei genomweite CRISPR-basierte Screens mit ICS-gemessener RBM20-WT- oder -R634Q-Lokalisierung als Messwerte in HeLa-Zellen durchgeführt. Ich konnte TNPO3 als den wichtigsten Kernimporter von RBM20 identifizieren. Ich konnte zeigen, dass der Verlust von TNPO3 zu einer zytoplasmatischen Fehllokalisierung des RBM20-WT-Proteins in iPSC-CMs führt, ähnlich wie es bei RS-Domänen-Mutationen beobachtet wurde. Dies zeigt einmal mehr, dass der pathogene Phänotyp hauptsächlich von der zellulären Lokalisierung abhängt. Zusammen mit Dr. Marta Rodríguez-Martínez habe ich mit Hilfe von biochemischen Verfahren und Massenspektrometrie gezeigt, dass RBM20 mit TNPO3 interagiert und dass diese Interaktion durch RS-Domänen-Mutationen gestört wird. Die am stärksten fehllokalisierende RBM20 Mutantenproteine zeigten auch gleichzeitig den stärksten Verlust der Interaktion mit TNPO3. Da TNPO3 immer noch in der Lage ist, mutiertes RBM20 mit RS-Domäne bis zu einem gewissen Grad zu binden, stellte ich die Hypothese auf, dass es als Startpunkt für die Entwicklung therapeutischer Strategien dienen könnte, die darauf abzielen, diese Interaktion zu stabilisieren und den Kernimport wiederherzustellen.

Um diese Hypothese zu testen, habe ich zunächst TNPO3 in iPSC-CMs eingebracht, welche RBM20 Varianten mit RS-Domän Mutationen aufweisen. Ich konnte zeigen, dass die zusätzliche Verfügbarkeit von TNPO3 in diesen Zellen die nukleäre

Lokalisierung wiederherstellte und zusätzlich auch die Spleiß-Aktivität von RBM20 proportional anstieg. Als nächstes testete ich die Verabreichung von AAv9-Tnpo3 in Mausmodellen, die die Patientenmutationen P633L und R634Q nachahmen. Ich fand heraus, dass die Überexpression von Tnpo3 in vivo das alternative Spleißen von unterschiedlichen RBM20-Zielen wiederherstellte und die Ejektionsfraktion des Herzens proportional verbessert, ohne systemische Veränderungen der Genexpression oder des alternativen Spleißens zu verursachen. Diese Ergebnisse zeigen, dass die Verbesserung des nukleären Imports von RBM20-Varianten entweder durch die Hochregulierung von TNPO3 oder durch andere Mittel als vielversprechende therapeutische Strategie dienen kann.

Insgesamt habe ich in diesem Promotionsprojekt den molekularen Mechanismus der RBM20-Fehllokalisierung bei DCM entschlüsselt. Darüber hinaus habe ich gezeigt, dass die Wiederherstellung der nukleären Lokalisierung von RBM20-Varianten eine praktikable Strategie zur Rettung beider Krankheitsphänotypen darstellt: Auflösung der schädlichen zytoplasmatischen Granula und Wiederherstellung des Spleißmangels.

Table of Contents

Chapter 1. Introduction

1.1. Splicing machinery of a cell

Splicing is a process whereby removing introns from a pre-mRNA results in the formation of a mature RNA. It has two modes: constitutive and alternative splicing. Constitutive splicing is defined as a process of removal introns from a pre-mRNA, followed by joining all exons together to produce a mature mRNA that then gets transported to the cytoplasm and translated by ribosomes. Alternative splicing (AS), on the contrary, is a process when an exon or a combination of exons can be removed from a mature transcript, or when an intron or a combination of introns can stay retained in a mature mRNA. Alternative splicing can be tissue-specific, and its dysregulation can cause various diseases such as cancer[1], neurogenerative or cardiovascular diseases[2,3], making it an attractive target for therapeutic interventions. In addition, AS is seen as a source of phenotypic biodiversity, playing an essential role in the evolution of eukaryotes[4]. For example, based on the data from 2017, the human genome contains 19,950 protein-coding genes, which is less than the mouse genome that encodes for 22,018 protein-coding genes. However, due to AS, lower number of genes in humans produces a larger number of mature protein-coding transcripts: 80,087 for human and 52,382 for mouse[5].

Splicing is mediated by a molecular machine called the spliceosome which is composed of small nuclear (sn)RNAs and proteins that together form small nuclear ribonucleoprotein (snRNP) complexes. In addition, many other proteins and splicing factors that are not a part of the core snRNPs associate with the spliceosome and are required for splicing to occur. About 99 % of all introns are spliced-out by the major spliceosome complex (Fig. 1.1, left panel), and the remaining 1 % by the minor spliceosome (Fig. 1.1, right panel).

At the splicing site, the 5' site of a major class intron, or splice donor site, is recognised by its nearly invariant canonical sequence “GU” followed by the less canonical “ACGU”. The 3' site of a major class intron, or splice acceptor site, consists of three conservative elements: a branch point sequence (BPS), a polypyrimidine tract, and a nearly invariant canonical “AG” at the end. From the 3' of the splice acceptor site, the

BPS is characterised by a canonical “A” located approximately 18-40 nucleotides towards 5’ direction, followed by a sequence of 15-20 polypyrimidine residues, most frequently “U”, that is adjacent to the 3’ splice site (Fig. 1.1, adapted from[2,3,6,7]).

Splicing is a two-step trans-esterification reaction. It starts with the U1 snRNP recognizing and tightly binding the 5’ splice site (5’ss) in a base-pairing manner, and splicing factor SF1 (or mBPP) binding to BPS, forming the pre-spliceosome, or Complex E. Once U2 snRNP tightly binds to the BPS by base-pairing, displacing SF1, Complex A is formed. This is followed by the recruitment of the U4/U5/U6 snRNP-triplex to join the U1 and U2, forming the pre-catalytic complex B. Next, upon several rearrangements, involving other players like RNA helicases, and removing of U1 and U4 snRNPs from the complex, the active spliceosome (complex B*) catalyses the first trans-esterification reaction. The 3’ hydroxyl group of the branch point adenosine performs a nucleophilic attack of the 5’ss phosphate group, freeing the 5’exon, and forming the complex C that contains a lariat intron. Complex C then catalyses the second trans-etherification step, where the 3’ss phosphate group is attacked by the hydroxyl of the freed 5’ exon, ligating two exons. The excised lariat intron together with U2, U5, and U6 is called P complex. At last, snRNPs get released in an ATP-dependent manner to catalyse more splicing reactions (Fig. 1.1, left panel).

The general principle of splicing mediated by the minor spliceosome is similar to the one described above for the major one. However, canonical sequences differ: 5’ss contains “AU” instead of “GU”, and 3’ss “AC” instead of “AG”. These divergent dinucleotides are recognised and bound by U11 and U12, which form complex A instead of U1 and U2. Then, to form complex B, U4 and U6 are replaced by U4atac and U6atac, and only U5 snRNP stays common between the two spliceosomes (Fig. 1.1, right panel).

1.1.1. Alternative splicing and its regulation

Due to alternative splicing – a process when different exons of a gene can be included or excluded in any possible combinations - each gene can give rise to a multitude of transcripts, which drives organism complexity and proteomic diversity. There are five major types of alternative splicing (Fig. 1.2): exon skipping (SE) – when an alternative exon is not included in the mature transcript; mutually exclusive exons (MXE) – when only one of the two exons gets included in the mature transcript; intron retention (RI) – when an intron, or a part of an intron, gets included in the mature transcript; and alternative 5' or 3' splice sites (A5SS and A3SS, respectively) – when alternative splice sites are used, resulting in longer or shorter exons being included in the mature transcript. In addition to alternative splicing, alternative promoter usage and alternative polyadenylation may result in different transcripts from 5' and 3' sites, although these processes are not mediated by the spliceosome.

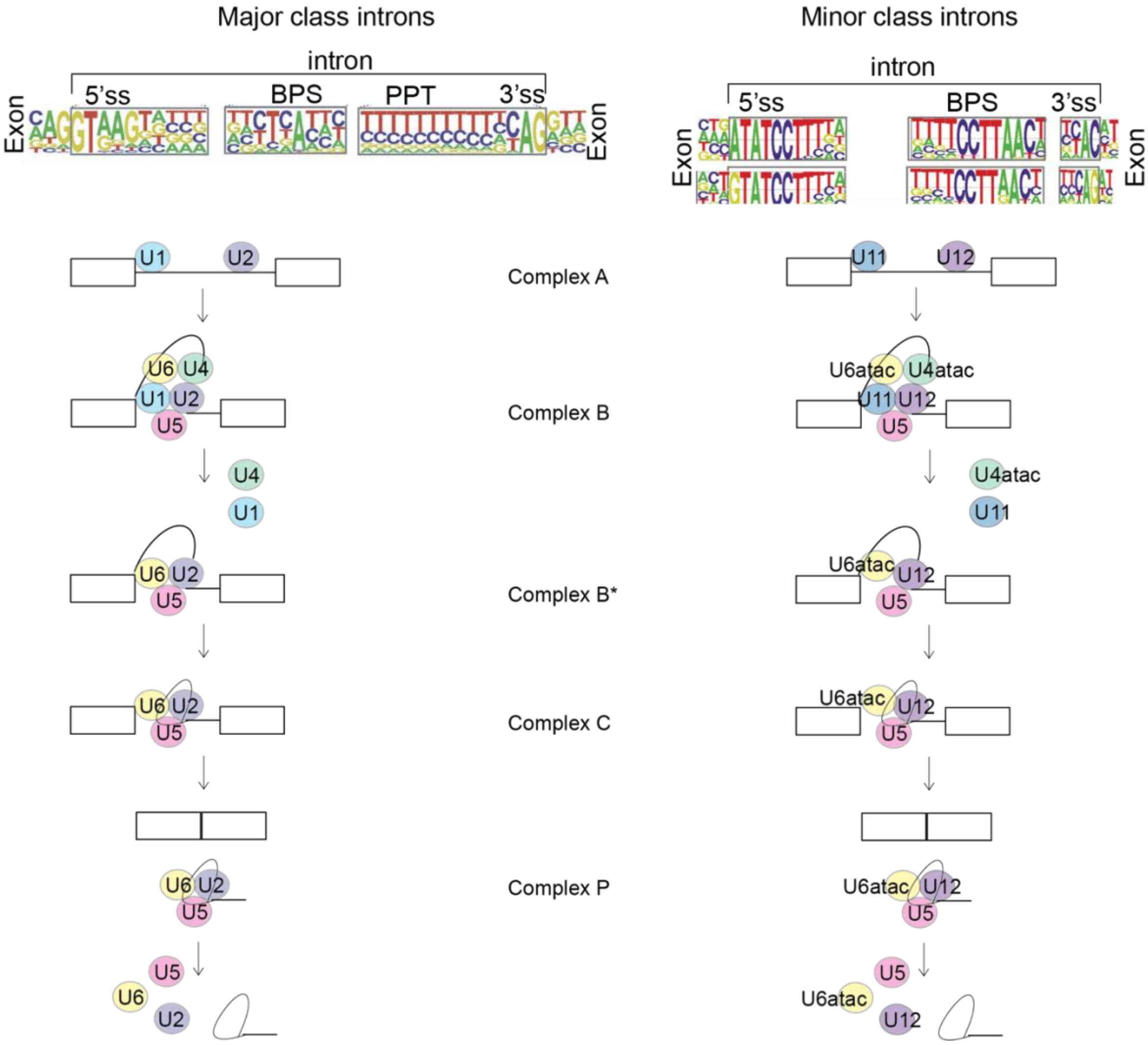


Figure 1.1. Schematic representation of splicing steps for major and minor introns. Adapted from [2,3,6,7].

Alternative splicing is regulated by *cis*-regulatory sequences, *trans*-acting factors, as well as transcriptionally, post-transcriptionally, and epigenetically. *Cis*-regulatory sequences control whether a particular intron will be recognised by the spliceosome.

For example, strong, or canonical, 5'ss and 3'ss can be spliced by the spliceosome and core splicing factors without the need for additional factors. Weaker 5'ss and 3'ss exhibit lower affinity to the spliceosomal snRNPs and therefore need additional *trans*-factors to recruit it. Exonic splicing enhancers (ESE) and silencers (ESS), as well as intronic spicing enhancers (ISE) and silencers (ISS) are recognised by *trans*-factors to enhance or silence splicing of an exon or an intron, respectively, by recruiting or supressing the spliceosome on these sites. Mutations in such cis-regulatory sequences can cause splicing defects of a certain transcript which can be found in various diseases[3].

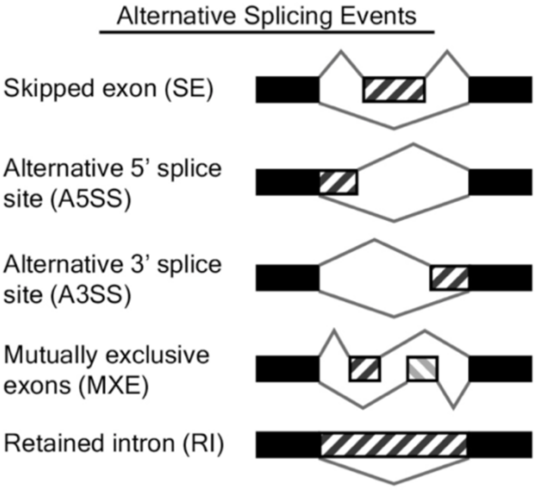


Figure 1.2. Types of alternative splicing events. Copied from [8].

Trans-acting factors are RNA binding proteins (RBPs) that can specifically recognise ESE, ESS, ISE, or ISS, and activate or supress splicing of a certain exon or intron. They are involved in both constitutive and alternative splicing. Their expression is often tissue specific, which is the main mechanism controlling transcript isoform diversity across different tissues and organs. Mutations in *trans*-factors can affect splicing of many downstream genes, and are therefore involved in a large number of human diseases[3]. The two most well-described categories of RBPs that play an essential regulatory role in splicing are serine-arginine (SR)-protein family and nuclear ribonucleoprotein (hnRNP) family. SR-proteins are characterised by the presence of an Arg/Ser (RS)-rich domain with at least 50 amino acids and more than 40% Arg/Ser content, and at least one RNA-recognition domain (RRM)[9]. RS-rich domains are responsible for protein-protein interaction between an SR-protein and other players in splicing regulation, and RRM domains recognise the *cis*-sequence of an exon or an intron and bind to it. Binding of SR proteins to a target exon usually leads to its inclusion, whereas binding to flanking exons or introns can lead to skipping of the target exon. Binding of SR proteins is usually either competitive or cooperative – one can either repress or enhance binding of another. In addition to approximately twenty SR-proteins, there are many more SR-related proteins that also contain RS-domain but do not fall under the strict criteria to be classified as SR-proteins. The combination of different RBPs, including SR- and SR-related proteins, as well as hnRNPs in the spliceosome determines its affinity to a certain exon, regulating the alternative splicing of this exon[2,9].

Splicing of the vast majority of transcripts occurs co-transcriptionally. Therefore, the process of transcription has been shown to affect the splicing efficiency[7]. Transcriptional machinery has been shown to recruit many RPBs, and, specifically, RNA Polymerase II (Pol II)[2]. In addition, the rate of transcription by Pol II affects the rate at which weak and strong splicing sites get available, therefore impacting the choice of a splice site to be used. Thus, genome accessibility also plays a role in regulation of alternative splicing. For example, it has been shown that DNA methylation, as well as histone modifications leading to increased nucleosome density are often more enriched in exons than in their flanking introns[2]. Lastly, it has also been shown that RNA modifications, such as N6-methyladenosine (m6A), affect the secondary structure of premature RNA and therefore also increase its accessibility for interactions with RBPs, thereby affecting alternative splicing[2].

1.2. Nuclear import pathways

Many cellular processes such as gene transcription, chromatin organization, DNA replication, DNA repair, and pre-mRNA splicing occur in the nucleus, which requires regulators and mediators of these processes to be imported there from the cytoplasm, where they get initially translated. The nucleus is separated from the rest of the cell by the nuclear envelope (NE) which consists of two phospholipidic membranes about 30 nm apart. The outer nuclear membrane is continuous with the endoplasmic reticulum, while the inner membrane is associated with the lamina – a network of intermediate filaments that serve as chromosome attachment sites and shield the nucleus. Macromolecules can selectively enter the nucleus only via a large complex of proteins that forms an aqueous channel across two membranes – nuclear pore complex (NPC).

There are between 2000-5000 NPCs per nucleus of a vertebrate cell, with numbers varying between species and tissues[10]. NPC is composed of about 30 different proteins called nucleoporins, each of which is present in several copies. Small molecules like salts, nucleotides, and proteins less than 60 kDa can pass through the NPC via passive diffusion. Larger proteins get through the NPC with the help of nuclear receptors called importins, exportins, or biportins, depending on the direction of transport mediated by them.

1.2.1. Importins and their recognised NLS

Nuclear transport receptors belong to β-karyopherin family of proteins and are called Kaps. There are 20 Kaps identified in human cells, of which 10 are importins, 5 are exportins, 3 are biportins, and a function of the remaining two is still unknown[11]. Importins recognize their cargos by the presence of a specific nuclear localization signal (NLS) that is defined as an amino acid sequence that is necessary and sufficient for the nuclear import. Likewise, exportins recognise a sequence called nuclear export signal (NES). Each importin and exportin has its specific NLS and NES that they recognise. NLSs and NESs are located within the coding sequence of a protein, and remain intact after a round of nuclear/cytoplasmic shuttling, which allows such rounds to occur more than once.

The first ever discovered NLS was found in the simian virus 40 (SV40), and is composed of only seven amino acids: PKKKRKV. This strong NLS is first recognised by Importin alpha (Impα$_{1-8}$), that serves as an adaptor protein that then binds to Importin beta (Imp-β or Kap-β1 or TNPO1). In such heterodimeric complex, the cargo then gets transported through the NE, after which the complex is disassembled, and the importins get recycled back to the cytoplasm. This NLS is classified as a classic NLS. Classic NLSs are sub-divided into two sub-classes: monopartite and bipartite. Monopartite ones are usually short sequences (4-8 amino acids) with at least four positively charged amino acids like arginine (R) or lysine (K). The characteristic sequence of a monopartite classic NLS is usually defined as K (K/R) X (K/R), where X is any amino acid[12]. Bipartite NLSs are usually composed of two clusters of 2-3 positively charged residues that are about 9-12 residues apart, with the consensus sequence R/K(X)$_{10-12}$KRXK[12]. Examples of cargos transported by Impα/Impβ heterodimer with classical NLSs are several histones (H1, H2A, H2B, H3, H4), some HIV-proteins, and cyclin B1[10].

Many nuclear proteins have non-classical NLSs, that do not have much similarity with any of the classical ones. One example of such NLSs is the proline-tyrosine category (PY-NLS). Proteins with such NLSs usually have hydrophobic or basic N-terminal residues, and C-terminal R/K/H$(X)_{2\text{-}5}$PY motifs[12]. Cargos with such motifs are usually recognised by Kap-β2 or TNPO2 transporter, and among its targets are several hnRNPs, Cyclin T1, FUS, cytomegalovirus protein UL79, and others.

Besides these two well-studied NLSs, there are more types that are less well-studied. For example, the isoleucine-lysine NLS (IK-NLS) that has a consensus sequence K-V/I-X-K-X1–2-K/H/R[11]. It is supposed that in humans, such NLS is recognised by IPO5 (Importin-5 or Kapβ3). Another example is RS-rich regions in various RNA binding proteins and splicing factors, usually recognised by TNPO3. Other non-classical NLSs may be recognised by other importin proteins including Importin-4, -9, -7, -8, and -13.

Nuclear transporters mediate transport of their bound cargos by binding the phenylalanine-glycine (FG) repeats of nucleoporins located in the central pore of the NPC. The directionality of the transport mediated by transportins is controlled by RAN-GTPases. Chromatin-bound factor RCC1 is responsible for exchanging guanine nucleotides in the nucleus, keeping nuclear RAN in the GTP-bound state. In the cytoplasm, RAN is kept in the GDP-bound state by cytoplasmic RAN GTPase-activating protein (RANGAP) and RAN-binding protein RANBP1[11] that mediate GTP hydrolysis. Binding to RAN-GTP or RAN-GDP induces conformational changes of the transporter, increasing or decreasing its affinity to the cargo. For example, nuclear importins recognise and bind their cargos in the cytoplasm, and, after the cargo is transported to the nucleus, the importin binds to RAN-GTPases and releases the cargo. On the contrary, exportins can recognise their cargos in the nucleus in the RAN-GTP-bound state, and upon GTP hydrolysis in the cytoplasm, they release their cargos[11].

In addition to Importin-mediated nuclear import, there are a few other, non-conventional nuclear import pathways utilizing other mechanisms. For example, cytoskeletal-associated transport, mediated by microtubules or specifically dynein; calmodulin-mediated transport that is independent of importins and RAN-GTPases and that solely relies on calcium (Ca) mobilization; direct interaction with the NPC proteins without importins; passive diffusion of small proteins; glyco-dependent nuclear import, and potentially others[13].

1.2.2 Nuclear importins in the prevention of pathogenic protein mislocalization and aggregation

Many human diseases are linked to protein mislocalization, and especially, defects in nuclear import. For example, cytoplasmic mislocalization of p53 mediates progression of several cancers[14]. Moreover, several neurodegenerative disorders including amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), and multisystem proteinopathy (MSP), are associated with cytoplasmic mislocalization and aggregation of RNA-binding proteins (e.g. TDP-43 and FUS)[15]. Intriguingly, it was shown that, importin Kapβ2 can inhibit polymerization and liquid droplet granule formation of RBPs containing the PY-NLS that it recognizes[15]. For example, fibrilization of several variants of FUS non-impacting its PY-NLS was fully abolished by Kapβ2. For another FUS variant impacting the PY-NLS, Kapβ2 inhibited fibrilization by about 50%, while for a FUS variant with a nonsense mutation resulting in PY-NLS not being present, Kapβ2 had no effect on fibrilization[15]. Similarly, Impα and Kapβ1 abolished fibrillization of TDP-43 variants by recognizing their cNLS and chaperoning them[15]. Fibrils and liquid droplet granules formed by pathogenic variants in these and other RBPs were disaggregated and dispersed by addition of importins recognizing their NLSs[15]. In addition to the chaperone effect of these importins, they also restored nuclear localization of pathogenic RBP variants, restoring their physiological function, which makes them promising targets for therapies[15]. Later on, Transportin-3 was also shown to prevent aggregation and restore nuclear localization of a nuclear-speckle protein SRRM2[16]. In conclusion, more and more studies describe the novel function of Kap-proteins. Besides import of their cargos to the nucleus, they also display a chaperone-like function to prevent aggregation of their cargos in the cytoplasm, which makes them attractive targets for therapy. Although several compounds inhibiting activity of Kap-proteins already exist, compounds to stimulate their activity are yet to be developed[16].

1.2.3. Transportin-3 (TNPO3) function in the import of SR proteins into the nucleus

SR-proteins that regulate alternative splicing (see above) are usually transported to the nucleus by TNPO3 (also known as Transportin-3 or Transportin-SR). As early as in 1999, it was shown that Transportin-SR specifically recognizes and directly binds to the RS-domains that serve as NLSs, which was for the first time shown for ASF/SF2 and SC35[17]. This process is also controlled by RAN-GTPases, and is ATP-dependent[17]. A highly similar importer (authors call it “almost identical”) was identified later in the same year to be importing E2 transactivator and ASF and was named Transportin-SR2[18]. The authors showed that this transportin can directly interact with phosphorylated RS-domain, however, hyperphosphorylation by overexpression of SPRK1 abolished nuclear import[18]. It was also shown that the N-terminus is important for RAN-GTPase-mediated release of a bound cargo after its nuclear import, and for recycling of the transportin. The authors supposed that, the previously identified[17] Transportin-SR can transport its cargos in phosphorylation-independent manner, while Transportin-SR2 recognizes phosphorylated RS-domain as an NLS. It was later confirmed that Transportin-SR2 can target phosphorylated, but not unphosphorylated SR proteins to the nucleus[19]. In addition, it was demonstrated that, not only this transporter interacts with the NPC as other importins do too, but also targets its cargos to nuclear speckles[19]. Later on, it was clarified that, Transportin-SR[17] and Transportin-SR2[18] actually are alternatively-spliced products of the same gene in mammalian cells[20]. The Transportin-SR2 isoform is more ubiquitously expressed across cell types and tissues. As shown for ASF/SF2, both isoforms, although are able to bind its unphosphorylated RS-domain, have a preference to its phosphorylated version. However, for nuclear import of hTra2α, its phosphorylation did not influence the efficiency of binding for both isoforms. Only Transportin-SR2 required phosphorylation of hTra2β, while Transportin-SR did not. The authors, therefore, concluded, that phosphorylation may not be universally required for cargo recognition and nuclear import mediated by both Transportin-SR isoforms[20], further referred to as just TNPO3.

TNPO3 was shown to interact with more than a hundred of cargos in HEK 293T cells, the vast majority (> 70 %) of which contained a recognizable RS- or an RS-like domain[21]. The authors deciphered a structure of TNPO3 alone, in a GTP-bound state, as well as in complex with ASF/SF2. This lead to identification of the binding pocket of TNPO3 involved in the interaction with RS-domains[21]. Structural and functional studies of TNPO3 in complex with another SR-protein CPSF6 showed that its nuclear import is mediated by TNPO3 in a phosphorylation-independent manner[22]. Nuclear import of CIRBP was mediated by recognition of its RSY-rich domain by TNPO3 or of its RG/RGG-rich region by TNPO1, demonstrating that, some proteins may have more than one nuclear import mechanism[23]. Interestingly, it was demonstrated that, the presence of TNPO1 and/or TNPO3 prevents CIRBP to phase-separate and to be recruited to stress granules, validating findings discussed above. In addition, it was also shown for the first time, that methylation of Arg in the RGG-rich region of CIRBP decreased its interaction with TNPO1 and TNPO3[23], suggesting that post-translational modifications other than phosphorylation may play a role in TNPO3-mediated recognition and nuclear import.

TNPO3 was shown to be implicated in nuclear import of HIV-proteins, and, therefore, is studied as a potential drug target for HIV-infection[24]. In addition, *TNPO3* was shown to be involved in myogenesis[25], and its mutations are linked to myopathy[26] and muscle dystrophy[27]. In one report from 2011 it was also demonstrated that a mutation in TNPO3 was causative of a heart condition called dilated cardiomyopathy[28]. These facts indicate that disturbing TNPO3-mediated nuclear import can lead to human disease, and especially to muscle-related disorders. This, however, has not been extensively studied yet, and more TNPO3-related disorders could be discovered in the future.

1.3. Dilated cardiomyopathy

Cardiomyopathies are a heterogeneous class of disorders underlined by structural and functional abnormalities of the heart muscle leading to cardiac dysfunction in the absence of coronary artery disease or ischemia[29]. They are classified based on morphological and functional changes into four major groups: hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), restrictive cardiomyopathy (RCM), and arrhythmogenic right ventricular cardiomyopathy (ARVC)[30–32].

DCM is phenotypically distinguished from other cardiomyopathies by the characteristic left ventricular enlargement and reduced systolic function[33]. These are usually defined in the clinic by ejection fraction being less than 45%, and left ventricular (LV) end-diastolic diameter greater than 117% (greater than two standard deviations from the predicted value)[33,34]. DCM is diagnosed in around every 1 of 250 individuals[35]. It is the third most common cause of heart failure world-wide and is the leading indication for heart transplantation[36,37]. DCM patients often get diagnosed between 20 and 50 years of age[38], when the overt disease symptoms are already far progressed[36], which limits treatment options. Currently, DCM patients are generally treated in the same way as heart failure patients, including pharmacological and devise therapies, that are focused on managing symptoms and not curing the underlying cause[34,36,38]. Despite general improvement in life quality and expectancy for DCM patients in the past years, the mortality rates are still substantial: about 70% of death cases arise from pump failure, and the remaining 30% from sudden cardiac death caused by arrythmias[38].

Based on its cause, DCM is generally classified further into two main categories: acquired or idiopathic DCM – where no genetic determinant can be identified, and genetic or inherited or familial DCM – where it is caused by a mutation in one or more genes and is present in the family history[31,39].

1.3.1. Acquired DCM

In 50-70% of cases, a genetic cause of DCM is not determined, and such a case is characterised as acquired or idiopathic. Amongst the most common risk factors for acquiring DCM are[36–40]:

- Exposure to toxins, such as alcohol (risk is proportional to the extent and duration of ethanol intake; about 26-36% of DCM cases are associated with alcohol abuse in developed countries), recreational drugs (amphetamines, cocaine, ecstasy), steroids (anabolic or androgenic), cancer therapies such as anthracyclines (for example, doxycycline), carbon monoxide, cobalt, lead, mercury, and others. They act through cardiotoxicity, or, in other words, result in cardiomyocyte loss, leading to structural and functional changes in the heart and to fibrosis, which are mostly irreversible.

- Infectious: exposure to various infectious agents like viruses (including herpes, hepatitis C, HIV, adenovirus, as well as to the novel coronavirus Sars-CoV-2), bacteria, fungi, and parasites have been linked to DCM.
- Inflammation and autoimmune disorders: heart toxicity caused by any of the aforementioned agents may result in triggering the proinflammatory cytokines activating the immune response in the heart, a condition called myocarditis. This may eventually lead to chronic DCM in about 20-30% of cases. Fibrosis – another pathological characteristic of DCM in addition to LV dilatation – is often a consequence of cardiac inflammation.
- Metabolic imbalance. Deficiency of selenium, thiamine, zinc, copper, or carnitine, have been linked to higher risk of DCM. In addition, inborn errors in fatty acid oxidation, glycogen storage disorders, and other inborn metabolic conditions may contribute to DCM development.
- Pregnancy: in very rare cases, heart failure may develop in the last trimester of pregnancy, and is thought to be mediated by oxidative stress resulting in conversion of prolactin to an angiostatic factor. It is also associated with older age in women.
- Sex differences: males are overall at higher risk of DCM development. It may be associated with sex hormones binding to androgen and oestrogen receptors of vascular cells, myocytes, or fibroblasts, as well as through sex hormones binding and affecting immune cells.
- Combined factors: all the described factors are not mutually exclusive: a combination of risk factors may increase the overall risk even further.

In case all possible causes of DCM have been excluded, such DCM case is classified as idiopathic. However, current advances in genetic screening of patients and identification of more and more genes that may be involved in DCM progression reduces the frequency of such diagnosis every year.

1.3.2. Familial DCM

Familial DCM accounts for 30-50% of all DCM cases[41]. As a result of the disease heterogeneity and high prevalence, variants in many genes (more than 100) have been linked with stronger or weaker confidence to familial DCM to date and are being evaluated in clinics[34,41]. However, a more recent study re-evaluated genetic contribution of 56 of these genes by analysing a cohort of 2,538 DCM patients and confirmed only 12 genes to have a strong evidence of being DCM-causal[42]. Another, even more recent study[43], curated 51 genes classified as DCM-causing using the ClinGen gene curation scoring framework[44]. This framework curated and classified all published scientific and clinical evidence for a gene to be causal of a disease into definitive, strong, moderate, limited, no reported evidence, or conflicted evidence. Based on this classification, only 12 genes (*BAG3, DES, DSP, FLNC, LMNA, MYH7, PLN, RBM20, SCN5A, TNNC1, TNNT2, TTN*) were classified to have definitive or strong implications in DCM progression, and 7 genes were classified as moderate (*ACTC1, ACTN2, JPH2, NEXN, TNNI3, TPM1, and VCL*)[43]. All of the 12 definitive or strong genes, as well as 5 out of 6 moderate genes, have autosomal dominant inheritance, with only one moderate gene (*JPH2)* presenting a semidominant way of inheritance.

Functionally, DCM-causing genes can be classified in the following categories:

- Variants in sarcomeric genes affect cardiomyocyte contractile force generation and represent the most frequent cause of familial DCM (*ACTC1, MYH7, TNNC1, TNNI3, TNNT2, TPM1, TTN)*[37,41,43]. These genes encode for components of sarcomere – actin, myosin, troponin, tropomyosin, and titin. Titin is the largest protein expressed in the human body that spans half of the sarcomere length and controls passive stiffness[34]. Truncating variants of *TTN* alone account for about 20-25% of all familial DCM cases[39,41] affecting cell contractility and shifting cardiac metabolism[36].
- Laminopathies caused by *LMNA* mutations resulting in defects in nuclear envelope account for about 5-8% of all familial DCM cases[34,37]. They are characterised by a severe phenotype with malignant arrythmias.

- Mutations in the alternative splicing regulator *RBM20* account for about 3% of inherited DCM cases and result in an early-onset DCM with frequent malignant arrythimas[45,46].
- Variants in cytoskeleton proteins like filamins, dystrophin, and desmin encoded by genes *FLNC, VCL, DES* affect contractile force transmission[37,43] and can lead to muscular dystrophies in addition to DCM.
- Mutations in Actinin-2 (*ACTN2*) – a Z disk gene that serves as an anchor for myofibrillar actin, and sarcoplasmic reticulum gene mutations (*PLN)* - affecting Ca-handling, both directly impact contractility[35,39].
- Deficits in cellular adhesion result in desmosomal cardiomyopathies are caused by variants in *DSP*[37].
- Mutations in the main sodium channel protein encoded by *SCN5A* account for about 2% of familial DCM cases and result in arrhythmogenic phenotype[37,39].
- Variants of junctional membrane gene *JPH2* have been classified as moderate based on the evidence to cause DCM[43].
- Co-chaperone and an anti-apoptotic heat shock protein encoded by *BAG3* has also been linked to familial DCM, resulting in LV disfunction[37,43].

This implicates a wide range of cellular processes that, when impacted, can lead to DCM progression. Altogether, given the high heterogeneity of phenotypical representation and causative determinants of DCM, development of personalised therapy approaches is urgently needed to improve current scarce curative treatment options that are based on the "one fits all" approach[37]. For that, knowledge of precise molecular mechanisms causing the disease progression in each case, as well as efficient classification of patients based on such mechanisms, is a requisite. Recent advances in functional genomics and high-throughput phenotyping methods make up for outstanding tools for deciphering the exact molecular mechanisms behind every cause of DCM.

1.4. RBM20-mediated DCM

The RNA binding motif protein 20 (*RBM20*) gene is located on chromosome 10 and encodes for an RNA-binding protein that regulates alternative splicing. The gene is composed of 14 exons, producing a 134 kDa protein that contains P-rich, L-rich, RS-rich, RNA recognition motif (RRM), E-rich, and two zinc finger (ZnF) domains (Figure 1.3, copied from[47] who adapted it from[48])[49]. According to the Human Protein Atlas and other studies[45,50], it is predominantly expressed in the heart muscle tissue, however lower expression levels are also detected in small intestine and seminal vesicle.

For the first time, missense mutations in *RBM20* were identified in 2009 in two families with an early-onset and highly severe DCM[45]. A year later, the first list of five missense mutations got extended by other four missense mutations[46]. Later on, more RBM20 variants have been linked to familial DCM, accounting for approximately 3% of all familial DCM cases[41,51]. To date, a total of twenty-two heterozygous missense and one heterozygous nonsense variant in *RBM20* have been linked to DCM progression[47], and the majority of them result in highly penetrant form of the disease with increased risk of arrythmia and sudden cardiac death[48]. Interestingly, many (9 out of 23) of identified mutations cluster in a conserved stretch of six amino acids located in the RS-rich region (P633-P638, Figure 1.3). Other pathogenically-relevant regions include RRM and E-rich regions[48].

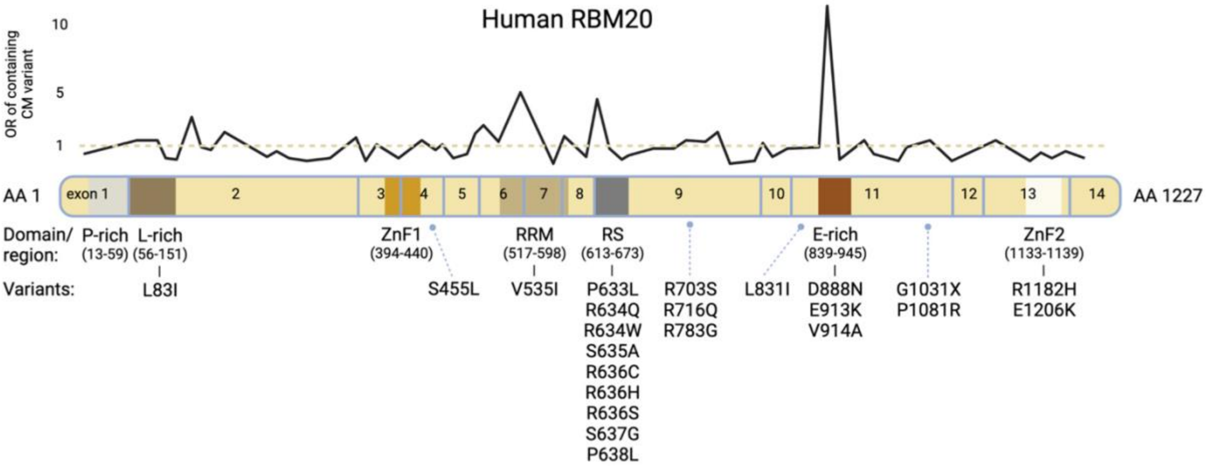


Figure 1.3. Schematic structure of RBM20 with indicated exons, domains, and DCM-causing variants. Black line illustrates the Odd Ratio (OR) value of a certain region in cardiomyopathy population versus general population (Genome Aggregation Database [gnomAD]). Copied from Koelemen et al., 2021[47], who adapted it from Parikh et al.[48], 2019.

1.4.1. RBM20-dependent alternative splicing

RBM20 is a *trans*-acting splicing regulator (see above) that has been shown with high confidence to induce exon skipping and regulate mutually exclusive exons located in more than 30 genes[47,50,52–55]. Using PAR-CLIP in HEK293 cells and rat cardiomyocytes, Maatz et al.[53] showed that it specifically recognizes the UCUU-core motif of its targets. It binds to intronic regions 100 bp downstream and 50 bp upstream of repressed exons, and promotes assembly of the early spliceosome, interacting with U1 and U2 components[53]. It is absent from the catalytically-active spliceosome[53]. Presumably, RBM20's role is to recognize the alternative exons essential for cardiomyocyte functioning, bind to the intronic regions in close proximity to them via its RRM-domain, and prevent the recruitment of U4/U5/U6 to stop complex B assembly and, thereby, repress splicing. Later on, in cell culture experiments with *TTN* exon 241-243 minigene reporter, a sequence between 102 and 910 bp of the upstream and downstream introns, respectively, from the regulated exon 242, was shown to be sufficient to repress its splicing[56]. The minimal genomic segment sufficient for RBM20-dependent splicing repression of an exon was determined to include the repressed exon and its downstream intronic region[56]. The observed binding to the downstream intron suggested that exon repression is achieved by interfering with intron definition, even before complex A is assembled. Presence of the UCUU core was experimentally confirmed to be a requirement for exon recognition by RBM20[56]. Altogether, it is clear that RBM20 represses its target exons at early stages of spliceosome assembly, binding to their flanking intronic regions containing the UCUU-core sequence.

So far, information regarding the necessity of different RBM20 domains for its splicing function is controversial. In one study, it was shown that only RRM-, RS-, and C-terminus of RBM20 together were sufficient for RBM20's splice regulation of *TTN* minigene in cell culture[56], suggesting the importance of the C-terminus in splicing repression. In turn, another study demonstrated that, deletion of RRM and both ZnF domains did not impact splicing of fluorescent *TTN* minigene reporter[57], while the RS- and E-rich domains were required for splicing repression. The consensus between these studies suggests that, RS- and E-rich regions are required for splicing function of RBM20, which is supported by the highest Odds Ratio (OR) of mutations in these domains to be causative of DCM[48] (Figure 1.3). The exact function of each of RBM20 domains in splicing repression, as well as the exact molecular consequences of each pathogenic point mutations, are yet to be determined.

Amongst RBM20 target genes are many genes that have also been linked to DCM progression with stronger or weaker confidence[43] (*TTN, MYH7, NEXN, LDB3, TNNT2, TPM1, DTNA*, *OBSCN, PDLIM3*), genes involved in signal transduction (*CAMK2D, CAMK2G*), calcium handling (*RYR2*), mitochondrial function (*IMMT*), and other processes essential for cardiomyocyte function[47,55]. Koelemen et al.[47] recently summarised core RBM20 target genes. Their splicing was consistently affected across *RBM20* perturbations (e.g. KO, S635A) and across model species (rat, human, mouse)[50,53,54]. These RBM20 target genes are further referred to as "core RBM20 targets".

1.4.2. Upstream regulation of RBM20

To date, surprisingly not much is known about transcriptional regulation of *RBM20* expression. RBM20 is predominantly expressed in the heart, therefore its expression is presumably regulated by cardiac-specific transcription factors or pathways. It has been shown that its expression gets activated during early cardiac development in mice[58]. So far, it is known that RBM20 expression and downstream *TTN* splicing are induced by insulin and thyroid hormone via the PI3K/Akt/mTOR signalling pathway[59,60]. This may be achieved either through activation of RBM20 phosphorylation, or by upregulating its translation[59]. Moreover, angiotensin II has been demonstrated to activate RBM20 expression acting through the same pathway on ETS Transcription Factor (ELK1) that can bind the promoter of *RBM20*[61]. Our lab previously showed that a small molecule all trans retinoic acid (ATRA) can upregulate RBM20 expression in iPSC-CMs[62], although this effect is likely indirect. *RBM20* expression was also positively correlated with Body Mass Index (BMI) which suggests that metabolic stimuli can impact this process[63]. Presumably, distinct metabolic state of striated muscle cells dictates *RBM20* expression.

1.4.3. Pathobiology of RBM20-mediated DCM

After identification of the first *RBM20* mutations to cause DCM in 2009[45] and 2010[46], the pathobiological mechanism remained unclear for another few years. In 2012, Guo et *al.*[50] demonstrated for the first time that RBM20 regulates splicing of multiple cardiac genes including *TTN*. In *Rbm20* knockout (KO) rats, as well as in a human patient with heterozygous S635A mutation, they showed the expression of a giant N2BA-G isoform of *TTN*, instead of N2BA and N2B that were expressed in healthy individuals and WT rats. Guo et *al.* identified for the first time a set of RBM20 target genes, splicing of which was impacted by both *RBM20* KO and the point mutation in rats and humans. In addition, the authors demonstrated that splicing was dose-dependent on RBM20 level, with heterozygous KO rats exhibiting an intermediate splicing pattern between homozygous KO and WT rats[50]. RBM20 was found to co-localize with *TTN* pre-mRNA in nuclear RBM20 speckles. There, it represses *TTN* exons leading mostly to exon/intron skipping, and to a lower extent to intron retention, and exon shuffling[52]. Maatz et al.[53] validated 18 of RBM20 targets that showed RBM20-dependent alternative splicing and that were directly bound by RBM20. The authors also demonstrated that an S637A mutation in RBM20's RS-domain abolished interaction with other alternative splicing factors but not with core spliceosomal proteins[53]. This suggests that RBM20-mediated repression of complex B formation may need to be stabilised by other factors. Since the RS-domain is thought to be responsible for protein interactions, it is likely that it is also the one that gets bound by stabilizing factors. Therefore, pathogenic RS-domain variants may affect RBM20's splice regulatory activity due to decreased protein interaction ability.

Phenotypically, *Rbm20* KO rats showed lower cardiac output and reduced skeletal muscle performance, as well as abnormal myofibril arrangement, Z line streaming, and lipofuscin deposits observed by electron microscopy[64]. Analysis of biopsies derived from human patients with P638L mutation using transmission electron microscopy (TEM) revealed atypical sarcomere geometry: sarcomeres were elongated and appeared thinner upon RBM20 deficiency[58]. Later on, this finding was also confirmed by TEM analysis of patient-derived iPSC-CMs bearing a R636S mutation: sarcomere length was significantly higher, and sarcomere width was significantly lower in R636S-mutant iPSC-CMs compared to WT[65]. In addition to abnormal sarcomere structure, Ca^{2+} handling was also disturbed upon both knockdown (KD) of *RBM20,* and R636S mutation in iPSC-CMs[58,65]. Similar observations regarding sarcomere disorganisation and Ca^{2+} cycling abnormalities were made based on patient-derived iPSC-CMs harbouring a S635A mutation[66]. β-adrenergic stimulation of R636S iPSC-CMs with norepinephrine (NE) resulted in disorganisation of sarcomeres characterised by punctate sarcomeric α-actinin distribution not observed in WT cells[65]. Pre-treatment with β-antagonist, carvedilol, and L-type Ca^{2+} channel blocker, verapamil, rescued the observed effects of NE-stress[67]. In iPSC-CMs with P633L and R634Q mutations, contractility was decreased, while Ca^{2+} upstroke time was increased, in addition to splicing deregulation[62]. Defects in Ca^{2+} handling upon *Rbm20* KO were confirmed *in vivo*: myocytes from mice lacking *Rbm20* displayed action potential prolongation, increased L-type calcium current, Ca^{2+} overload, and higher chance of spontaneous Ca^{2+} releases, altogether leading to arrythmias[54]. Verapamil treatment of primary mouse myocytes decreased spontaneous Ca^{2+} releases[54]. Moreover, *Rbm20* KO mice displayed evidence of fibrosis indicating cardiac injury[54].

While sarcomere abnormalities are likely to be explained by *TTN* mis-splicing, the observed defects in Ca^{2+}-handling implicated in severe arrythmias associated with *RBM20* mutations could be linked with *CAMK2D* mis-splicing[54]. Patients with *TTN* mutations display less frequent arrhythmogenic burden compared to RBM20 patients[54]. Other downstream genes that are affected by *RBM20* mutations are likely to impact more cellular processes contributing to the disease phenotype, explaining why RBM20-DCM is so malignant.

Moreover, molecular mechanism of RBM20-DCM is likely different when different parts of RBM20 are mutated. For example, an E913K mutation in RBM20's glutamate-rich region was shown to decrease the protein level without impacting its localization, leading to haploinsufficiency of splicing, and resulting in stiffer CMs[68]. A nonsense mutation G1031X resulted in a similar splicing haploinsufficiency phenotype, although located downstream of RS- and E-rich domains[57]. Interestingly, I538T mutation in the RRM-domain identified in a sudden death patient did not cause DCM, and neither did mice with such a mutation develop any DCM phenotype, despite displaying affected splicing of *TTN* and other RBM20 targets[69]. It was shown that a combination of RRM- and RS-domains of RBM20 is required for its nuclear retention, and, more specifically, the sequence between 491–658 amino acids positions[70]. In addition, mutations in RBM20's RS-domain result in protein mislocalization to the cytoplasm in a form of RNP granules, that potentially worsen the disease phenotype (see below)[71–80]. Pathogenic mutations outside of the RS-domain did not impact RBM20's nuclear localization[76]. These suggest that, some mutations in *RBM20*, and specifically mutations outside of the RS domain, result in haploinsufficiency of alternative splicing, resulting in a similar loss-of-function phenotype as observed in *Rbm20* KO model animals. This leads to mis-splicing of many important for cardiac function genes, causing structural and functional abnormalities of cardiac myocytes. On the other hand, RS-domain mutations, besides leading to haploinsufficiency of splicing, may have gain-of-function effects due to their aberrant localization.

1.4.4. RBM20 mislocalization

RBM20 mislocalization was for the first time described in detail in 2020[71,72]. For the R636S variant in a pig model and in hiPSC-CMs[71], it was shown that this variant mislocalised to the cytoplasm and formed ribonucleoprotein (RNP) granules. The authors showed that RBM20-granules in the cytoplasm are phase-separated membranelles organelles, originating from liquid droplet fusion, with the core structure similar to stress granules. They demonstrated that these granules are linked to the cytoskeleton, i.e. they co-localised with F-actin and sarcoplasmic ACTC1 filaments. These granules fused with stress-granules upon metabolic stress induction with sodium arsenate. The authors hypothesise that RBM20 normally contributes to a cytoskeleton-linked liquid condensate network, mediating the normal flow of mRNA from the nucleus to the cytoplasm. Accumulation of RNP granules caused by RBM20 variants and their interaction with cytoskeleton disrupts this flow that causes cellular stress and early proliferation arrest. The latter impedes cardiomyocyte differentiation process and actin polymerization dynamics. This previously uncharacterised gain-of-function phenotype caused by R636S mutation further worsens the disease phenotype, in addition to the well described before splicing deficiency and its pathobiology[71].

By comparing *RBM20* KO and S637A-knockin (KI) mice, it was for the first time proven that the point mutation leads to a more severe cardiac dysfunction than the protein KO[72]. The authors demonstrated that, while splicing dysfunction was similar between the KO and KI mice, point mutation resulted in higher incidence of atrial fibrillation, fibrosis, left ventricular dilatation, electrophysiological abnormalities, and overall poorer survival rates. In addition, the authors detected changes in gene expression enriched for skeletal muscle and DCM pathways in S637A mice, unlike in KO mice. These clearly demonstrate that, S637A variant, in addition to classical loss of splicing that is modelled by the protein KO, possesses a detrimental gain-of-function. The authors hypothesised that the observed perinuclear localization of S637A variant and formation of granule-like structures is the cause of gain-of-function phenotype[72].

A year later, another study compared phenotypes of hiPSC-CMs with RBM20-R636S and *RBM20* KO[73]. The authors showed that electrophysiological parameters of point mutant cells differed greater from the WT cells than those of KO cells. In addition, contractility of engineered heart tissues (EHTs) was more affected upon point mutation than upon KO. Based on eCLIP (enhanced crosslink and immunoprecipitation) analysis, the authors found that, unlike normal preferential binding of RBM20-WT to intronic regions, RBM20-R636S preferably bound to 3'UTRs. They also found that, the direct mRNA targets of RBM20-WT and RBM20-R636S were mostly distinct. Amongst enriched binding motifs for RBM20-R636S were those of other RBPs binding to 3'UTR of transcripts in stress granules. In addition, the authors found that cells with RBM20-R636S and RBM20-KO exhibited distinct alternative splicing alterations from the WT cells. They identified transcripts that were specifically dysregulated upon point mutation and not upon KO. These observations indicate transcripts that are potentially unique to this variant, and mis-splicing of which is not explained by nuclear depletion of RBM20 alone. The authors found that RBM20-R636S co-localizes with a marker of P-bodies DDX6 in the absence of cellular stress, and hypothesised that RBM20 plays a role in mRNA stabilization in P-bodies, binding to their 3'UTRs[73]. This provides the first molecular evidence that, RBM20-R636S can bind to a distinct set of transcripts in the cytoplasm, and, potentially, impact alternative splicing of distinct transcripts in the nucleus, in addition to the effects caused by nuclear depletion studied in KO models.

Later on, it was shown that RBM20-S639G also forms perinuclear granules and leads to a severe DCM phenotype in mice (human S637G)[74]. The authors demonstrated that this variant forms RNP granules at the nuclear periphery, similar to what was shown previously, and affects expression of genes involved in protein binding, regulation of protein translation, and stress granule components. The authors presume that this variant may create a microenvironment that promotes the condensation of RNA-protein complexes.

Given the results of the studies described above, it became apparent that, mutations in the RS-domain result in RBM20 mislocalization to the cytoplasm which results in worsening of the disease phenotype beyond the one seen upon loss-of-function of RBM20. In addition, it was previously shown that the RS-domain is necessary for nuclear localization of RBM20[57,70]. However, to date, the mechanism of RBM20 nuclear import, and the participation of the RS-domain in this process, has remained unclear.

It has been proposed that phosphorylation of S635 and S637 plays an essential role in RBM20 nuclear import[57]. Using an anti-phospho-RBM20 antibody, the authors showed that, S635 and S637 are normally phosphorylated in RBM20-WT ectopically expressed in HeLa cells. This coincided with mislocalization of RBM20 upon mutation of these residues, leading to an assumption that phosphorylation is essential for nuclear import. In addition, it was shown that kinases SPRK1, CLK1, and AKT2 phosphorylate RSRSP stretch[81]. Later on, by middle-down mass spectrometry (MS), sixteen serine phosphorylation sites were identified in rat RBM20, with two of them located in the RSRSP stretch[75]. Only mutations of these two resulted in complete loss of splicing measured by TTN-reporter assay, while the mutations of the other phospho-S to Ala had only a slight impact on *TTN* splicing. However, phospho-mimicking mutations of RSRSP stretch (S638D and S640D in rat) also resulted in mislocalization and granule formation, suggesting that phosphorylation itself might not be the major driver of nuclear import[75]. Moreover, it was shown that, phosphomimicking mutations of ten Ser in the human RBM20's C-terminus after the RSRSP stretch to Asp (D10) result in less nuclear and more cytoplasmic localization, without losing the ability to interact with spliceosomal components. In turn, phospho-dead mutations of the same ten Ser in the C-terminus to Ala (A10) do not affect localization and splicing as much as D10[78]. This indicates that, phosphorylation of the C-terminus of RBM20 could be regulatory to export RBM20 from the nucleus. The authors also observed that, for 8 out of 10 Ser in the C-terminus, a Pro is located in +1 position, which suggests that CDK12 could be the responsible regulatory kinase. The authors suggest that, phosphorylation of the C-terminus might be regulatory of RBM20 localization, which is in line with previous studies showing that C-terminus together with the RS-domain are for splicing repression of a TTN minigene[56,57]. Inhibiting this phosphorylation with, for example, kinase inhibitors, could be an interesting therapeutic avenue to restore nuclear localization of RBM20 variants[78]. Altogether, while phosphorylation of the C-terminus appears to be inhibitory of nuclear localization, the exact role S635 and S637 phosphorylation in RSRSP stretch in nuclear import is yet to be determined.

1.5. Aims

Despite accumulating evidence that cytoplasmic mislocalization of RSRSP-RBM20 variants is detrimental, the exact molecular causes and consequences of this pathobiological process are still poorly understood. In particular, it has remained unknown what is the general mechanism of RBM20 nuclear import, and why mutations in the RSRSP stretch disrupt this process. In addition, it has been unclear whether RSRSP mutants confer pro-aggregative qualities to RBM20 or whether granule formation is only due to the import failure. It was shown before in HeLa cells that forcing nuclear localization of S635A/S637A variants by fusing them with the SV40 NLS could restore splicing of *TTN* minigene[57]. However, it hasn't been addressed for other RS-domain variants, for other RBM20 splicing targets, and in more physiological systems such as hiPSC-CMs or *in vivo*.

Therefore, the current work had the following aims:

1 – Characterise a recently discovered[62] P633L variant adjacent to the RSRSP stretch in terms of its cytoplasmic mislocalization and splicing deficiency in iPSC-CMs;

2 – Address whether PRSRSP-RBM20 variants retain their splice regulatory activity if localised to the nucleus;

3 – Identify factors that are required for RBM20 nuclear import;

4 – Identify factors that may retain mutated RBM20 in the cytoplasm;

5 – Explore possible ways to restore nuclear localization and dissolve cytoplasmic granules of RBM20 variants *in vitro* and *in vivo*.

Chapter 2. Materials and Methods

Some parts of the materials and methods section were adapted from Kornienko et al., *Nat Commun*, 2023[80].

2.1. Cell culture

2.1.1. Induced pluripotent stem cells (iPSCs)

In this project, I used iPSCs generated by Briganti et al.[62], who obtained WT iPSCs from the Stanford Cardiovascular Institute Biobank, and introduced single point mutations via CRISPR/Cas9 technology (see details in[62]). The protocol for culturing of iPSCs was adapted from[82]. Cell culture dishes were coated for at least 1 h at room temperature with Vitronectin (VTN-N) Recombinant Human Protein, Truncated (Gibco, A14700), prior to plating of iPSCs. Cells were cultured in monolayer in the E8-Flex medium (Gibco A2858501) and passaged two times a week using Versene solution (Gibco, 15040066). Cells were passaged in the following way: after one time rinsing with PBS, they were incubated in Versene solution (Gibco 15040066) at room temperature (RT) for 5-10 minutes. Then, Versene was aspirated, cells were flushed with fresh E8-Flex medium, pipetted up and down several times, and re-plated at the desired concentration. For routine passaging, cells were usually split in 1:8 to 1:10 dilution. RevitaCell (1:100) supplement (Gibco, A2644501) was added to the culture medium when cells were being thawed or FACS sorted. Cryopreservation medium was supplemented with 1 % RevitaCell and 10 % DMSO. The study protocol of iPSCs was approved by the EMBL Ethics committee.

2.1.2. Cardiomyocyte differentiation

Prior to initiation of cardiomyocyte differentiation, I cultured iPSCs in a monolayer as described above. I induced cardiac differentiation by modulating Wnt signaling as previously described[83]. In brief (see Fig. 2.1), iPSCs were plated on to vitronectin-coated 6-well plates at about 5% confluency (day -4) to reach about 70-80 % 4 days post plating. On Day 0, the medium was exchanged to 3 ml per well of RPMI 1640 (Gibco 21875034) with addition of B27-insulin supplement (Gibco, A1895601) (RPMI B27-ins) and 4 µM CHIR 99021 (LC Laboratories C-6556) dissolved in DMSO. On day 1, the medium was diluted by addition of 3 ml of RPMI B27-ins. On day 3, the medium was changed to 3 ml of fresh RPMI B27-ins with 2 µM Wnt-C59 inhibitor (Tocris 5148) dissolved in DMSO. On Day 5 and Day 7, the medium was exchanged to 3 ml of fresh RPMI B27-ins without any additions. On Day 9, or as soon as the first beating cells were detected, the medium was exchanged to RPMI with B27 supplement (Gibco, 17504044) (RPMI B27+ins). On Day 11, the medium was exchanged to 4 ml of RPMI 1640 with no glucose (Gibco, 11879020) with addition of 0.1% Sodium DL-lactate (L4263-100ml, Sigma). On Day 14, the medium was changed back to RPMI B27+ins. After at least day 16, wells with beating iPSC-CMs were passaged every two weeks until harvested for downstream analyses. For every passaging, freshly coated VTN-N plates were used. The passaging was done in the following way. Cells were washed once with PBS, incubated in 500 µl (per one well of a 6-well plate) of TrypLE Select Enzyme (10X) (Gibco A1217701) at 37°C for 10-15 minutes, and flushed with 2 ml of the Passaging medium. The Passaging medium (or P-medium) is composed of RPMI B27+ins supplemented with 10 % knockout serum replacement (Gibco 10828010) and 1,6 µM Thiazovivin (Stem Cell technologies, 72252). Detached cells resuspended in the P-medium were then pelleted at 350 g for 5 minutes, resuspended in fresh P-medium, and plated on to pre-coated VTN-N plates. Cells were passaged at 1:1 ratio, since they are non-dividing. The P-medium was exchanged 1-2 days after passaging to the RPMI B27+ins, after which it was exchanged twice per week until I performed downstream analyses or until the next passaging.

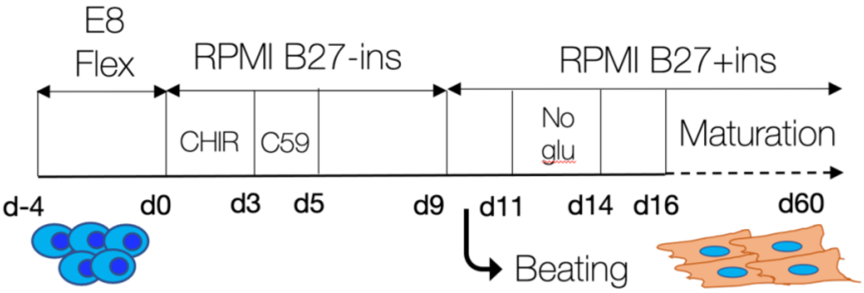


Figure 2.1. Overview of the differentiation protocol adapted from[83]**.**

2.1.3. HeLa and HEK293FT culture

I cultured HeLa Tet::Cas9 cells (described in and obtained from)[84] and HEK293FT (Thermo Fisher Scientific) in DMEM with high glucose medium (Gibco, 11965084) supplemented with 10 % FBS Supreme (Pan Biotech, P30-3031), 1 % Sodium Pyruvate (Gibco, 11360070), and 1 % Penicillin-Streptomycin (Gibco, 15140122). No plate coating was done for culturing of these cells. For passaging of these cells, they were rinsed once with PBS, incubated with Trypsin-EDTA (0.25%) (Gibco 25200056) for 5 minutes at 37°C, and flushed with the fresh culture medium. Cells were then collected to a falcon tube, diluted to the desired concentration, and plated on to the new plates or dishes. Cryopreservation was done in the culturing medium supplemented with 10 % DMSO.

2.2. Lentivirus production

For lentivirus production, I cultured HEK293FT cells as described above. About 24 hours before transfection, I plated them at about 40 % confluency to 6-well plates, so that the next day, cells would reach 80-90 % confluency. Cells were then transfected using Lipofectamine 3000 reagent (Thermo Fisher Scientific) according to manufacturer's instructions with two lentiviral packaging plasmids (pMD2.G and psPAX2), and a plasmid carrying the gene of interest mixed at 1:1:1 ratio (per well of a 6-well plate I used 833 ng each to get 2.5 µg total). Cells were split in approximately 1:6 ratio (one well of a 6-well plate into a 10 cm tissue culture dish, or two wells of a 6-well plate into a 15 cm dish) 6 hours post transfection. Then, they were cultured for three days at 37°C, releasing virus to the culture medium. After three days post transfection, the culture medium containing produced lentivirus was collected, and first filtered through the 0.45 µm filter to remove cellular debris. The filtered cell culture supernatant was then incubated at 4°C for a minimum of 1 h up to overnight with 1:3 of its volume (e.g., 3 ml for 9 ml of supernatant) of LentiX concentrator (Takara/Clontech). This was followed by centrifugation at 4°C, 4000 g, for 45 minutes, to pallet the viral particles. Then, the supernatant was aspirated, and the viral pellets were resuspended in PBS (1:200 of the initial supernatant volume, e.g., pallet from 9 ml of supernatant from a 10-cm dish was resuspended in 180 µl of PBS). Prepared virus was either used immediately or aliquoted and stored at -80°C until usage.

2.3. Engineering of HeLa reporter cell lines

I cloned the TetO-eGFP-GGSG-FLAG-RBM20 plasmid via Gibson assembly of the eGFP-GSSG cDNA (amplified from another plasmid available in the lab), Flag-RBM20 cDNA (purchased from GenScript), as well as the fragment of TetO-lenti backbone (gift from Moritz Mall lab, DKFZ, Germany), that was amplified with the oligos listed in Table 2.1, purified via gel extraction kit (Qiagen), and incubated with Gibson assembly master mix (NEB) at 50°C for 1 h. Then, 5-alpha *E. coli* (NEB) were directly transformed with the product of Gibson assembly.

For this and all other cloning procedures, I used the following transformation protocol. Competent *E. coli* (NEB) were pre-thawed on ice for 20 minutes. For each transformation reaction, 50 µl of thawed bacteria were used. They were incubated on ice for 20 minutes with about 50 ng of DNA dissolved in about 2 µl of H_2O, then heat-shocked at 42°C for 30 seconds, and recovered on ice for 5 minutes. This was followed by incubation of the transformed cells at 37°C in the 950 µl of the LB-medium (prepared by EMBL Media Kitchen) with constant shaking for 1 h. After this, 100-200 µl of recovered cells were plated on to agar-plates containing the selection antibiotic, spread homogeneously, and incubated at 37°C overnight. Next day, single colonies were picked and transferred to LB-medium supplemented with the selection antibiotic (for ampicillin, the final concentration was 0.1 mg/ml). Liquid bacterial cultures were then incubated at 37°C for 6-12 hours with constant shaking. This was followed by extraction of plasmid DNA using Mini-prep kit (Qiagen). Plasmid DNA was always eluted in H_2O and its concentration was measured by Nanodrop (NanoPhotometer N60, IMPLEN).

I cloned the TetO-eGFP-GGSG-NLS-Flag-RBM20 via site directed mutagenesis-based insertion of the SV40 NLS sequence (ccaaaaaagaagagaaaggta) into the TetO-eGFP-GGSG-FLAG-RBM20 plasmid (see oligonucleotide sequences in Table 2.1) with GeneArt Site Directed Mutagenesis Kit (Thermo Fisher Scientific).

To introduce all single point mutations used in this study into the *RBM20* cDNA sequence, I used GeneArt Site Directed Mutagenesis Kit (Thermo Fisher Scientific) according to manufacturer's instructions, and oligonucleotides listed in Table 2.1.

To clone the pEFa-eGFP–Flag–RBM20-WT, P633L, R634Q, and R634Q-S635E-S637E, I replaced TetO- promoter in the corresponding TetO-eGFP-GGSG-FLAG-RBM20 plasmids by the EFa-promoter sequence from the Addgene #125592 plasmid. This was achieved by Gibson assembly (NEB) of the amplified TetO-backbone (see above), eGFP-GSSG-Flag-RBM20 cDNA from the TetO-plasmids (described above), and the pEFa promoter (see oligos in Table 2.1).

For generation of TetO-NLS-tagged RBM20-expressing HeLa reporter cells, wildtype HeLa Kyoto cells were co-transduced with two lentivirus solutions delivering TetO-eGFP-FLAG-NLS-RBM20-WT, -P633L, or -R634Q, as well as rtTA (Addgene #20342). Cells were then cultured in the presence of 2 μg/ml of Doxycycline (Sigma) for at least 7 days, single cell sorted for eGFP fluorescence marker into 96-well plates with FACS, and used for the experiment at least two weeks after the sort.

For generation of HeLa reporter cell lines with inducible Cas9 and constitutive eGFP-RBM20 expression, HeLa TetO::Cas9 (provided by Paul Blainey (Broad Institute), generated in Iain Cheeseman's lab (MIT), and described in [84]) were transduced with lentivirus delivering pEFa-eGFP–Flag–RBM20-WT, P633L, R634Q, or R634Q-S635E-S637E. Transduced cells were then single-cell sorted into 96-well plates via FACS three days post transduction (single cell sorting was done by FCCF at EMBL). Two weeks after single cell sorting, I analysed the established lines for their purity with FACS and ICS, and the most stable and used pure clonal lines for downstream applications e.g., CRISPR/Cas9-ICS screens (see below).
All plasmids were checked by Sanger sequencing done by Eurofins.

Table 2.1. Oligonucleotides used for cloning of plasmids to generate HeLa reporter cell lines

Name	Forward primer	Reverse primer
TetO-eGFP-GGSG-Flag-RBM20: eGFP cDNA	agcctccgcggccccgaattcgccaccatggtgagcaagg	ttgtagtcaccagaaccacccttgtacagctcgtccatgc
TetO-eGFP-GGSG-Flag-RBM20: Flag-RBM20 cDNA	gggtggttctggtgactacaaggacgacgatgacaaggtg	atagtcttctctagaggatcctcagagcttttcctttcg
TetO-eGFP-GGSG-Flag-RBM20: TetO-lenti backbone	gatcctctagagaagactat	aattcggggccgcggaggct
N-NLS-tag site directed mutagenesis	ccaaaaaagaagagaaaggtagactacaaggacgacgatga	tacctttctcttcttttttggaccagaaccacccttgtaca
P633L site directed mutagenesis	tatggcccagaaaggctgcggtctcgtagtccg	cggactacgagaccgcagcctttctgggccata

R634Q site directed mutagenesis	gcccagaaaggccgcagtctcgtagtccggt	accggactacgagactgcggcctttctgggc
R634Q-S635E site directed mutagenesis	ccagaaaggccgcaggaacgtagtccggtgagc	gctcaccggactacgttcctgcggcctttctgg
R634Q-S635E-S637E site directed mutagenesis	aggccgcaggaacgtgaaccggtgagccggtca	tgaccggctcaccggttcacgttcctgcggcct
pEFa-eGFP-RBM20: eGFP-GSSG-RBM20-cDNA	ggtgtcgtgacgcgggatccgccaccatggtgagcaaggg	atagtcttctctagaggatcctcagagcttttcctttcg
pEFa-eGFP-RBM20: backbone from the TetO-plasmids	gatcctctagagaagactat	cactcgagccggcgccaaagtggatctctgctgtccctgt
pEFa-eGFP-RBM20: promoter	acagggacagcagagatccactttggcgccggctcgagtg	cccttgctcaccatggtggcggatcccgcgtcacgacacc

2.4. Cell treatments

2.4.1. siRNA knokdown

For siRNA transfections, I cultured HeLa Tet::Cas9 cells until they reached about 20 % in the normal culture medium but without Penicillin Streptomycin. I transfected them with Lipofectamine RNAiMAX reagent (Thermo Fisher Scientific) and 20 nM of either non-targeting control siRNA (D-001810-02), or siRNA targeting *TNPO3* (L-019949-01), targeting *DDI2* (84301), targeting *WTAP* (9589), or targeting *PPP5C* (5536) – all from Dharmacon, Horizon Discovery Group. I exchanged medium to the normal HeLa culture medium described above 24 hours post transfection, and then cells were incubated at 37°C until harvesting for further analyses, usually 72 hours post transfection.

For siRNA transfections of differentiated iPSC-CMs, I cultured them at 70-90 % confluency in the normal RPMI B27+ins medium, and transfected with LipofectamineStem reagent (Thermo Fisher Scientific) and 100 nM of either non-targeting control siRNA or siRNA targeting *TNPO3* (see references above). I exchanged medium to the fresh RPMI B27+ins media 24 hours post transfection. Four days after the first transfection, I transfected the cells again in the same way as described above. Three days after the second transfection, I harvested cells harvested for downstream analyses.

2.4.2. TNPO3 overexpression in iPSC-CMs

I purchased the *eGFP-TNPO3* cDNA from Addgene (167590), and produced lentivirus using the purified plasmid as described above. I cultured iPSC-CMs at 70-90 % confluency in the normal RPMI B27+ins medium, and transduced them with 1:1000 medium volume of the concentrated virus (see above). Medium was exchanged with the fresh RPMI B27+ins on the next day after transduction, and then once more three days later. A week after transduction, I either fixed cells for further microscopy analysis of RBM20 localization (see below), or bulk FACS-sorted for further RNA extraction (see below), to directly compare gene expression and splicing between eGFP-TNPO3 positive and negative cells.

2.4.3. Overexpression of RBM20 variants in RBM20-deficient iPSC-CMs

I cloned and packaged the constructs delivering TetO-eGFP-FLAG-RBM20-WT, -R634Q, or NLS-R634Q into lentiviral particles, as described above. I cultured iPSC-CMs with S635FS mutation generated and described by Briganti et al.[62] as described above at 70-90 % confluency in the RPMI B27+ins medium. I transduced them with 1:1000 medium volume of the concentrated virus (as described above) delivering one of the TetO-RBM20 constructs, as well as rtTA (Addgene 20342). On the next day, the medium was exchanged by the fresh RPMI B27+ins with addition of 2 μg/ml of Doxycycline (Sigma). Then, three days later, the medium was exchanged again the RPMI B27+ins with Doxycycline, and after a total of one week post transduction, I harvested the cells for downstream analyses.

2.4.4. Knockout of TNPO3 in iPSC-CMs

iPS-cell lines stably expressing Cas9 were engineered by Dr. Dominik Lindenhofer and Moritz Kueblbeck in our lab. They used a transgene carrying a CAG-Cas9-P2A-dTomato cassette and inserted it into the AAVS1 locus of WTC-11 cells (Coriell Institute for Medical Research – GM25256) as was preciously described in[85]. Prior to my experiments, they also functionally tested the cells for Cas9 activity using a set of control gRNAs described in[85], and read out dTomato signal using flow cytometry, as well as assessed chromosomal integrity using the Infinium CoreExome-24 v1.4. Kit (Illumina).

For my experiments, I used the engineered RBM20-WT iPSCs with constitutive Cas9, and differentiated them into iPSC-CMs as described above. I plated cells onto 24-well microscopy plates with the first passaging. Next day after passaging, they were transfected with a mix of three sgRNAs targeting *TNPO3* (500ng of each per well, see sequences in Table 2.2), using LipofectamineStem reagent (Thermo Fisher Scientific). One day later, the medium was exchanged by the fresh RPMI B27+ins. Four days after the first transfection, I repeated the transfection again in the same way as the first one, with exchanging the medium 24 hours after. Three days after the second transfection, I fixed the cells and used them for the IF staining and the downstream microscopy analyses (see below).

Table 2.2. sgRNA oligonucleotides for *TNPO3* knockout in iPSC-CMs

Targeted gene	Forward	Reverse
TNPO3-KO-1	CACCGAGTCCGAAGCAGCTTCATGT	AAACACATGAAGCTGCTTCGGACTC
TNPO3-KO-2	CACCGGAAAGAAGGCAACCCACCCT	AAACAGGGTGGGTTGCCTTCTTTCC
TNPO3-KO-3	CACCGAAGGCGATCTAAGAACACTG	AAACCAGTGTTCTTAGATCGCCTTC

2.4.5. Leptomycin B treatment

I cultured HeLa reporter cells with TetO-RBM20 in the presence of 2 µg/ml Doxycycline (Sigma) for at least a week prior to the experiment. I plated cells on microscopy plates, and treated with 50 ng of Leptomycin B (Sigma) dissolved in DMSO, which was followed by fixation and staining (see below) after 5 or 20 hours. Control cells were treated with the same volume of DMSO.

2.5. Cell staining

For FACS and ICS, I used viability staining to enable sorting of live cells only. For that, harvested and resuspended in PBS cells were stained with 50 µg/ml final concentration of DAPI (Thermo Fisher Scientific), or one drop of DRAQ7 (Biostatus) per 300 µl of cell suspension. To stain the nuclei of live cells, I used 100 mM of DRAQ5 (Biostatus) and added to the cell suspension in PBS at room temperature in 1:50 ratio. Cells were further used for FACS or ICS at least 5 minutes after addition of DAPI or DRAQ5.

2.5.1. Alkaline phosphatase (AP) staining

I grew iPSCs in a monolayer (see above), washed them twice with DMEM-F12 (Gibco), and incubated in the AP staining solution, containing 1:500 of the AP live stain dye (Thermo Fisher Scientific, A14353) in DMEM-F12, for 20-30 minutes. Then, cells were washed twice more with DMEM-F12, and imaged on the widefield fluorescent microscope (see below) using FITC filter.

2.5.2. Antibody staining for ICS

To measure endogenous RBM20 localization with ICS, as well as to sort cells based on this parameter, I harvested iPSC-CMs as described above, then washed them once with PBS, and resuspended them in a fixing solution containing 4% formaldehyde (FA) in PBS. This was followed by incubation at room temperature (RT) for 10 minutes. Then, cells were washed once with PBS, resuspended in a permeabilising solution containing with 0.1 % Triton-X100 (Merck) in 1 % BSA (Sigma) in PBS, and incubated for 5 minutes at RT. Then, after one time washing with PBS, I resuspended cells in a staining solution containing a 1:100 dilution of the anti-RBM20 (ab233147, Abcam, see table 2.3) antibody in 1% BSA in PBS, and incubated them for 1 hour at RT, followed by another wash with PBS. Then, I resuspended cells in the secondary antibody staining solution containing a 1:500 dilution of AlexaFluor488 goat anti rabbit antibody (Invitrogen, see table 2.3) in 1 % BSA and incubated for 30 minutes at 4°C in the dark. Finally, cells were washed once more with PBS, and resuspended in the final staining solution containing PBS with DAPI and DRAQ5 at the concentrations derived above. After at least five minutes, I took these stained cells for ICS.

2.5.3. Antibody staining for microscopy

For microscopy analyses, I cultured cells either in glass bottom plates, or on coverslips, under otherwise normal cell culture conditions as described above. For staining, cells were rinsed one time with PBS, and incubated in the fixing solution containing 4 % FA in PBS for 10 minutes at RT. Then, after being rinsed once with PBS, they were incubated with the permeabilizing solution containing 0.5 % Triton-X100 (Merck) in PBS for 5 minutes at RT. Then, to block any potential nonspecific antibody binding sites, I incubated them in a blocking solution containing 2 % BSA (Sigma) in PBS for 1 hour at RT. Then, cells were stained with the primary antibody solution containing a 1:250 dilution of primary antibodies in 1 % BSA in PBS at 4°C overnight (see the list of antibodies used in Table 2.3). In order to wash away the residual antibody solution, I then washed the cells 3 times with 2 % BSA in PBS for at least 5 minutes each time at RT. Then, they were incubated with fluorescent secondary antibody solution containing AlexaFluor antibodies (Invitrogen, see Table 2.3) at 1:500 dilution each in 1 % BSA in PBS for 1 hour at RT in the dark. After this, cells were washed two more times with 2 % BSA in PBS at RT for at least five minutes each time in the dark. Then, I incubated cells initially cultured on the microscopy plates in the nuclei-staining solution containing 2 µg/ml Hoechst 33258 (Invitrogen) in PBS for 10 minutes at RT in the dark. I incubated cells cultured on coverslips with just PBS for 10 minutes at RT in the dark for coverslips. Microscopy plates were then washed once with PBS and stored at 4°C in the dark until imaged. Coverslips were mounted with ProLong Gold antifade reagent with DAPI (Invitrogen) and stored at 4°C in the dark until imaged.

Table 2.3. Antibodies used in the study.

Antibody	Brand and reference	Use
anti-TNPO3	InvitrogenMA5-37991	WB, 1:1000
anti-GAPDH	Abcam ab9485	WB, 1:1000
anti-MOV10	Sigma PLA0195	WB, 1:000; Microscopy, 1:250
anti-RBM20	Abcam ab233147	WB, 1:1000; Microscopy, 1:250; ICS, 1:100
anti-H3	Abcam ab176842	WB, 1:1000
anti-RBM20	Invitrogen PA5-58068	Microscopy, 1:250
anti-alpha-Actinin	Abcam ab9465	Microscopy, 1:250
anti-TNPO3	Abcam ab54353	Microscopy, 1:250
anti-p62	Sigma P0067-25UL	Microscopy, 1:250
Goat-anti-rabbit-HRP	Abcam ab97051	WB, 1:10,000

AlexaFluor488 goat-anti-rabbit	Invitrogen A32731	Microscopy, 1:500
AlexaFluor488 goat-anti-mouse	Invitrogen A11001	Microscopy, 1:500
AlexaFluor568 goat-anti-rabbit	Invitrogen A11011	Microscopy, 1:500
AlexaFluor568 goat-anti-mouse	Invitrogen A11004	Microscopy, 1:500

2.6. FACS and ICS

Prior to FACS and ICS, I prepared and stained cells. Then I filtered the prepared cell suspensions through a 35 µM cell strainer to avoid clumping, and kept them on ice until processed.

Both singe and bulk FACS were done by the Flow Cytometry Core Facility (FCCF). For single cell FACS, they sorted cells based on the desired fluorophore expression into 96-well plates containing warm culture media with antibiotics (Penicillin Streptomycin, Gibco), one cell per well. For bulk FACS, they sorted cells into 1.5 ml microcentrifuge tubes containing warm culture media with antibiotics, based on the desired fluorophore expression. For all sterile sorts, BD FACSAria™ Fusion was used, with a 100 µm sort nozzle.

For routine checking of transfection/transduction efficacies, as well as for optimising experimental conditions, BD LSRFortessa™ was used, and was operated by me.

Image-enabled cell sorting (ICS) is based on the BD CellView™ Imaging Technology as previously described[86]. ICS experiments were performed with a 100 µm sort nozzle, with the piezoelectric transducer driven at 34 kHz, automated stream setup by BD FACSChorus™ Software, and a system pressure of 20 psi. All sorts were performed in purity mode. The sorter was operated by Dr. Marta Rodríguez-Martínez and myself, with input from the FCCF and Dr. Daniel Schraivogel. For correlation-based sorts and measurements, cells were stained with DRAQ5 as described above, and for each cell, a Pearson correlation coefficient R was calculated based on the overlap between the RBM20 and DRAQ5 channels.

Flow cytometry and ICS data were analysed by me using FlowJo_v10.7.1_CL software.

2.7. Microscopy

Widefield fluorescence microscopy analysis was performed by me using Zeiss Cellobserver microscope equipped with an AxioCam camera, using Plan-APOCHROMAT 20x NA0.8 Air DIC2 or LD Plan-NEOFLUAR 40x NA 0.6Air Ph2 correction collar 0-1.5 objectives.

I performed confocal microscopy analysis using Zeiss LSM900 microscope equipped with 405 nm – 5 mW, 488 nm – 10 mW, 561 nm – 10 mW, 640 nm – 5 mW lasers, using Objective Plan-Apochromat 40x/0.95 Corr M27 air (FWD=0.25mm) objective, and 3 Gallium Arsenide Phosphid-PMT (GaAsP-PMT) for fluorescence detection, standard PMT as transmission detector.

2.7.1. Co-localization analysis

For colocalization analysis of RBM20 and DAPI, I used the Fiji (v.2.1.0/1.53c) plugin Coloc2 that quantifies Pearson correlation coefficient R. For confocal images, Z-stack images were max projected, and fluorescent channels were split. Area covering at least five cells was selected in RBM20 channel, and was used as ROI/mask for quantification of its correlation with DAPI channel.

2.8. RNA extraction

In all cases, RNA concentration was determined using Qubit High Sensitivity RNA kit (Thermo Fisher Scientific), according to the manufacturer's instructions. All RNA exctraction was done by me.

2.8.1. Live cells

Cells cultured in tissue culture dishes were either directly washed 3 times with PBS and lysed by adding the TRIzol reagent (Invitrogen) while still attached to the culture dish, or harvested, washed once with PBS, and lysed by resuspending the pallet in the

TRIzol reagent. Then, cells in TRIzol were transferred to 1.5 ml reaction tubes. Direct-zol RNA Miniprep Plus Kit (Zymo Research) was used for RNA extraction and DNAseI treatment, according to the manufacturer's instructions.

2.8.2. Fixed cells

For RNA extraction from fixed iPSC-CMs post ICS or FACS, at least 5000 cells per sample were used. They were pelleted by centrifugation at 500 g for 3 minutes and resuspended in 16 µl of a 1:16 proteinase K in PKD buffer (Qiagen), incubated at RT for 5 minutes, briefly palleted down, and incubated at 56°C for one hour for reverse crosslinking. The solution was then resuspended in 100 µl of TRIzol LS Reagent (Thermo Fisher Scientific). After this, 20 µl of chloroform was added to the TRIzol-sample. Sample was phase separated by vigorous shaking and centrifugation at 12,500 r.p.m. for 5 min at RT. The aqueous phase (about 40 µl) was then collected from each sample to a new sample tube, and mixed with 75.5 µl of isopropanol with addition of 1:150 of glycoblue (Invitrogen) as a coprecipitant. The RNA samples were then dehydrated at -80°C for 24-36 hours. Then, precipitated RNA was pelleted by centrifugation at 4°C at maximum speed for 15 minutes. The supernatant was aspirated and discarded, and the pellet was washed once with 70% ethanol, let air dry, and resuspended in 8 µl of nuclease-free water.

2.8.3. Tissue

A piece of left ventricle was homogenised in 600 µl PBS in tubes containing metallic beads with a Fastprep homogenizer with three 15 s runs at maximum velocity. The homogenised tissue was then spun down, and the pellet was then resuspended in TRIzol (Invitrogen). RNA extraction and DNAse I treatment were then performed as above, using the Direct-zol RNA Miniprep Plus Kit (Zymo Research), according to the manufacturer's instructions.

2.9. Quantitative RT-PCR

For quantitative reverse transcription polymerase chain reaction (RT-qPCR) I synthetised cDNA using the SuperScript IV (Thermo Fisher Scientific) according to the manufacturer's instruction, with addition of 0.5 mM of each dNTP (NEB), 2.5 µM oligo-dT (Thermo Fisher Scientific), 1.25 µM random hexamer primers (Invitrogen), 5 mM DTT (Thermo Fisher Scientific), 2 u/µl RNAse inhibitor (Invitrogen), 1X SSIV buffer (Thermo Fisher Scientific), and 10 u/µl SSIV RT (Thermo Fisher Scientific), per each reaction. I used at least 10 ng and a maximum of 1 µg of total RNA per reaction. For RNA samples from the same replicate of the same experiment, RNA input was equalised across samples.

I performed a one-step qPCR reaction (95°C for 10 minutes, 40 cycles of [95°C 15 seconds, 60°C 1 minute]) using SYBR Green Master Mix (Thermo Fisher Scientific), and primers listed in Table 2.4. The Applied Biosystems StepOnePlus Real-Time PCR System (272006365), and StepOne Software v2.3 were used for qPCR. I quantified delta-delta cT (ΔΔCt) values versus *GAPDH* as a housekeeping gene, and versus a control sample for each experiment. All qPCR plots display fold change values (FC, 2^-ΔΔCt).

Table 2.4. Oligonucleotides used for RT-qPCR.

Name	Forward	Reverse
SOX2	ggtgtgaaccatgagaagtatga	gagtccttccacgataccaaag
OCT4	gagagaaagaagaggagagagaaag	gccgccgatgattgttattatt
KLF4	gccaacacagccatctactat	ccacatcaaagtcaggcatttc
VIM	tcgccttgctgattgtctatt	aattggccgagatccttcttc
hsGAPDH	tcgacagtcagccgcatctt	ctccgaccttcaccttcccc
hsTTN spliced-in	aattcaactgggggttctttcac	ctgagggatgtaaagttagaagatgc
hsTTN spliced-out	gaattccacatgaggagctttcac	ctgagggatgtaaagttagaagatgc
hsRBM20	cagagggagagggacatgttcc	ggagggctgtgggaagagctgc
hsTNPO3	agcaagtcactagtgctgagg	cagtgcaggagtgtgagcta
hsIMMT spliced-in	acctgcactttcagaagaagc	ttttcctgttgtgcaaggcg
hsIMMT spliced-out	tgcactttcagctctagcca	agaattgtccatggcggctt
mRbm20	tccattcccagagggagaggga	gagaggggcctgggggac
mTnpo3	gagagtgtcagacctggtgaag	agagaaccgcttctgtcacctc
mGapdh	aacagcaactcccactcttc	cctgttgctgtagccgtatt
mTtn unspliced	cagaaggaagagttcacacgc	aaccgggggttctttcac
mTtn spliced	cagaaggaagagttcacacgc	attccacatgaggagctttcac
mTtn RT-PCR	gtccacgaggaatgggagga	ttgtcacaggaacaggaatc

2.10. RNA sequencing

2.10.1 Library preparation and sequencing

Prior to library preparation, I routinely checked RNA quality for all samples with the RNA pico or nano Bioanalyzer (Agilent) kit, depending on the RNA concentration.

I prepared RNA sequencing libraries for RNA extracted from fixed iPSC-CMs, using 1-10 ng of RNA as input. To prepare libraries based on ultra-low-input and highly degraded RNA (RNA Integrity Number, RIN < 7), I used SMARTer RNA-Seq Kit v3 - pico (Takara Bio), according to the manufacturer's instructions. The fragmentation step was omitted. After completing the protocol, I checked prepared libraries with attached unique dual index barcodes for each sample again on Bioanalyzer (Agilent). After ensuring the quality of prepared libraries, I pooled them together at equimolar concentrations, with six libraries per pool. Each pool was then sequenced individually, with final cDNA sample concentrations of 8-10 ng/ml for each pool. A 2.1 pM solution of each pool was loaded on the Illumina sequencer NextSeq 500 and submitted to the Genomics Core Facility (GeneCore). They sequenced each pool bi-directionally, generating ~500 million paired-end reads, each 75 bases long. They then performed demultiplexing of the obtained reads based on the unique dual barcodes into separate fastq files. After demultiplexing, they transferred data back to me for analysis. Each sample had between 50,000,000-100,000,000 reads in total. Quality control of the sequencing data was done using FASTQC (v0.11.5).

Library preparation for HeLa cells was done based on RNA extracts from live cells (see above). Barcoded stranded mRNA-seq libraries were prepared by the GeneCore from 150 ng of high quality total RNA samples that I prepared and submitted. They used the NEBNext Poly(A) mRNA Magnetic Isolation Module and NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs (NEB), Ipswich, MA, USA) implemented on the liquid handling robot Beckman i7. Obtained libraries that passed the QC step were pooled in equimolar amounts; 2.1 pM solution of this pool was loaded on the Illumina sequencer NextSeq 500 and sequenced bi-directionally, generating ~500 million paired-end reads, each 75 bases long.

Library preparation for mouse left ventricle tissues was done based on RNA extracts from tissue (see above). Individually barcoded 'stranded' mRNA-seq libraries were prepared by the GeneCore from high quality total RNA samples (prepared by me, ~100 ng/sample) using the New England Biolabs NEBNext RNA Ultra II Kit with 14 PCR cycles implemented on the liquid handling robot Beckman i7. Obtained libraries that passed the QC step were pooled in equimolar amounts; 650 pM solution of this pool was loaded on the Illumina sequencer NextSeq 2000 and sequenced bi-directionally with the P3 kit, generating ~1,300 million of sequence pairs (pair-end reads), each 110 bases long. After demultiplexing, they transferred data back to me for analysis. Each sample had between at least 50,000,000 reads in total. Quality control of the sequencing data was done using FASTQC (v0.11.5).

2.10.2. Data analysis

All the data analysis was done by me, as follows.

For human samples, obtained demultiplexed pair-end reads were aligned to Homo_Sapiens.GRCh38.101 using STAR[87] (v2.7.5c), and bam files were sorted by coordinate using samtools (v.1.9). Read count files were generated with featureCounts[88] v1.6.4, summarizing read numbers covering each gene. Both raw reads and read count files are deposited at GEO (GSE220833).

For mouse samples, obtained demultiplexed pair-end reads were aligned to Mus_musculus.GRCm39.109 using STAR[87] (v. 2.7.9a). Read count files were generated with featureCounts[88] v. 2.0.6, summarizing read numbers covering each gene. The mouse RNA-sequencing data has not been published yet.

To perform differential gene expression (DEG) analysis in pairwise comparisons between genotypes, the DeSeq2[89] (v. 1.36.0) R package was used with the design expression ~ genotype, according to the published vignette. The Benjamin & Hochberg (BH) method was used as a p value adjustment method. Criteria for considering a gene as differentially expressed were: log2 of its expression fold change greater than 0.5, and the adjusted p-value less than 0.1, for human cells. For DEG analysis in mice, a gene was considered differentially expressed if log2 of its expression fold change greater than 1, and the adjusted p-value less than 0.05. Pathway enrichment analysis for lists of differentially expressed genes was performed using Metascape[90].

Global alternative splicing changes were analysed in pairwise comparisons to the RBM20-WT expressing cells or RBM20-WT-*Ctr* mice using rMATS turbo[91] (v.4.1.1) based on the junction counts (JC), according to the published manual. rMATS classifies splicing events into exon skipping (SE), intron retention (RI), mutually exclusive exons (MXE), alternative 3' splice site (A3SS), and alternative 5'splice site (A5SS). Significant alternative splicing events were determined as events with the absolute values of their inclusion level differences compared to the respective WT controls greater than 0.1, and with the false discovery rate less than 0.01.

To assess splicing differences specifically in known and previously validated RBM20 target genes, I took genes listed and summarised by Koelemen et al.[47]. This list consists of 45 genes, splicing of which was consistently affected across RBM20 perturbations (KO, S635A) and across model species (rat, human, mouse) based on several studies[50,53,54]. I refer to this list throughout this dissertation as "core RBM20 targets". In order to calculate absolute percentage of spliced-in values (PSI-values) for all exons and their parts across samples, without restricting to only pairwise comparisons to RBM20-WT control cell lines, I used pipeline described in[92] as follows. First, for each exonic part (annotation was done based on DEXSeq[93], v.1.42.0), numbers of inclusive and exclusive reads were identified. A read is classified as inclusive, if it includes the exon or the exonic part of interest. A read is classified as exclusive, if it spans both the upstream and downstream exons or their parts, but does not include any parts of the exon of interest. For each exonic part, inclusive and exclusive reads were counted directly from the BAM files based on STAR output as described in[92]. The PSI-values are then quantified per each exonic part as ratios of inclusive reads to the sum of inclusive and exclusive reads. I then performed a Student's T-test to determine mis-spliced exons of the core RBM20 targets. The criteria for such classification were: p-value < 0.05, and the absolute value of differences between means of PSI-values greater than 0.1, for comparisons indicated in the Results chapter. Calculated for iPSC-CM experiments PSI values are deposited at GEO (GSE220833).

2.11. CRISPR/Cas9 gRNA library design and cloning

The genome-wide CRISPR/Cas9 gRNA libraries used in this study were designed as described by Schraivogel et al.[86]. Briefly, the library is composed of 18,408 sgRNAs targeting all protein-coding genes according to the Consensus Coding Sequence Database[94]. It is built of six independent sub-libraries, with one sgRNA per gene in each of them[86]. In addition to targeting genes, each of these sub-libraries also delivers 118 targeting and 487 nontargeting sgRNA controls, same for all sub-libraries. The library was cloned into the CROPSeq-guide(F+E)-Puro backbone harborouring puromycin-resistance gene for selection of successfully transduced cells, by Dr. Daniel Schraivogel, as described in[86]. The plasmid library was sequenced previously[86] to check the gRNA representation without any natural selection. This sequenced plasmid library from[86] was used as a plasmid stock for the lentivirus generation for this work. The sequencing data were also used as a reference for gRNA representation at the plasmid stage for the downstream data processing (see below).

2.12. Cloning of individual gRNAs

To introduce individual gene knockouts and test hits from the CRISPR ICS screens, I, together with my intern Maral Azodi and master student Brigit Tunaj, synthesised gRNAs as two short oligos with flanking sequences resembling the Esp3I sticky ends of CROPSeq-guide(F+E)-Puro vector (fwd 5'- CACCG[N20], rev 5'-AAAC[N20-reverse complement]), see sequences in Table 2.5.

First, the oligos (10 mM each) were phosphorylated and annealed using 1 U/µl T4 PNK (NEB), and 1X T4-ligase buffer (NEB) in a thermocycler with the following program: 37°C 30 min, 65°C 20 min, 95°C 5 min, ramp down to 25°C at 5°C/min. Esp3I-digested (NEB) and gel-extracted (Qiagen) CROPSeq-guide(F+E)-Puro backbone (25 ng) was then ligated with the phosphorylated and annealed oligos (1 µl) using 1 µl of T4 ligase (NEB), and 1X T4 ligase buffer (NEB) in total volume of 10 µl, for 10 minutes at RT, inactivated for 10 min at 65 °C followed by transformation into the NEB Stable competent *E. coli* (NEB), according to the manufacturer's instructions. We prepared and concentrated lentivirus delivering these sgRNAs for cell transductions as described above.

Table 2.5. Oligonucleotides used for cloning of individual sgRNAs

Targeted gene	Forward	Reverse
TNPO3	CACCGGGGATGTGTGCAAACACTGG	AAACCCAGTGTTTGCACACATCCCC
CLDN14	CACCGCGATTGTCTTTGTAGGCAGC	AAACGCTGCCTACAAAGACAATCGC
ADAMTS16	CACCGGCAGGCACCTGCGTGCGCCA	AAACTGGCGCACGCAGGTGCCTGCC
GALE	CACCGGAGAAGGTGCTGGTAACAGG	AAACCCTGTTACCAGCACCTTCTCC
SLC29A2	CACCGGGGCGTGATAAAGTACCCCA	AAACTGGGGTACTTTATCACGCCCC
CEBPB	CACCGATGCTGGGTCCCAGGCCACC	AAACGGTGGCCTGGGACCCAGCATC
UBQLNL	CACCGGAGTGGCACAGATATGGCTC	AAACGAGCCATATCTGTGCCACTCC
TRIM33	CACCGGACACACGGCGCAGGTGTCC	AAACGGACACCTGCGCCGTGTGTCC
PMM2	CACCGATCGGACTTTGAGAAAGTGC	AAACGCACTTTCTCAAAGTCCGATC
TRIM24	CACCGGAGGACAACGCAGAAGCCAA	AAACTTGGCTTCTGCGTTGTCCTCC
IPPK	CACCGCAAAGATGCCCGGAGCCCCG	AAACCGGGGCTCCGGGCATCTTTGC
TTN	CACCGGTGCAGATCTCCTTTAGCGA	AAACTCGCTAAAGGAGATCTGCACC
XPO6	CACCGCACCTTATCGACAAGTCGCA	AAACTGCGACTTGTCGATAAGGTGC
AKT2	CACCGCATCGAGAGGACCTTCCACG	AAACCGTGGAAGGTCCTCTCGATGC
CLK1	CACCGTACACTCAAGGATGTGAACC	AAACGGTTCACATCCTTGAGTGTAC
SPRK1	CACCGATGGAGAAAGAGTCGGGCCC	AAACGGGCCCGACTCTTTCTCCATC
LMNA	CACCGCCATGGAGACCCCGTCCCAG	AAACCTGGGACGGGGTCTCCATGGC
CCDC12	CACCGACTCACCACCTCCTCGATGA	AAACTCATCGAGGAGGTGGTGAGTC
CEBPB	CACCGATGCTGGGTCCCAGGCCACC	AAACGGTGGCCTGGGACCCAGCATC
COPS2	CACCGGGCCAAGATGTCTGACATGG	AAACCCATGTCAGACATCTTGGCCC
DBR1	CACCGAAGAGTTACCCTATGGTGGC	AAACGCCACCATAGGGTAACTCTTC
DDI2	CACCGGTGTAACATAATGAGACTGG	AAACCCAGTCTCATTATGTTACACC
DDX21	CACCGCCATCATGTTTACAGCGGGA	AAACTCCCGCTGTAAACATGATGGC
EXOSC8	CACCGAATCTCCTGTAATACTCCAG	AAACCTGGAGTATTACAGGAGATTC
FANCM	CACCGGAAATTGTACATGACCACGG	AAACCCGTGGTCATGTACAATTTCC
FOXJ1	CACCGGAGAAGGACGAACCAGGCAA	AAACTTGCCTGGTTCGTCCTTCTCC
GCN1	CACCGGGTCCGGCGACAGGACGGAA	AAACTTCCGTCCTGTCGCCGGACC
GLDN14	CACCGCGATTGTCTTTGTAGGCAGC	AAACGCTGCCTACAAAGACAATCGC
HNRNPH	CACCGGCTCGGCCGATGAAGTGCAG	AAACCTGCACTTCATCGGCCGAGCC
IL3	CACCGGGCGGACCAGGAGTTGGAGC	AAACGCTCCAACTCCTGGTCCGCCC
METTL14	CACCGGGGGTTGGACCTTGGAAGAG	AAACCTCTTCCAAGGTCCAACCCCC
NPEPPS	CACCGTTGTGATAGGGACCATCCAC	AAACGTGGATGGTCCCTATCACAAC
OPA1	CACCGCAAGCTCGAATATTCCCAAG	AAACCTTGGGAATATTCGAGCTTGC
PPP5C	CACCGCATCAAGGGTTATTACCGCC	AAACGGCGGTAATAACCCTTGATGC
RBM39	CACCGTGGCGGCAAGAATTCGACCA	AAACTGGTCGAATTCTTGCCGCCAC
SLSF3	CACCGGCAACCCACACACTTCGGAG	AAACCTCCGAAGTGTGTGGGTTGCC

SNAPC4	CACCGGACACTGGGAGCCAGACCGA	AAACTCGGTCTGGCTCCCAGTGTCC
WDR46	CACCGTATGTGCGAGATCAACGTCA	AAACTGACGTTGATCTCGCACATAC
WTAP	CACCGACAGCAGGAGTCTGCACGCA	AAACTGCGTGCAGACTCCTGCTGTC

2.13. Pooled and individual CRISPR perturbations

For the pooled perturbations in CRISPR-ICS screen experiments, I plated HeLa Tet::Cas9 pEFa-eGFP-RBM20-WT or -R634Q with a density of 750,000 cells per 15 cm tissue culture dish, and cultured for three days until a confluency of 40 % was reached (6,000,000 cells per 15 cm dish). In order to get the coverage of at least 500 cells per each sgRNA in the library with about 20% infection rate, $60x10^6$ of cells were transduced per each genome-wide library (ten 15-cm dishes per library). I used 25 µl of 100X concentrated virus per plate, to achieve the low infectivity rate at which only single qRNA is integrated per cell. On the next day after transduction, I treated resuspended cells with 2 µg/ml of Puromycin (Thermo Fisher Scientific) and replated them back to the same tissue culture dishes. The day after, to wash away dead cells, I exchanged the medium to fresh Puromycin-containing culture medium. Roughly 20% of cells were successfully transduced as judged by puromycin resistance. These cells were further kept in culture. After letting the cells grow for three days, I was further splitting them every three days in the presence of 2 µg/ml Puromycin (Thermo Fisher Scientific) and 2 µg/ml Doxycycline (Sigma) to activate Cas9, for the next week. Next, the culturing medium was switched back to the one containing only puromycin (no doxycycline), and I cultured cells at normal conditions (see above) until they were harvested for ICS (at least 17 days after the transduction day). To keep the sgRNA coverage at > 500X, cells infected with the same library from all plates were pooled together after trypsinization at each splitting, and 1,500,000 of cells were plated per each new 15-cm dish, with seven dishes per each library. Three days prior to harvesting for ICS, I plated fifteen 15-cm dishes with 1,500,000 cells each for each library, and used all of them for sorting to achieve enough coverage in each sorting bin. Sorting was done with Dr. Marta Rodríguez-Martínez.

For testing hits of the CRISPR ICS screen experiments, we performed individual CRISPR perturbations together with my students Maral Azodi and Brigit Tunaj. For that, cells were cultured in 6-well plates until 40 % confluency (150,000 cells per well), and 10 µl of 50X concentrated virus delivering a sgRNA with puromycin resistance was added per each well. Twenty-four hours later, transduced cells were resuspended in the medium containing 2 ng/µl Puromycin (Thermo Fisher Scientific), and plated back to the same wells. Next day, to wash away dead cells, the medium was exchanged to a fresh Puromycin-containing medium. When successfully cells reached 90-100 % confluency, they were passaged twice a week in the medium containing 2 µg/ml Puromycin (Thermo Fisher Scientific) and 2 µg/ml Doxycycline (Sigma) for the first week, followed by another week of passaging with the medium containing only Puromycin-containing. After at least a total of 16 days post transduction, we harvested cells downstream analyses.

2.14. ICS-based CRISPR screens

After transducing cells with genome-wide libraries and preparing them as described above, I harvested them for ICS. Their viability was assessed by DAPI staining, and nuclei were stained with DRAQ5 (see details above). Samples were kept on ice at all times from the moment of harvest until genomic DNA isolation from sorted populations. Cells were sorted by Dr. Marta Rodríguez-Martínez and me in bulk using ICS as described before[86] with the following modifications. We sorted them in batches of 100,000 cells to separate 1.5 ml tubes, and the sorting purity was checked for each of them prior to pooling. Populations from the eGFP-DRAQ5 correlation parameter were selected based on the ranged gates comprising the 7 % of cells with the lowest or highest correlation values. For each batch of the sorted cells, an input sample containing the same number of cells as present in the sorted bins was collected. Cells from pooled sorted samples and the unsorted inputs were palleted by centrifugation for 8 min at 400 g at 4 °C, and then pellets were either frozen at -20 °C or stored on ice until genomic DNA extraction. Per each library, we collected one million of sorted cells that were pooled into a single tube for genomic DNA preparation.

2.15. Isolation of genomic DNA, library preparation, and sequencing

I isolated genomic DNA from the sorted cells and the input samples using NEB Monarch genomic DNA purification kit (New England Biolabs), including the RNase treatment and elution in 50 µl elution buffer, according to manufacturer's instructions. I measured DNA concentration using Qubit High sensitivity dsDNA kit (Thermo Fisher Scientific), according to the manufacturer's instructions.

For library preparation, I performed the first PCR reaction (PCR1) with 125-525 ng of gDNA (per reaction, 6 reactions per library), 1.5 µl of 10 µM pU6 fwd, 1.5 µl of 10 µM pLTR-CROP-rev (see sequences in[86]), and 25 µl KAPA HiFi Hotstart Readymix (Roche) in 50 µl total volume. A total of six 50 µl reactions were set up per each sgRNA sub-library per each sorted or input sample using the total amount of genomic DNA (gDNA) recovered. Cycling conditions for PCR1 were: one cycle at 95 °C for 3 min; 24 cycles at [98 °C for 20 s, 67 °C for 15 s, 72 °C for 15 s]; one cycle at 72 °C for 1 min; and cooling to 4 °C. I pooled PCR reactions of the same template (same sublibrary) of the same sample after PCR1 (six PCR products into one of total volume 300 µl) and the pooled product was then size selected and purified with 0.8X volume of AMPure XP (Beckman) with two 80 % ethanol washes, and elution in 40 µl water. Concentrations were then measured with Qubit High sensitivity dsDNA kit (Thermo Fisher Scientific), according to the manufacturer's instructions.

For the second PCR reaction (PCR2), I used 10 ng of the PCR1 per reaction, with a total of 6 reactions per sub-library per sample, 5 µl of 3 µM CROPseq_libQC_i5_s:n staggered primer[95] (different primer for each reaction for one sample), 5 µl of 3 µM CROPseq_i7:n barcoded primer (same for all reactions for one sample, but unique to every sample), and 25 µl KAPA HiFi Hotstart Readymix (Roche) in 50 µl total volume. I used the same PCR2 primers as in[86]. Cycling conditions for PCR2 were: one cycle at 95 °C for 3 min; 8 cycles at [98 °C for 20 s, 67 °C for 15 s, 72 °C for 15 s]; one cycle at 72 °C for 1 min; and cooling to 4 °C. After this, I pooled PCR2 products of the same sub-library of the same sample, size selected and purified them as above, followed by the elution in 40 µl of H_2O. Concentrations were then measured with Qubit High sensitivity dsDNA kit (Thermo Fisher Scientific). The qualities of the prepared sequencing libraries were checked using DNA 1000 Bioanalyzer (Agilent). Libraries that yielded an expected single product around 300 bp were then used for Illumina sequencing.

Libraries that passed the quality control were then pooled in equimolar ratio (nine libraries per pool) and sequenced using an Illumina NextSeq 500 (75 bp, single end mode) in high output mode with 8 reads to read out the i7 barcode, and 67 reads on Read1 to read through the stagger sequence and identify the gRNA. PhiX spike-in was used to diversify the libraries. Sequencing was performed by the GeneCore.

2.16. CRISPR screen data analysis

I quantified the abundance of sgRNA from the sequenced reads of each sub-library for each sorted or input sample using MAGeCK (v 0.5.9) tool with default parameters[96], as described in[86]. I accounted for potential differences in sequencing depth by normalization of raw gRNA counts to the median count of the targeting control gRNAs in the respective sample. Multiplication with the median count of targeting controls across all samples was used to scale the normalised counts. I evaluated the screen quality using the pipeline and published codes from[86] – they are based on the dropout of essential genes in input cell populations compared to the plasmid library (sequenced previously in[86]) with reference core- and non-essential gene lists described previously in[97]. The R package "ROCR"[98] was used for all precision-recall-curves generation. For hit calling, I used the MAUDE (v. 0.99.4)[99] package for R, with targeting controls serving as a reference for MAUDE analysis. False discovery rate values and cutoffs are indicated for each individual plot in the Results chapter.

2.17. Whole cell extract preparation and cell fractionation

For whole cell extracts, cells were pelleted and lysed in NP-40 lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% (v/v) NP-40. This was done by Dr. Marta Rodríguez-Martínez for all western blots of HeLa cells, and me for iPSC-CMs. All protein extraction buffers contained PhosSTOP (Sigma-Aldrich, 04906837001) and Protease Inhibitor Cocktail (Sigma-Aldrich, 05056489001). After lysing for at least 20 minutes on ice, lysates were spun down for 20 minutes at 4 °C at a maximum speed. Supernatants were then used for either co-immunoprecipitation or western blotting (see below).

Cell fractionation of cultured cells was performed by Dr. Marta Rodríguez-Martínez as follows. First, cell pellets were resuspended in a double of their volume of hypotonic buffer (10 mM HEPES pH 7.5, 10 mM KCl, 1.5 mM $MgCl_2$), and incubated on ice for 15 min. Then, they were homogenised with 20 strokes using a loose pestle. This resulted in extraction of the cytoplasmic fraction leaving nuclei and insoluble cellular compartments intact. The latter were pelleted at 3,900 rpm for 15 min and supernatant collected as soluble cytoplasmic fraction, which was corrected to 10% (v/v) glycerol, 3 mM EDTA, 0.05% (v/v) NP-40 and 150 mM NaCl final concentration. The remaining nuclear and insoluble fraction pellet was resuspended in chromatin digestion buffer (20 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 150 mM NaCl, 0.1% (v/v) NP-40 and 125 U benzonase (MerckMillipore, 70746-4) and incubated for 1 h at 4°C. To ensure extraction of all nuclear soluble and insoluble proteins, as well as of insoluble cytoplasmic components, NaCl concentration was then increased to 500 mM and samples were incubated on ice for 30 minutes. Then, the salt concentration was diluted back to 150 mM NaCl by addition of high salt dilution buffer (20 mM HEPES pH 7.9, 3 mM EDTA, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 500 mM NaCl and 0.1% (v/v) NP-40), and the samples were centrifugated at 20,000 g for 20 min at 4°C. The resulting supernatant was then kept as nuclear and insoluble fraction.
All preparation of protein samples was done by either Dr. Marta Rodriguez Martinez or myself (indicated for each experiment).

2.18. Co-immunoprecipitation

Co-immunoprecipitation was done by Dr. Marta Rodríguez-Martínez as follows: 1 mg of the whole cell extracts, cytoplasmic or nuclear and insoluble fractions were incubated with 30 µl of GFP-Trap® Magnetic Particles M-270 (ChromoTek) for at 4°C for 3 h. Then, she washed the beads 5 times in IP wash buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.5, 1.5 mM $MgCl_2$, 3mM EDTA, 10% (v/v) glycerol, 0.1% (v/v) NP-40, phosphatase inhibitors (PhosSTOP, Sigma-Aldrich, 04906837001) and protease inhibitor cocktail (Sigma-Aldrich, 05056489001) and eluted in 30 µl of laemmli buffer with 100 µM DTT.

2.19. Western blotting

For western blotting, protein lysates, intracellular fractions, or co-IP products were first desaturated by heating at 95°C for 5-10 minutes in the laemmli buffer with 100 µM DTT. Then, 10–100 µg denatured protein/lane was run and separated on 4%–12% or 3-8% NuPage gels (Invitrogen). Separated proteins from the gel were transferred to Trans-Blot Turbo Mini 0.2 µm Nitrocellulose (biorad 1704158) using Trans-Blot Turbo Transfer System. Membranes were then blocked in 5% (w/v) skimmed milk in PBS-T (PBS, 0.1% (v/v) Tween20) for 1 h at RT, followed by an incubation with primary antibody solutions (see table 2.3) in 5% (w/v) skimmed milk in PBS-T overnight at 4°C with constant shaking. An anti-GAPDH antibody was used as a loading control. After primary antibody incubation, membranes were washed several times in PBS-T at RT with constant shaking, and then incubated with HRP-conjugated secondary antibody in 5% (w/v) skimmed milk in PBS-T for 1h at RT with constant shaking. Then, after three more washes in PBS-T, membranes were visualised using SuperSignal West Dura Chemiluminescent Substrate ECL reagent (Thermo Fisher Scientific, 34075) and Bio-Rad ChemiDoc Touch (Software v. 2.3.0.07).

All western blotting was done by either Dr. Marta Rodríguez-Martínez or myself.

2.20. Mass spectrometry

LC-MS/MS analysis was performed based on the prepared co-IP and fractionation samples (see above) by Dr Per Haberkant from the Proteomics Core Facility (PCF), see details in[80]. Acquired raw data were analysed by Dr. Frank Stein from the PCF, see details in[80]. After receiving statistics for each identified protein in each possible comparison, I performed analyses of the processed data as follows. All proteins that displayed a false discovery rate (fdr) smaller 0.05 and a fold-change of at least 50 % in comparison to the no bait control, were considered as interactors of RBM20 in a given sample. Common between WT and the mutant variants, as well as unique for mutant variants only, interactors were then analysed for pathway enrichment with Metascape[90]. All plots were prepared using R programming language.

2.21. AAV9 production and injection

I cloned pCMV-*Tnpo3* into the AAV9 packaging backbone (derived from Addgene 137177, gift from the Genetic and Viral core facility, EMBL Rome, Italy) by performing a BglII and AgeI (NEB) digest of the backbone, and amplifying the murine *Tnpo3* cDNA (GenScript) with the oligos fwd: tcacttttttcaggttggaggtaccatggagggagccaa; rv: cagaggttgattatcagatcggatccctatcgaaacaacc. This was followed by Gibson assembly of the two fragments (NEB), according to manufacturer's instructions.
Production of the rAAV with the serotype 9 rAAV containing pCMV-Tnpo3 cDNA was done fully by the Genetic & Viral Engineering Facility at EMBL Rome (see details in[79]). Viral titers were determined by qPCR using primers within the CMV promoter using primers listed in[80].

2.22. Mouse handling and treatments

Mouse lines with RBM20-WT, -P635L-Hom, and -R636Q-Hom were generated previously[79]. All animal maintenance and treatments were done by Laboratory Animal Resources (LAR), and especially by Frank Diego Montoya Castillo. Experimental design was done by me. Briefly, we used strains in C57BL/6J genetic background that were maintained in individually ventilated plastic cages (Tecniplast) in an air-conditioned (temperature 22 °C ± 2 °C, humidity 50% ± 10%) and light-controlled room (illuminated from 07:00 to 19:00 h). Mice were fed 1318 P autoclavable diet (Altromin, Germany) ad libitum. All animal care and procedures performed in this study conformed to the EMBL Guidelines for the Use of Animals in Experiments and were reviewed and approved by the Institutional Animal Care and Use Committee (IACUC).

Mice were treated with 10^{12} VG of AAV9 delivering either *Tnpo3* cDNA (see above), or *pCAG-iCre-T2A-eGFP* (produced by the Genetic and Viral core facility, EMBL Rome, Italy) as a negative control, diluted in 100 µl of PBS. Unless otherwise specified, mice were injected in the tail vein at 4-week-old of age. Animals were humanely sacrificed according to the protocol approved by the IACUC procedure by LAR technicians, or by the lab of Dr. Maarten M. G. van den Hoogenhof. I then received pieces of tissues and further processed them for RNA or protein extraction. All echocardiograms were measured by Dr. Maarten M. G. van den Hoogenhof.

2.23. Isolation of primary mouse cardiomyocytes for imaging

For extraction of primary mouse cardiomyocytes for imaging, microscopy slides were coated with 10 µl/ml of laminin (Gibco, 23017015) in PBS overnight. For CM isolation, we adapted the protocol from[100], and described in detail in[79]. I performed this procedure with the help from Frank Frank Diego Montoya Castillo (LAR), and Laura Schraft. I then used the slides for fixation and staining protocols as described above.

2.24. Statistical analysis and data visualization

I designed and performed all experiments in at least three biological replicates, unless otherwise specified in the figure legends or main text of the Results chapter. I quantified statistical significance either with Welch Two Sample t-test for pairwise comparisons, or with ANOVA with Tukey's HSD or Bonferroni post-tests for multiple comparisons, unless otherwise specified in the figure legends.

I used the package ggplot2 (v.3.4.0) from R Studio (v. 4.2.1) to prepare plots of all figures throughout the manuscript. Heatmaps were plotted using pheatmap package (v. 1.0.12).

2.25. Data availability statement

All the sequencing (Raw and processed RNA-sequencing and ICS screen) data from human cells discussed in this dissertation are published at GEO and are available for public (GSE220833). The discussed in this dissertation proteomics data are published at PRIDE and are available to public (PXD038790).

Chapter 3. Results

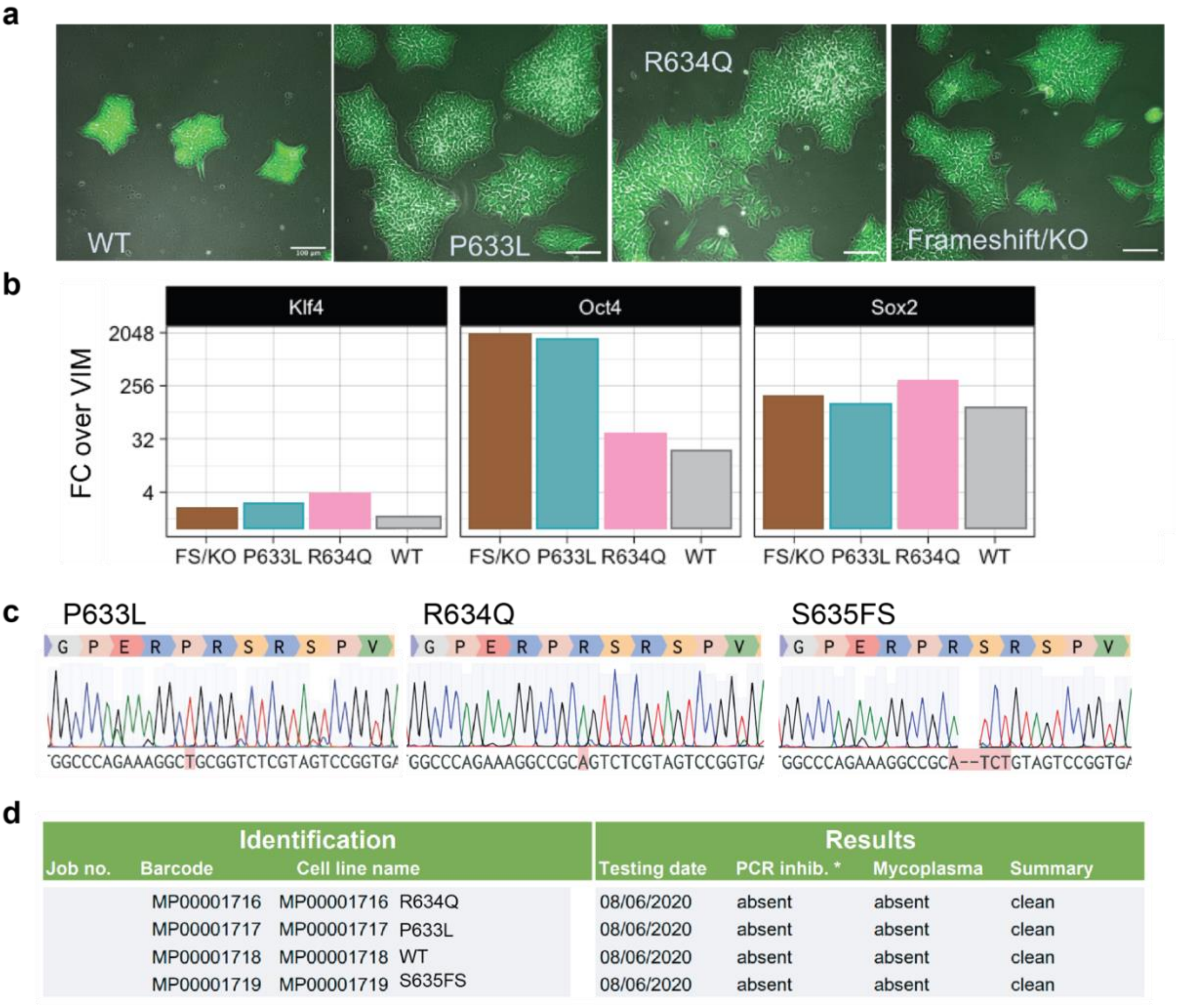
3.1. Splice regulatory function of RBM20 variants is proportional to their nuclear localization

3.1.1. Quality control of iPSCs

For my experiments, I used human induced pluripotent stem cells (hiPSCs) that were engineered and characterised by Briganti et al.[62]. Prior to differentiating them into iPSC-derived cardiomyocytes (iPSC-CMs) and starting downstream experiments, I confirmed their pluripotency by analysing expression of pluripotency markers Oct4, Sox2, Klf4 by qRT-PCR, as well as performing alkaline phosphatase staining as an additional pluripotency test (Figure 3.1 a, b). Both of these indicated pluripotent state of the cells. In addition, I confirmed their *RBM20* mutations by Sanger sequencing, and ensured the absence of mycoplasma contamination (Fig. 3.1 c, d, both done using Eurofins prepaid services). Altogether, these analyses demonstrated that these iPSCs were in a good shape for downstream experiments.

3.1.2. Assessment of splicing activity and mislocalization of RBM20 variants in iPSC-CMs

To assess RBM20-P633L and -R634Q mislocalization, I differentiated iPSCs generated by Briganti et al.[62] into iPSC-derived cardiomyocytes (iPSC-CMs) using the protocol adapted from[83] (see methods). After maturation for at least 6 weeks post first passaging of the beating cells, I performed immunofluorescence staining for endogenous RBM20, sarcomeric α-actinin, and DAPI to stain the nuclei (Fig. 3.2 a). While RBM20-WT displayed solely nuclear localization, RBM20-R634Q formed cytoplasmic granules and was mostly absent from the nucleus. Interestingly, RBM20-P633L exhibited intermediate mislocalization phenotype: despite the presence of cytoplasmic granules, a large proportion of the protein still localised to the nucleus. To quantify this effect, I used Coloc2 plugin for ImageJ (see methods), that calculates Pearson correlation coefficient between signal from two selected channels (Fig. 3.2 b). The high degree of RBM20-WT co-localization with DAPI concludes its nuclear localization. Co-localization of both RBM20-P633L and -R634Q with DAPI was significantly different from RBM20-WT. However, RBM20-P633L displayed co-localization values significantly higher than those of RBM20-R634Q (Fig. 3.2 b), confirming its milder mislocalization. The same pattern was observed when using image-enabled cell sorting (ICS) to quantify co-localization of RBM20 with DRAQ5 (Fig. 3.2 c, see methods), which I operated together with Dr. Marta Rodríguez-Martínez.



Identification			Results			
Job no.	Barcode	Cell line name	Testing date	PCR inhib. *	Mycoplasma	Summary
	MP00001716	MP00001716 R634Q	08/06/2020	absent	absent	clean
	MP00001717	MP00001717 P633L	08/06/2020	absent	absent	clean
	MP00001718	MP00001718 WT	08/06/2020	absent	absent	clean
	MP00001719	MP00001719 S635FS	08/06/2020	absent	absent	clean

Figure 3.1. Quality control of iPSCs. (a) Alkaline phosphatase life stain, green fluorescence indicates activity of the enzyme marking pluripotent stem cells (PSCs). Scale bar = 100 µm, shown is an overlay of FITC channel with the phase contrast. (b) Expression of PSC markers analysed by qRT-PCR. Expression is normalised to *GAPDH*, represented is fold change (FC, 2^-ΔΔCt) versus expression of a fibroblast marker *VIM* in a corresponding cell line. (c) Sanger sequencing of the mutation hotspot. The wildtype (WT) amino acid sequence is displayed on top of the DNA sequencing peaks. Sanger sequencing was done by Eurofins. (d) Mycoplasma check report created by Eurofins.

I then analysed RBM20 gene and protein expression in the differentiated lines. *RBM20* gene expression measured by qPCR was not different amongst these cell lines (Fig. 3.2 d). In addition, differences in RBM20 protein level that I measured by western blotting, were in line with differences in loading measured by GAPDH level (Fig. 3.2 e). These served as a control that, localization and splicing changes are not dependent on the amount of RBM20 expressed. Next, I assessed splice regulatory activity of these RBM20 variants. For that, I used *TTN* as one of the main RBM20 splice targets (see introduction). I used primers that span exon-exon junctions of *TTN* exons 241 and 243, excluding alternative exon 242. This *TTN* isoform will be further referred to as *TTN* spliced isoform. Strikingly, almost no *TTN* spliced isoform was detected in RBM20-R634Q cells, unlike RBM20-P633L that displayed only a reduction, and not complete absence of this isoform, compared to RBM20-WT cells (Fig. 3.2 f).

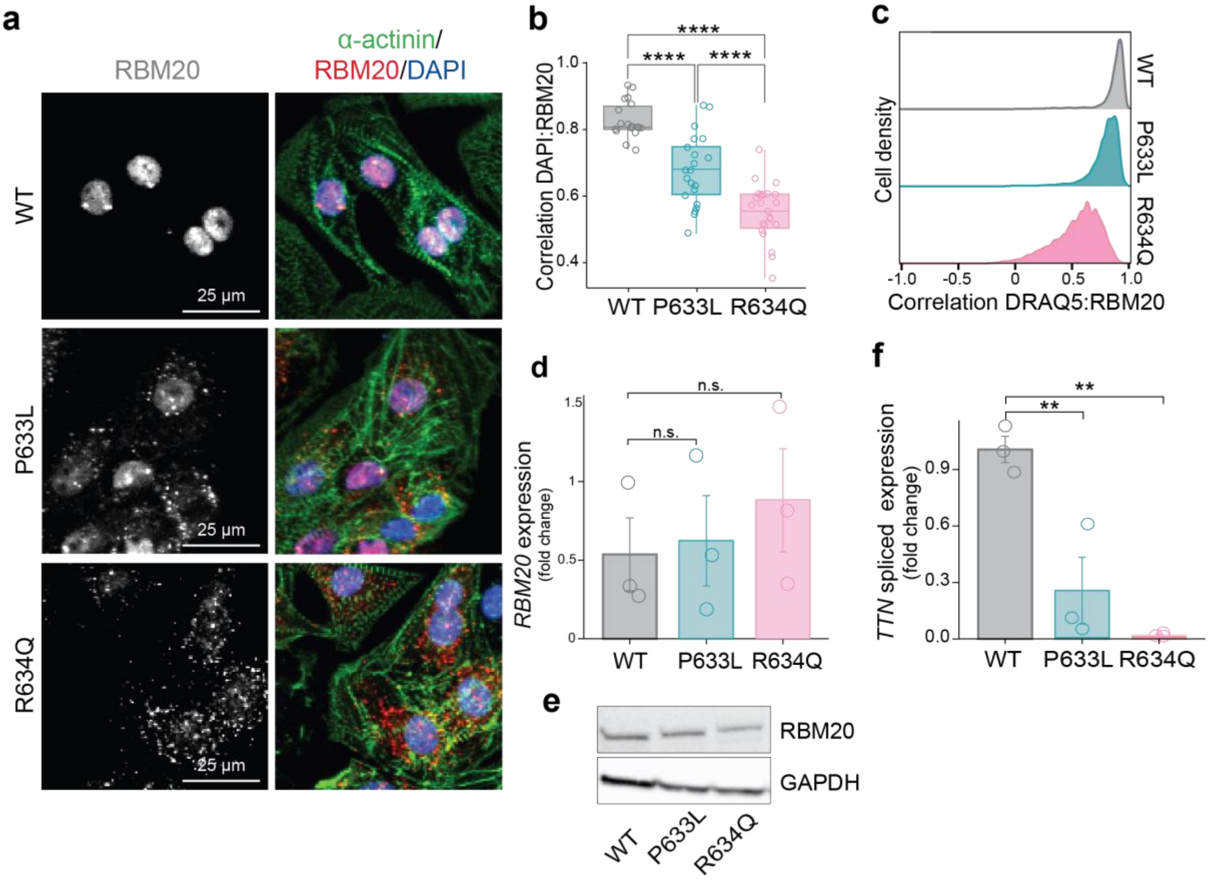


Figure 3.2. Splice deficiency of RBM20 variants is proportional to their misclocalization in iPSC-CMs. (a) Representative images of iPSC-CMs after immunofluorescence staining with anti-RBM20 and anti-α-actinin antibodies (b) Quantification of RBM20 nuclear localization based on co-localization with DAPI (from confocal microscopy images illustrated at panel a). Data are combined from images of cells from three independent batches of iPSC-CM differentiations. Each dot is a Pearson correlation coefficient R for at least five cells. (c) ICS-based analysis of DRAQ5:RBM20 correlation in iPSC-CMs. (d) Expression of *RBM20*-WT, -P633L, or -R634Q in iPSC-CMs analysed with qPCR, three biological replicates. Data is normalised to *GAPDH* expression and displayed as fold change versus the first replicate of the WT line, with standard errors indicated. (e) Western blot analysis of RBM20 protein level in iPSC-CMs. GAPDH serves as a loading control. (f) Expression of *TTN* isoform with spliced-out exon 242 in iPSC-CMs analysed with qPCR. Data is normalised to *GAPDH* expression and displayed as fold change versus the first replicate of RBM20-WT line (standard errors are indicated). Three independent iPSC-CM differentiations were used as biological replicates. Ns = not significant, * - $p < 0.05$, ** - $p < 0.01$, *** - $p < 0.001$, **** - $p < 0.0001$, one-way ANOVA with Tukey's HSD post-test (two-sided).

These results demonstrate that, RBM20-P633L is only mildly mislocalised to the cytoplasm, with a proportion of the protein still localizing to the nucleus. This coincides with mild splicing deficiency, where *TTN* spliced isoform expression is reduced compared to WT cells but not completely absent. These are not explained by differences in RBM20 expression levels. Based on these results, I hypothesised that nuclear-localised RBM20-P633L might retain its splice regulatory activity.

3.1.3. Cells with nuclear-localised RBM20-P633L are similar to RBM20-WT in their gene expression and splicing patterns

To directly test whether nuclear-localised RBM20-P633L indeed retains its splice regulatory activity, I used ICS to isolate cells with nuclear- or cytoplasmic-localised RBM20 and performed RNA sequencing of the sorted populations. I differentiated three independent batches of iPSCs with RBM20-WT, -P633L, and -R634Q, and after two months of maturation, I harvested them, fixed, and stained with an anti-RBM20 antibody for ICS (see methods). Then, together with Dr. Marta Rodríguez-Martínez, we sorted RBM20-P633L cells based on the correlation between DRAQ5 and RBM20 into the top 7% highest correlation displaying the nuclear phenotype (P633L-nuc, Fig. 3.3 a), and the bottom 7% of the correlation distribution displaying the cytoplasmic phenotype (P633L-cyt, Fig. 3.3 a). To exclude potential impacts of cell fixing and staining on RNA quality and consequent gene expression and splicing analyses, I processed RBM20-WT and RBM20-R634Q cells in the same way as the P633L-RBM20, and sorted only positive for the RBM20 staining cells. I then subjected these sorted cells to RNA extraction and RNA sequencing library preparation, followed by sequencing which was done by the GeneCore.

After I processed raw sequencing reads (see methods), I first compared differences between the sorted populations on the gene expression level. According to the principal component analysis (PCA), gene expression in the cells with P633L-nuc was more similar to WT than to P633L-cyt (Fig. 3.3 b). I then performed differential gene expression analysis with DeSeq2[89] (see methods for details). I considered a gene to be differentially expressed if the log2 of its expression fold change was greater than 0.5, and if the adjusted (BH) p-value was less than 0.1. I identified 1,415 differentially expressed genes in P633L-cyt compared to WT. Strikingly, there were only 50 differentially expressed genes in P633L-nuc compared to WT. For differentially expressed genes, I then performed gene set enrichment analysis using Metascape[90]. Down regulated genes in both P633L-cyt and R634Q compared to WT were mostly involved in cardiac muscle development and function (Fig. 3.3. d, f). Up-regulated genes in P633L-cyt and R634Q were enriched for cell-cell adhesion, and other processes less related to cardiomyocyte function (Fig. 3.3 e, g). I then analysed gene expression for the core RBM20 targets - 45 genes, splicing of which was consistently affected across RBM20 perturbations (KO, S635A) and across model species (rat, human, mouse) based on several studies[50,53,54] (see methods). Interestingly, gene expression differences to WT cells for the core RBM20 targets were either unchanged or down-regulated in P633L-cyt and, to a lesser degree, in R634Q, but not in P633L-nuc (Fig. 3.3 h). The reason why RBM20-R634Q does not impact gene expression as severely as P633L-cyt could be that a small proportion of this variant still reaches the nucleus (Fig. 3.2. a), unlike for P633L-cyt that is enriched for only cytoplasmic-localised protein. These data demonstrate that, cytoplasmic localization of RBM20-P633L impacts expression of many genes crucial for cardiomyocyte development and function, including some of the core RBM20 target genes. This is not the case if RBM20-P633L is localised to the nucleus, suggesting that this aberrant gene expression the consequence of mislocalization rather than of the mutation itself.

Next, I examined alternative splicing differences. I performed analysis of global splicing changes using rMATS[91] in pairwise comparisons to RBM20-WT cells. I considered an event to be mis-spliced if its inclusion level difference was greater than 0.1 and the false discovery rate less than 0.01, compared to WT cells. Cells with P633L-cyt had more mis-splicing events (522) than those with R634Q (307), and especially, those with P633L-nuc (234, Fig. 3.4 a). Similar reduction of mis-splicing events in P633L-nuc was observed when analysing only core RBM20 targets (Fig. 3.4 b). From the total mis-splicing events, 125 were shared between P633L-cyt and R634Q, unlike P633L-nuc that shared only 50 events with P633L-cyt and 34 with R634Q (Fig. 3.4 c).

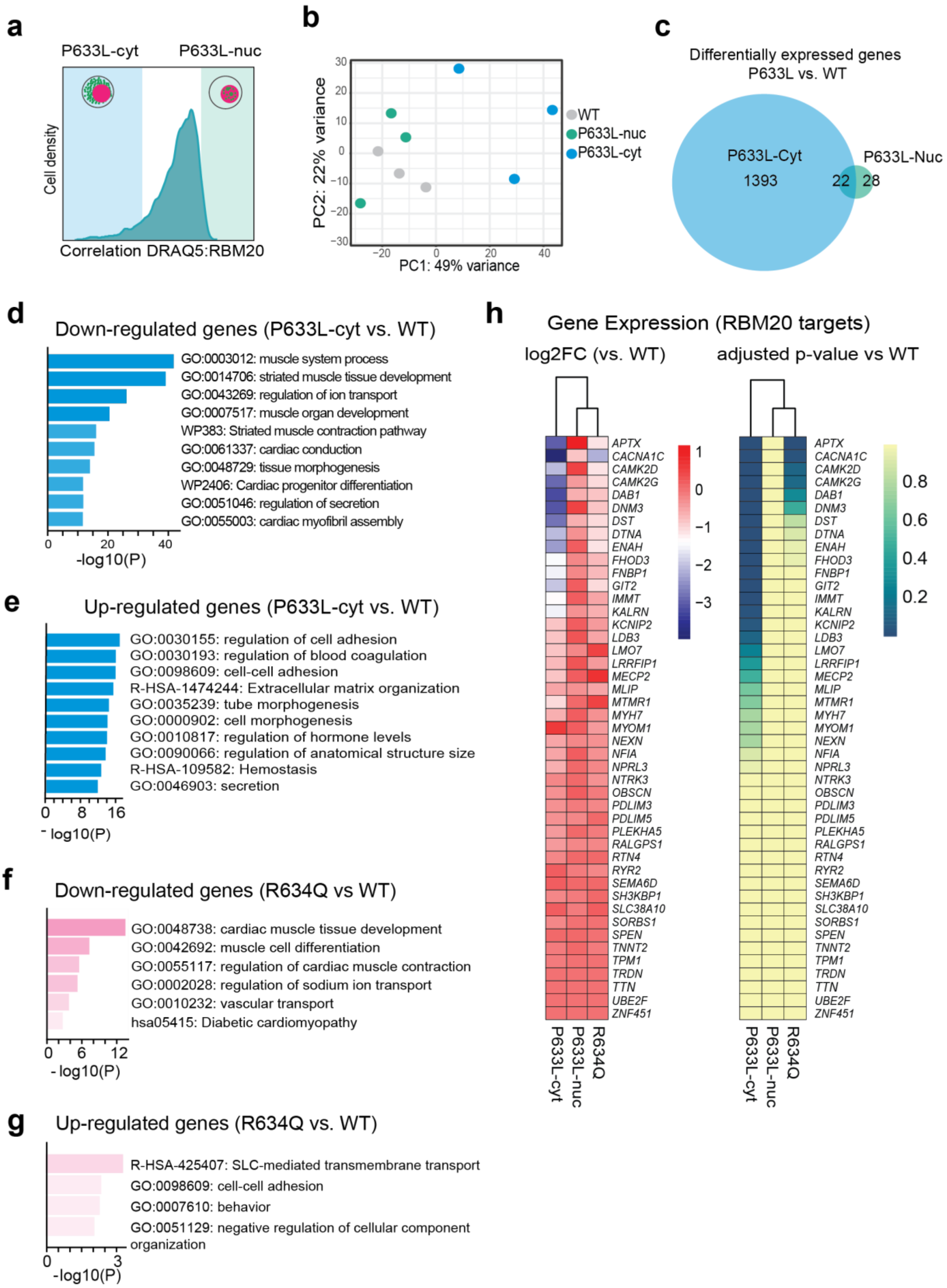


Figure 3.3. Gene expression differences between cells with nuclear and cytoplasmic RBM20-P633L. (a) ICS gating strategy. (b) PCA based on read counts per gene, in RBM20-WT, -P633L-nuc, and -P633L-cyt iPSC-CMs, three biological replicates. (c) Numbers and a relative overlap of significantly (log2 of expression fold change greater than 0.5, adjusted (BH) p-value less than 0.1) differentially expressed genes in P633L-nuc and P633L-cyt in pairwise comparison to WT-cells. (d)

Gene set enrichment analysis for down-regulated in P633L-cyt genes compared to WT. (e) Gene set enrichment analysis for up-regulated in P633L-cyt genes compared to WT. (f) Gene set enrichment analysis for down-regulated in R634Q genes compared to WT. (g) Gene set enrichment analysis for up-regulated in R634Q genes compared to WT. (e) Log2 fold changes and adjusted p-values for expression of the core RBM20 targets compared to WT cells. Three biological replicates for each cell line were used for the RNA-sequencing experiment.

To assess percentage of spliced-in (PSI) values for the sorted and unsorted lines, I used the pipeline described in[92], see methods. PSI is the fraction of all reads that include the exon of interest from all reads that either include or exclude this exon (see methods). After quantifying PSI values for all exonic parts of all exons (see methods), I selected only exons located in the core RBM20 targets for further analysis. For those, I performed the student's t-test for three replicates of PSI values in WT versus R634Q cells. Then, I filtered for exonic parts that differed between WT and R634Q (delta PSI > 0.1, p-value < 0.05). PSI values for these exonic parts are depicted on Fig. 3.4 d. The majority of exonic parts that were mis-spliced in R634Q were spliced-in, while in WT cells they were predominantly spliced-out, which is consistent with the known function of RBM20 as a splicing repressor. Strikingly, these exons were mostly spliced-in in P633L-cyt, and out in P633L-nuc (67% AS events restored in P633L-nuc from those mis-spliced in P633L-cyt). I then confirmed this finding by qPCR analysis of *TTN* splicing (Fig. 3.4 e). These results demonstrate that, splice regulatory function of nuclear localised RBM20-P633L is similar to that of RBM20-WT, especially when analysing core RBM20 targets. This indicates that, RS-domain variants do not lose their splice regulatory activity but rather their ability to be imported to the nucleus.

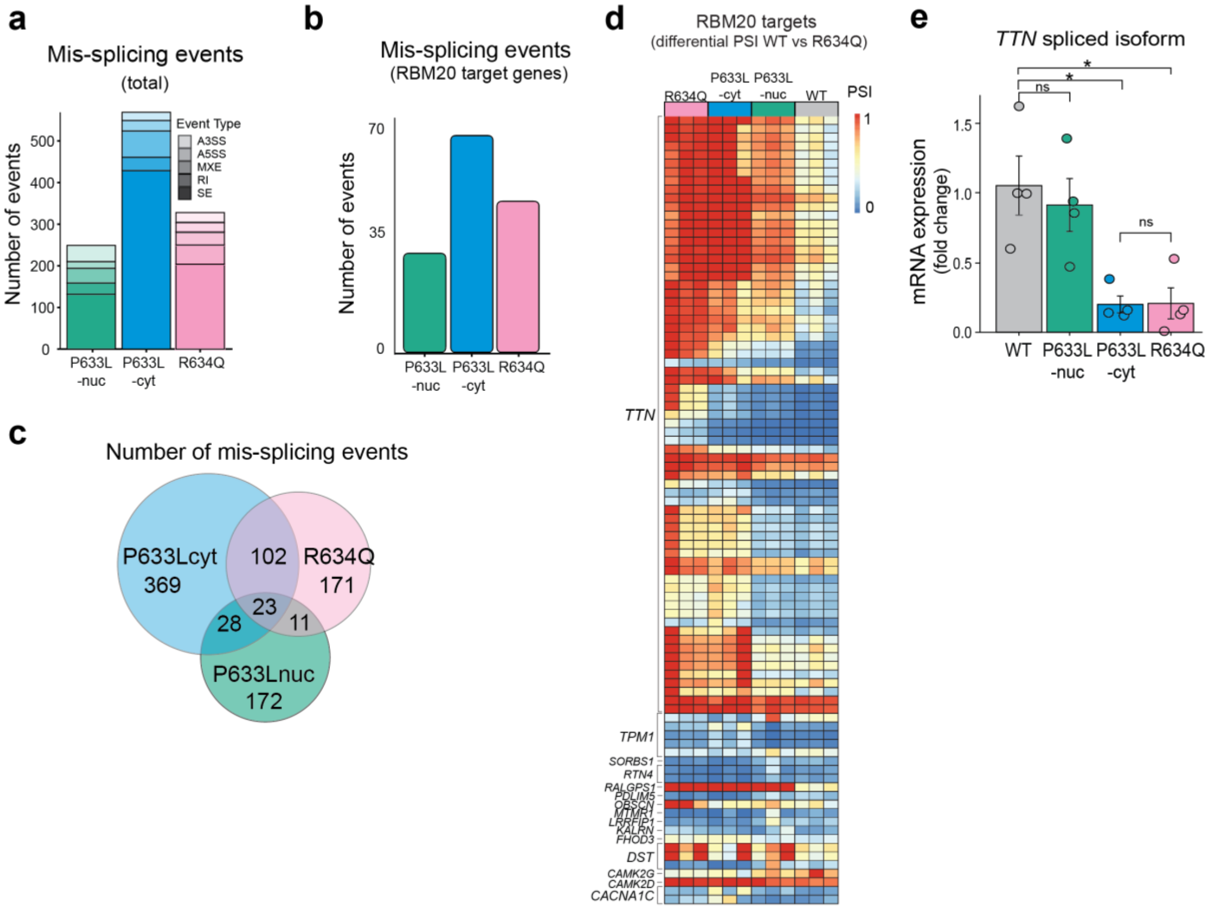


Figure 3.4. Alternative splicing differences between cells with nuclear and cytoplasmic RBM20-P633L. (a) Total number of differential splicing events (inclusion level differences greater than 0.1 and false discovery rate less than 0.01, rMATS[91]) in pairwise comparisons to RBM20-WT-cells. SE = exon skipping, RI = intron retention, MXE = mutually exclusive exons, A3SS and A5SS are alternative 3' and 5' splice sites, respectively. (b) Numbers of differential splicing events from panel (a) that are located in the core RBM20 targets. (c) Numbers and relative overlaps of all differential splicing events from panel (a). (d) Percentage of spliced-in (PSI) values (calculated according to[92]) of exonic parts located in the core RBM20 targets that were different (delta PSI > 0.1, p-value < 0.05, t-test) between RBM20-WT and RBM20-R634Q cells. Three biological replicates for each cell line were used for the RNA-sequencing experiment. (e) qPCR analysis of *TTN* exon 242 spliced-out isoform expression. Data is normalised to *GAPDH* and is displayed as fold change versus the first replicate of the RBM20-WT line, two biological replicates with two technical replicates each, standard errors are indicated. Ns = not significant, * - $p < 0.05$, one-way ANOVA with Tukey's HSD post-test (two-sided).

3.1.4. iPSC-CMs expressing RBM20-R634Q tagged with the SV40-NLS are similar to RBM20-WT expressing cells in their gene expression and splicing patterns

I then went on to test whether the observed WT-like functionality of the nuclear-localised RBM20-P633L is specific to this variant, or whether nuclear-localised RBM20-R634Q exhibits similar WT-like properties. To test this, I used differentiated iPSC-CMs harbouring the S635FS mutation as a cardiac cell line shown not to express functional RBM20[62]. I then transduced them with lentiviral plasmids I produced expressing TetO:eGFP-tagged RBM20-WT, -R634Q, or -R634Q cDNA tagged with an SV40 NLS sequence (see methods). I then either fixed the cells for microscopy analysis, or FACS-sorted only eGFP-positive cells for RNA-sequencing. I observed that overexpressed eGFP-tagged RBM20-WT localised to the nucleus (Fig. 3.5 a). RBM20-R634Q localised to the cytoplasm in a form of granules of the same appearance as in the endogenous cell lines (Fig. 3.5 a, Fig. 3.2 a). Importantly, the SV40-NLS tag restored nuclear localization of RBM20-R634Q and resulted in disappearance of cytoplasmic granules (Fig. 3.5 a). This demonstrated that, up-regulating nuclear import by involving possibly a different import mechanism to the one that is native to RBM20, alleviates aberrant cytoplasmic localization phenotype.

Next, I analysed gene expression differences between RBM20-WT expressing cells and RBM20-R634Q / -NLS-R634Q -expressing cells. I used DeSeq2 with BH correction to identify genes that were significantly changed between the cell lines (see methods). I identified 1,751 differentially expressed genes in R634Q-expressing cells, and only 115 in NLS-R634Q (Fig. 3.5 b). Similar to what was observed before for cytoplasmic P633L and endogenous R634Q, down-regulated genes in R634Q-expressing cells were impacting muscle development and contraction pathways (Fig. 3.5 c). When comparing gene expression differences to WT-expressing cells for the core RBM20 targets (see explanation in methods), they were changed in R634Q-expressing cells more than in NLS-R634Q-expressing cells (Fig. 3.5 d).

These results, combined with those described in Fig. 3.3, suggest that, cytoplasmic localization of RBM20 variants impacts gene expression, in particular of genes involved in muscle physiology and the core RBM20 targets. Whether this is the prime consequence of RBM20 mislocalization, or a secondary effect of mis-splicing, is currently unknown. However, this might be one of the effects associated with detrimental gain-of-function of cytoplasmic RBM20 variants.

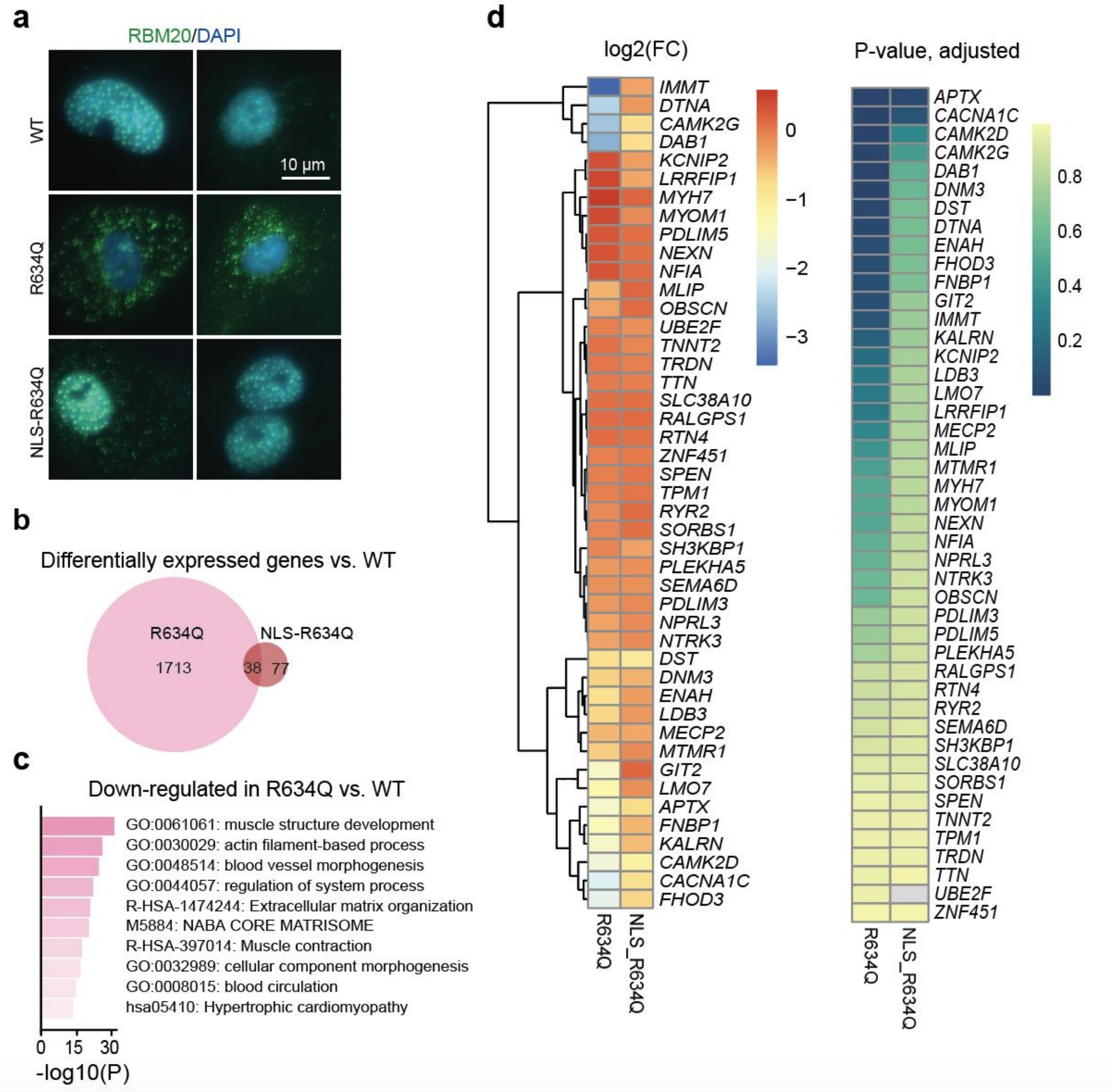


Figure 3.5. Gene expression differences between cells with RBM20-R634Q and -SV40-NLS-R634Q. (a) Microscopy analysis of S635FS iPSC-CMs expressing TetO-eGFP-RBM20-WT, -R634Q, or SV40 NLS-R634Q (b) Numbers and relative overlap of significantly (log2 of expression fold change greater than 0.5, adjusted (BH) p-value less than 0.1) differentially expressed genes in R634Q and NLS-R634Q expressing cells in pairwise comparison to WT-expressing cells. (c) Gene set enrichment analysis for down-regulated genes in R634Q-expressing cells compared to WT-expressing cells. (d) Log2 fold changes and adjusted p-values for expression of the core RBM20 targets compared to WT cells. Three biological replicates for each cell line were used for the RNA-sequencing experiment.

I then assessed alternative splicing in S635FS cells expressing eGFP-RBM20-WT, -R634Q, or NLS-R634Q. First, I performed global splicing analysis with rMATS[91] in pairwise comparisons to WT-expressing cells (see methods). I identified fewer mis-splicing events in cells expressing NLS-R634Q than normal R634Q, and, in particular, exon skipping – the major type of AS regulated by RBM20 – was reduced from 536 to

277 by NLS-tagging (Fig. 3.6 a). Even a more pronounce reduction in the number of AS events upon NLS-tagged tagging of R634Q was observed when analysing only core RBM20 targets (8 instead of 24, Fig. 3.6 b). There were relatively few AS events that were shared between R634Q- and NLS-R634Q-expressing cells (148 out of 676 and 437, respectively, Fig. 3.6 c). These indicate that, NLS-R634Q-expressing cells exhibit an AS pattern more similar to that in WT-expressing cells, than R634Q-expressing cells do. To assess percentage of spliced-in (PSI) values, similar to above, I used the pipeline described in[92] (see methods). I performed the same analysis as for P633L-nuc and -cyt, described in Fig. 3.4.: I selected only exonic parts in the core RBM20 targets, and performed the student's t-test for three replicates of PSI values in WT- versus R634Q-expressing cells. Then, I filtered for exonic parts that differed between WT and R634Q (delta PSI > 0.1, p-value < 0.05) and plotted them as Fig. 3.6 d. Same as before, the vast majority of exonic parts that were mis-spliced in R634Q were spliced-in, while in WT cells they were predominantly spliced-out. However, most of the alternative spliced-in exons (80 %) were restored and spliced-out in NLS-R634Q expressing cells with respect to those mis-spliced in R634Q (Fig. 3.6 d). These results support that the splice regulatory function of nuclear localised R634Q is similar to that of RBM20-WT, especially when analysing core RBM20 targets. Similar effect was discussed above for nuclear-localised P633L-RBM20. Altogether, these results propose that restoring nuclear localization of RS-domain RBM20 variants may also restore their splice regulatory function.

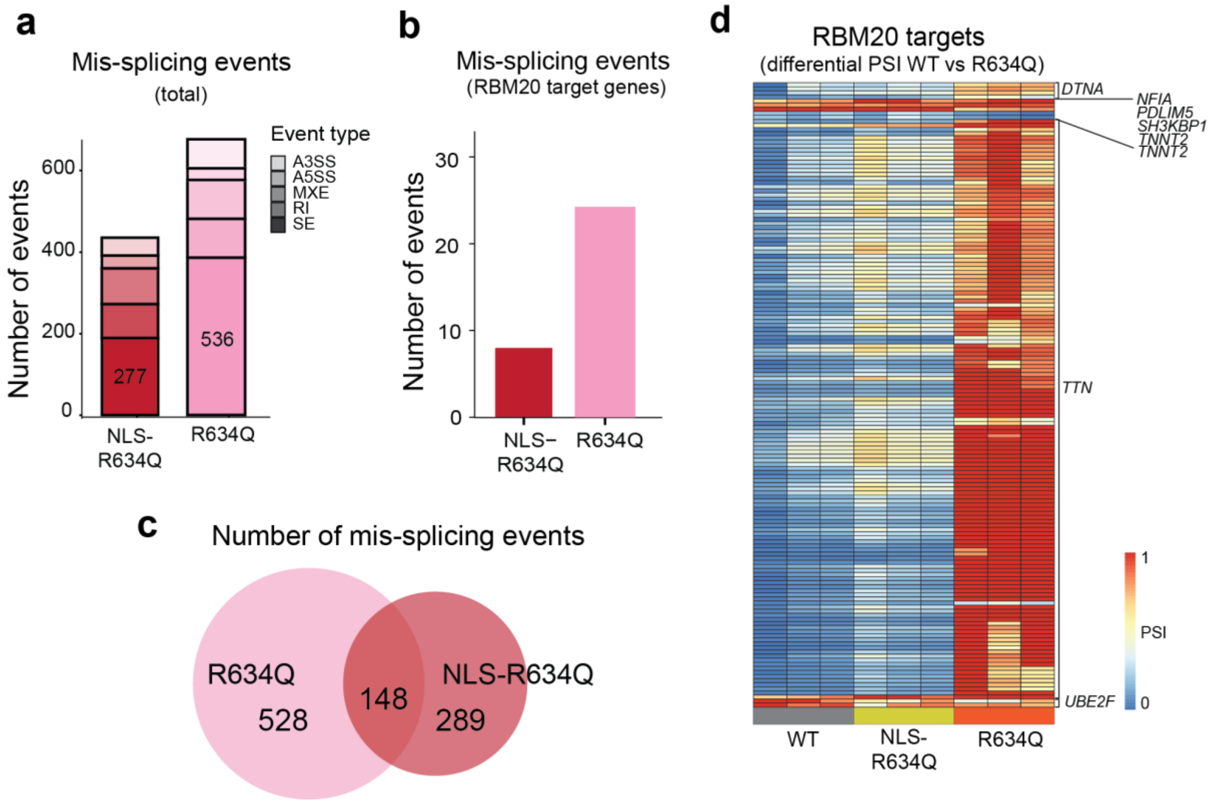


Figure 3.6. Alternative splicing differences between cells with RBM20-R634Q and -SV40-NLS-R634Q (a) Total number of differential splicing events (inclusion level differences greater than 0.1 and false discovery rate less than 0.01, rMATS[91]) in pairwise comparisons to RBM20-WT-expressing cells. SE = exon skipping, RI = intron retention, MXE = mutually exclusive exons, A3SS and A5SS are alternative 3' and 5' splice sites, respectively. (b) Numbers of differential splicing events from panel (a) that are located in the core RBM20 targets. (c) Numbers and relative overlaps of all differential splicing events from panel (a). (d) Percentage of spliced-in (PSI) values (calculated according to[92]) of exonic parts located in the core RBM20 targets that were different (delta PSI > 0.1, p-value < 0.05, t-test, two-sided) between RBM20-WT and RBM20-R634Q expressing cells. Three biological replicates for each cell line were used for the RNA-sequencing experiment.

3.1.5. Assessment of splicing activity and mislocalization of RBM20 variants in HeLa reporter cell lines

While iPSC-CMs are a suitable cell line model to study RBM20 biology in its native environment where it is endogenously expressed, their propagation for high throughput experiments like genome-wide CRISPR screens or proteomics is extremely challenging. The differentiation protocol is time- and reagent-consuming, and not 100% efficient. Moreover, they are post-mitotic cells and therefore it is impossible to generate more iPSC-CMs once they are differentiated. Due to these reasons, I generated cell lines with ectopic expression of RBM20-tagged with eGFP in human HeLa cells, where no endogenous RBM20 is expressed. I used cells already expressing doxycycline-inducible Cas9 (TetO::Cas9)[84] and introduced N-terminally tagged with eGFP-RBM20 via lentiviral transduction and generation of single-cell derived lines that displayed the most stable and bright eGFP-RBM20 expression (see methods). I generated four stable lines: RBM20-WT, -P633L, -R634Q, and -R634Q/S635E/S637E (RSS). The latter was at first generated to test whether phosphomimicking mutations of the serine residues 635 and 637 could rescue mislocalization of RBM20-R634Q, but in turn was used as a control cell line where RBM20 localization was the most cytoplasmic (see below).

To assess the validity of this model for the study of RBM20 nuclear transport, I first assessed eGFP-RBM20 localization in HeLa cells via confocal microscopy (Fig. 3.7 a), and quantified its co-localization with DAPI or DRAQ5 in microscopy (Fig. 3.7 b, see methods) or ICS (Fig. 3.7 c), respectively. RBM20-WT localised solely to the nucleus, same as in iPSC-CMs. Likewise, RBM20-P633L exhibited intermediate mislocalization, while RBM20-R634Q mostly localised to the cytoplasm (Fig. 3.7 a-c). I quantified the degree of correlation between RBM20-WT, -P633L, and -R634Q localization (based on microscopy) between iPSC-CMs and HeLa, and found a very strong linear correlation (Fig. 3.7 d). The latter supports that, most likely, the transport mechanism of RBM20 is preserved between the cell lines, and that the HeLa model is well-suited for this study. Interestingly, the RSS variant displayed the most severe mislocalization, with almost no protein detected in the nucleus (Fig. 3.7 a-c). This indicated that, phospho-mimicking did not rescue mis-localization, in line with the previous study[75]. Therefore, I used the RSS variant further as the most severely mislocalised control.

I then assessed splicing regulation in the reporter lines. Because *TTN* is not expressed in HeLa cells, I used another splicing target of RBM20 – *IMMT*. I used primers that bind exon-exon junctions between exons 5 and 7, excluding alternative exon number 6 that was characterised previously[55]. I found that, *IMMT* splicing correlated with the mislocalization severity in HeLa cells (Fig. 3.7 e,f). This once again supports that splicing function of RBM20 variants is likely preserved in HeLa cells.

To further test whether enhancing nuclear localization of RBM20 variants can also rescue splicing in HeLa cells, I overexpressed RBM20 variants tagged with SV40 NLS (Fig. 3.8). Similar to what I observed for iPSC-CMs, NLS-tagged P633L and R634Q variants localised to the nucleus, did not exhibit cytoplasmic granules, and expressed spliced *IMMT* isoform at the levels similar to those detected in RBM20-WT expressing cells (Fig. 3.8).

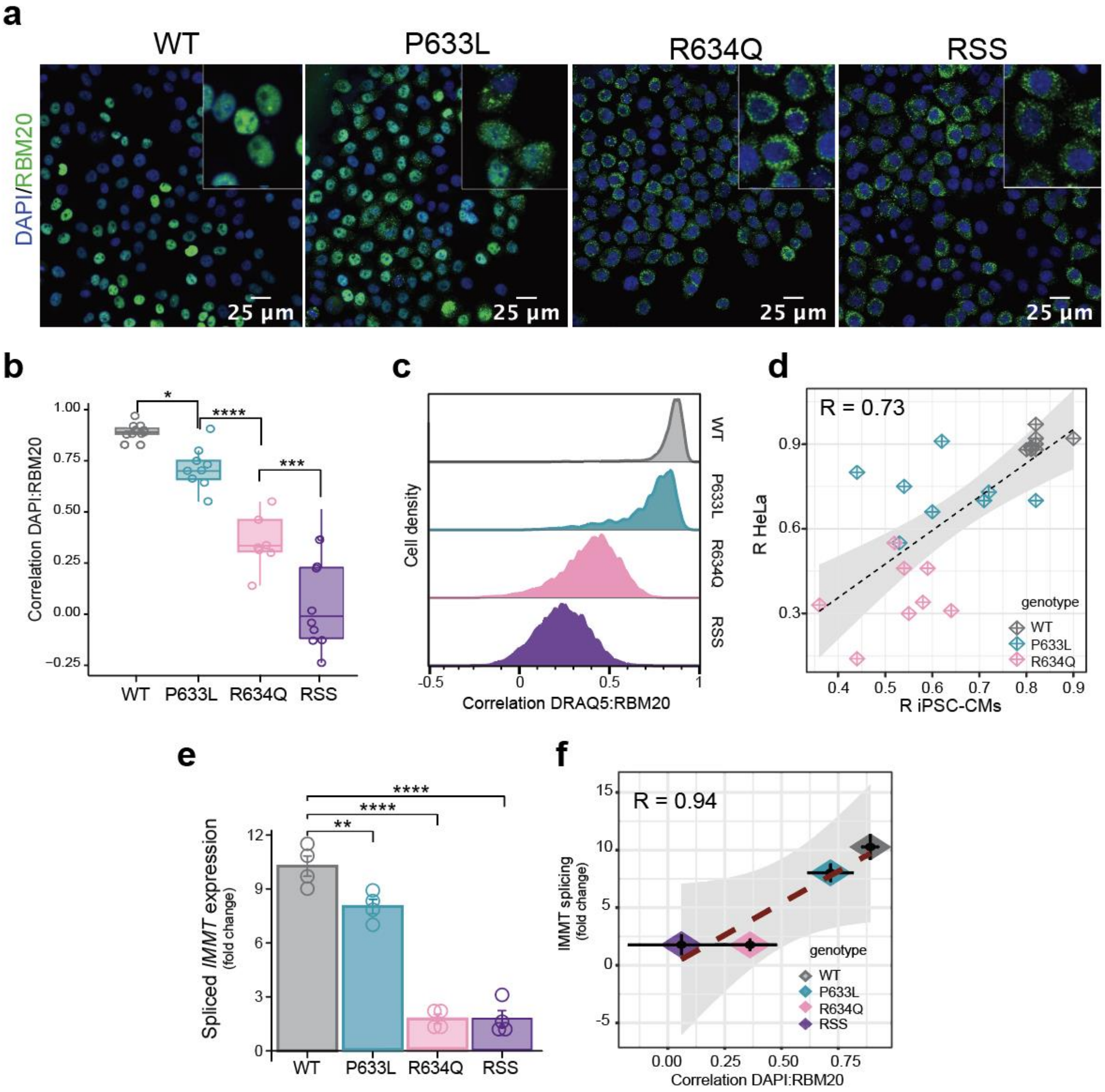


Figure 3.7. Splice deficiency of RBM20 variants correlates with their misclocalization in HeLa. (a) Representative confocal microscopy images of eGFP-RBM20 localization in HeLa, maximum projections. (b) Quantification of DAPI:RBM20 co-localization based on the data displayed in the panel (a). Each dot represents a Pearson correlation coefficient R for at least five cells (c) ICS-based analysis of eGFP-RBM20 co-localization with DRAQ5 in HeLa reporter cell lines. (d) Pearson correlation coefficient R values for DAPI:RBM20 correlation in iPSC-CMs (x axis) and HeLa (y axis) are presented as dots, each dot stands for at least five cells analysed. Linear regression between the two is displayed as a dashed line, and regression standard error is displayed in grey. (e) qPCR analysis of *IMMT* exon 6 spliced-out isoform expression. Data is normalised to *GAPDH* and is displayed as fold change versus the first replicate of the RBM20-WT line, two biological replicates with two technical replicates each, standard errors are indicated. (f) Means of *IMMT* splicing fold change (data from panel (e), y axis) and means of Pearson correlation coefficient R for DAPI:RBM20 correlation (data from panel (b), x axis). Linear regression between the two is displayed as a dashed line, and regression standard error is displayed in grey. Standard errors are indicated as black solid lines. Ns = not

significant, * - $p < 0.05$, ** - $p < 0.01$, *** - $p < 0.001$, **** - $p < 0.0001$ one-way ANOVA with Tukey's HSD post-test (two-sided).

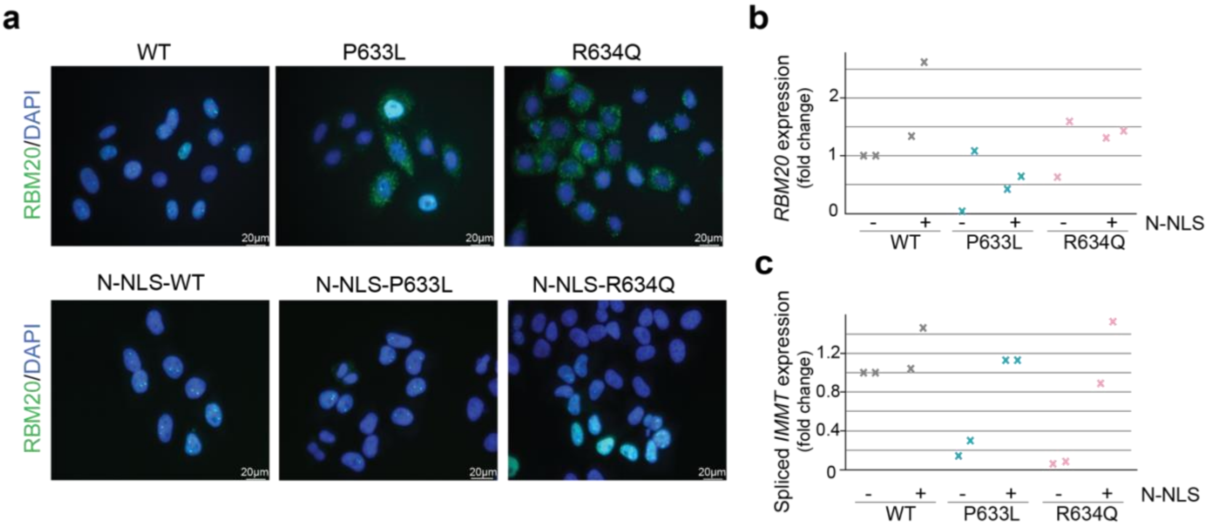


Fig. 3.8. SV40 NLS-tagging of RBM20 variants in HeLa rescues their localization and splice regulatory function. (a) Microscopy analysis of TetO-eGFP-RBM20 localization with or without RS domain mutations, with or without the SV40 NLS tag. (b) qPCR analysis of *RBM20* and (c) *IMMT* exon 6 spliced-out isoform expression. Data are normalised to *GAPDH* and displayed as fold change versus the first replicate of the RBM20-WT line, two biological replicates.

3.1.6. Intracellular localization and not an RS-domain mutation itself affects protein-protein interactors of RBM20

To investigate protein interactors of WT or mutated RBM20 in the nucleus and the cytoplasm, I worked closely with Dr. Marta Rodríguez-Martínez and the Proteomics Core Facility (PCF). I prepared HeLa reporter cells with constitutive expression of eGFP-tagged RBM20-WT, -P633L, or R634Q. HeLa cells with no eGFP-RBM20 expression were used as a negative control, where no bait is expressed. Marta and I harvested the cells, and Marta performed intracellular fractionation (protocol described in methods and in[80]). This was followed by pulling down of RBM20 and its interactors using GFP-antibody magnetic beads. We then submitted these samples for in-solution digest and mass spectrometry (MS) performed by PCF. Raw MS data was analysed by PCF, and I took over from there.

I classified a protein interactor to be specific to a certain variant in a certain fraction if its false discovery rate was less than 0.02 and a fold-change of at least 50 %, compared to the same intracellular fraction of the negative no bait control. The majority of identified interactors were shared between WT, P633L, and R634Q cells in the nucleus (66 out of 98, 116, and 139, respectively, Fig. 3.9 a). Common interactors in the nucleus were enriched for categories involved in protein folding, splicing regulation, and mRNA metabolism (Fig. 3.9 b). Only 17 proteins mildly lost their interaction with RBM20-P633L and -R634Q compared to -WT in the nucleus (Fig. 3.9 c), four of which were also identified before by Maatz et *al.*[53] to be enriched with WT protein (SRP14, RBM14, RBMX, and RBM15). Only one of them (RBMX) is a component of the spliceosome[101]. These results suggest that RBM20 variants can still interact with the majority of spliceosomal proteins if localised to the nucleus.

Next, I investigated interactors that were common and uncommon between WT and the two variants in the cytoplasm. I observed again that the majority of identified interactors for RBM20-WT were also identified to interact with RBM20-P633L and -R634Q variants (Fig. 3.9 a). However, I identified many proteins to be specifically binding to the mutant variants and most of them to both RBM20-P633L and -R634Q (Fig. 3.9 a). The common for RBM20-P633L and -R634Q interactors were enriched in mRNA metabolism, protein folding, and stress response, which is consistent with the proposed[73] localization of RS-domain variants to processing bodies (P-bodies, Fig. 3.9 d). I observed many proteins involved in spliceosomal complex to be gaining the interaction with the variants in the cytoplasm (Fig. 3.9 c). This could indicate that, RBM20 in the cytoplasm sequesters other components of the splicing machinery, potentially disrupting mRNA processing of other targets, in addition to the splicing defects observed in *RBM20* KO models. In addition, MOV10 and PUM1 gained interaction with the variants in the cytoplasm (Fig. 3.9 c), in agreement with the previous study[78]. Common interactors in the cytoplasm were enriched for spliceosomal proteins and proteins involved in mitochondrial organization (Fig. 3.9 e). Importantly, since RBM20-WT presents predominantly nuclear localization, in this experiment I likely underestimated its cytoplasmic interactors. Despite this, I identified two proteins – TNPO3 and CHD1 – to be losing their interaction with the RS-domain variants in the cytoplasm (Fig. 3.9 c). This provided the first insight that, TNPO3 might be responsible for RBM20 nuclear import and that RS-domain mutations prevent RBM20 variants to be imported by TNPO3.

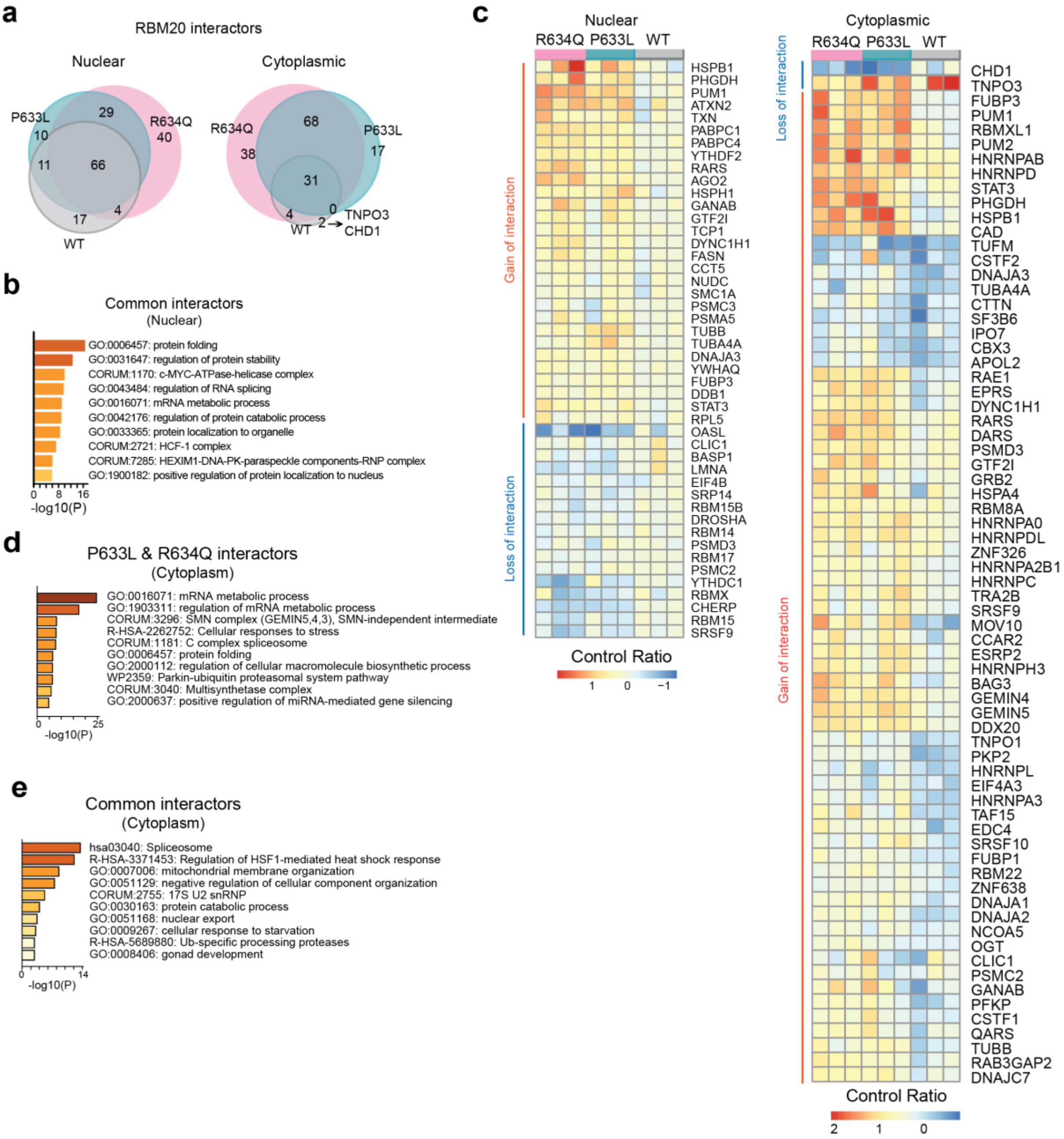


Figure 3.9. Analysis of RBM20 protein interactors in the nucleus and the cytoplasm. (a) Numbers and relative overlaps of interactors enriched with WT or RS-domain mutated RBM20 in the nucleus and the cytoplasm. A protein was considered as an enriched interactor if its fdr was less than 0.02 and a fold-change of at least 50 %, compared to the same intracellular fraction of the negative no bait control. (b) Enrichment analysis of the common interactors in the nucleus done with Metascape[90]. (c) Fold changes of interactors that were enriched with only WT but not with both of the RS-domain variants (loss of interaction), or that were enriched with both of the variants but not with WT (gain of interaction) in the nuclear or cytoplasmic fractions. The co-IP mass spectrometry experiment was done using three biological replicates for each of the fractions. (d) Enrichment analysis of the interactors specific for the RS-domain variants in the cytoplasm. (e) Enrichment analysis of the common interactors in the cytoplasm for RBM20-WT and both of the RS-domain mutant variants.

3.1.7. RS-domain RBM20 variants are retained in the cytoplasm

In order to assess whether faulty localization of RS-domain RBM20 variants is due to nuclear import defect or up-regulation of nuclear export, I treated HeLa reporter cell lines with the inhibitor of nuclear export – Leptomycin B (LMB). I used HeLa cells expressing RBM20-P633L or -R634Q under TetO promoter. I treated cells with either 50 nM of LMB or its solvent DMSO, and fixed them for microscopy 5- or 20-hours post treatment. As a control for nuclear shuttling dependent on nuclear export, I used staining with anti-p62 antibody. I did not see any impact of LMB treatment on RBM20-P633L or -R634Q localization (Fig. 3.10). While p62 protein, as expected, accumulated the nucleus after inhibition of nuclear export, RBM20 signal was unchanged. These data demonstrate that, RBM20 variants are retained in the cytoplasm upon inhibition of nuclear export mediated by exportin 1. This result evidences that mutations in RBM20's RS-domain likely impact nuclear import of RBM20, and not nuclear export, or that another nuclear export mechanism may be up-regulated for RS-domain RBM20 variants.

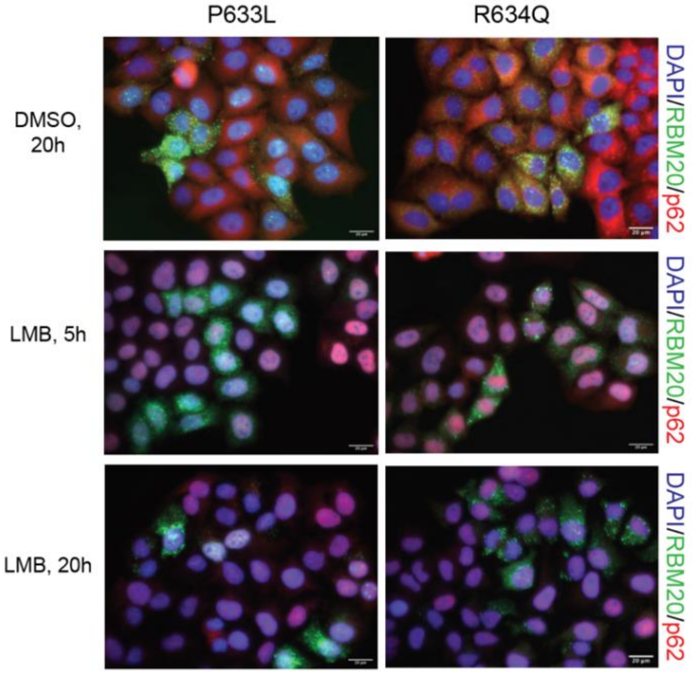


Figure 3.10. RBM20 variants remain in the cytoplasm upon inhibition of nuclear export by Leptomycin B (LMB), unlike p62.

Taken together, these data underly the importance of deciphering the molecular mechanism of RBM20 nuclear import, and identifying potential ways to restore it for RS-domain variants. The latter could both alleviate detrimental cytoplasmic granules, and restore alternative splicing.

3.2. Genome-wide CRISPR/Cas9 screens identify TNPO3 as the main nuclear importer of RBM20

3.2.1. Pooled genome-wide CRISPR/Cas9 screens

To enable the discovery of RBM20 nuclear import regulators in an unbiased way, I performed two genome-wide CRISPR/Cas9 knockout screens coupled with image enabled cell sorting as a readout (see methods and Schraivogel et al.[86]). For that, I made use of generated Tet:Cas9 eGFP-RBM20-WT and eGFP-RBM20-R634Q HeLa reporter cell lines, characterised in the previous chapter (see Figure 3.7.). For both screens, I applied genome-wide CRISPR/Cas9 libraries that were designed, produced, and provided by Dr Daniel Schraivogel (Figure 3.11). I used the ICS technology to isolate cells with cytoplasmic or nuclear RBM20 localization absed in the correlation between DRAQ5 and eGFP-RBM20, as described before.

I designed the first genome-wide screen using the RBM20-WT cell line (further referred to as "WT screen") to identify genes that are essential for nuclear import of wild type RBM20. The idea was that if we knock-out a gene that is involved in controlling the nuclear import process, RBM20-WT will then accumulate in the cytoplasm. With this aim, I transduced the reporter cell line with six guide RNA libraries targeting 18,408 protein-coding genes, 100 cells per gRNA coverage, at a low infectious rate (to enable 1 gRNA per cell, see methods, Figure 3.11 a). I selected cells with gRNA expression by culturing in the presence of puromycin for at least 17 days prior to ICS. I activated Cas9 by addition of doxycycline, and to enable stable KO, cells were cultured in the presence of doxycycline for a total of seven days (see methods, Figure 3.11 a). At least 17 days post transduction, one million of cells with 7 % highest and 7 % lowest DRAQ5:RBM20 correlation were sorted using ICS for each genome-wide library, operated by Dr. Marta Rodríguez-Martínez and myself. Together with the unsorted input sample, gRNA representation in each of the fractions was then identified by amplicon sequencing (see methods). I prepared sequencing libraries using genomic DNA extracted from the sorted cells (see methods), and sequencing was then done by the GeneCore. Data processing and analysis was done by me, with the help of Dr. Benedikt Rauscher whose example code I used to perform my own analysis. Plasmid libraries were sequenced previously[86] and already published data was used here for the analyses. The median count per gRNA for the sorted and input samples was, as intended, above 500 for all libraries besides upper and lower sorted samples form library 6, which were very slightly below 500 and above 400 (Figure 3.11 b). To account for differences in sequencing depth, I further normalised the counts to the median of the targeting controls. As can be seen from the distribution of the normalised counts for each sample (Figure 3.11. c, d), there are differences between the plasmid samples and the other samples for the gold standard essential genes[97]. No selection pressure was observed for nontargeting controls (Figure 3.11. d). To further visualize how well core essential and nonessential genes can be separated in each of the sub-libraries, I created a precision-recall curve plot (Fig. 3.11. e), and calculated the areas under the curves (Fig. 3.11. f). It can be clearly seen that essential and non-essential genes can be separated with high confidence based on the gRNA counts from all six genome-wide libraries (Area under the precision-recall curves > 0.85 for all libraries), indicating that the knockouts were efficiently introduced in the pooled screen experiment.

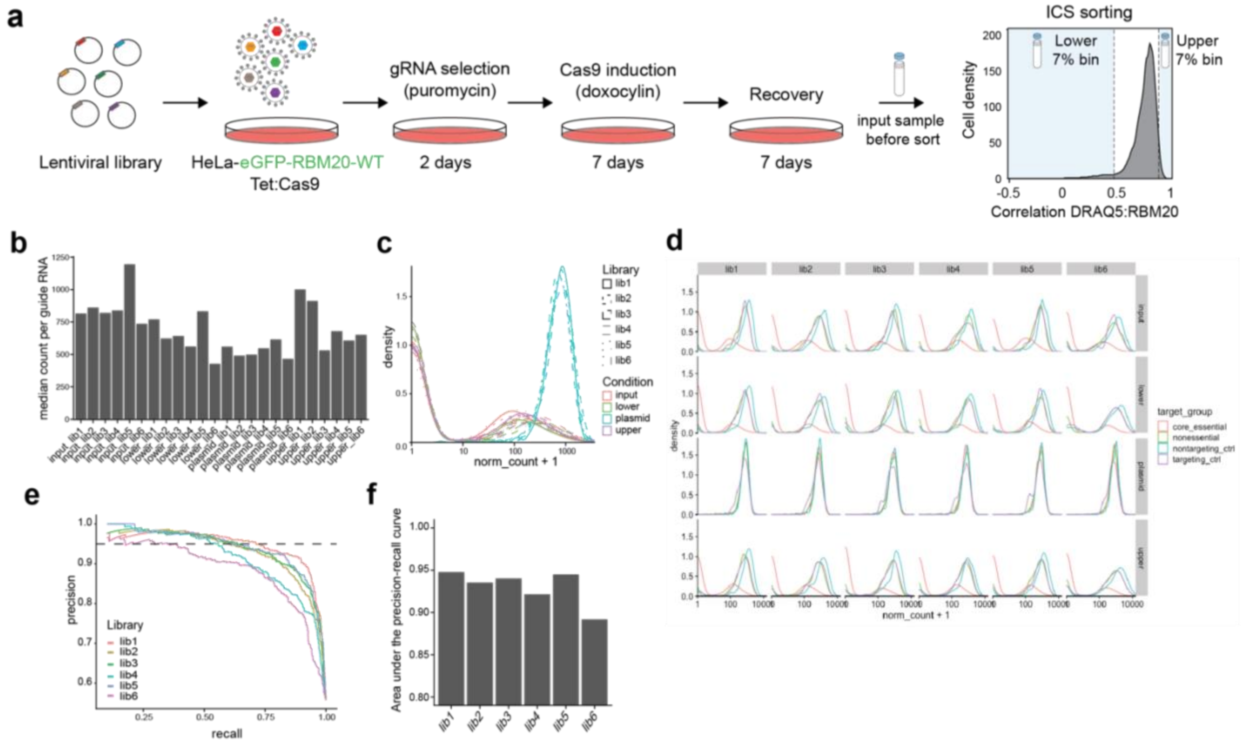


Figure 3.11. Genome-wide ICS CRISPR-screen to identify genes that are essential for nuclear import of RBM20-WT (WT screen). (a) Schematic outline of the genome-wide ICS screen with RBM20-WT. Six genome-wide sub-libraries were used to transduce HeLa cells expressing eGFP-RBM20-WT and Tet::Cas9, with 100 cells per gRNA initial coverage. Cells were sorted based on the correlation between RBM20 and DRAQ5 into 7% upper and 7% lower fractions at final coverage of 500 cells per gRNA per sorted bin. Unsorted input samples were collected too. (b). Median read count per gRNA for each of the six genome-wide libraries and each sample. (c) The distributions of normalised counts per library demonstrate abundance changes of gRNAs targeting core essential genes in the unsorted and sorted samples compared to the plasmid library. (d) The distributions of normalised

counts per library show abundance changes of gRNAs targeting essential genes in the unsorted and sorted samples compared to the plasmid library. Core-essential and non-essential gene annotations are based on[97]. (e) Precision-recall-curves indicate how well reference core essential and non-essential genes can be separated based on gRNA abundance changes in the unsorted samples after Cas9 induction compared to the plasmid library. (f) Areas under the precision recall curves shown in panel (e).

Next, I performed hit calling, using MAUDE[99], and comparing upper and lower sorted fractions to the unsorted input sample, and using plasmid library as a negative control. I identified only one hit with FDR < 0.01 (TNPO3), and ten hits with FDR < 0.05, that were depleted from the upper fraction and enriched in the lower fraction – essential for nuclear import, or positive regulators of nuclear import (Figure 3.12 a, b). I detected 56 genes to be enriched in the upper fraction with FDR < 0.01 (negative regulators of RBM20 nuclear import, Fig. 3.12 a, b). For the further validation, I selected only positive regulators of RBM20 nuclear import, either based on their low FDR scores (*TNPO3, CLDN14*, *GALE*, *ADAMTS16*, *SLC29A2*, *CEBPB*, *UBQLNL*, *TRIM33*, *PMM2*, *TRIM24*, *IPPK*, *XPO6*), or based on prior indications about their relevance for RBM20 function[50,81,102] (*TTN*, *AKT2*, *SPRK1*, *CLK1*, *LMNA*). I did not further analyse the potential negative regulators for two reasons. First, since RBM20-WT localization is solely nuclear, genes that may potentially make it "more nuclear" were not of direct relevance for the scope of this experiment aimed at identifying essential for nuclear import genes. Second, according to previous experiments[86], cellular shape amongst other parameters can influence the sorting decision of ICS, making these potential negative regulators likely to be, at least in part, experimental noise (e.g. mitotic cells). For the top hits, as well as for *RBM20* itself as a control, I plotted normalised read counts in the sorted, unsorted, and plasmid samples (Fig. 3.12 c). I detected a strong depletion of *RBM20*-targeting gRNAs in all libraries for sorted samples, as expected, since a KO of *RBM20* would result in a loss of the eGFP signal used for sorting. In addition, I observed the strongest effect on the gRNA representation for *TNPO3*, which was also the only hit with FDR < 0.01, validating the hit calling strategy.

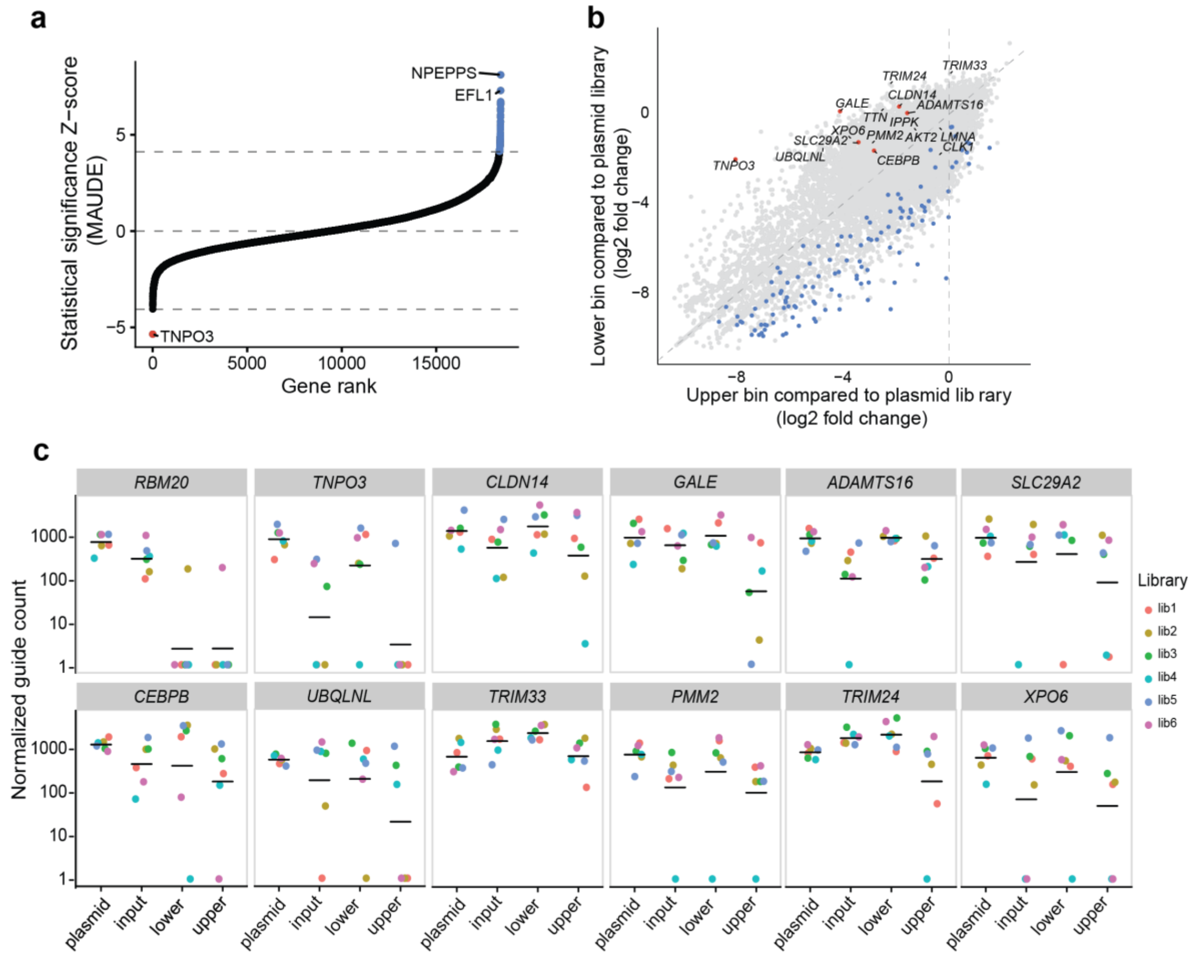


Figure 3.12. Hits of the WT screen. (a) Gene ranking based on statistical significance of hit calling with MAUDE[99]. The horizontal dashed lines indicate an FDR of 1%. Positive/negative regulators with FDR <1% are marked in red and blue, respectively. (b) Scatter plot visualizing gRNA abundance changes in the upper (x axis) and lower (y axis) sorted bins compared to the plasmid library. Red and blue dots indicate positive and negative regulators, respectively (FDR < 5% according to MAUDE). Labelled are positive regulators selected for further analyses. (c) Normalised read count per gRNA of the selected hits in all samples for all libraries.

I designed the second screen with RBM20-R634Q cell line (further referred to as "rescue screen") to identify genes that could potentially rescue the aberrant localization of mutant RBM20 (i.e. genes, KO of which could enhance nuclear import). The experiment was conducted in the same way as the WT screen with the following modifications. I applied only five, instead of six, genome-wide libraries to RBM20-R634Q Tet::Cas9 cells. After puromycin selection and Cas9 induction, Dr. Marta Rodríguez-Martínez and I sorted cells based on their DRAQ5:RBM20 correlation into 4% upper and 10% lower fractions, and collected the input samples for each of the libraries, as before (Fig. 3.13 a). We sorted only 4% upper this time to be more stringent in detecting the real effects of gene KOs, since RBM20-mutant localization displays stronger heterogeneity than that of RBM20-WT. The median read number per gRNA was similar to that of the WT screen (>500, Fig. 3.13 b). The detected dropouts of essential genes in the sorted and unsorted samples again indicated that gene KOs were successful in the pooled screen (Fig. 3.13 c). The precision-recall curve analysis (Fig. 3 d, e) indicated that the core essential and non-essential genes can be clearly classified based on the experiment.

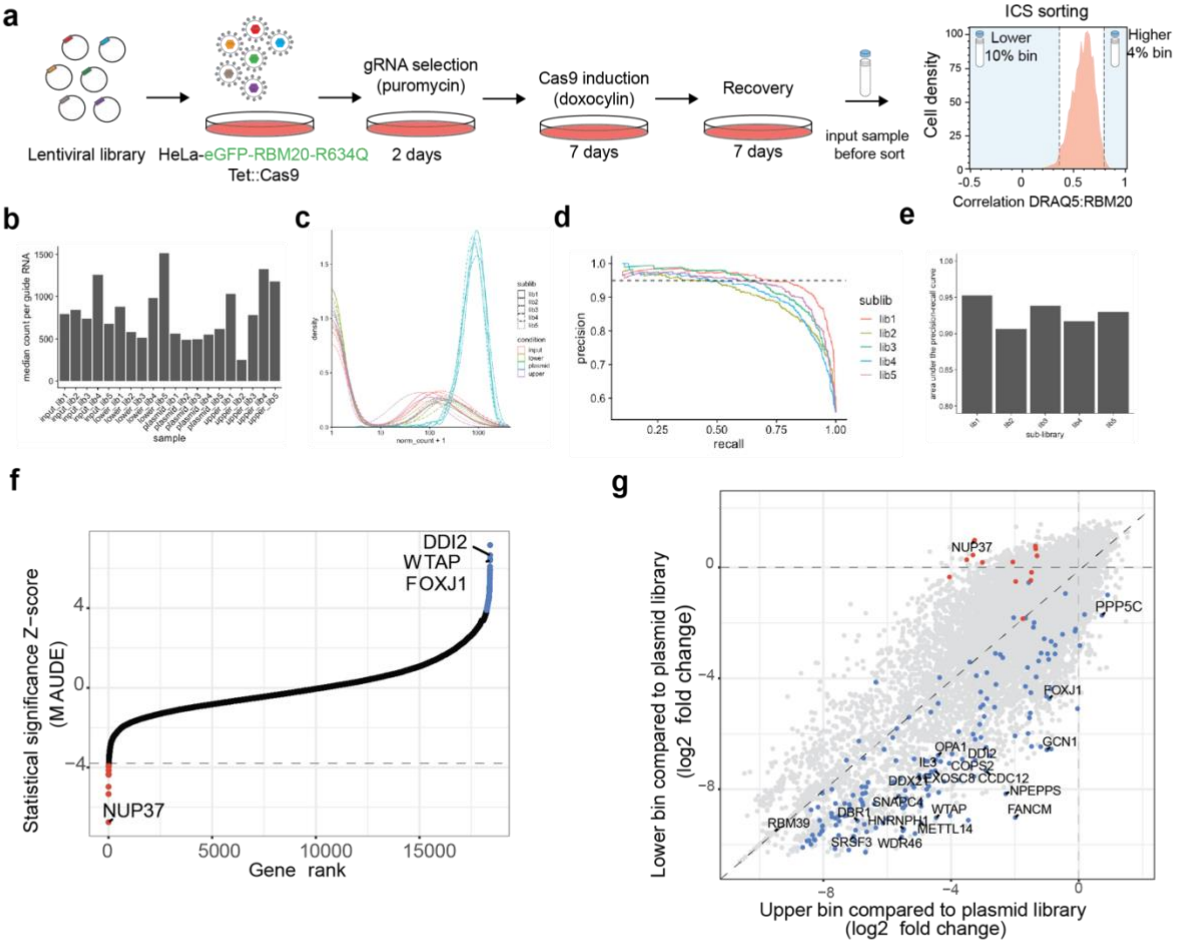


Figure 3.13. Genome-wide ICS CRISPR-screen to identify genes that may retain RBM20-R634Q in the cytoplasm (Rescue screen) and its hits. (a) Schematic representation of the Rescue screen. (b) Median read counts per gRNA for each library and each sample. (c) Density of the normalised counts per library for core essential genes based on[97]. (d) Precision-recall curves for separation of core essential and non-essential genes based on gRNA abundance changes in the unsorted samples input compared to the plasmid library. (e) Areas under the precision-recall curves shown in panel (d). (f) Gene ranking based on statistical significance of hit calling with MAUDE[99]. The horizontal dashed lines indicate an FDR of 1%. Positive/negative regulators with FDR <1% are marked in red and blue, respectively. (b) Scatter plot visualizing gRNA abundance changes in the upper (x axis) and lower (y axis) sorted bins compared to the plasmid library. Red and blue dots indicate positive and negative regulators, respectively (FDR < 5% according to MAUDE). Labelled are positive and negative regulators selected for further analyses.

Next, I performed hit calling with MAUDE[99] as described above. I identified 151 genes, knockout of which resulted in an increase of RBM20-R634Q nuclear localization (enriched in the upper and depleted in the lower fractions, FDR < 0.01, Fig. 3.13 f, g). These genes can be potential targets for restoring nuclear localization of RBM20 RS-domain variants. For the further analyses and validations that are discussed further, I focused on 21 genes (20 negative and 1 positive regulators of the nuclear import). I selected them based on their FDR score (lowest from the detected hits), and excluding potential false positive hits based on the previous knowledge. Altogether, results of these two genome-wide ICS screens provide an important insight into regulation of RBM20 nuclear import, and testing other detected hits (all deposited at GEO: GSE220833), that were not further tested by me in my PhD project, may reveal more details of the molecular mechanism regulating RBM20 nuclear import, and/or identify more potential targets for rescuing this process for RBM20 RS-domain variants.

In order to test whether identified in these two screens genes are also relevant for cardiomyocytes – cells with native RBM20 expression – I performed RNA sequencing of the RBM20-WT and -R634Q HeLa reporter cell lines, and compared expression of all selected hits in HeLa to iPSC-CMs (dataset analysed in Fig. 3.3, 3.4). I normalised read counts using DeSeq2[89] and combined them from three replicates of RBM20-WT and RBM20-R634Q cell lines (Fig. 3.14). I was able to detect expression of all but two (*CLDN14*, *FOXJ1)* selected hits from both screens (17 from the WT screen and 21 from the Rescue screen) in iPSC-CMs. I considered a gene to be expressed if its normalised read counts were above 50 on average for both cell lines and for all replicates. I observed high variability in gene expression levels for most of selected genes between HeLa and iPSC-CMs - which is to be expected due to completely different origins of these cells. The only insignificant (adjusted p-value > 0.01) differences were observed for four genes, namely *PMM2, DDI2, SNAPC4*, and *RBM39*. As expected, normalised read counts for *TTN* were below 15 in all replicates of HeLa cells (Fig. 3.14 b). Altogether, the presence of detectable gene expression of all but two (out of 38) selected hits confirms the relevance of identified candidates for regulators of RBM20 transport in iPSC-CMs.

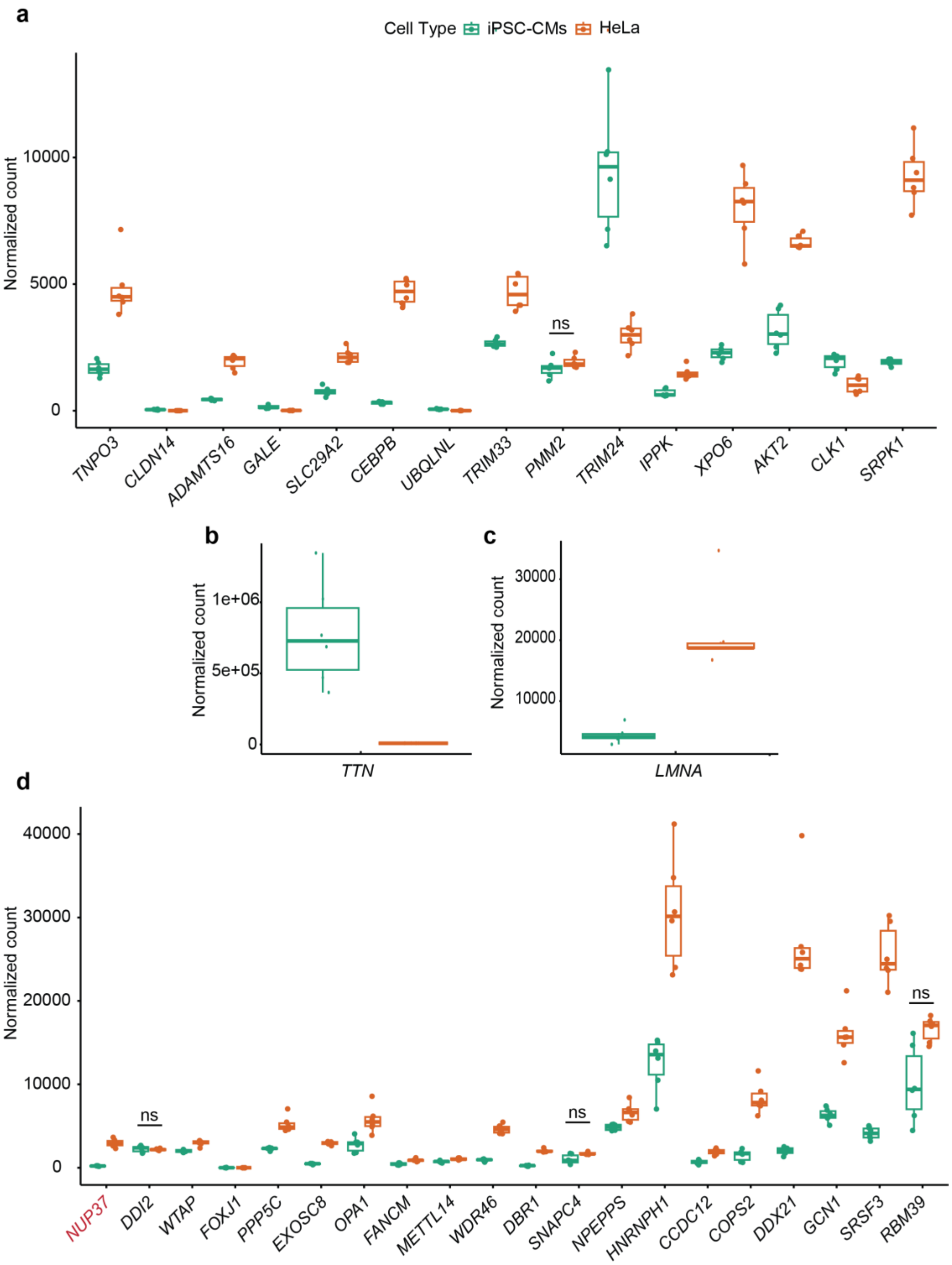


Figure 3.14. Normalised read counts (DeSeq2) in iPSC-CMs and HeLa. (a-c) for positive hits (from Fig. 3.12), and (d) for NUP37 and negative hits (from Fig. 3.13). Only four out of 38 genes were changed insignificantly between HeLa and iPSC-CMs (labelled ns). Only two tested hits (CLDN14 and FOXJ1) were expressed at the level below 50 normalised counts in iPSC-CMs, and TTN was expressed at the level below 50 in HeLa. Normalised read counts for *TTN* (b) and *LMNA* (c) are plotted separately, due to much higher level of their expression. N.s. = not significant, one-way ANOVA with Tukey's HSD post-test (two-sided).

3.2.2. Individual perturbations of the screen hits

To test whether identified in the pooled screens (described above) hits do indeed affect RBM20 localization, and how strong the impact of particular hits could be, I worked closely together with a master student in the lab who worked on her master thesis under my supervision - Brigit Tunaj - and Dr. Marta Rodríguez-Martínez (referred to as "we" in this chapter). All results of this chapter 3.2.2. were obtained as a result of the collective work.

We performed single gene knockouts in the same reporter cell lines as the ones used in the pooled screens. Brigit Tunaj cloned 38 lentiviral plasmids delivering sequences of the best gRNAs (according to the scores from the pooled screens) and puromycin resistance gene (see methods).

First, we focused on validating 17 selected hits from the WT screen. We prepared lentiviral vectors delivering these 17 hits, and applied them individually using the same transduction, puromycin selection, and Cas9 induction protocol, as I did for the pooled screen. Then, together with Dr. Marta Rodríguez-Martínez, we analysed the impact of these single KOs on RBM20 localization with ICS, and based on this I quantified the degree of RBM20 mislocalization upon each of the single KOs (Fig. 3.15 a). We observed that only one gene, *TNPO3*, had a dramatic impact on RBM20 localization, a few others (*TRIMM24, TRIMM33, CEBPB,* and *GALE*) mildly affected RBM20 nuclear import, and the rest had a minimal or no influence on RBM20 localization as measured by ICS (Fig. 3.15 a, b). Next, to further validate the most promising hits based on the ICS measurements, Brigit Tunaj analysed their impact on RBM20 localization using fluorescent microscopy and quantifying the co-localization between DAPI and RBM20 (Fig. 3.15 c, d). She identified only one highly statistically significant (p-value < 0.001, Fig. 3.15 d) shift in RBM20-WT localization that was caused by *TNPO3* KO.

Next, we moved on to test the selected hits of the Rescue screen. Brigit cloned 21 plasmids with the best gRNAs (according to their Z-scores from the pooled screen) and delivered them individually via lentiviral transduction to Tet::Cas9 RBM20-R634Q HeLa reporter cells. We used the same transduction, puromycin selection, and Cas9 induction protocol, as I did for the pooled screen. Then, we performed the same analyses of RBM20 localization as described above for the hits of the WT screen. We could see that several genes had a mild impact on RBM20-R634Q localization as measured by ICS (Fig. 3.16 a, b), fluorescent microscopy (Fig. 3.16 c, d), and even *IMMT* splicing as measured by qPCR (Fig. 3.16 e). In particular, knockouts of *DDI2, PPP5C, WTAP,* and *OPA1* resulted in mild, but significant shift in RBM20 localization. However, these effects were not as strong as the one observed for RBM20-WT upon *TNPO3* KO discussed above (Fig. 3.15). Notably, TNPO3 was one of the two proteins that lost their interaction with the RS-domain variants (Fig. 3.9). We did not detect the other one – CHD1 – in either of the CRISPR screens. Therefore, we further focused on studying TNPO3 in more detail.

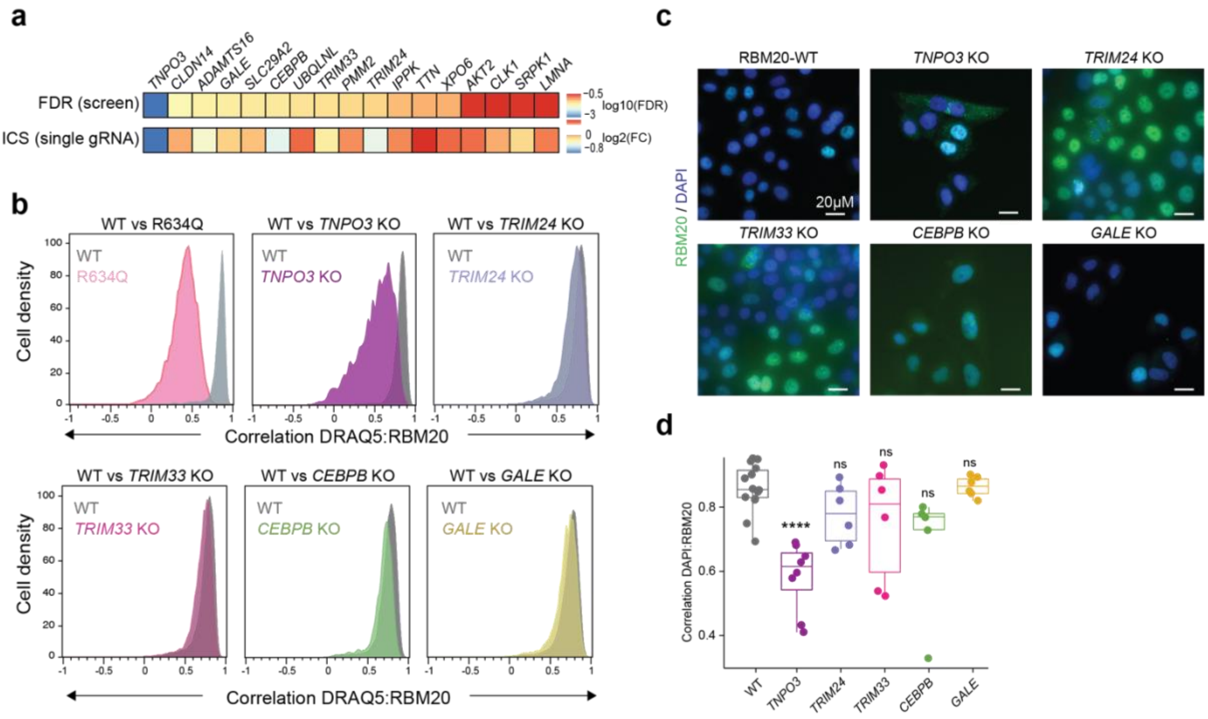


Figure 3.15. Single knockouts of the WT screen hits. (a) ICS-based quantification of the impact of single knockouts (KOs) of the selected hits (one gRNA per gene picked based on the strongest Z-score from the pooled screen) on RBM20 localization. The top row in the heatmap illustrates the $\log_{10}$(FDR) value for each selected gene from the pooled screen. The second row shows the standardised difference in quantified RBM20 localization between the knock-out (KO) and control (WT) cell populations. It is quantified as log2 of the ratio (cell fraction with Pearson coefficient (DRAQ5:RBM20) > 0.7 in the knockout): (cell fraction with Pearson coefficient (DRAQ5:RBM20) > 0.7 in the WT). So, for the WT cells, the ratio is 1 and log2 is 0; the lower the ratio is upon a certain KO, the more cytoplasmic RBM20 localization is for that KO. (b) ICS histograms for DRAQ5:RBM20 localization for selected gene KOs. (c) Fluorescent microscopy analysis of RBM20 localization upon the selected gene KOs (same genes as panel b). (d) DAPI:RBM20 correlation quantified based on fluorescent microscopy images shown in panel c, for selected gene KOs. Each dot represents a Pearson correlation coefficient R for at least five cells, n=3. N.s. = not significant, *** - $p < 0.001$, one-way ANOVA with Tukey's HSD post-test (two-sided).

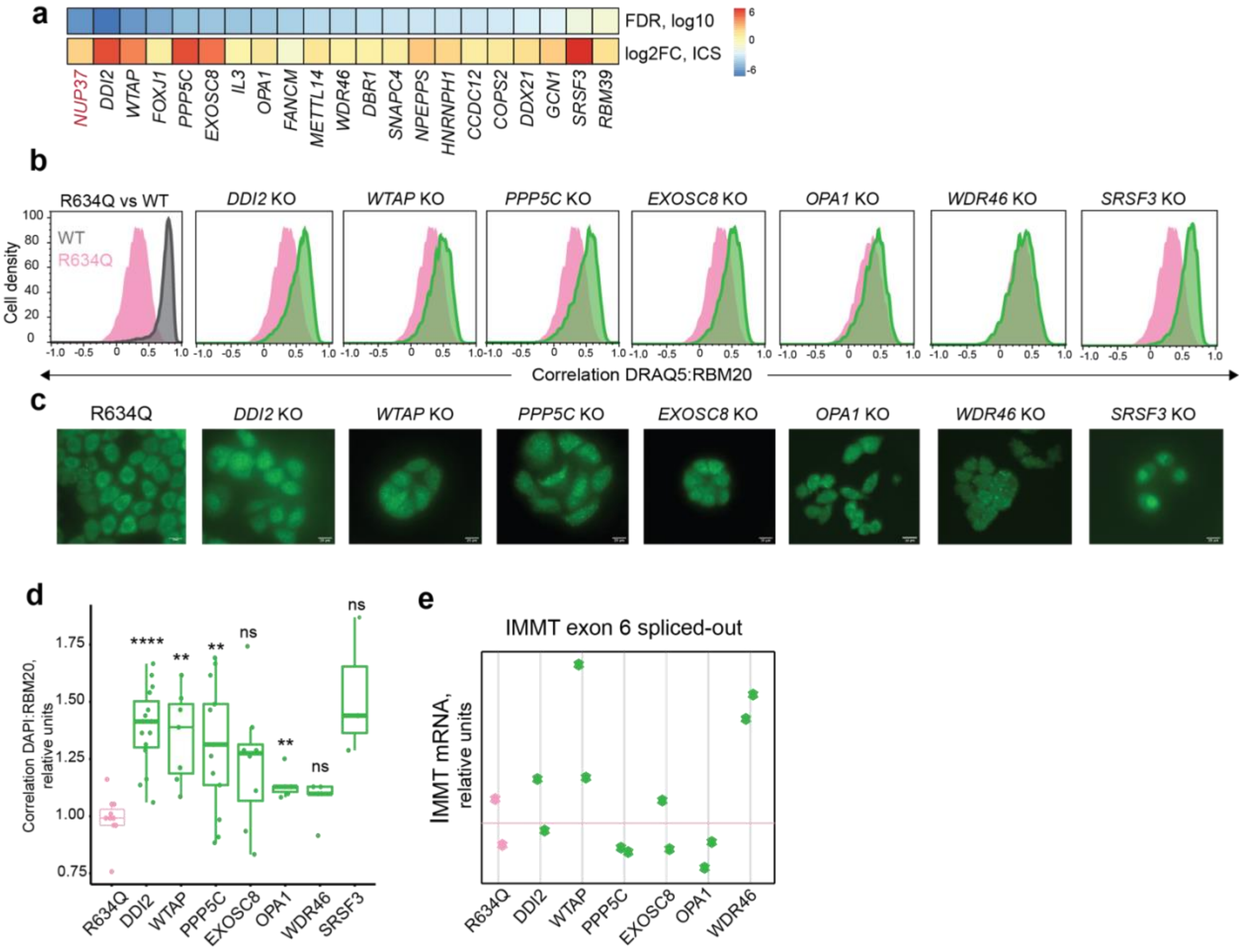


Figure 3.16. Single knockouts of the Rescue screen hits. (a) ICS-based quantification of the impact of single knockouts (KOs) of the selected hits (one gRNA per gene picked based on the strongest Z-score from the pooled screen) on RBM20 localization. The top row in the heatmap illustrates the $\log_{10}$(FDR) value for each selected gene from the pooled screen. The second row shows the standardised difference in quantified RBM20 localization between the knock-out (KO) and control (R634Q) cell populations. It is quantified as log2 of the ratio (cell fraction with Pearson coefficient (DRAQ5:RBM20) > 0.7 in the knockout) : (cell fraction with Pearson coefficient (DRAQ5:RBM20) > 0.7 in the WT). So, the for the R634Q cells, the ratio is 1 and log2 is 0; the higher the ratio is upon a certain gene KO, the more nuclear RBM20 localization is for that KO. (b) ICS histograms for DRAQ5:RBM20 localization for selected gene KOs. (c) Fluorescent microscopy analysis of RBM20 localization upon the selected gene KOs (same genes as panel b). (d) DAPI:RBM20 correlation quantified based on fluorescent microscopy images shown in panel c, for the selected gene KOs. Each dot represents a Pearson correlation coefficient R for at least five cells, n=3. (e) qPCR analysis of *IMMT* exon 6 splicing upon the indicated gene KOs. Data are normalised to *GAPDH* and are displayed as fold change versus the first replicate of the RBM20-WT line, two biological replicates, standard errors are indicated. N.s. = not significant, * - $p< 0.05$, ** - $p < 0.01$, *** - $p < 0.001$, **** - $p < 0.0001$, one-way ANOVA with Tukey's HSD post-test (two-sided).

We tested the effect of *TNPO3* siRNA knockdown on RBM20 localization in the reporter RBM20-WT and -mutant HeLa cell lines (Fig. 3.17). We found that, TNPO3 depletion significantly affected nuclear localization of WT, P633L, and even RBM20-R634Q, as measured by ICS (Fig. 3.17 a) and confocal microscopy (Fig. 3.17 b, c). RSS localization did not change, as it was already fully cytoplasmic (Fig. 3.17 b, c) as I also showed before (Fig. 3.7). We confirmed that these effects were indeed caused by TNPO3 depletion by western blot analysis (Fig. 3.17 d). These results were the first indication that TNPO3 is responsible for localizing not just RBM20-WT, but also RBM20 variants to the nucleus, and that its effectiveness is mutation-dependent. We further confirmed this hypothesis by re-analysing our mass spectrometry data of co-immunoprecipitation experiment (Fig. 3.9). We found that, the stability of the TNPO3-RBM20 interaction decreased in line with the severity of RBM20 mislocalization, as observed by Western blot (Fig. 3.17 e) as well as by mass spectrometry analyses (Fig. 3.17 f). This effect could not be explained by the potential differences in the amounts of TNPO3 expressed in different cell lines – they appeared to constant between WT and mutant RBM20 cell lines (Fig. 3.17 e).

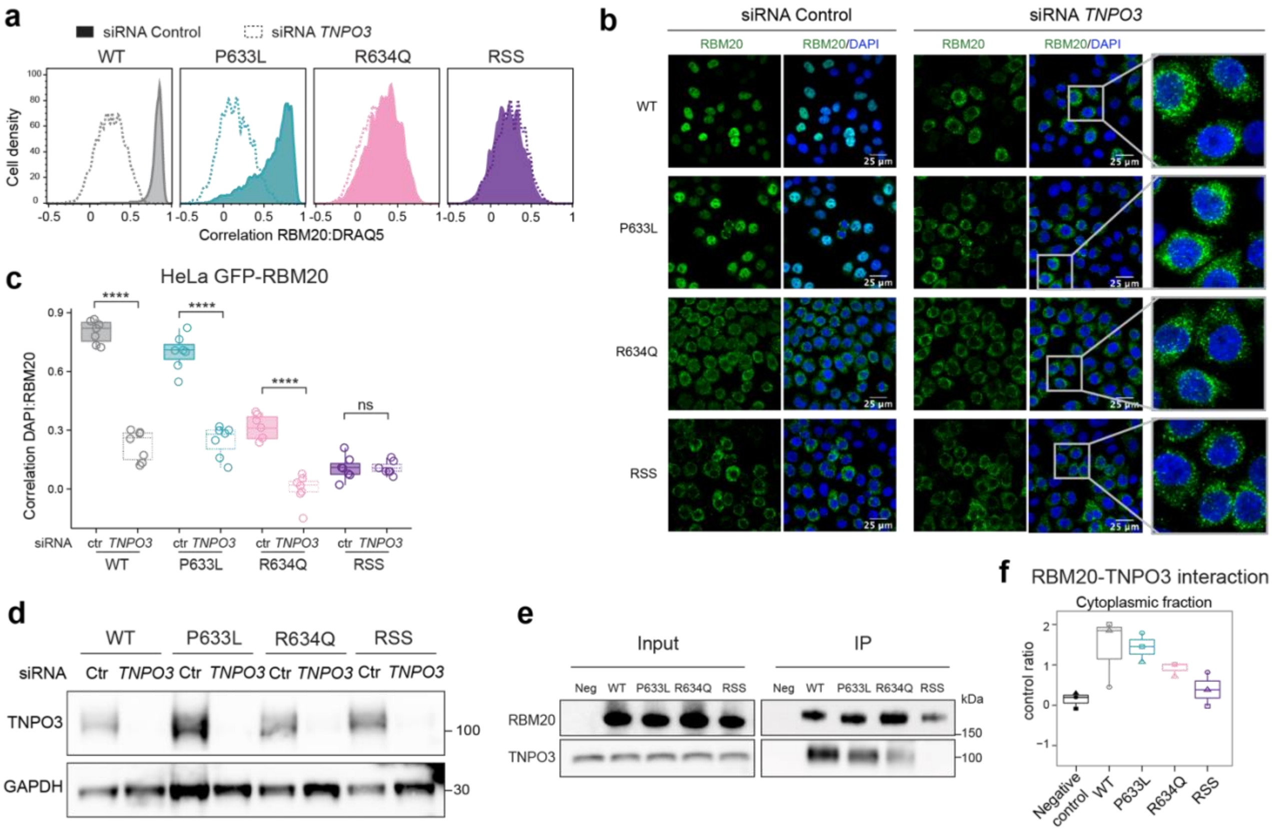


Figure 3.17. The severity of RBM20 mislocalization is proportional to the severity of loss-of-interaction with TNPO3. (a) ICS analysis of RBM20 localization in RBM20-WT, -P633L, -R634Q, and -RSS HeLa lines, transfected either with control or targeting *TNPO3* siRNA. (b) Confocal microscopy images of RBM20-WT, -P633L, -R634Q, and -RSS HeLa lines, transfected either with control or targeting *TNPO3* siRNA. (c) Quantification of DAPI:RBM20 correlation based on the images illustrated in the panel b. Each dot represents a Pearson correlation coefficient for at least five cells. (d) Western blot analysis of TNPO3 levels in RBM20-WT, -P633L, -R634Q, and -RSS HeLa lines, transfected either with control or targeting *TNPO3* siRNA. GAPDH was used as a loading control. (e) Western blot analysis of RBM20 and TNPO3 protein levels in the cytoplasmic fraction of HeLa, and their co-immunoprecipitation with eGFP-RBM20 (Neg = negative no bait control, experiment described in Fig. 3.9). (f) Quantification of TNPO3 peptides identified by mass-spectrometry in the cytoplasmic fractions of co-immunoprecipitants with indicated cells, normalised to Neg, n=3

(experiment described in Fig. 3.9). N.s. = not significant, **** - $p < 0.0001$, one-way ANOVA with Tukey's HSD post-test (two-sided).

Further, a postdoctoral fellow in the lab – Dr. Kai Fenzl – used AlphaFold2 and MutaBind2 programs to structurally model the direct interaction between RBM20 and TNPO3. He concluded that, RBM20's RS-domain was predicted to be directly interacting with the inner pocket of TNPO3, and that RS-domain mutations destabilised this complex in proportion to the severity of RBM20 mislocalization (see in[80]).

Altogether, we (my master student Brigit Tunaj, Dr. Marta Rodriguez Martinez, and I) were able to validate that TNPO3 is the major RBM20 nuclear importer. Mutations in RBM20's RS-domain affect the direct interaction with TNPO3, which causes the mislocaization. The severity of mislocalization is proportional to the severity of loss-of-interaction with TNPO3. In addition, we showed that, other factors (like DDI2, PPP5C, and WTAP) may have an impact on retaining mutant RBM20 in the cytoplasm too, however their effect on restoring RBM20 localization is relatively minor, compared to the necessity of TNPO3. Therefore, in the following chapters I focus on the impact of TNPO3 on RBM20 in iPSC-CMs and mice, and do not follow up on the other potential regulators of RBM20 nuclear import.

3.2.3. TNPO3 is essential for RBM20 nuclear import in iPSC-CMs

As it is clear from the results of the previous chapter, TNPO3 is essential for nuclear import of RBM20 in HeLa. However, HeLa reporter cell lines, although represent a very useful tool to study RBM20 regulation in high throughput, do not endogenously express RBM20. Therefore, to ultimately confirm the results discussed above, I went on to study the necessity of TNPO3 for RBM20 nuclear import in iPSC-CMs.

First, I performed a siRNA knockdown (KD) of *TNPO3* in RBM20-WT and RBM20-P633L iPSC-CMs (Figure 3.18). Unlike for HeLa, one-time transfection with siRNA was not enough to efficiently deplete TNPO3 in iPSC-CMs, therefore I performed two consecutive transfections 4 days apart (see methods). First, I analysed RBM20 localization with confocal microscopy (Fig. 3.18 a, b), and ICS (Fig. 3.18 c). Given that siRNA was only transiently transfected, and there was no selection for transfected cells, the observed heterogeneity of the RBM20 localization shift was to be expected: cells with efficient *TNPO3* KD are expected to have RBM20 mislocalised, while cells without *TNPO3* KD are not. Indeed, I could see that, upon transfection with *TNPO3* siRNA, both RBM20-WT and RBM20-P633L localization was significantly shifted towards more cytoplasmic in a fraction of iPSC-CMs (Fig. 3.18 a-c). Importantly, RBM20-WT also formed granules in the cytoplasm (Fig. 3.18). This result confirmed that TNPO3 is essential for RBM20 nuclear import in iPSC-CMs, too. The observed differential mislocalization of RBM20 variants could not be explained by potential differences in *TNPO3* expression, according to the RNA sequencing data (Fig. 3.18 d), and qPCR (Fig. 3.18 e). KD of *TNPO3* in RBM20-WT and -P633L iPSC-CMs (Fig. 3.18 e) also resulted in down-regulation of *TTN* (Fig. 3.18 f) and *IMMT* (Fig. 3.18 g) spliced-out isoforms expression in both cell lines. Altogether, KD of *TNPO3* resulted in mislocalization and decrease in splice regulation of both RBM20-WT and -P633L in iPSC-CMs. These results validate the previous data from HeLa cells (Fig. 3.17) concluding that, TNPO3 is the main nuclear importer of RBM20 and that the TNPO3-RBM20 interaction is not fully abolished by the P633L variant. However, the heterogeneity of the observed phenotype (Fig 3.18 c) left unanswered the question of whether this is simply due to difficulties with establishing a homogeneous KD, or whether some cells might have alternative nuclear import mechanism.

Generally, homogenous KD and KO of *TNPO3* are challenging because *TNPO3* is an essential gene. It has been classified as a pan-essential gene, i.e. absence of *TNPO3* led to loss of fitness or cell death in multiple normal tissues and cell lineages[103,104]. This generally excludes the possibility of establishing a stable and homogeneous KO in any cell system, including iPSC-CMs. Therefore, to address the question of whether *TNPO3* is the only nuclear importer of RBM20 and that there are no bypassing mechanisms, I performed a *TNPO3* KO followed by cell fixation done early enough to capture cells with *TNPO3* KO before strong essentiality effects would lead to their complete loss from the culture. I then quantified the endogenous TNPO3 signal and the localization of endogenous RBM20 from microscopy data (Fig. 3.19).

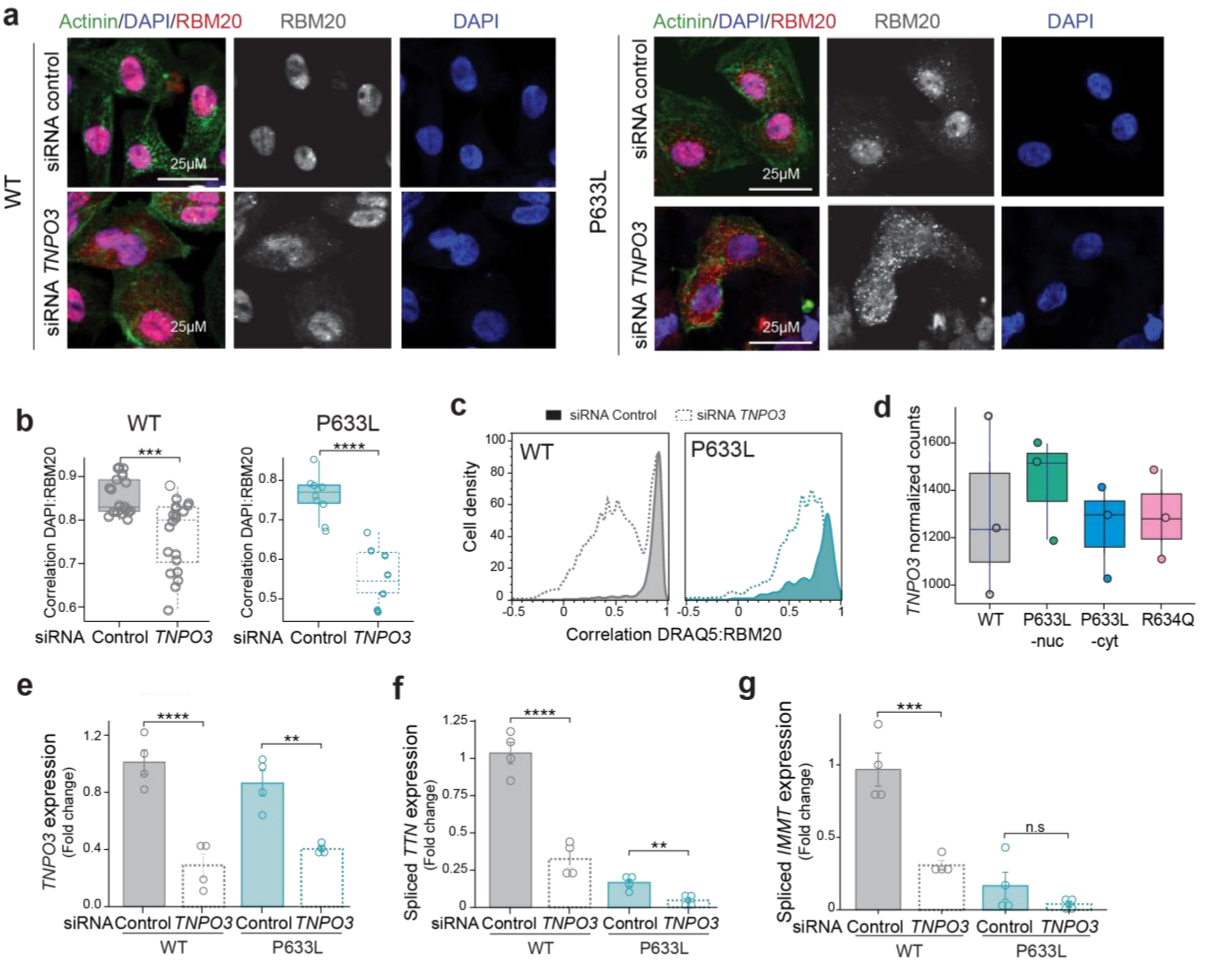


Figure 3.18. Knockdown of *TNPO3* in iPSC-CMs. (a) Representative confocal microscopy images of RBM20-WT or –P633L iPSC-CMs transfected with control or *TNPO3* siRNA. (b) DAPI:RBM20 correlation calculated based on the data demonstrated in the panel a. Each dot represents Pearson coefficient for at least five cells, n=3. (c) ICS-measured DRAQ5:RBM20 correlation in WT or P633L iPSC-CMs transfected with control or *TNPO3* siRNA. (d) Normalised read counts (DeSeq2[89]) for *TNPO3* expression in RBM20-WT, -P633L-nuc, -P633L-cyt, and -R634Q iPSC-CMs demonstrate that differential localization is not explained by differences in *TNPO3* expression. (e) qPCR analysis of *TNPO3*, (f) *TTN* exon 242 spliced-out isoform, and (g) *IMMT* exon 6 spliced-out isoform expression in RBM20-WT or -P633L iPSC-CMs with control or *TNPO3* siRNA normalised to *GAPDH* (mean fold change versus the RBM20-WT with control siRNA, with standard errors, two biological replicates with two technical replicates each). N.s. = not significant, * - $p < 0.05$, ** - $p < 0.01$, *** - $p < 0.001$, **** - $p < 0.0001$, one-way ANOVA with Tukey's HSD post-test (two-sided).

I used an iPSC line with constitutive Cas9 expression and RBM20-WT (kind gift from Dr. Dominik Lindenhofer and Moritz Kueblbeck), differentiated it to iPSC-CMs, and transfected it with a mix of three *TNPO3* targeting sgRNAs. As before, I performed a double-transfection of the sgRNA mix, four days apart. I then performed microscopy analysis 3 days after the last transfection. I stained the cells with anti-RBM20 and anti-TNPO3 antibodies (see methods). I observed only minor background signal in unstained and non-transfected cells (Fig. 3.19 a). I did not see any correlation between

RBM20 localization and TNPO3 levels in cells that were not transfected with sgRNAs (Fig. 3.19 b, d).

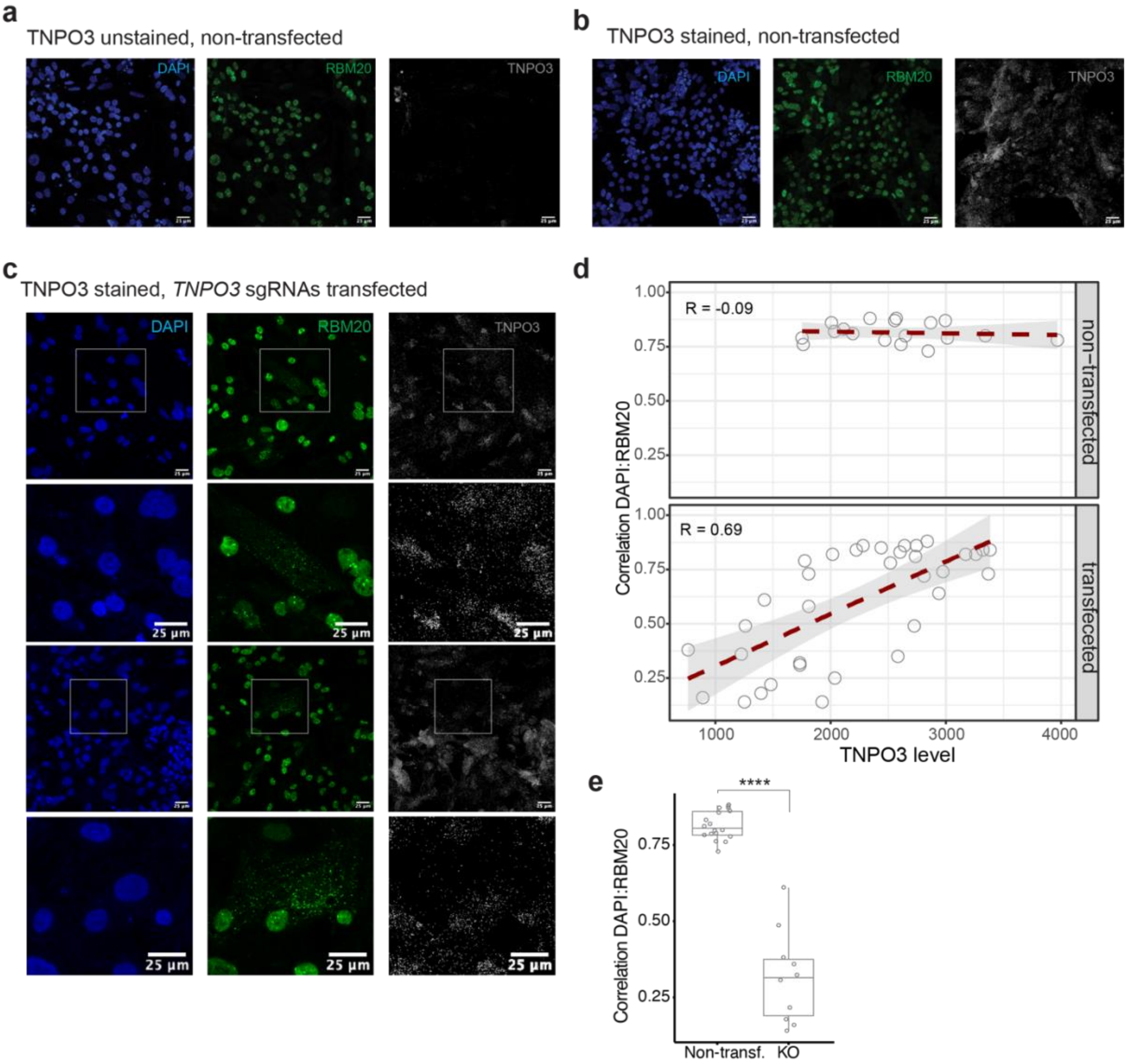


Figure 3.19. *TNPO3* knockout in iPSC-CMs. (a). A representative confocal microscopy image of iPSC-CMs stably expressing Cas9, not transfected with sgRNAs, and not stained with anti-TNPO3 antibody (background signal) (b) A representative confocal microscopy image of same iPSC-CMs as panel a, but stained with anti-TNPO3 antibody. (c) Representative confocal microscopy images of same iPSC-CMs as panel a, but stained with anti-TNPO3 antibody, and transfected with *TNPO3* targeting sgRNAs. (d) Quantification of correlation between RBM20 localization (DAPI:RBM20 co-localization) and TNPO3 expression in non-transfected and transfected with sgRNAs targeting *TNPO3* iPSC-CMs (data from panels b and c, respectively). Dashed line shows the linear regressions between the two, and regression standard errors are displayed in grey. Correlation coefficient = -0.09, p-value = 0.71 for non-transfected cells, and correlation coefficient = 0.69, p-value = 4.79e-06 for transfected cells. Each dot represents one cell. (e) DAPI:RBM20 co-localization comparison between non-transfected and KO cells (division is based on the average pixel intensity of TNPO3 staining more or less than 1750, respectively). **** - $p < 0.0001$ Student's T-test.

In contrast, I observed a strong correlation between TNPO3 levels and DAPI:RBM20 colocalization upon transfection with sgRNAs (Fig. 3.19 c, d). Next, I classified cells based on TNPO3 expression levels compared to non-transfected cells. I denoted them as "non-transfected" if the average pixel intensity per cell was greater than the lowest level observed in non-transfected cells, and "KO" if it was lower. I then directly compared RBM20 localization between these two populations (Fig. 3.19 e) and observed a statistically significant (p-value < 0.0001) difference. This result strongly suggests that there are no additional bypassing mechanisms of RBM20 nuclear import in iPSC-CMs.

In conclusion, results of this chapter suggest that, TNPO3 is the main nuclear importer of RBM20. RS-domain mutations in RBM20 destabilise the direct interaction between TNPO3 and RBM20, causing cytoplasmic mislocalization. Other factors like DDI2, PPP5C, WTAP, and OPA1 might have a mild impact on retaining mutant RBM20 in the cytoplasm, however, these results need to be further validated in cardiomyocytes. Next chapter will focus on testing TNPO3 as a potential therapeutic target for RBM20-DCM.

3.3. Up-regulation of TNPO3 rescues aberrant localisation of RBM20 and restores RBM20-mediated alternative splicing *in vitro* and *in vivo*

I hypothesised that enhancing the interaction between TNPO3 and pathogenic RBM20 variants could be used as a promising therapeutic strategy to relocalise RBM20 to the nucleus, thereby dissolving cytoplasmic granules and restoring RBM20-dependent alternative splicing regulation. This may be achieved either by increasing the availability of the transporter, or by directly stabilizing the interaction with a small molecule. Here, to provide the first proof-of-principle, I used the approach of overexpressing *TNPO3*.

3.3.1. Overexpression of TNPO3 in iPSC-CMs

To start, I tested whether overexpression of TNPO3 could rescue localization, thereby rescuing splicing in iPSC-CMs with RBM20-P633L and -R634Q variants. For that, I transduced differentiated RBM20-WT, -P633L, and -R634Q iPSC-CMs with lentiviral vector delivering eGFP-tagged *TNPO3* cDNA. A week later, I imaged transduced cells (Fig. 3.20 a-c), and sorted eGFP-positive cells for further RNA extraction and qPCR analysis of RBM20 targets (Fig. 3.20 d, e). It is apparent that in cells with high eGFP-TNPO3 signal, RBM20 localization is predominantly nuclear (Fig. 3.20 a, c). Of note, both variants were able to form the two characteristic nuclear foci usually observed for RBM20-WT, indicating once again their functionality upon nuclear relocalization. In addition, I observed a strong correlation between the intensity of TNPO3 signal and nuclear localization of RBM20-P633L and -R634Q (Fig. 3.20 b). These results demonstrate that supplying more transporter can directly enhance nuclear import of RS-domain RBM20 variants. Moreover, alternative splicing of *TTN* (Fig. 3.20 d) and *IMMT* (Fig. 3.20 e) was restored proportionally to the degree of nuclear relocalization. TNPO3 overexpression in RBM20-P633L iPSC-CMs resulted in restoration of nuclear localization to the level observed in WT iPSC-CMs, which coincided with restoration of *TTN* and *IMMT* splicing to the levels statistically indistinguishable from those in WT iPSC-CMs (Fig. 3.20 c-e). In turn, TNPO3 overexpression in RBM20-R634Q iPSC-CMs restored RBM20 localization, as well as *TTN* and *IMMT* splicing only partially (Fig. 3.20 c-e). These differences between the degrees of localization rescue for these two variants can be explained by the observed previously differences in their natural affinity to TNPO3 and the initial amount of protein found in the cytoplasm. While P633L only mildly affected TNPO3-RBM20 interaction, R634Q had a greater impact on the interaction stability (Fig. 3.17). Nonetheless, the observed partial restoration of nuclear localization for RBM20-R634Q could still beneficial for rescuing the disease phenotype. Therefore, I went on to testing this strategy further in our in vivo RBM20-DCM mouse models.

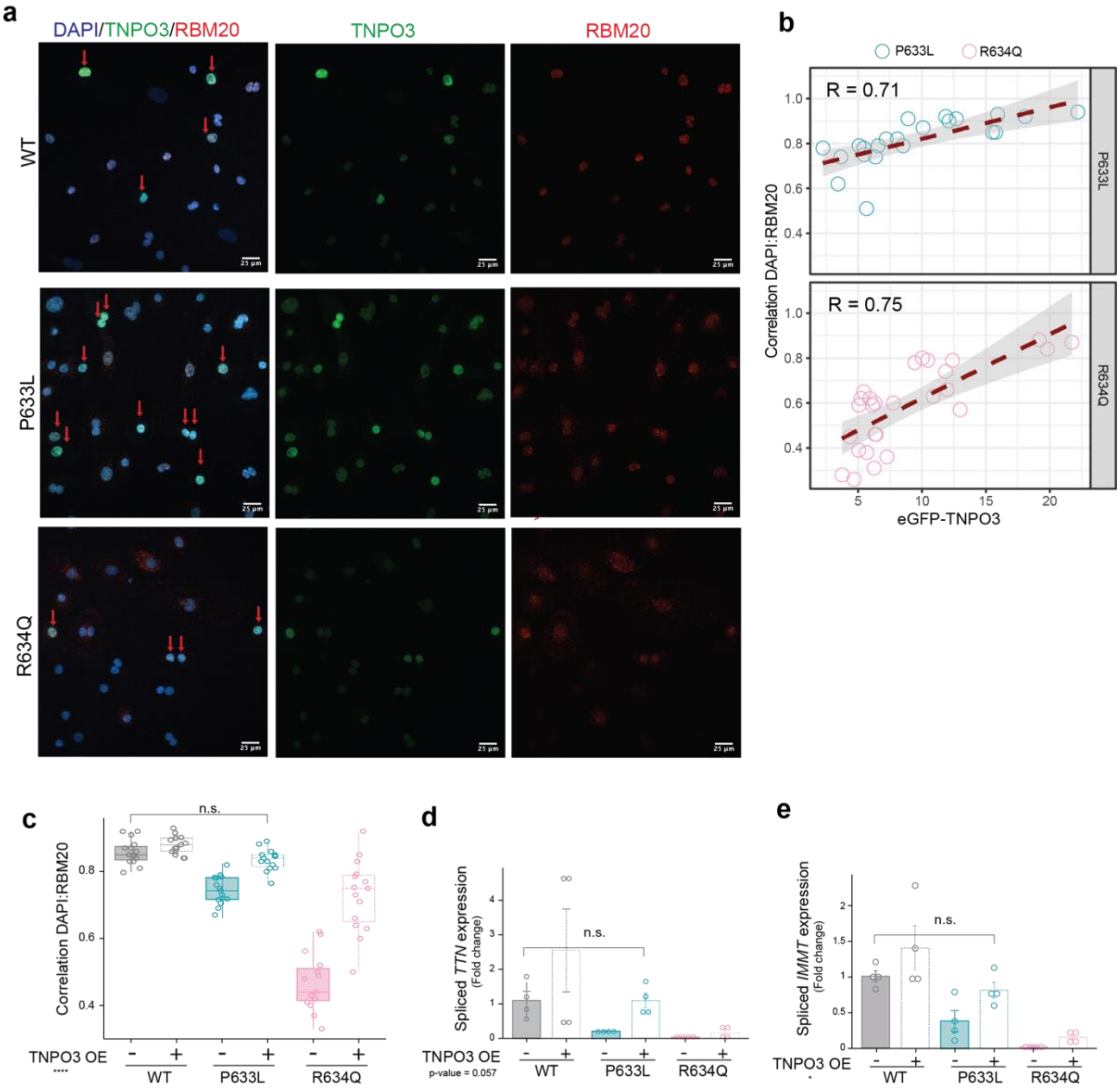


Figure 3.20. Overexpression of TNPO3 in iPSC-CMs restores nuclear localization and proportionally restores splicing of RBM20 targets. (a) Representative confocal microscopy images of RBM20-WT, -P633L, and -R634Q iPSC-CMs transduced with eGFP-tagged TNPO3. Red arrows point at successfully transduced cells (considered as eGFP positive). (b) Quantification of correlation between RBM20 localization (measured by DAPI:RBM20 correlation, y axis) and eGFP-TNPO3 signal intensity (x-axis) in RBM20-P633L and -R634Q iPSC-CMs transduced with eGFP-TNPO3 (sample images are shown in panel a). (c) Quantification of DAPI:RBM20 co-localization, division is based on the intensity of TNPO3 signal based on images demonstrated in panel a. Each dot represents a Pearson coefficient for at least five cells, n=3. TNPO3 overexpression effect's p-value < 0.0001. (d) *TTN* exon 242 spliced-out and (e) *IMMT* exon 6 spliced-out isoform expression in iPSC-CMs transduced with eGFP-TNPO3. Cells were FACS sorted based on the intensity of eGFP signal. *TNPO3* overexpression effect's p-value = 0.057 and 0.015, respectively. Isoforms expression was normalised to *GAPDH* and displayed as fold change versus the WT line (first replicate) without *TNPO3* OE (means with standard errors, two biological replicates with two technical replicates for each). For panels (c), (d), (e), TNPO3's overexpression effect's p-values were calculated with Two-way ANOVA with two-tailored Tukey's HSD post-test; comparison of WT vs P633L + TNPO3 - two-tailored t-test.

3.3.2. Overexpression of TNPO3 in vivo

In order to find the optimal in vivo experimental conditions, I first performed a pilot experiment with P635L$^{+/+}$ (homozygous, HOM, P633L in human) mice, generated and characterised by Grosch et al.[79]. I selected mice of two different age groups (5 weeks old, referred to as "young", and 11 weeks old, referred to as "old"), and tested two different concentrations of AAV9 virus delivering murine *Tnpo3* cDNA per age group (0.5x10^{12} and 1x10^{12} viral genomes (v.g.) for young mice; 1x10^{12} and 2x10^{12} v.g. for old miceI cloned the plasmid, and the virus production and titration were done by the EMBL Genetic and Viral Engineering Facility. All mouse handling and treatments were done by the EMBL Laboratory Animal Resources (LAR). After four weeks post injection of young and old mice with either PBS or AAv9-*Tnpo3*, together with Frank Diego Montoya Castillo and Laura Schraft, we performed the cardiomyocyte extraction protocol described in Grosch et al.[79] for further analysis of RBM20 localization. I also snap froze a piece of the left ventricle (LV) for RNA extraction, followed by qPCR analysis of *Tnpo3* expression. All mice showed a detectable increase of *Tnpo3* expression in the LV, although, the levels were higher in young mice even for lower viral concentrations (Fig. 3.21 a). I further performed an immunofluorescence staining for RBM20 in isolated primary mouse cardiomyocytes from RBM20-WT, and -P635L mice injected with either PBS or *Tnpo3* cDNA (Fig. 3.21b). I could see the expected nuclear localization of RBM20-WT and perinuclear granule appearance in RBM20-P635L mice injected with PBS, in line with what was previously observed[79]. In RBM20-P635L mice injected with AAv9-*Tnpo3*, RBM20 localization got restored in a fraction of cells (Fig. 3.21b). These results of the pilot experiment demonstrated that, AAV9-*Tnpo3* injection at chosen concentrations results in successful elevation of *Tnpo3* levels in mouse hearts, and that this is more efficient in younger mice. The up-regulation of *Tnpo3* expression in vivo resulted in partial re-localization of RBM20, which I considered promising enough to continue the experiment with more mice and more downstream analyses.

For the next experiment, I used six mice per genotype (RBM20-WT, -P635L, and -R636Q, all HOM) per condition (1x10^{12} v.g. of AAV9-control (pCAG-iCre-T2A-eGFP), further referred to as Ctr or AAV9-*Tnpo3*, further referred to as Tnpo3) all at four weeks old (Fig. 3.22 a). After injection of the AAV9 at LAR, animals were sent to our collaborator Dr. Maarten van den Hoogenhof at the Uniklinik Heidleberg. He measured echocardiograms (ECGs) for these mice every four weeks, and, after 12 weeks total, when mice were 16 weeks old, he performed the last ECGs measurements, humanely sacrificed the animals, and collected their heart tissues. He snap-froze the tissues, and prepared cryosections for future imaging. Then, he sent me all collected tissue samples for further analyses. While immunocytochemistry analysis of cryosections is still work in progress, I finished RNA-sequencing analysis based on RNA extracted from the LV of these mice.

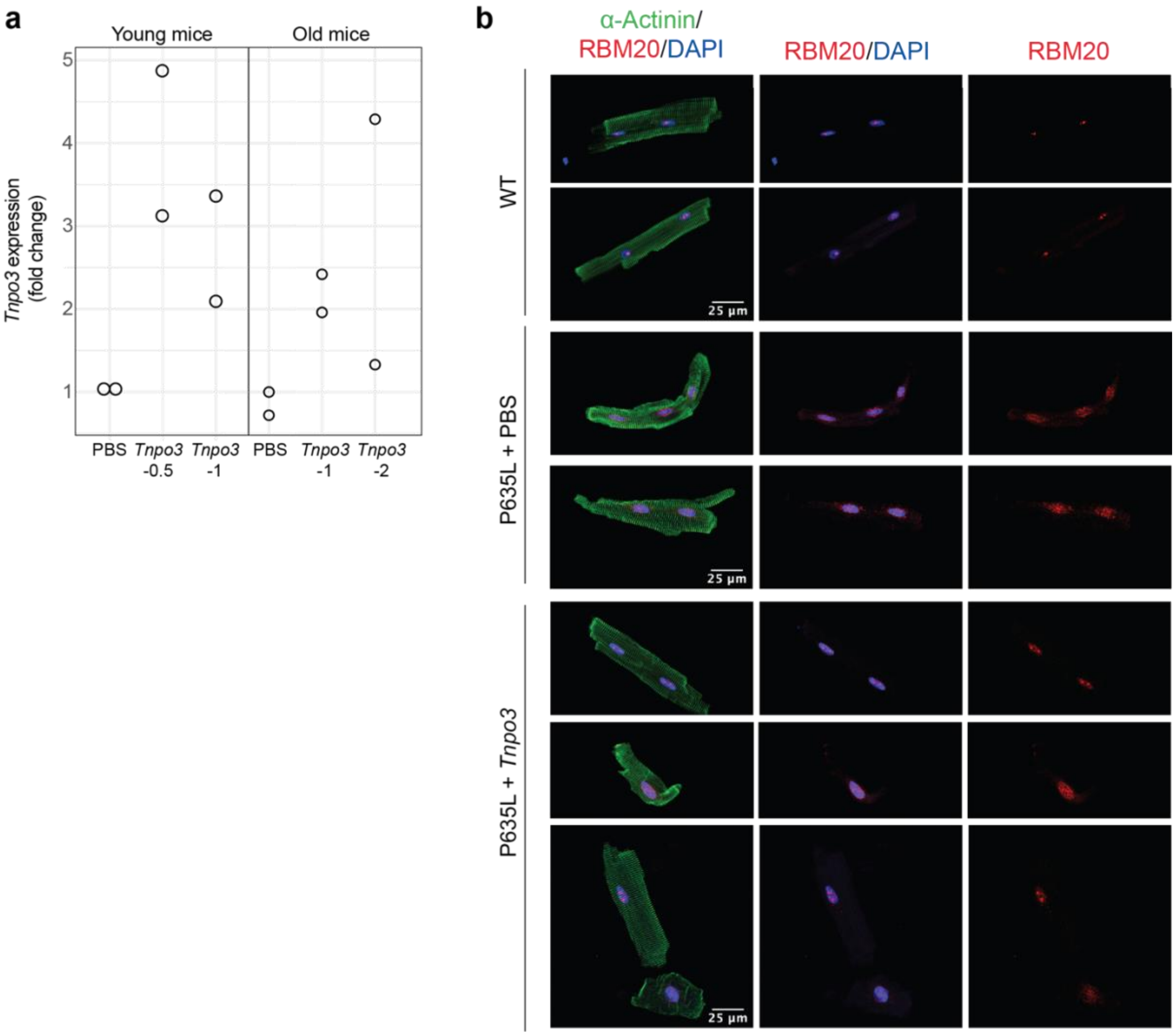


Figure 3.21. ***Tnpo3*** **expression and RBM20 localization in AAV9-Tnpo3-treated RBM20-P635L mice: pilot experiment.** (a) qPCR analysis of *Tnpo3* expression in the LV of mice injected with PBS, 0.5x10^{12}, 1x10^{12}, 2x10^{12} (labelled as *Tnpo3*-0.5, -1, and -2, respectively) viral genomes (v.g.) at 5 weeks old (young) or 11 weeks old (old), 4 weeks post injection. Data is normalised to *Gapdh* and displayed as fold change versus the PBS-injected young mice. Each dot represents a mean of two technical replicates per mouse. (b) Immunofluorescence analysis of RBM20 localization in isolated primary cardiomyocytes from RBM20-WT mice, and RBM20-P635L HOM mice injected with either PBS or with AAV9-Tnpo3 (1x10^{12} v.g., young mice).

First, to check whether there is indeed an increase in *Tnpo3* mRNA levels in the hearts of these mice, I performed qPCR analysis (Fig. 3.22 b). I found that, *Tnpo3* expression was significantly up-regulated in hearts of mice injected with AAV9-*Tnpo3* compared to hearts of injected with AAV9-*Ctr* mice (Fig. 3.22 b). Then, I selected four mouse RNA samples with the highest levels of *Tnpo3* expression as measured by qPCR for each genotype (WT, P635L, or R636Q) and each treatment type (Ctr or Tnpo3), and submitted them for RNA sequencing to the GeneCore Facility at EMBL. After processing the reads (see Methods), I first performed PCA analysis to see how well sample groups separate. All RBM20-WT mice clustered together, independently of the treatment (Fig. 3.22 c). In turn, RBM20-P635L mice clearly separated into two clusters: Tnpo3-treated clustered closer to WT, while Ctr-treated clustered closer to R636Q. Mice with RBM20-R636Q did not show a clear separation into two clusters depending on the treatment, and appeared the furthest from the WT mice (Fig. 3.22 c). This result demonstrates that, WT and R636Q mice clearly separate based on their gene expression which is not strongly impacted by *Tnpo3* overexpression. P635L mice display an intermediate phenotype between WT and R636Q, and overexpression of *Tnpo3* makes their gene expression more similar to that of WT mice. This is in line with all the accumulated data showing that P633L (P635L) mutation is less severe in terms of molecular and physiological DCM phenotypes than other pathogenic mutations in the PRSRSP stretch (including R636Q).

I then performed differential gene expression analysis. I did pairwise comparisons to RBM20-WT Ctr mice using DeSeq2[89]. I considered a gene to be differentially expressed if its log2FC was greater than 1, and p-adjusted < 0.05. Of note, *Tnpo3* was highly and significantly up-regulated in all *Tnpo3*-treated mice compared to WT Ctr. I found that, *Tnpo3* overexpression decreased the overall number of differentially expressed genes (DEGs) both in P635L and R636Q mice (Fig. 3.22 d). This effect was stronger for P635L mice, in line with the PCA analysis results: the number of DEGs went down from 1,338 to 457 upon Tnpo3 treatment, while for R636Q mice it decreased from 1,762 to 981 (Fig. 3.22 d). Strikingly, I identified only 3 genes to be differentially expressed between WT-Ctr and Wt-Tnpo3 mice (Fig. 3.22 d), with the strongest one being *Tnpo3*. This suggests that Tnpo3 treatment on its own does not lead to strong changes in gene expression profile. I further analysed whether DEGs overlap between the genotypes and the treatments (Fig. 3.22 e). It appeared that many

DEGs were shared between P635L and R636Q Ctr mice (732), and many of them were also shared between all mice (207). There were some genes unique to a specific genotype only, shared between Ctr and Tnpo3 mice (99 for P635L and 198 for R636Q). Importantly, there were only 9 genes that were differentially expressed in both P635L and R636Q Tnpo3 mice and not in Ctr mice. Only one of these 9 was also differentially expressed in WT Tnpo3 mice as well – *Tnpo3* itself. These results conclude that, *Tnpo3* overexpression in vivo does not have a systematic impact on overall gene expression. However, it reduces the differences in gene expression installed by RBM20 RS-domain mutations, making RBM20-mutant mice more similar to WT-mice.

Next, I performed differential splicing analysis in Tnpo3 treated mice compared to RBM20-WT Ctr mice with rMATS[91]. I considered an event to be differentially spliced if its FDR was less than 0.01 and if its |delta PSI| was more than 0.1. I identified 224 differential splicing events in WT Tnpo3 mice compared to WT Ctr mice (Fig. 3.23 a). This was substantially less than in P635L Ctr (819) and R636Q Ctr (929) mice. Tnpo3 treatment in RBM20-mutant mice decreased the number of mis-splicing events to 611 for P635L mice, and 620 for R636Q mice (Fig. Fig. 3.23 a). Importantly, there was a very little overlap (17 events) between all mice treated with Tnpo3 compared to WT Ctr mice (Fig. 3.23 b). This provides evidence that, similar to gene expression, *Tnpo3* overexpression does not lead to systemic changes in the splicing profile. I then looked closer at alternative splicing of core RBM20 targets (Fig. 3.23 c,d). In P635L mice, Tnpo3 treatment reduced the number of mis-splicing events in RBM20 targets from 45 to 23. In R636Q mice, the treatment did not change the overall number of mis-spliced events. In WT mice, the treatment did not lead to mis-splicing of RBM20 targets (only 5 events identified, Fig. 3.23 d). Importantly, based on the average PSI values for exons of RBM20 targets, P635L-Tnpo3 mice clustered together with WT-Tnpo3 mice, while P635L-Ctr and R636Q mice clustered together (Fig. 3.23 c). For many exons of RBM20 targets, *Tnpo3* overexpression restored their splicing to the levels seen in WT mice (Fig. 3.23 c). I then validated this result by qPCR analysis of *Ttn* splicing (Fig. 3.23 e) and confermed that, *Tnpo3* overexpression indeed restored *Ttn* splicing in P635L mice to the same levels as observed in WT mice. *Ttn* splicing did not change upon the treatment in WT mice, while in R636Q mice, although it was enhanced to the levels observed in P635L Ctr mice, it did not reach the WT levels (Fig. 3.23 e).

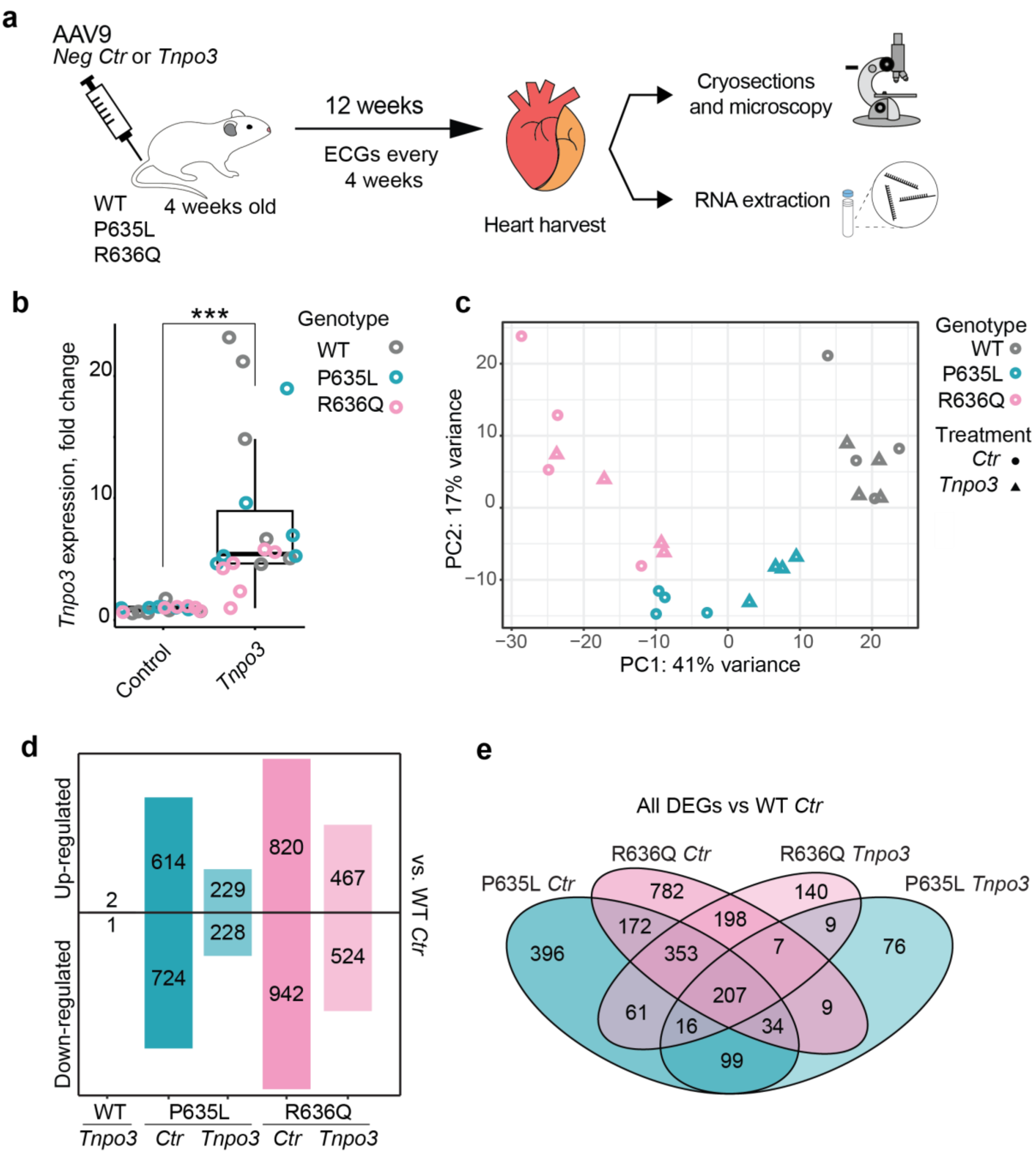


Figure 3.22. TNPO3 overexpression in RBM20-WT, -P635L, and -R636Q mice: effects on the gene expression. (a) Schematic representation of the experiment (b) qPCR analysis of *Tnpo3* expression in the left ventricles of RBM20-WT, -P635L, or -R636Q mice treated with either AAV9-pCAG-iCre-T2A-eGFP (*Ctr)* or AAV9-Tnpo3 (*Tnpo3*). Data is normalised to *Gapdh* and displayed as fold change versus one of the RBM20-WT Ctr mice. Each dot represents a mean of two technical replicates per mouse. (c). PCA analysis based on the normalised read counts (DeSeq2[89]) per gene, genome-wide. (d) Numbers of differentially expressed genes (DEGs, log2(fold change) > 1 and p-adj < 0.05, DeSeq2[89]) up- or down-regulated compared to RBM20-WT Ctr mice. (e) Venn diagram demonstrating numbers and relative overlaps of DEGs (up-and down-regulated combined) in RBM20-P635L or -R636Q mice either *Ctr*- or *Tnpo3*-treated.

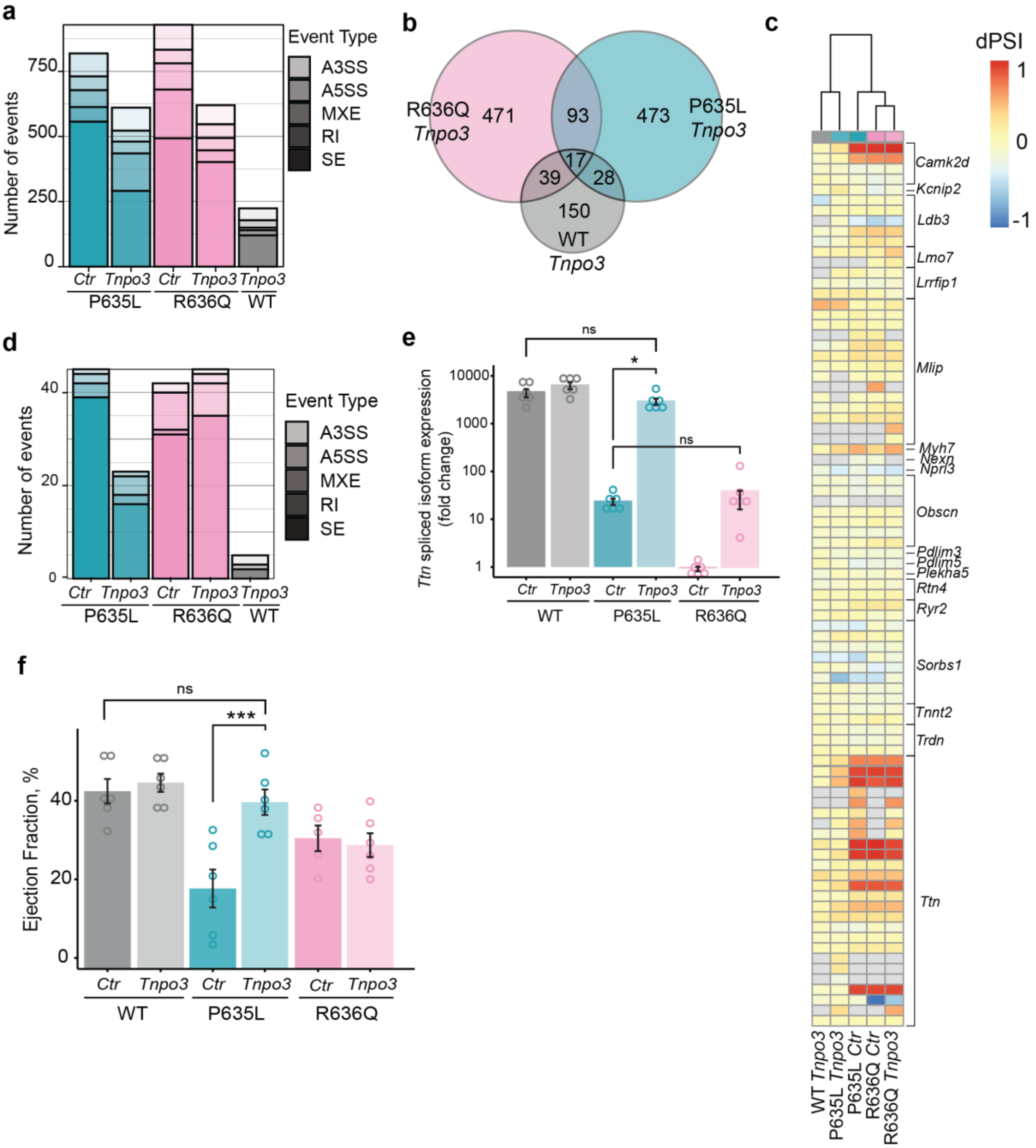


Figure 3.23. TNPO3 overexpression in RBM20-WT, -P635L, and -R636Q mice: effects on alternative splicing and ejection fraction. (a) Numbers of alternative splicing events detected with rMATS[91] in left ventricles of RBM20-WT, -P635L, or -R636Q mice treated with either AAV9-pCAG-iCre-T2A-eGFP (*Ctr)* or AAV9-Tnpo3 (*Tnpo3*), in pairwise comparisons to RBM20-WT Ctr mice. SE = exon skipping, RI = intron retention, MXE = mutually exclusive exons, A3SS and A5SS are alternative 3' and 5' splice sites, respectively. (b) Numbers and relative overlaps of alternative splicing events (all categories combined) in mice treated with *Tnpo3*. (c) Average (for four replicates) delta PSI (dPSI) values versus those of RBM20-WT Ctr mice for all alternative splicing events in the core RBM20 targets detected in at least one of the pairwise comparisons. Grey colour indicates that this event was not detected as alternatively spliced by rMATS. Positive dPSI values indicate that this exon was more included in the sample compared to the RBM20-WT Ctr mice, and negative – more excluded. (d) Numbers of alternative splicing events located in the core RBM20 targets detected with rMATS[91] in pairwise comparisons to RBM20-WT Ctr mice. SE = exon skipping, RI = intron retention, MXE = mutually exclusive exons, A3SS and A5SS are alternative 3' and 5' splice sites, respectively. (e) qPCR analysis of *Ttn* spliced isoform expression in the LV of RBM20-WT, -P635L, or -R636Q mice treated with either *Ctr* or *Tnpo3*. Data is normalised to *Gapdh* and displayed as fold change versus one of the

RBM20-WT Ctr mice. (f) Ejection fraction (EF) measured by echocardiogram 12 weeks after treatment of RBM20-WT, -P635L, or -R636Q mice with either *Ctr* or *Tnpo3* (done by Dr. Maarten van den Hoogenhof). One-way ANOVA with two-tailored Tukey's HSD post-test was used for quantifying statistical significance in (e) and (f). Ns = not significant, * - $P < 0.05$, ** - $P < 0.01$, *** - $P < 0.001$, **** - $P < 0.0001$.

Lastly, I display results of ECGs performed by Dr. Maarten van den Hoogenhof as Fig. 3.23 f. As can be seen, Tnpo3 treatment restored cardiac ejection fraction (EF) in P635L mice to the levels seen in WT mice, in agreement with the results for splicing restoration. Tnpo3 treatment did not affect EF of R636Q mice (Fig. 3.23 f). This is in line with the little effect on splicing restoration and could be due to lower affinity to TNPO3 and stronger mislocalization of this variant, as shown before (Fig. 3.2, 3.16). These results are very encouraging and provide the first evidence that up-regulating RBM20 nuclear import by *Tnpo3* overexpression can restore RBM20 localization, proportionally restoring splicing and ejection fraction. Currently, I am performing similar analyses for *Tnpo3* overexpression in R636Q$^{+/-}$ (heterozygous, HET) mice. Although I do not have these data ready to present yet, they preliminary show that, *Tnpo3* overexpression can restore EF in R636Q HET mice. This demonstrates that, cardiac insufficiency measured by EF, is proportional to the severity of mislocalization and mis-splicing, and that restoring mislocalization and mis-splicing can proportionally restore cardiac function in vivo. Next steps in these analyses are to further analyse the R636Q HET mice, and quantify the impact of *Tnpo3* overexpression on localization of RBM20 variants in vivo.

In summary, the in vivo data demonstrate that, *Tnpo3* overexpression does not lead to systematic effects on gene expression or alternative splicing but can rescue mis-splicing and restore cardiac phenotype in RBM20 RS-domain mutant mice.

Chapter 4. Discussion and future directions

Protein mislocalization and aggregation are known hallmarks of many disorders, including neurodegenerative diseases and cardiomyopathies. Developing targeted preventive and curative therapies for human disease generally requires understanding of the detailed molecular mechanisms underlying disease progression. Until now, it has been challenging to systematically study protein mislocalization in human disease with high throughput methods. Although microscopy-based readouts can provide detailed characterization of protein mislocalization phenotypes, their use for functional genomics experiments is limited to array-based screens in a multi-well format, with one perturbation per well. Achieving genome-wide scale with an array screen with microscopy-based readout is challenging if at all possible. The recently developed ICS technology[86] provides an outstanding tool to combine image-based readout with pooled genetic screens and allows for genome-wide coverage without investing a substantial amount of time or resources. In my PhD work, I applied this technology for the first time to decipher the mechanism of protein localization in human disease. This resulted in the discovery of a new potential therapeutic target for DCM that I further validated in vitro and in vivo. I believe that, applying a combination of the ICS-based readout with functional genomics, transcriptomics, and proteomics to study mechanisms of protein mislocalization in other human diseases is a promising way towards understanding mechanisms of aberrant protein mislocalization in human disease and can help identifying novel therapeutic targets.

In this project, I addressed the mechanism of mislocalization of the cardiac splice factor RBM20, mutations of which result in severe forms of dilated cardiomyopathy. To date, all identified pathogenic mutations in RBM20 have been shown to result in alternative splicing deficiency, demonstrating the importance of correct isoform expression in the physiology of the human heart. In addition to splicing deficiency, a subset of mutations located in the RS domain of RBM20 have been shown to result in aberrant accumulation of RBM20 granules in the cytoplasm[71–77,79]. This aberrant localization was shown to cause a more severe disease phenotype than a full protein KO in vitro and in vivo[72–74]. However, the functional cause underlying these phenotypes has remained elusive. In my project, I demonstrated that mislocalizing mutations in the PRSRSP stretch are splice-competent if relocalised to the nucleus. I demonstrated this by analysing alternative splicing of the core RBM20 targets in cells with physiological (cells with RBM20-P633L-nuclear isolated with ICS, Fig. 3.4) or induced (cells overexpressing SV40 NLS-tagged RBM20-R634Q, Fig. 3.6) nuclear localization of RS-domain RBM20 variants. Importantly, our collaborators also showed that, NLS-tagging of all other pathogenic RSRSP-stretch variants can rescue *TTN* minigene splicing[80]. These findings demonstrate that, restoring nuclear localization of RS-domain RBM20 variants could be a promising therapeutic way to a) restore splicing deficiency and b) dissolve detrimental cytoplasmic granules. One question that remains unanswered is whether relocalised to the nucleus RS-domain variants may bind to new mRNAs and induce their splicing, which may result in undesired gain-of-function effects. In my work, I based my conclusions on either not fully relocalised (cells with RBM20-P633L-nuclear isolated with ICS, Fig. 3.4), or not endogenously expressed (cells overexpressing SV40 NLS-tagged RBM20-R634Q, Fig. 3.6) RBM20 variants. Therefore, to fully address potential gain-of-function of RS-domain variants in the nucleus, and endogenous system with clear nuclear relocalization is needed. At the moment, a postdoctoral fellow in the lab – Dr. Kai Fenzl, and I, are working together with LAR to generate a new mouse model with endogenously-tagged RBM20-R636Q with an NLS. Therefore, by comparing alternative splicing profiles in RBM20-WT, -NLS-WT, -R636Q, and -NLS-R636Q mice we hope to address this question in an in vivo endogenous system in the nearest future.

In this project, I identified TNPO3 as the main factor responsible for RBM20 nuclear import and showed that RS domain mutants lose this interaction proportionally to the severity of their mislocalization phenotype. Although I identified other factors that may retain mutant RBM20 in the cytoplasm (like DDI2, PPP5C, WTAP, OPA1), their impact on RBM20 localization was relatively minor compared to TNPO3. In addition, I did not test the effect of these factors on RBM20 localization in iPSC-CMs, which is the next step in this analysis. Nonetheless, based on the data collected during my project, TNPO3 appears to be the sole nuclear importer of RBM20. For many SR- and SR-like proteins, binding to TNPO3 is phosphorylation-dependent[18,19,21], however, it has also been shown that some bind TNPO3 regardless of the phosphorylation status of their RS-domain[22]. In the RBM20's RS domain predicted to bind TNPO3, it has been shown that serines S635 and S637 are normally phosphorylated[57,75,78], however, the essentiality of this phosphorylation for nuclear import via TNPO3 is yet to be understood. Although mutations of these residues result in a severe mislocalization phenotype[72,74,81], it has not been confirmed that phosphorylation itself and not the absence of a mutation is required for nuclear import. Importantly, phospho-mimicking mutations did not rescue the aberrant phenotype[75], which may indicate that phosphorylation plays a complex role in RBM20 cellular transport mechanism. In addition, in my project I tested both pooled and individual KOs of kinases AKT2, CLK1, and SPRK1 previously shown to phosphorylate RBM20[81] (Fig. 3.15 a). Loss of these kinases did not impact RBM20 localization, suggesting that either they are interchangeable, and loss of one kinase can be compensated by activation of another, or that the RBM20-TNPO3 interaction is phosphorylation-independent.

My study had one factor limiting identification of all factors regulating RBM20 import. For CRISPR/Cas9 screens done in my project I used HeLa cell lines and not CMs, where RBM20 is normally expressed. HeLa cells and CMs are two very distinct cell types, and many genes expressed in CMs are not detected in HeLa. Therefore, it may be that more factors regulating RBM20 transport remained unidentified, including those controlling post-translational modifications of RBM20 or secondary shuttle mechanisms. In the future, performing a targeted CRISPR/Cas9 ICS screen using endogenous RBM20 in, for example, iPSC-CMs, could help identifying additional players in RBM20 nuclear-cytoplasmic shuttling. They may act via changing the affinity of TNPO3 to RBM20 depending on post-translational modifications of the latter. Nevertheless, the data collected during my PhD project, including effects of TNPO3 expression on RBM20 localization in iPSC-CMs and in vivo, leave no doubt that TNPO3 is the main nuclear importer of RBM20.

Mutations in TNPO3 have been linked to a few myopathies[26,105], and TNPO3-dependent nuclear import of SRSF1 was shown to impact cellular myogenesis[25], illustrating the importance of nucleoplasmic transport of splicing factors for muscle tissue functioning. Interestingly, in one medical report from 2011[28], an R870Q mutation in TNPO3 was linked to a familial DCM, however, seemingly, this study has not been followed up on. In the future, it would be of interest to establish a pathobiological mechanism caused by that mutation and its relation to RBM20-mediated DCM.

One important question about RBM20 mislocalization that remains to be addressed is the exact nature and toxic function of RBM20 granules. Here, I demonstrated that, upon loss of TNPO3, RBM20-WT formed cytoplasmic granules of the same appearance as RS-domain mutants (Fig. 3.17 b, 3.18). In addition, together with Dr. Marta Rodríguez-Martínez, we showed that, similar to RS-domain variants (shown before in[78]), in the absence of TNPO3, RBM20-WT interacts with MOV10 in the cytoplasm[80]. Structural predictions of RS-domain variants[80] identified no structural rearrangements caused by the mutations compared to RBM20-WT, indicating that RS-domain mutations do not confer pro-aggregative qualities to RBM20. These results provide strong evidence that cytoplasmic aggregation of RS mutants is the result and not the cause of the protein's inability to interact with TNPO3 and to be transported into the nucleus. Future studies should address the pathobiological mechanism of cytoplasmic RBM20 granules. It has been shown that, mutant RBM20 binds to 3'UTRs of transcripts that are not classified as core RBM20 targets in the cytoplasm, and co-localised with P-body markers[73]. These results lead to a speculation that mislocalised RBM20 variants may impact gene expression and/or splicing of a set of genes, therefore affecting cardiomyocyte function. In addition, based on results of my co-IP experiments, I identified proteins involved in spliceosome machinery (e.g. SRSF9, SRSF10, HNRNPH3, HNRNPA3, HNRNPA0, HNRNPC) interacting with RBM20 in the cytoplasm (Fig. 3.9). This may indicate that, mislocalised RBM20 potentially sequesters other components of the splicing machinery and thereby affects additional steps of mRNA processing. Of note, all these experiments cannot distinguish between effects of mislocalization versus mutation. Ideally, to address RBM20 gain-of-function, one would generate a model with RBM20-WT localizing to the nucleus or to the cytoplasm (e.g. by fusing RBM20-WT with NES), and compare it with a model of mutant RBM20 localizing to the nucleus (e.g. by fusing RBM20-mut with NLS) and to the cytoplasm. This will help identifying direct effects of a mutation presence versus the effect of mislocalization. Detailed analysis of proteins interacting with nuclear- or cytoplasmic- WT- or mutated-RBM20 in an endogenous system (e.g. cardiomyocytes) would help assigning the exact nature and composition of these granules, which may hint on their pathobiological mechanism. Addressing the exact localization of RBM20 cytoplasmic granules within a cell, e.g., via combining immune-labelling with electron microscopy (EM) might also give an indication of their function. At the moment, I am working closely with the Electron Microscopy Core Facility

(EMCF) at EMBL to perform such analysis. I hope that this joint effort will soon provide a new insight into the nature of RBM20 cytoplasmic granules.

In this study, I used in vitro and in vivo models to demonstrate that overexpression of TNPO3 can enhance nuclear transport of RS-domain RBM20 variants, restore their splicing deficiency and result in aggregates dissolution. This is in line with previous studies which demonstrated that up-regulation of other importins can restore nuclear localization and dissolve aggregates formed by mutated RNA-binding proteins linked to neurodegenerative diseases[15]. This suggests that targeting the interaction between a nuclear import receptor and its mislocalised cargo protein could be used as a therapeutic strategy, not just in RBM20-DCM, but in other diseases as well. Importantly, restoration of alternative splicing upon *Tnpo3* overexpression in vivo resulted in restoration of heart pumping function as measured by ECG (Fig. 3.23). Even though I have not quantified the degree of relocalization and granule dissolution for this experiment yet, restoration of splicing is a good indication of localization restoration. At the moment, the relative contribution of granule dissolution versus rescue of splicing deficiency into restoration of heart function upon *Tnpo3* overexpression in vivo is unclear. To address this, we are currently characterizing a newly generated mouse strain where we crossed the severe RBM20-R636Q mice with RBM20-KO mice. These new RBM20-R636Q$^{+/KO}$ mice are supposed to have half the number of granules of RBM20-R636Q$^{+/+}$, with a minimal change in splicing activity. We hope that this experiment will help establishing the relative contribution of granules vs. mis-splicing into the disease phenotype, therefore identifying the more crucial pathway to target for therapies.

One important reservation for using *Tnpo3* overexpression in therapy is the risk of side effects caused by up-regulating nuclear import TNPO3 targets other than RBM20, resulting in unwanted alternative splicing and/or other pathway up-regulation. Importantly, I did not observe any systemic changes in gene expression or alternative splicing caused by *Tnpo3* overexpression in vivo (Fig. 3.22, 3.23). These results indicate that, using *Tnpo3* gene therapy could be safe. However, more thorough assessment of in vivo *Tnpo3* overexpression consequences is needed, including life-long monitoring of all organ function and well-being of the animals, as well as molecular analysis of transcriptome and proteome changes in different organs targeted by AAVs. Should all these additional tests demonstrate the safety of *Tnpo3* gene delivery in vivo, this strategy could be further considered for large animal and human clinical trials.

Based on results of this project, it appears that restoring nuclear localization of RBM20 RS-domain variants could be a promising therapeutic strategy worth investigating further. In addition to supplying full-length *TNPO3* cDNA or mRNA, or TNPO3 protein, other strategies should be tested in future studies. For example, deciphering the exact structure of RBM20-TNPO3 complex in the presence and absence of RS-domain mutations will allow running an in silico screen including millions of compounds to identify potential small molecules to stabilize the complex. Shortlisted compounds can then further be tested in cell and animal models. In addition, knowledge on post-translational modifications important for the complex formation may allow developing molecules targeting these modification pathways. Finding ways to stimulate other transport pathways, e.g., by adding other NLS to RBM20 gene or protein sequence may be beneficial too, as I demonstrated here for the SV40 NLS (Fig. 3.5). Lastly, dissolving cytoplasmic granules on their own, without restoring nuclear localization, could serve as a potential therapeutic strategy as well, however this needs further investigation.

Altogether, in my PhD project I identified the molecular mechanism of pathogenic RBM20 mislocalization, which has implications for developing new therapeutic strategies targeted at restoring nuclear localization of RBM20 to treat DCM.

Bibliography

1. Stanley, R. F. & Abdel-wahab, O. Dysregulation and therapeutic targeting of RNA splicing in cancer. *Nat. Cancer* **3**, 536–546 (2022).
2. Van Den Hoogenhof, M. M. G., Pinto, Y. M. & Creemers, E. E. RNA Splicing regulation and dysregulation in the heart. *Circ. Res.* **118**, 454–468 (2016).
3. Montes, M., Sanford, B. L., Comiskey, D. F. & Chandler, D. S. RNA Splicing and Disease: Animal Models to Therapies. *Trends Genet.* **35**, 68–87 (2019).
4. Wright, C. J., Smith, C. W. J. & Jiggins, C. D. Alternative splicing as a source of phenotypic diversity. *Nat. Rev. Genet.* **23**, 697–710 (2022).
5. Breschi, A., Gingeras, T. R. . & Guigó, R. Comparative transcriptomics in human and mouse. *Nat. Rev. Genet.* **18**, 425–440 (2017).
6. Akinyi, M. V. & Frilander, M. J. At the Intersection of Major and Minor Spliceosomes: Crosstalk Mechanisms and Their Impact on Gene Expression. *Front. Genet.* **12**, 1–10 (2021).
7. Gehring, N. H. & Roignant, J. Y. Anything but Ordinary – Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends Genet.* **37**, 355–372 (2021).
8. Park, J. W., Tokheim, C., Shen, S. & Xing, Y. *Identifying Differential Alternative Splicing Events from RNA Sequencing Data Using RNASeq-MATS. Deep Sequencing Data Analysis* (2013).
9. Twyffels, L., Gueydan, C. & Kruys, V. Shuttling SR proteins: More than splicing factors. *FEBS J.* **278**, 3246–3255 (2011).
10. Cautain, B., Hill, R., De Pedro, N. & Link, W. Components and regulation of nuclear transport processes. *FEBS J.* **282**, 445–462 (2015).
11. Wing, C. E., Fung, H. Y. J. & Chook, Y. M. Karyopherin-mediated nucleocytoplasmic transport. *Nat. Rev. Mol. Cell Biol.* **23**, 307–328 (2022).
12. Lu, J. *et al.* Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Commun. Signal.* **19**, 1–10 (2021).
13. Wagstaff, K. M. & Jans, D. A. Importins and beyond: Non-conventional nuclear transport mechanisms. *Traffic* **10**, 1188–1198 (2009).
14. Fabbro, M. & Henderson, B. R. Regulation of tumor suppressors by nuclear-cytoplasmic shuttling. *Exp. Cell Res.* **282**, 59–69 (2003).
15. Guo, L. *et al.* Nuclear-Import Receptors Reverse Aberrant Phase Transitions of RNA-Binding Proteins with Prion-like Domains. *Cell* **173**, 677-692.e20 (2018).
16. Odeh, H. M., Fare, C. M. & Shorter, J. Nuclear-Import Receptors Counter Deleterious Phase Transitions in Neurodegenerative Disease: Nuclear-import receptors combat deleterious phases. *J. Mol. Biol.* **434**, 167220 (2022).
17. Kataoka, N., Bachorik, J. L. & Dreyfuss, G. Transportin-SR, a nuclear import receptor for SR proteins. *J. Cell Biol.* **145**, 1145–1152 (1999).
18. Lai, M. C., Lin, R. I., Huang, S. Y., Tsai, C. W. & Tarn, W. Y. A human importin-β family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J. Biol. Chem.* **275**, 7950–7957 (2000).

19. Lai, M. C., Lin, R. I. & Tarn, W. Y. Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10154–10159 (2001).

20. Yun, C. Y., Velazquez-Dones, A. L., Lyman, S. K. & Fu, X. D. Phosphorylation-dependent and -independent nuclear import of RS domain-containing splicing factors and regulators. *J. Biol. Chem.* **278**, 18050–18055 (2003).

21. Maertens, G. N. *et al.* Structural basis for nuclear import of splicing factors by human Transportin 3. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2728–2733 (2013).

22. Jang, S. *et al.* Differential role for phosphorylation in alternative polyadenylation function versus nuclear import of SR-like protein CPSF6. *Nucleic Acids Res.* **47**, 4663–4683 (2019).

23. Bourgeois, B. *et al.* Nonclassical nuclear localization signals mediate nuclear import of CIRBP. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 8503–8514 (2020).

24. Brass, A. L. ; *et al.* Identification of host proteins required for HIV infection through a functional genomic screen. *Science (80-.).* **319**, 921–927 (2008).

25. Costa, R. *et al.* Morphological study of TNPO3 and SRSF1 interaction during myogenesis by combining confocal, structured illumination and electron microscopy analysis. *Mol. Cell. Biochem.* **476**, 1797–1811 (2021).

26. Vihola, A., Palmio, J., Danielsson, O. & Penttil, S. Novel mutation in TNPO3 causes congenital limb-girdle myopathy with slow progression. *Neurol. Genet.* **0**, 1–9 (2019).

27. Id, F. D. W. *et al.* The mutation of Transportin 3 gene that causes limb girdle muscular dystrophy 1F induces protection against HIV-1 infection. 1–22 (2019).

28. Li, D. *et al.* Whole Exome Sequencing Identifies a Mutation in Transportin 3 as a Cause of Dilated Cardiomyopathy. *Late-Breaking Basic Sci. Abstr. From Am. Hear. Assoc. Sci. Sess. 2011* **109**, 56–57 (2011).

29. Elliott, P. *et al.* Classification of the cardiomyopathies: A position statement from the european society of cardiology working group on myocardial and pericardial diseases. *Eur. Heart J.* **29**, 270–276 (2008).

30. Hershberger, R. E., Morales, A. & Siegfried, J. D. Clinical and genetic issues in dilated cardiomyopathy: A review for genetics professionals. *Genet. Med.* **12**, 655–667 (2010).

31. McKenna, W. J., Maron, B. J. & Thiene, G. Classification, epidemiology, and global burden of cardiomyopathies. *Circ. Res.* **121**, 722–730 (2017).

32. Hauh Watkins, Houman Ashrafian, C. R. Inherited cardiomyopathies. *N. Engl. J. Med.* **364**, 1643–1656 (2011).

33. Pinto, Y. M. *et al.* Proposal for a revised definition of dilated cardiomyopathy, hypokinetic non-dilated cardiomyopathy, and its implications for clinical practice: A position statement of the ESC working group on myocardial and pericardial diseases. *Eur. Heart J.* **37**, 1850–1858 (2016).

34. McNally, E. M. & Mestroni, L. Dilated cardiomyopathy: Genetic determinants and mechanisms. *Circ. Res.* **121**, 731–748 (2017).

35. Hershberger, R. E., Hedges, D. J. & Morales, A. Dilated cardiomyopathy: The complexity of a diverse genetic architecture. *Nat. Rev. Cardiol.* **10**, 531–547 (2013).

36. Verdonschot, J. A. J., Hazebroek, M. R., Ware, J. S., Prasad, S. K. & Heymans, S. R. B. Role of

Targeted Therapy in Dilated Cardiomyopathy: The Challenging Road Toward a Personalized Approach. *J. Am. Heart Assoc.* **8**, 1–18 (2019).

37. Orphanou, N., Papatheodorou, E. & Anastasakis, A. Dilated cardiomyopathy in the era of precision medicine: latest concepts and developments. *Heart Fail. Rev.* **27**, 1173–1191 (2022).
38. Schultheiss, H. P. *et al.* Dilated cardiomyopathy. *Nat. Rev. Dis. Prim.* **5**, (2019).
39. Weintraub, R. G., Semsarian, C. & Macdonald, P. Dilated cardiomyopathy. *Lancet* **390**, 400–414 (2017).
40. Japp, A. G., Gulati, A., Cook, S. A., Cowie, M. R. & Prasad, S. K. The Diagnosis and Evaluation of Dilated Cardiomyopathy. *J. Am. Coll. Cardiol.* **67**, 2996–3010 (2016).
41. Haas, J. *et al.* Atlas of the clinical genetics of human dilated cardiomyopathy. *Eur. Heart J.* **36**, 1123–1135 (2015).
42. Mazzarotto, F. *et al.* Reevaluating the Genetic Contribution of Monogenic Dilated Cardiomyopathy. *Circulation* 387–398 (2020) doi:10.1161/CIRCULATIONAHA.119.037661.
43. Jordan, E. *et al.* Evidence-Based Assessment of Genes in Dilated Cardiomyopathy. *Circulation* **144**, 7–19 (2021).
44. Strande, N. T. *et al.* Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am. J. Hum. Genet.* **100**, 895–906 (2017).
45. Brauch, K. M. *et al.* Mutations in Ribonucleic Acid Binding Protein Gene Cause Familial Dilated Cardiomyopathy. *J. Am. Coll. Cardiol.* **54**, 930–941 (2009).
46. Li, D. *et al.* Identification of novel mutations in RBM20 in patients with dilated cardiomyopathy. *Clin. Transl. Sci.* **3**, 90–97 (2010).
47. Koelemen, J., Gotthardt, M., Steinmetz, L. M. & Meder, B. Rbm20-related cardiomyopathy: Current understanding and future options. *J. Clin. Med.* **10**, (2021).
48. Parikh, V. N. *et al.* Regional Variation in RBM20 Causes a Highly Penetrant Arrhythmogenic Cardiomyopathy. *Circ. Hear. Fail.* **12**, 1–9 (2019).
49. Watanabe, T., Kimura, A. & Kuroyanagi, H. Alternative splicing regulator RBM20 and cardiomyopathy. *Front. Mol. Biosci.* **5**, 1–11 (2018).
50. Guo, W. *et al.* RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat. Med.* **18**, 766–773 (2012).
51. Refaat, M. M. *et al.* Genetic variation in the alternative splicing regulator RBM20 is associated with dilated cardiomyopathy. *Hear. Rhythm* **9**, 390–396 (2012).
52. Li, S., Guo, W., Dewey, C. N. & Greaser, M. L. Rbm20 regulates titin alternative splicing as a splicing repressor. *Nucleic Acids Res.* **41**, 2659–2672 (2013).
53. Maatz, H. *et al.* RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. *J. Clin. Invest.* **124**, 3419–3430 (2014).
54. Van Den Hoogenhof, M. M. G. *et al.* RBM20 mutations induce an arrhythmogenic dilated cardiomyopathy related to disturbed calcium handling. *Circulation* **138**, 1330–1342 (2018).
55. Zhu, C. *et al.* Single-molecule, full-length transcript isoform sequencing reveals disease-associated RNA isoforms in cardiomyocytes. *Nat. Commun.* **12**, 1–9 (2021).
56. Dauksaite, V. & Gotthardt, M. Molecular basis of titin exon exclusion by RBM20 and the novel

titin splice regulator PTB4. *Nucleic Acids Res.* **46**, 5227–5238 (2018).

57. Murayama, R. *et al.* Phosphorylation of the RSRSP stretch is critical for splicing regulation by RNA-Binding Motif Protein 20 (RBM20) through nuclear localization. *Sci. Rep.* **8**, 1–14 (2018).
58. Beraldi, R. *et al.* Rbm20-deficient cardiogenesis reveals early disruption of RNA processing and sarcomere remodeling establishing a developmental etiology for dilated cardiomyopathy. *Hum. Mol. Genet.* **23**, 3779–3791 (2014).
59. Zhu, C., Yin, Z., Tan, B. & Guo, W. Insulin regulates titin pre-mRNA splicing through the PI3K-Akt-mTOR kinase axis in a RBM20-dependent manner. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1863**, 2363–2371 (2017).
60. Zhu, C. *et al.* RBM20 is an essential factor for thyroid hormone-regulated titin isoform transition. *J. Mol. Cell Biol.* **7**, 88–90 (2015).
61. Cai, H. *et al.* Angiotensin ii influences pre-mRNA splicing regulation by enhancing RBM20 transcription through activation of the MAPK/ELK1 signaling pathway. *Int. J. Mol. Sci.* **20**, (2019).
62. Briganti, F. *et al.* iPSC Modeling of RBM20-Deficient DCM Identifies Upregulation of RBM20 as a Therapeutic Strategy. *Cell Rep.* **32**, 108117 (2020).
63. Lennermann, D., Backs, J. & van den Hoogenhof, M. M. G. New Insights in RBM20 Cardiomyopathy. *Curr. Heart Fail. Rep.* **17**, 234–246 (2020).
64. Guo, W., Pleitner, J. M., Saupe, K. W. & Greaser, M. L. Pathophysiological defects and transcriptional profiling in the RBM20 -/- rat model. *PLoS One* **8**, 1–11 (2013).
65. Wyles, S. P. *et al.* Modeling structural and functional deficiencies of RBM20 familial dilated cardiomyopathy using human induced pluripotent stem cells. *Hum. Mol. Genet.* **25**, 254–265 (2016).
66. Streckfuss-Bömeke, K. *et al.* Severe DCM phenotype of patient harboring RBM20 mutation S635A can be modeled by patient-specific induced pluripotent stem cell-derived cardiomyocytes. *J. Mol. Cell. Cardiol.* **113**, 9–21 (2017).
67. Wyles, S. P. *et al.* Pharmacological Modulation of Calcium Homeostasis in Familial Dilated Cardiomyopathy: An In Vitro Analysis From an RBM20 Patient-Derived iPSC Model. *Clin. Transl. Sci.* **9**, 158–167 (2016).
68. Beqqali, A. *et al.* A mutation in the glutamate-rich region of RNA-binding motif protein 20 causes dilated cardiomyopathy through missplicing of titin and impaired Frank-Starling mechanism. *Cardiovasc. Res.* **112**, 452–463 (2016).
69. Yamamoto, T. *et al.* I536T variant of RBM20 affects splicing of cardiac structural proteins that are causative for developing dilated cardiomyopathy. *J. Mol. Med.* **100**, 1741–1754 (2022).
70. Filippello, A., Lorenzi, P., Bergamo, E. & Romanelli, M. G. Identification of nuclear retention domains in the RBM20 protein. *FEBS Lett.* **587**, 2989–2995 (2013).
71. Schneider, J. W. *et al.* A ribonucleoprotein-granule pathway to heart failure in human RBM20 cardiomyopathy gene-edited pigs. *Nat. Med.* **26**, 1788–1800 (2020).
72. Ihara, K. *et al.* A missense mutation in the RSRSP stretch of Rbm20 causes dilated cardiomyopathy and atrial fibrillation in mice. *Sci. Rep.* **10**, 1–14 (2020).

73. Fenix, A. M. *et al.* Gain-of-function cardiomyopathic mutations in RBM20 rewire splicing regulation and re-distribute ribonucleoprotein granules within processing bodies. *Nat. Commun.* **12**, 1–14 (2021).
74. Wang, C. *et al.* RBM20S639G mutation is a high genetic risk factor for premature death through RNA-protein condensates. *J. Mol. Cell. Cardiol.* **165**, 115–129 (2022).
75. Zhang, Y. *et al.* RBM20 phosphorylation and its role in nucleocytoplasmic transport and cardiac pathogenesis. *FASEB J.* **36**, (2022).
76. Gaertner, A. *et al.* Cardiomyopathy-associated mutations in the RS domain affect nuclear localization of RBM20. *Hum. Mutat.* **41**, 1931–1943 (2020).
77. Nishiyama, T. *et al.* Precise genomic editing of pathogenic mutations in RBM20 rescues dilated cardiomyopathy. *Sci. Transl. Med.* **14**, 1–15 (2022).
78. Vieira-Vieira, C. H., Dauksaite, V., Sporbert, A., Gotthardt, M. & Selbach, M. Proteome-wide quantitative RNA-interactome capture identifies phosphorylation sites with regulatory potential in RBM20. *Mol. Cell* **82**, 2069-2083.e8 (2022).
79. Grosch, M. *et al.* Striated muscle-specific base editing enables correction of mutations causing dilated cardiomyopathy. *Nat. Commun.* **14**, 3714 (2023).
80. Kornienko, J. *et al.* Mislocalization of pathogenic RBM20 variants in dilated cardiomyopathy is caused by loss-of-interaction with Transportin-3. (2023) doi:10.1038/s41467-023-39965-6.
81. Sun, M. *et al.* SR Protein Kinases Regulate the Splicing of Cardiomyopathy- Relevant Genes via Phosphorylation of the RSRSP Stretch in RBM20. *Genes (Basel)* **13**, 1526 (2022).
82. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nat. Methods* **8**, 424–429 (2011).
83. Burridge, P. W. *et al.* Chemically defined and small molecule-based generation of human CMs. *Nat Methods* **11**, 855–860 (2014).
84. Feldman, D. *et al.* Optical Pooled Screens in Human Cells. *Cell* **179**, 787-799.e17 (2019).
85. Esk, C. *et al.* A human tissue screen identifies a regulator of ER secretion as a brain-size determinant. *Science (80-.).* **370**, 935–941 (2020).
86. Schraivogel, D. *et al.* High-speed fluorescence image – enabled cell sorting. *Science (80-.).* **320**, 315–320 (2022).
87. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
88. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
89. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
90. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, (2019).
91. Shen, S. *et al.* MATS: A Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. *Nucleic Acids Res.* **40**, 1–13 (2012).
92. Schafer, S. *et al.* Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI). *Curr. Protoc. Hum. Genet.* **87**, 11.16.1-11.16.14 (2015).
93. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data.

Genome Res. **22**, 2008–2017 (2012).

94. Pujar, S. *et al.* Consensus coding sequence (CCDS) database: A standardized set of human and mouse protein-coding regions supported by expert curation. *Nucleic Acids Res.* **46**, D221–D228 (2018).
95. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
96. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
97. Hart, T. *et al.* Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3 Genes, Genomes, Genet.* **7**, 2719–2727 (2017).
98. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: Visualizing classifier performance in R. *Bioinformatics* **21**, 3940–3941 (2005).
99. De Boer, C. G., Ray, J. P., Hacohen, N. & Regev, A. MAUDE: Inferring expression changes in sorting-based CRISPR screens. *Genome Biol.* **21**, 1–16 (2020).
100. Ackers-Johnson, M. *et al.* A Simplified, Langendorff-Free Method for Concomitant Isolation of Viable Cardiac Myocytes and Nonmyocytes from the Adult Mouse Heart. *Circ. Res.* **119**, 909–920 (2016).
101. Cvitkovic, I. & Jurica, M. S. Spliceosome database: A tool for tracking components of the spliceosome. *Nucleic Acids Res.* **41**, 132–141 (2013).
102. Zahr, H. C. & Jaalouk, D. E. Exploring the crosstalk between LMNA and splicing machinery gene mutations in Dilated Cardiomyopathy. *Front. Genet.* **9**, 1–14 (2018).
103. Dempster, J. M. *et al.* Chronos: a cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. *Genome Biol.* **22**, 1–23 (2021).
104. Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
105. Gamez, J. *et al.* Autosomal dominant limb-girdle muscular dystrophy: A large kindred with evidence for anticipation. *Neurology* (2001) doi:10.1212/WNL.56.4.450.