

Global identification of hnRNP A1 binding sites for SSO-based splicing modulation

Bruun, Gitte H; Doktor, Thomas K; Borch-Jensen, Jonas; Masuda, Akio; Krainer, Adrian R; Ohno, Kinji; Andresen, Brage S

Published in:
BMC Biology

DOI:
[10.1186/s12915-016-0279-9](https://doi.org/10.1186/s12915-016-0279-9)

Publication date:
2016

Document version:
Final published version

Document license:
CC BY

Citation for published version (APA):
Bruun, G. H., Doktor, T. K., Borch-Jensen, J., Masuda, A., Krainer, A. R., Ohno, K., & Andresen, B. S. (2016). Global identification of hnRNP A1 binding sites for SSO-based splicing modulation. *BMC Biology*, 14(1), 54. <https://doi.org/10.1186/s12915-016-0279-9>

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use

This work is brought to you by the University of Southern Denmark.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk

RESEARCH ARTICLE

Open Access



Global identification of hnRNP A1 binding sites for SSO-based splicing modulation

Gitte H. Bruun¹, Thomas K. Doktor¹, Jonas Borch-Jensen¹, Akio Masuda², Adrian R. Krainer³, Kinji Ohno² and Brage S. Andresen^{1*} 

Abstract

Background: Many pathogenic genetic variants have been shown to disrupt mRNA splicing. Besides splice mutations in the well-conserved splice sites, mutations in splicing regulatory elements (SREs) may deregulate splicing and cause disease. A promising therapeutic approach is to compensate for this deregulation by blocking other SREs with splice-switching oligonucleotides (SSOs). However, the location and sequence of most SREs are not well known.

Results: Here, we used individual-nucleotide resolution crosslinking immunoprecipitation (iCLIP) to establish an in vivo binding map for the key splicing regulatory factor hnRNP A1 and to generate an hnRNP A1 consensus binding motif. We find that hnRNP A1 binding in proximal introns may be important for repressing exons. We show that inclusion of the alternative cassette exon 3 in *SKA2* can be significantly increased by SSO-based treatment which blocks an iCLIP-identified hnRNP A1 binding site immediately downstream of the 5' splice site. Because pseudoexons are well suited as models for constitutive exons which have been inactivated by pathogenic mutations in SREs, we used a pseudoexon in *MTRR* as a model and showed that an iCLIP-identified hnRNP A1 binding site downstream of the 5' splice site can be blocked by SSOs to activate the exon.

Conclusions: The hnRNP A1 binding map can be used to identify potential targets for SSO-based therapy. Moreover, together with the hnRNP A1 consensus binding motif, the binding map may be used to predict whether disease-associated mutations and SNPs affect hnRNP A1 binding and eventually mRNA splicing.

Keywords: hnRNP A1, iCLIP, Splicing splice-switching oligonucleotides (SSOs), Pseudoexons, Alternative splicing, Splicing silencer, Cross-linking immunoprecipitation (CLIP), RNA-seq, Surface plasmon resonance imaging (SPRI)

Background

Splicing of pre-mRNA is a key step in gene expression, and alternative splicing, the production of multiple mRNA isoforms from one gene, is a key mechanism to expand the diversity of the human proteome. Stringent regulation of splicing is crucial, since missplicing may lead to the production of nonfunctional or malfunctioning mRNA isoforms. Splicing regulatory proteins regulate splicing by binding to *cis*-acting splicing regulatory elements (SREs). A fine balance between negative and positive SREs determines the splicing fate of exons. Up to one third of all disease-causing mutations may disrupt

pre-mRNA splicing [1, 2]. Mutations that affect pre-mRNA splicing may directly or indirectly affect the recognition of the splice sites or the SREs. Alternatively, mutations may create or disrupt SREs and in this way disturb the complex regulatory network of splicing.

A well-studied example of a disease caused by disruption of an SRE is spinal muscular atrophy (SMA). Patients with SMA lack a functional *SMN1* gene and depend on the *SMN2* gene to survive. *SMN2* is highly homologous to *SMN1*, but a single nucleotide difference in *SMN2* exon 7 (c.840C > T) simultaneously disrupts a splicing enhancer and introduces a splicing silencer, which binds heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) [3–7]. Hence, *SMN2* has a high degree of exon 7 skipping, which results in a nonfunctional protein. In patients with MCAD deficiency a mutation in exon 5 (c.362C > T) disrupts a splicing enhancer and

* Correspondence: bragea@bmb.sdu.dk

¹Department of Biochemistry and Molecular Biology and The Villum Center for Bioanalytical Sciences, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

Full list of author information is available at the end of the article



induces exon skipping. Interestingly, a silent polymorphism in a flanking hnRNP A1-binding splicing silencer can protect against exon 5 skipping [8].

A promising approach to treat diseases caused by aberrant splicing is the use of antisense oligonucleotide-mediated splicing modulation [9, 10]. One way to do this is to block SREs by complementary antisense oligonucleotides (also called splice-switching oligonucleotides, SSOs) to make them inaccessible to splicing regulatory proteins and thereby redirect splicing. One of the most promising SSO-based therapies is targeting of the hnRNP A1-binding splicing silencer N1 in *SMN2* intron 7 to correct *SMN2* splicing [11–14]. However, identification of SREs which are suitable for this type of therapy is not easy. Mutations that disrupt or create new SREs are difficult to identify, since the binding sites for most splicing regulatory proteins have mainly been characterized by *in vitro* studies. Thus, there is limited knowledge of the *in vivo* binding sites of splicing regulatory proteins; moreover, splicing regulatory proteins may compete and cooperate for binding in an unpredictable manner [15]. To identify *in vivo* binding sites, a number of RNA binding proteins have been subjected to crosslinking and immunoprecipitation (CLIP) analysis [16–24].

Here, we performed individual-nucleotide resolution crosslinking immunoprecipitation (iCLIP) analysis of the multifunctional RNA binding protein hnRNP A1. hnRNP A1 is involved in many RNA processing events including both constitutive and alternative splicing [25, 26]. Traditionally, hnRNP A1 has been considered a splicing repressor, though it may in some cases also stimulate splicing. hnRNP A1 may repress splicing by antagonizing the function of positive splicing factors such as the SR proteins [27–29]; alternatively, hnRNP A1 may sterically block binding of SR proteins or members of the spliceosome to prevent splice site recognition [30]. Furthermore, hnRNP A1 proteins are able to multimerize after initial binding to a high-affinity site and spread across an exon to inhibit its inclusion [31]. Protein-protein interactions between different hnRNP A1 molecules may also promote either exon inclusion by looping out introns to reduce intron size and stimulate splice site pairing, or cause exon skipping by looping out exons to inhibit splice site recognition [32–34].

Diseases associated with deregulated expression of hnRNP A1 include several cancers [35–38] and Alzheimer's disease [39]. In addition, a number of diseases are caused by mutations or single nucleotide polymorphisms (SNPs) which create or abolish hnRNP A1 binding SREs. For example, a missense mutation in the *ETFDH* gene improves an hnRNP A1-binding splicing silencer, thereby inducing exon 2 skipping and degradation of the ETFDH protein in patients with multiple acyl-CoA dehydrogenation deficiency [40]. In the *PAH* gene, a synonymous mutation

creates an hnRNP A1-binding splicing silencer, thereby causing aberrant splicing and phenylketonuria (PKU) [41]. Also, a SNP affecting an hnRNP A1-binding SRE plays a potential role in cardiovascular disease risk by altering the alternative splicing of *HMGCR* [42]. Furthermore, a synonymous SNP in exon 11 of *ACADM* disrupts an hnRNP A1-binding SRE, which improves exon inclusion and may influence fatty acid oxidation capacity [43].

Despite the fact that hnRNP A1 is a well-studied splicing regulatory factor, transcriptome-wide *in vivo* binding of hnRNP A1 has previously only been carried out as part of a study comparing binding of several hnRNP proteins. In that study, crosslinking and immunoprecipitation followed by high-throughput sequencing (HITS-CLIP) analysis was performed in HEK293T cells [17]. In the present study, we have examined the transcriptome-wide binding of hnRNP A1 by a newer method, iCLIP coupled with high-throughput sequencing [44] in a different cell type, namely HeLa cells. Our main goal is to create a more comprehensive catalogue of hnRNP A1 binding sites to enable identification of disease-associated binding sites. In this study, we increase the number of CLIP-identified hnRNP A1 binding sites in the human transcriptome approximately 20 times. This dramatically increases the number of identified hnRNP A1-binding SREs, which can serve as targets for SSO-based therapy. While some binding sites overlap between the two studies, many seem to be cell type- or method-specific. Furthermore, we hypothesize that, by combining the hnRNP A1 binding map with the validated hnRNP A1 binding motif, one will be able to better predict the effect of sequence variations on splicing. We find that hnRNP A1 binding in the proximal intron is important for hnRNP A1-mediated splicing repression, and that targeting these binding sites with SSOs can improve exon inclusion. This will be useful for correction of splicing of constitutive exons which have been inactivated by mutations in SREs or to shift between alternatively spliced exons. We selected pseudoexons as a model of constitutive exons, which have been inactivated by mutations that either abolish a splicing enhancer or create a splicing silencer, and show that blocking of downstream hnRNP A1 binding sites improves exon inclusion. Furthermore, we show that we can shift the alternative splicing of *SKA2* exon 3 towards more inclusion when blocking an identified hnRNP A1-binding splicing silencer downstream of the 5' splice site.

Thus, as a proof of concept, we show here that hnRNP A1 binding sites identified by iCLIP immediately downstream of the 5' splice site can be blocked by SSOs to activate splicing of exons.

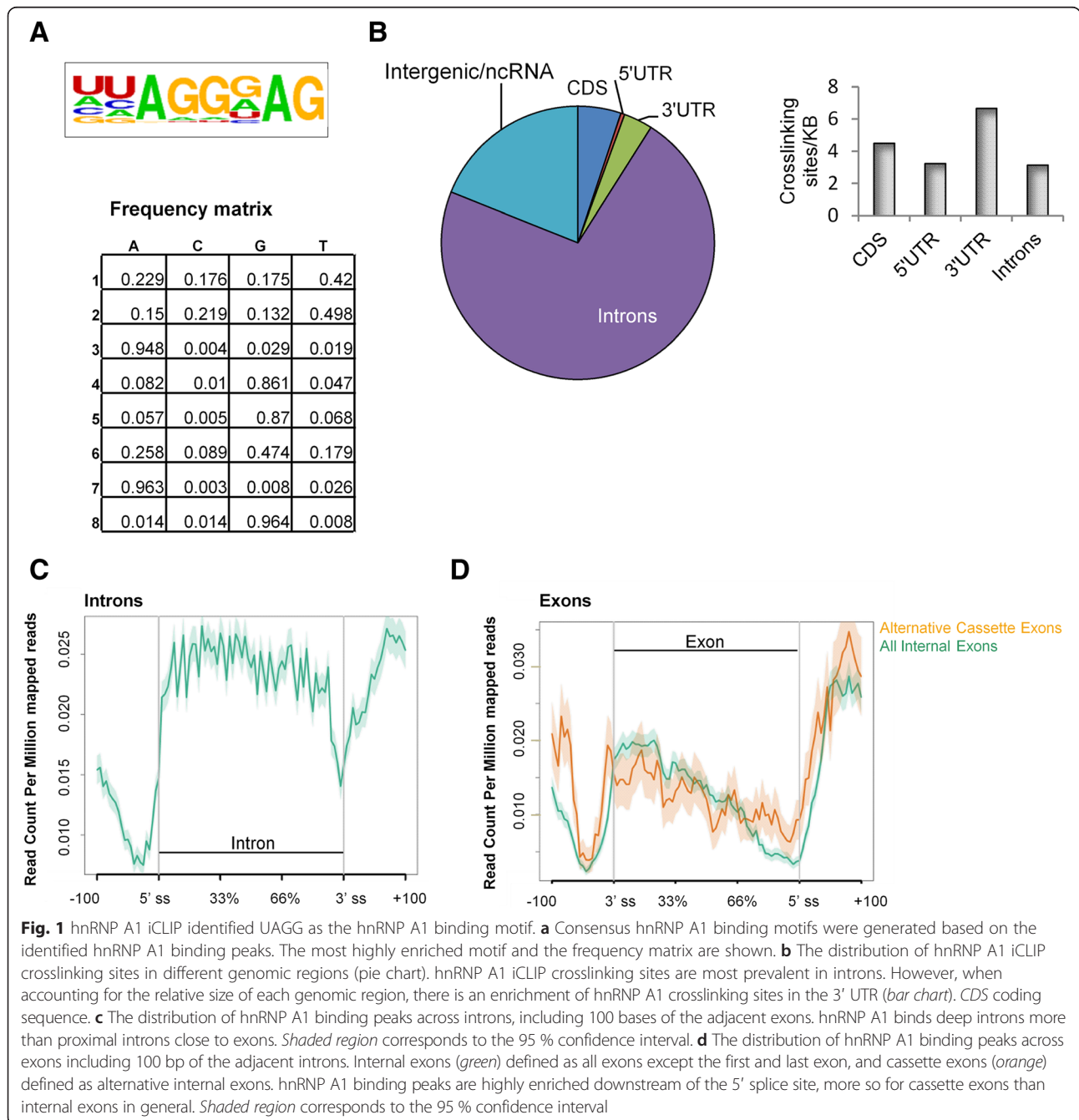
Results

Transcriptome-wide analysis of hnRNP A1 binding sites

To analyze the transcriptome-wide binding of hnRNP A1 to protein-coding genes, we performed iCLIP [44] in

HeLa cells with inducible expression of T7-tagged hnRNP A1 (Additional file 1: Figure S1). The T7-tagged hnRNP A1 was previously described [45], and was demonstrated to behave similarly to endogenous hnRNP A1 [46]. After filtering, removal of PCR duplicates, and adaptor removal, a total of 5,115,079 reads uniquely aligned to the genome (Additional file 2: Table S1). Due to the nature of the iCLIP protocol, the protein binding site is assumed to be at the 5' end of the sequencing read [18, 47]. In some cases crosslinking will induce deletions; thus, the protein binding site is assumed to be at

the deletion site [48–50]. We tested our dataset with the iCLIPro analysis tool [51], which confirmed that our iCLIP reads do cluster around the 5' end of the reads (Additional file 3: Figure S2). In total, we identified 40,670 hnRNP A1 binding peaks in 6864 genes (Additional file 4). We searched for enriched motifs in the hnRNP A1 binding peaks and, confirming the validity of our iCLIP approach, the most highly enriched motif has high similarity to the SELEX-based motif UAGGGA/U [25] and to the previous HITS-CLIP generated motifs [17] (Fig. 1a).



The hnRNP A1 crosslinking sites are predominantly located in introns; however, when accounting for the total length of the analyzed RNA regions, hnRNP A1 crosslinking is enriched within exonic regions, especially in the 3' UTR (Fig. 1b). Generally, hnRNP A1 binds uniformly across introns, though close to exons hnRNP A1 binding is reduced (Fig. 1c). hnRNP A1 may thus silence introns by packaging pre-mRNA into hnRNP particles and repress splicing of the vast number of pseudoexons present in introns [52, 53], while exons are left free for splice site recognition. Close to the 3' splice site the binding of hnRNP A1 increases (Fig. 1d); this may reflect a low affinity towards the polypyrimidine tract region immediately upstream of the 3' splice site but high-affinity binding near the 3' splice site, which resembles the hnRNP A1 binding motif [25]. We observe slightly increased hnRNP A1 binding in the immediate flanking intronic regions (± 100 bases) of alternatively spliced exons (cassette exons) when compared to this region in all internal exons (Fig. 1d), suggesting that hnRNP A1 is involved in regulation of alternative splicing by binding in introns close to cassette exons. Furthermore, intronic regions surrounding cassette exons are far more conserved than those flanking constitutive exons, indicating that these sequences are important for regulation of alternative splicing [54]. A total of 240 hnRNP A1 binding peaks were identified inside or within 50 bp of cassette exons (Additional file 4).

Blocking of the splice sites may be a major mechanism of hnRNP A1-mediated exon repression

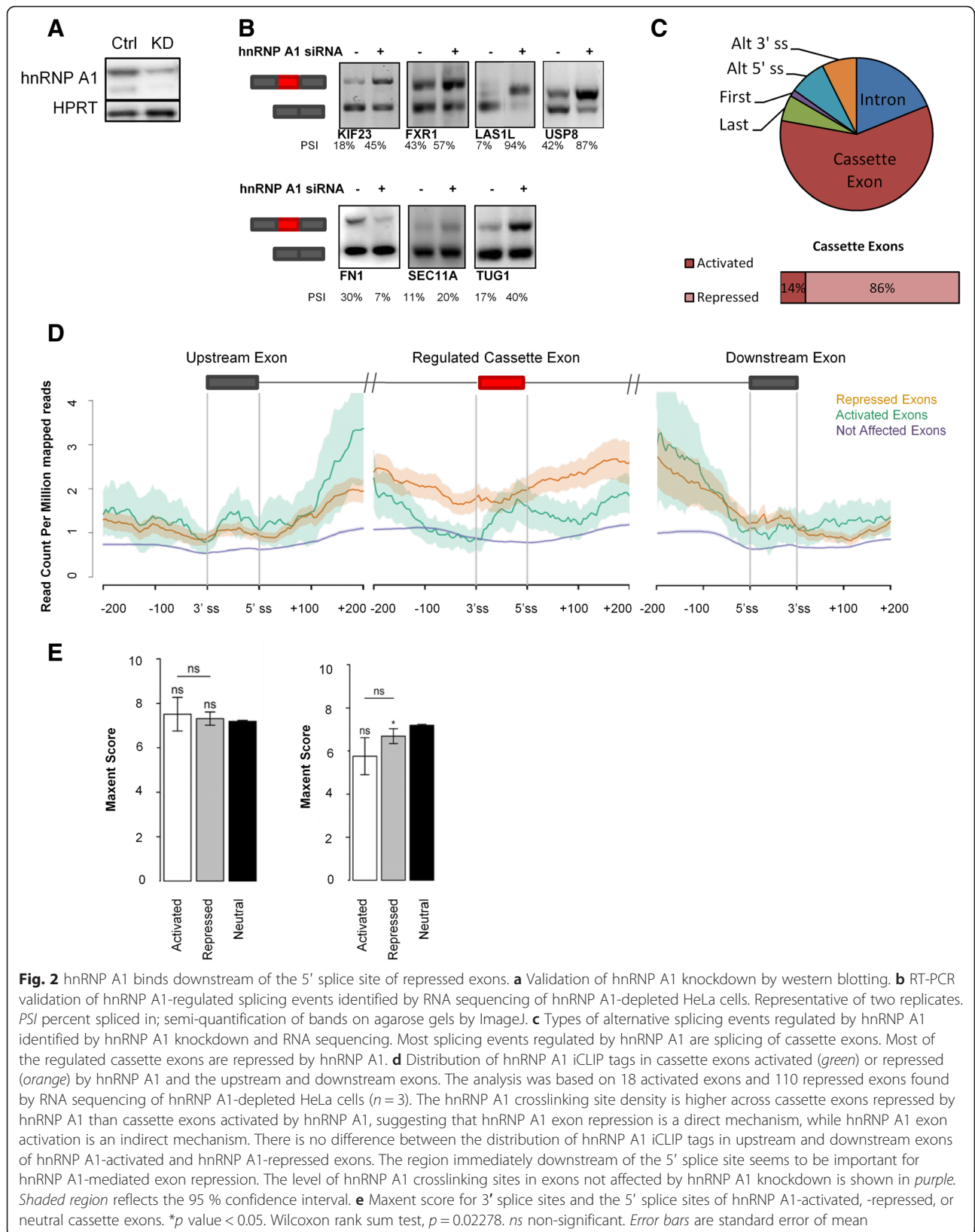
We speculated that the enrichment of hnRNP A1 binding near cassette exons would be further enriched near exons actually regulated by hnRNP A1. Therefore, we knocked down hnRNP A1 in HeLa cells and analyzed the samples by RNA sequencing to identify splicing events regulated by hnRNP A1 (Fig. 2a, b). Consistent with the previous hnRNP A1 CLIP study [17], we find that hnRNP A1 predominantly regulates cassette exons, and in line with the established role of hnRNP A1 as a splicing repressor, we observed that hnRNP A1 preferentially represses cassette exon inclusion (Fig. 2c, Additional file 4); 18 exons were activated and 110 exons were repressed by hnRNP A1. To identify if there is any correlation between the effect of hnRNP A1 knockdown and the location of hnRNP A1 binding sites, we mapped the density of hnRNP A1 iCLIP crosslinking sites across hnRNP A1-regulated exons and the upstream and downstream exons (Fig. 2d). The density of hnRNP A1 crosslinking sites across the repressed exons is higher than across activated exons, especially in the proximal introns. This further supports the idea that binding of hnRNP A1 in introns close to splice sites represses exon inclusion by, for instance, steric blocking of the splice

sites. Inside exons, hnRNP A1 may repress exon inclusion by antagonizing the binding and function of positive splicing regulatory proteins such as the SR proteins. The sequence composition might affect the ability of hnRNP A1 to polymerize along the exon as previous models of hnRNP A1 repression suggest, and to investigate this we examined the GC content in the affected exons and in 100-bp regions surrounding the exons (Additional file 5: Figure S3). We found that the GC content is significantly decreased in repressed exons relative to unaffected exons while there was no significant difference in the activated exons, indicating that there are more hnRNP A1 binding motifs in repressed than activated exons, consistent with a model in which hnRNP A1 activates exon splicing indirectly and not by binding directly to sites near the exon.

Since cassette exons generally have weaker splice sites than constitutive exons, hnRNP A1 may generally interfere more with the recognition of the splice site of cassette exons, since U1 snRNP and U2AF are less capable of competing with the binding of hnRNP A1 at these splice sites. We hypothesized that exons regulated by hnRNP A1 would have even weaker splice sites. We therefore estimated splice site strength with the Maxent scoring algorithm [55], and found that among the repressed exons the 5' splice sites were weaker than in unaffected exons (Wilcoxon rank sum test, $p = 0.02278$). The activated exons also showed a trend towards lower score, although not to a statistically significant extent (Wilcoxon rank sum test, $p = 0.05643$) (Fig. 2e). There was no significant change in the strength of the 3' splice site, indicating that the 5' splice site may be a primary target of regulation by hnRNP A1.

iCLIP identifies hnRNP A1 binding sites in disease-associated genes

We wanted to use our hnRNP A1 iCLIP binding map to identify binding sites in disease genes. A number of the hnRNP A1-regulated cassette exons identified by RNA sequencing were found in disease-associated genes (Additional file 2: Table S2). Furthermore, we identify hnRNP A1 binding sites in a number of known disease-associated genes (Additional file 4). Splicing of the mutually exclusive *PKM* exons 9 and 10 determines the metabolic switch towards lactate production and the Warburg effect in cancer cells [56–60]. We observed increased inclusion of *PKM* exon 9 and increased skipping of exon 10 after hnRNP A1 knockdown, and a large number of hnRNP A1 reads at the previously identified hnRNP A1 binding sites in *PKM* exon 9 and intron 9, consistent with the proposed regulatory role of hnRNP A1 in *PKM* alternative splicing (Additional file 6: Figure S4A). Interestingly, we identify an hnRNP A1 binding site in exon 10 that overlaps a previously reported splicing



silencer. This silencer is normally masked by an adjacent splicing enhancer, and only when it was inserted in a different context did it function as a silencer [60].

The role of hnRNP A1 in regulating splicing of *SMN2* exon 7 has been extensively studied. However, the high similarity of *SMN1* and *SMN2* (more than 99 % identical nucleotides) makes it impossible to uniquely map all the reads in the two transcripts. In our iCLIP analysis, we filtered away non-unique reads, so, initially, we obtained only few reads mapping to *SMN1* and *SMN2*. Thus, to identify all reads in *SMN1* and *SMN2*, we also examined the non-unique reads. We identified hnRNP A1 cross-linking sites overlapping all the functionally important motifs previously demonstrated to regulate *SMN2* exon 7 splicing by hnRNP A1 binding. (1) We identified the hnRNP A1 binding peak in intron 7 position +100 in *SMN2*, which is an *SMN2*-specific hnRNP A1 binding site (Fig. 3a) [61]. (2) We identified one hnRNP A1 crosslinking site in *SMN2* exon 7 (and not in *SMN1*) at the silencer motif created by the *SMN2*-specific mutation (Fig. 3b, ESS), supporting the idea that the *SMN2* exon 7 mutation both disrupts an SRSF1 binding splicing enhancer and creates an hnRNP A1-binding splicing silencer [7, 62]. (3) We identified the hnRNP A1-binding splicing silencer that overlaps the 3' splice site of both *SMN1* and *SMN2* exon 7 (Fig. 3b, ISS) [63]. Interestingly, we observe the hnRNP A1 crosslinking site only in *SMN2* consistent with our previously proposed model, which suggests that binding of hnRNP A1 to this splicing silencer is synergistic with binding of hnRNP A1 to the other splicing silencers in *SMN2* that inhibit exon 7 inclusion. (4) The hnRNP A1 binding intronic silencer N1 [12] is identical in *SMN1* and *SMN2* and consequently it was not identified in the initial analysis, but manual analysis did identify two hnRNP A1 crosslinking sites at this well-known hnRNP A1-binding splicing silencer (Additional file 6: Figure S4B). This exemplifies that it is a general problem that most sequencing reads from highly similar regions cannot be uniquely aligned. Altogether, our results support that hnRNP A1 generally binds more efficiently to the *SMN2* transcript than to the *SMN1* transcript, and therefore splicing of *SMN2* exon 7 is inhibited.

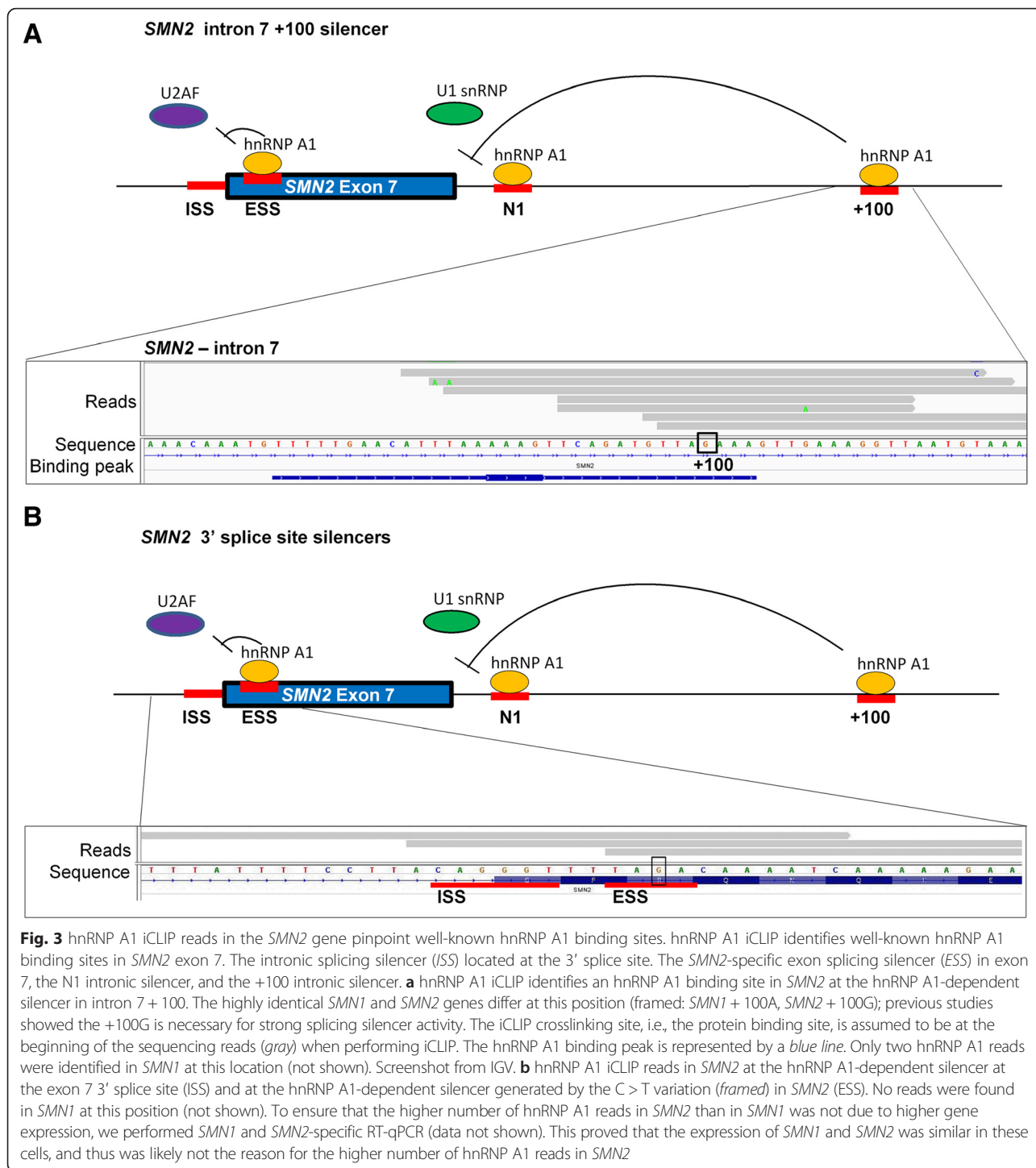
We suggest that our hnRNP A1 iCLIP binding map can also be used to predict if pathogenic mutations will affect hnRNP A1 binding. We find an enrichment of hnRNP A1 iCLIP reads in the long exons (exon 11) of *BRCA1* and *BRCA2* (Additional file 7: Figure S5A, B), indicating that hnRNP A1 is particularly important for the splicing of these exons. The hnRNP A1 binding peaks in these exons co-localize with a number of unclassified or pathogenic SNPs, and we found two SNPs in both *BRCA1* and *BRCA2* which disrupt the identified hnRNP A1 binding motif (Additional file 7: Figure S5C),

suggesting that these SNPs affect hnRNP A1 binding. We performed RNA oligonucleotide-affinity chromatography of HeLa nuclear extracts followed by western blotting with hnRNP A1 specific antibody, and indeed we observed that three of the SNPs (rs398122533, rs80357105, and rs80357458) reduced hnRNP A1 binding and thus they could potentially affect splicing.

Blocking of iCLIP-identified hnRNP A1 binding sites improves splicing of cassette exons

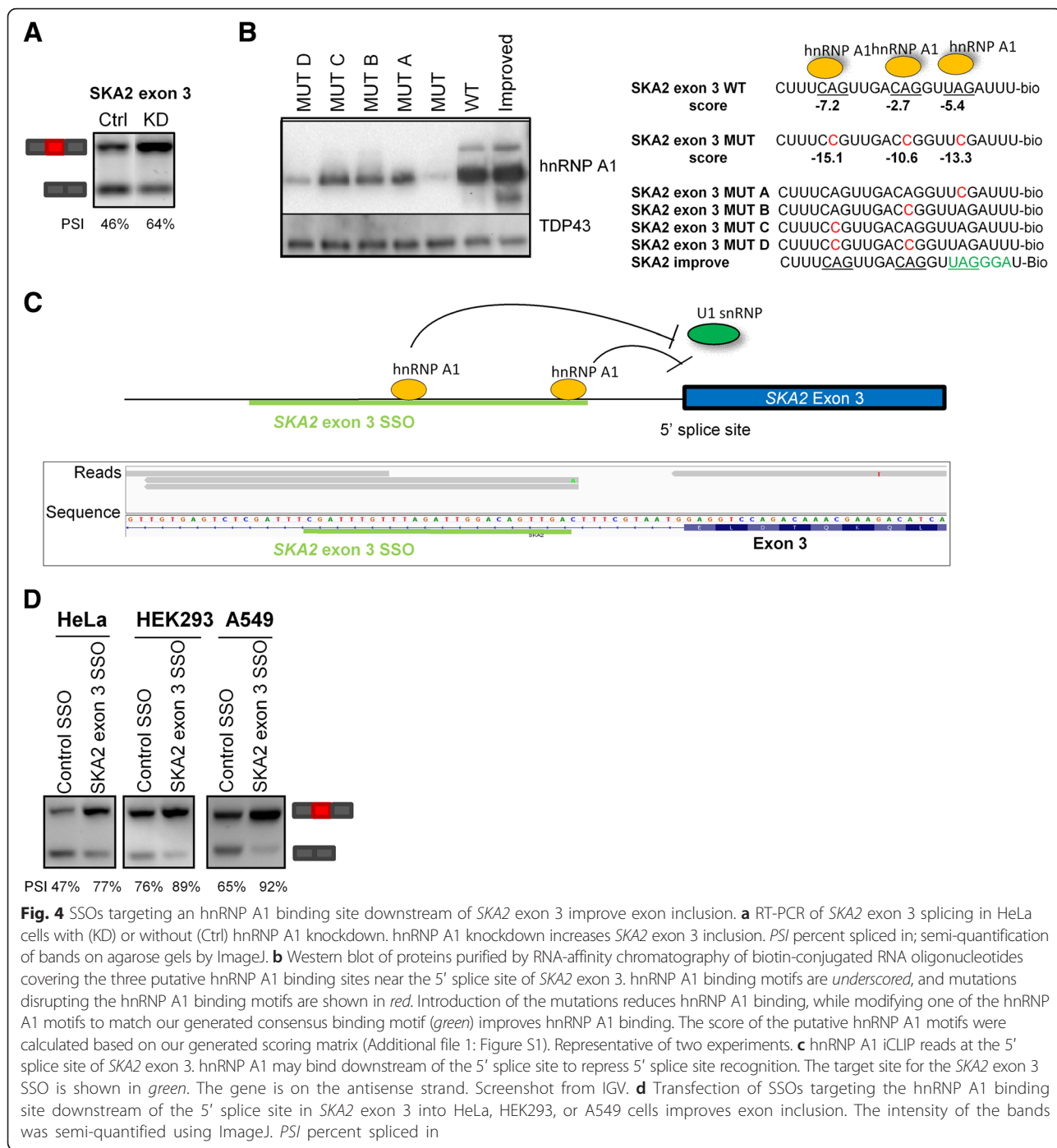
Since we observed that hnRNP A1 binding to sites close to the splice sites may be important mechanisms of hnRNP A1-regulated splicing repression, we hypothesized that SSO-mediated blocking of these sites could be used to activate exons. Since SSOs targeting introns close to the 3' splice site may interfere with the recognition of the polypyrimidine tract, it would likely be easier to activate exons by blocking hnRNP A1 binding sites close to the 5' splice site. Furthermore, since we observed a significant lower strength of the 5' splice site in hnRNP A1-regulated exons, but not in the strength of the 3' splice site, compared with unaffected exons (Fig. 2), the 5' splice site may be a primary target of regulation by hnRNP A1. Thus modulation of 5' splice site strength by blocking a nearby hnRNP A1 binding motif would have a greater effect on exon splicing. The efficient correction of *SMN2* exon 7 inclusion obtained by SSO-mediated blocking of the hnRNP A1-binding silencer N1 flanking the 5' splice site in *SMN2* supports this hypothesis [14].

To identify potential alternative cassette exons, where exon inclusion could be improved by SSO-mediated blocking of hnRNP A1 binding sites, we examined the distribution of hnRNP A1 iCLIP crosslinking sites in cassette exons identified by hnRNP A1 knockdown. Two iCLIP crosslinking sites were located 11 nucleotides downstream of the 5' splice site of the alternative exon 3 in the *SKA2* gene, and this region covers five potential hnRNP A1 binding sites containing the CAG/UAG trinucleotide (Fig. 4). Furthermore, the inclusion of *SKA2* exon 3 was higher in HeLa cells depleted of hnRNP A1 than in control cells (Fig. 4a). *SKA2* is part of the SKA complex which binds microtubules and is necessary for correct cell division in human cells [64]. *SKA2* expression is upregulated in lung cancer and involved in the progression of the disease, but it is not known which of the alternatively spliced *SKA2* transcripts (with or without exon 3) are, or if they are both, upregulated in lung cancer [65]. To confirm binding of hnRNP A1 downstream of the 5' splice site of *SKA2* exon 3, we performed RNA oligonucleotide-affinity chromatography of HeLa nuclear extracts followed by western blotting (Fig. 4b and Additional file 8: Figure S6A). We used RNA oligonucleotides in which the important UAG/CAG trinucleotide is



inactivated by a 2A > C mutation to UCG/CCG. According to our hnRNP A1 binding motif (Fig. 1a), this change is expected to disrupt hnRNP A1 binding, and we have previously demonstrated that the 2A > C mutation of the UAG/CAG motif disrupts hnRNP A1 binding [8, 40, 63]. We identified hnRNP A1 motif-specific binding of hnRNP A1 to the *SKA2* exon 3 oligonucleotides and found that

hnRNP A1 binds in an additive way to the motifs present in this region. Confirming the validity of our identified hnRNP A1 binding motif, we find that mutating the UUAGAUUU motif to UUAGGGAU, which resembles our hnRNP A1 consensus binding motif, increases hnRNP A1 binding. The binding of hnRNP A1 to the region downstream of the exon 3 5' splice site was also



confirmed by surface plasmon resonance imaging (SPRi) (Additional file 8: Figure S6B). To further examine the mechanism of hnRNP A1-mediated repression of SKA2 exon 3, we constructed SKA2 minigenes containing exon 2, exon 3, and part of exon 4 and shortened intervening introns (see Methods section) (Additional file 9: Figure S7). Unfortunately, we had to dramatically reduce the length of introns 2 and 3 (intron 2 was reduced from 11,785 nt to 1305 nt and intron 3 was reduced from 6973 nt to

1200 nt) to allow construction of the minigene. This resulted in very efficient, nearly complete, exon 3 inclusion in the minigene context, probably due to the reduction of intron sizes, which is known to improve exon inclusion. Although, exon 3 skipping from the minigene was minimal compared to the endogenous SKA2 gene, we did observe that disruption of the hnRNP A1 binding motifs by mutating the important UAG/CAG to UCG/CCG results in complete exon inclusion. This is consistent with the

proposed model suggesting that hnRNP A1 binds near the 5' splice site of *SKA2* exon 3 and represses exon inclusion.

To test our hypothesis that exons can be activated by blocking hnRNP A1 binding sites adjacent to the 5' splice site, we designed an SSO covering three hnRNP A1 binding sites near the 5' splice site of *SKA2* exon 3. Transfection of this SSO into HeLa cells, HEK293 cells, and the lung cancer cell line A549 significantly improves *SKA2* exon 3 inclusion endogenously (Fig. 4c, d) and in the minigene (Additional file 9: Figure S7C). Furthermore, we show that a dose-response curve exists between the dose of SSO and the amount of *SKA2* exon 3 inclusion in HeLa cells (Additional file 9: Figure S7D). Altogether, this supports the idea that blocking of iCLIP-identified hnRNP A1 binding sites downstream of the 5' splice site is a feasible way to improve splicing of alternative exons.

SSO-mediated blocking of hnRNP A1 binding sites improves inclusion of inactivated exons

To further examine the potential of blocking hnRNP A1 binding sites near the 5' splice sites of weak or inactivated exons, we used pseudoexons as a model of constitutive exons inactivated by mutations in SREs [3–8, 40, 41]. A group of pseudoexons in the *PCCA*, *GLA*, *FGB*, *CFTR*, *ATM*, *MTRR*, and *COL4A5* genes are activated by disease-causing point mutations, which either disrupt a splicing silencer or create a splicing enhancer [53, 66–72]. We examined the presence of hnRNP A1 iCLIP binding peaks between the 5' splice site and 100 bp downstream in these seven pseudoexons. We identified an hnRNP A1 binding peak downstream of the *MTRR* pseudoexon (approximately 11 bases from the 5' splice site in intron 6) (Fig. 5a). The region downstream of the 5' splice site of the *MTRR* pseudoexon contains three potential hnRNP A1 binding sites, and when introducing the 2A > C mutation at all these sites (MUT A), the binding of hnRNP A1 was significantly reduced (Fig. 5b). When analyzing oligonucleotides containing mutations at only one or two of these sites (MUT B, MUT C, and MUT D), we observed hnRNP A1 binding only when the motif closest to the 5' splice site was intact. This was confirmed by SPRi analysis. We observed that hnRNP A1 bound with approximately nine times higher affinity to the wild-type *MTRR* sequence than to the MUT A sequence (Additional file 10: Figure S8). However, when leaving only the proximal hnRNP A1 binding site intact (MUT B), the affinity for hnRNP A1 was similar to the wild-type sequence. We could not detect binding of hnRNP A1 to MUT C and MUT D oligonucleotides using SPRi; thus, in contrast to the hnRNP A1 binding sites near *SKA2* exon 3, the binding sites in *MTRR* were not additive, as the proximal hnRNP A1 binding site seems to be required for hnRNP A1 binding. This site could thus function as the nucleating site of

hnRNP A1 binding from which polymerization to lower affinity sites can be initiated. Confirming the importance of hnRNP A1 in *MTRR* pseudoexon repression, hnRNP A1 knockdown in HeLa cells increased *MTRR* pseudoexon inclusion approximately 5 times compared to control cells (Fig. 5c).

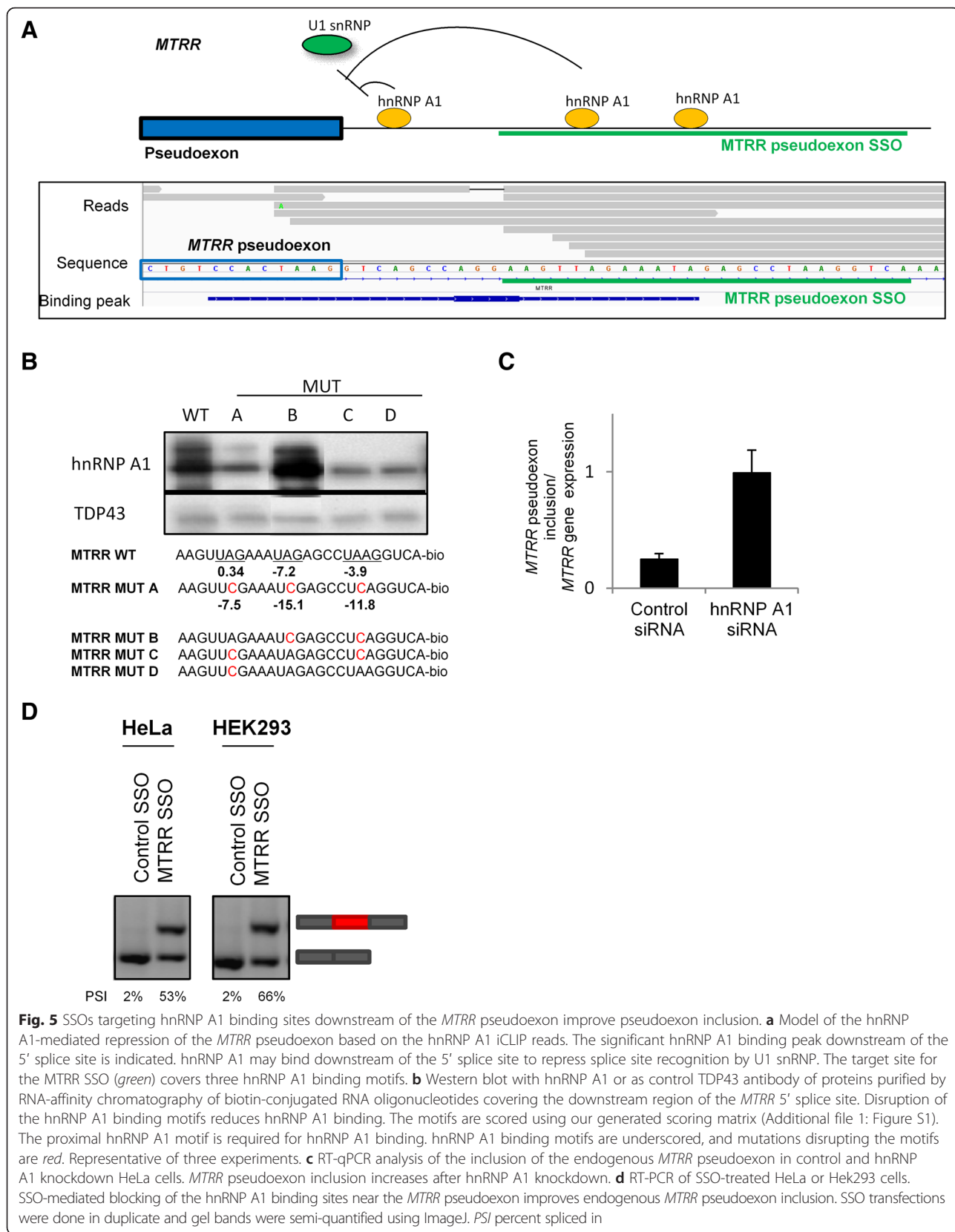
We also identified an hnRNP A1 binding peak approximately 70 bases from the 5' splice site of the *COL4A5* pseudoexon in intron 6. We examined the binding of hnRNP A1 to sequences covering the putative hnRNP A1 binding site near the *COL4A5* pseudoexon. The introduction of the 2A > C mutation in the *COL4A5* oligonucleotide reduced hnRNP A1 binding as shown by both RNA oligonucleotide-affinity chromatography and SPRi analysis (Additional file 11: Figure S9). The *COL4A5* pseudoexon inclusion was, however, not improved after hnRNP A1 knockdown in HeLa cells (Additional file 12: Figure S10A).

To examine if the inclusion of the *MTRR* and *COL4A5* pseudoexons could be improved by blocking the hnRNP A1 binding sites downstream of the 5' splice site, we designed SSOs blocking the putative hnRNP A1 binding sites and transfected HeLa and HEK293 cells (Fig. 5d, Additional file 12: Figure S10B and S10C). Transfection of the SSOs dramatically improved inclusion of the endogenous *MTRR* pseudoexon. This is a direct effect of the SSO treatment, as we demonstrate a dose-response relationship between SSO concentration and *MTRR* pseudoexon inclusion (Additional file 13: Figure S11). This suggests that binding of hnRNP A1 downstream of the *MTRR* pseudoexon directly represses pseudoexon inclusion, and that this splicing silencer is a major determinant of pseudoexon repression, which can be antagonized by SSO treatment to enable considerable levels of inclusion. Consistent with the results of siRNA-mediated knockdown of hnRNP A1, SSO-mediated blocking of the *COL4A5* hnRNP A1 binding site did not improve *COL4A5* pseudoexon inclusion, indicating that the hnRNP A1 binding site near the *COL4A5* pseudoexon is not a major determinant for *COL4A5* pseudoexon repression.

Taken together, our data corroborate our hypothesis that iCLIP can be used to pinpoint important regulatory hnRNP A1 binding sites close to the 5' splice site, and that splicing of alternatively spliced cassette exons or constitutive exons weakened from point mutations in SREs can be corrected by SSO-mediated blocking of the identified hnRNP A1 binding sites.

Discussion

Approximately 10–15 % of all disease-causing mutations are located in splice sites, and an estimated 25 % of exonic disease-causing mutations also affect pre-mRNA splicing [73]. As an increasing number of patients will be subject to gene sequencing for diagnostic and therapeutic purposes, there is a growing demand to be able to



predict the consequences of genetic variation. We used iCLIP to identify the transcriptome-wide binding sites of hnRNP A1 in HeLa cells and identified more than 40,000 hnRNP A1 binding sites across the transcriptome. iCLIP thus provides a map of occupied *in vivo* binding sites, and using the identified hnRNP A1 binding sites, we derived an hnRNP A1 consensus binding motif (Fig. 1a). Together, this enables us to evaluate the effect of genetic variation on hnRNP A1 binding and, as we show here, identify target sites for SSOs to redirect splicing.

The hnRNP A1 binding motif that we generate resembles the *in vitro* SELEX-based motif and the motif reported from the previous HITS-CLIP study [17]. Also previously identified functional hnRNP A1 binding sites are consistent with our consensus motif [57, 74–76], including the UAGGGA motif created by a c.158A > G mutation in the *ETFDH* gene which creates an hnRNP A1 binding site [40]. When we analyzed the generated *ETFDH* hnRNP A1 binding motif, UAGGGA, along with the wild-type sequence UAAGGA by SPRi using recombinant hnRNP A1, we found the mutant motif to bind hnRNP A1 with approximately six times higher affinity than the wild-type motif, confirming that UAGG is a strong hnRNP A1 motif (Additional file 14: Figure S12).

Overall, this hnRNP A1 iCLIP analysis in HeLa cells is consistent with the previous hnRNP A1 HITS-CLIP study using HEK293T cells [17] (Additional file 2: Tables S3 and S4; Additional file 4). We find similar distributions of hnRNP A1 binding sites across the transcriptome, with the 3' UTR being the most enriched region. This probably reflects some of the non-splicing functions of hnRNP A1. Furthermore, after hnRNP A1 knockdown in both HeLa and HEK293T cells, cassette exons are the most deregulated splicing event. In particular, hnRNP A1 mainly represses cassette exon inclusion. Nineteen (15 %) of the cassette exons we identify in this study to be regulated by hnRNP A1 were also detected in HEK293T cells, including KIF23 and LAS1L (Additional file 4). We identified approximately 20 times more binding peaks than the previous HITS-CLIP study (Additional file 2: Table S3). This may be explained by the higher sensitivity of our iCLIP approach compared to the HITS-CLIP-based approach and the higher number of aligned reads in our study. For comparison we analyzed the HITS-CLIP hnRNP A1 dataset [17] in a similar way as we analyzed our own data. This resulted in identification of 1432 binding peaks, which is still less than 20 times the number we obtained (Additional file 2: Table S3). Of these 1432 binding peaks, 328 (23 %) overlapped with our hnRNP A1 iCLIP binding peaks. Altogether, this suggests that *in vivo* hnRNP A1 binding sites are cell-type specific, and that CLIP analysis in different cell lines will significantly increase the number of known hnRNP A1 binding sites in the transcriptome. Thus, there may still be important disease-associated

hnRNP A1 binding sites in genes which we do not detect in HeLa cells because of, e.g., low expression levels. However, we believe that this iCLIP analysis can be used as a guide to identify potential hnRNP A1 binding sites, especially if one includes not only the significant iCLIP binding peaks but also individual reads and clusters of reads.

The binding profile of hnRNP A1 is consistent with several of the reported hnRNP A1 splicing mechanisms. hnRNP A1 preferentially binds introns, likely reflecting the inhibitory role of hnRNP A1 in splice site recognition [31, 33, 77] (Fig. 1b, c). hnRNP A1 may thus be important in suppressing the usage of the numerous cryptic splice sites in introns [78]. Interactions between intronic hnRNP A1 proteins may also promote exon skipping by looping out alternative exons [32, 33]. Alternatively, interactions with other hnRNP A/B proteins bound in the same intron will loop out or condense the intron to improve exon recognition [34]. In hnRNP A1-mediated repressed exons, we observe a high density of hnRNP A1 crosslinking sites in the immediate flanking introns compared to both activated exons and unaffected exons (Fig. 2d). This suggests that these are important regions for hnRNP A1-mediated splicing repression and that hnRNP A1 represses exons directly by sterically blocking U1 snRNP or U2AF recognition of the splice sites. We observe weaker 5' splice sites in hnRNP A1-repressed cassette exons than in unaffected exons, indicating that the 5' splice site is the target for hnRNP A1-mediated exon repression. Generally, the hnRNP A1-repressed exons have a high density of hnRNP A1 crosslinking sites, indicating that hnRNP A1 also represses exons by other mechanisms, for example, by antagonizing the binding and function of SR proteins. In general, hnRNP A1-activated exons and their proximal introns have fewer hnRNP A1 crosslinking sites than repressed exons, suggesting that hnRNP A1 activates exons in an indirect way. Furthermore, hnRNP A1 may also regulate gene expression and splicing indirectly. hnRNP A1 iCLIP binding peaks were abundant in the pre-mRNAs encoding hnRNP proteins, in particular in the 3' UTR region (Additional file 15: Figure S13). We identified hnRNP A1 binding peaks in the 3' UTR of *HNRNP A2/B1*, where hnRNP A1 binding has been shown to induce alternative splicing, resulting in an mRNA transcript downregulated by nonsense mediated decay (NMD) [17, 29]. These results are also consistent with the findings from Huelga et al. A large number of iCLIP crosslinking sites were also present in hnRNP A1's own pre-mRNA. hnRNP A1 exists as two isoforms depending on the inclusion of exon 7B. The longer isoform binds more strongly to RNA, but with less capability to regulate alternative splicing [79]. Different elements in the intronic region surrounding exon 7B are implicated in *HNRNP A1* alternative splicing [33, 80, 81], and we identified hnRNP A1 binding peaks at the hnRNP

A1-dependent CE1 and CE4 splicing regulatory elements near exon 7B. Besides the hnRNP proteins, we find hnRNP A1 binding peaks in a large number of other splicing regulatory proteins, including many of the SR proteins such as SRSF1 and SRSF2 (Additional file 4). We identified 1141 genes (false discovery rate [FDR] <0.1) with altered expression after hnRNP A1 knockdown in HeLa cells (Additional file 4). An interesting observation was the finding that the oncogenic transcription factor c-Myc was upregulated in response to hnRNP A1 knockdown. This is consistent with the proposed regulatory feedback loop where c-Myc upregulates hnRNP A1 transcription and hnRNP A1 downregulates c-Myc by blocking the biogenesis of let-7a, which in turn downregulates c-Myc [82].

We believe our hnRNP A1 binding map can be used to identify important hnRNP A1 binding sites in disease-associated genes. We identify well-known hnRNP A1 binding sites in *SMN1/2* and *PKM*. However, some of these well-documented hnRNP A1 binding sites were covered by only a few reads. Other well-documented hnRNP A1 binding sites were also represented by relatively few reads, such as the previously reported hnRNP A1 binding site (CAGGGG at c.350–c.355) in an exonic splicing silencer in *ACADM* exon 5 [8]. Although this splicing silencer is functionally important, the fact that it is positioned close to a splicing enhancer may prevent strong hnRNP A1 binding and thus limit detection by CLIP methods. It is well known that splicing regulatory proteins compete for binding at closely spaced silencers and enhancers in a finely tuned balance; thus, the proteins that bind at the splicing enhancer may sterically block the adjacent splicing silencer. If a mutation disrupts the splicing enhancer, this will allow increased hnRNP A1 binding to the silencer and inhibit exon inclusion, as observed for *ACADM* exon 5 [8]. Thus, in regions with low read coverage, functional hnRNP A1 silencers may be represented by only a few reads, because hnRNP A1 competes with other proteins for binding in vivo. This may also be the reason why we only identify hnRNP A1 crosslinking at the 3' splice site of *SMN2* and not of *SMN1*, where it is antagonized by SRSF1 binding at the adjacent *SMN1*-specific enhancer.

As an increasing number of pathogenic genetic variants are found to affect mRNA splicing, the ability to reverse or compensate splicing will be of great value. Employing SSOs which block access to hnRNP A1-binding splicing silencers may be a way to shift between alternatively spliced isoforms or relieve the repression of a splicing-incompetent exon, as illustrated by the very successful upregulation of *SMN2* exon 7 inclusion by SSO-mediated blocking of the hnRNP A1-binding splicing silencer N1 [14]. Since SSOs targeting introns flanking the 3' splice site may interfere with the recognition

of the polypyrimidine tract, we expect that SSO-mediated blocking of the hnRNP A1 binding sites close to the 5' splice site will be a much more feasible way of activating exons. Here, we demonstrated that inclusion of the alternative cassette exon 3 in *SKA2* can be significantly increased by SSO-mediated blocking of an iCLIP-identified hnRNP A1 binding site downstream of the 5' splice site (Fig. 4d). However, we do not know whether this approach can be exploited therapeutically in the future, as the functional importance of the alternative splicing of *SKA2* exon 3 is still unknown.

Cassette exons are vulnerable exons with generally weak splice sites; therefore, they may be rather easily manipulated. Ideally, we would also like to use this approach to improve inclusion of constitutive exons which are skipped due to pathogenic mutations in SREs. However, since constitutive exons are normally included in the mRNA transcript, we cannot test the usability of splicing-improving SSOs on these exons. A group of pseudoexons which contain functional splice sites are inactivated by mutations in SREs, and thus these pseudoexons functionally resemble constitutive exons inactivated by mutations in SREs. We identified hnRNP A1 binding sites located close to the 5' splice site of a pseudoexon of this type in *MTRR* (Fig. 5a), and we show that SSO-mediated blocking of these hnRNP A1 binding sites dramatically increases exon inclusion (Fig. 5d). This provides a clear proof of the principle that iCLIP-identified hnRNP A1 binding sites can be targeted by SSOs to improve inclusion of inactivated exons. Blocking of the hnRNP A1 binding site downstream of the *COL4A5* pseudoexon did not, however, increase exon inclusion (Additional file 12: Figure S10C). The hnRNP A1 binding site downstream of the *COL4A5* pseudoexon is 70 bp from the 5' splice site, whereas the hnRNP A1 binding sites downstream of *SKA2* exon 3 and the *MTRR* pseudoexon are located only 11 bp from the 5' splice sites. This could indicate that SSO-mediated blocking of hnRNP A1 binding sites only improves exon inclusion when the target site is close to the 5' splice site. Also the hnRNP A1-binding N1 splicing silencer in *SMN2* exon 7 is located only 11 bp from the 5' splice site, supporting the idea that hnRNP A1 binding sites located immediately (10–40 bp) downstream of 5' splice sites are good targets for this SSO-based approach.

Conclusion

We show here that our hnRNP A1 binding map can be used to identify functional hnRNP A1 binding sites, and that hnRNP A1 binding sites immediately downstream of the 5' splice site can be blocked by SSOs to increase exon inclusion. This may be a useful therapeutic approach to compensate for the missplicing caused by pathogenic genetic variants in splicing regulatory

elements. Furthermore, in combination with our hnRNP A1 binding map, our generated hnRNP A1 binding motif will enable identification of hnRNP A1 binding sites which have been created or disrupted by genetic variants or mutations to cause aberrant splicing and disease.

Methods

iCLIP analysis

iCLIP was performed as described [44] with three biological replicates. Briefly, HeLa cells with stable inducible expression of T7-tagged hnRNP A1 [45] were cultured in RPMI 1640 with L-glutamine (Lonza) supplemented with streptomycin, penicillin, and glutamine. Cells were seeded in 10-cm dishes and induced with 1 μ g/ml doxycycline (Sigma-Aldrich) for 48 h before UV crosslinking. At the time of crosslinking, cells were approximately 90 % confluent. After cell lysis, samples were sonicated and RNase treated with RNase Cocktail™ Enzyme Mix (Life Technologies), a mixture of RNaseA and RNaseT1 diluted 1:10, 1:200, and 1:2000. T7-hnRNP A1 was immunopurified with Dynabeads Protein G (Life Technologies) coupled to anti-T7-antibody (69522-4, Merck Millipore). RNA-protein complexes were separated by size by SDS-PAGE and transferred to a nitrocellulose membrane from which RNA-protein complexes of an appropriate size (>20 kDa above protein size) were cut out. Subsequently, the protein was digested by Proteinase K (Fisher Scientific). cDNA was synthesized with Superscript II Reverse Transcriptase (Life Technologies) and purified on a 5 % TBE-Urea gel (Bio-Rad). The cDNA was circularized using CircLigase II ssDNA Ligase (Epicentre), relinearized with *Bam*HI restriction enzyme, and used as a template in 20–25 cycles of PCR amplification. After PCR amplification, DNA fragments below 100 bp were removed using Agencourt AMPure XP beads (Beckman Coulter). The amplified fragments were sequenced by high-throughput sequencing on a HiSeq 1500 (Illumina) to obtain 50-bp single reads. The 3' linker, reverse transcription primers, and PCR primers for amplification were as described previously [44].

Mapping of iCLIP reads

FASTQ reads were first demultiplexed according to the barcode sequence RRRRIIIRR, in which R is a random nucleotide and I is a part of the identifier sequence. The two random segments were then stitched together into a 5-bp random tag used for identifying PCR duplicates and saved within the read name. Following demultiplexing, the reads were quality trimmed and adapter trimmed such that any adapter sequence at the 3' end of the reads was removed, allowing up to one mismatch. Reads with a length of 20 bp or more were retained. For these steps, we used custom Perl scripts. Next, reads were mapped to the human genome (hg19) using the

Burrows-Wheeler Aligner (BWA) [83] (bwa aln parameters: -l 20 -n 2, bwa samse parameters: -n 10) and, using mapping positions together with the random tags, PCR duplicates were removed using a custom Perl script. Tags with up to one mismatch between them were considered identical. This method allows detection of PCR duplicates that contain sequencing errors, either in the random tag or within the fragment sequence. Finally, we removed non-uniquely aligned reads prior to downstream analyses.

iCLIP peak identification and motif analysis

Crosslinking sites were defined as the base immediately preceding the read start for reads without deletions, while the crosslinking site in reads containing deletions was defined as the site of the deletion. The mapping of crosslinking sites to the beginning of fragments was validated using iCLIPPro [51] to estimate the overall crosslinking profile of hnRNP A1 in our iCLIP experiments. The program was run on aligned fragments of lengths 20–35 bp compared to fragments of length 41 bp in windows of 300 bp containing at least 20 aligned fragments. To identify binding sites, we extended these crosslinking sites by 10 bp to either side, or in reads with deletions it was defined as ± 10 bases from the deletion site. Peak detection was performed on these regions using CLIPper [84]. We used the “superlocal” algorithm to account for local sequencing bias in 1-kb flanking regions, and the “random” algorithm to estimate *p* values within these windows. We set a threshold at FDR = 10 % to identify significant binding sites with full parameters: -s hg19 -premRNA -disable_global_cutoff -FDR = 0.1 -superlocal -threshold-method = random. To identify binding motifs, we used the HOMER analysis program [85] run in RNA mode on the identified binding sites with the exact sizes of the binding sites as the targets, and a random similarly sized set of regions as the background (full parameters: -rna -size given -len 6,7,8 -mis 3 -depth high -basic). To generate the scoring matrix we used the frequency matrix to compute log₂ scores of the frequencies relative to a background frequency of 0.25. We downloaded raw fastq files for the hnRNP A1 samples from the study of Huelga et al. [17] and mapped them with BWA as described for our own data. We removed PCR duplicates based on mapping position and then removed non-uniquely aligned reads. To identify binding sites, we performed peak detection using CLIPper directly on the mapped reads. We used BEDTools to identify overlapping binding sites. FDR values were calculated by adjusting *p* values using the Benjamini-Hochberg method.

Genomic distribution of binding peaks

We used the RSeQC [86] package to calculate genomic distributions and ngs.plot [87] to plot them in localized windows. Overall densities across genomic regions (Ensembl version 75) were calculated by normalizing to

the total combined length of the region in question (reads per kb), while plotted densities were normalized to the total number of reads (reads per million). The alignment of iCLIP reads to the human genome was visualized using the Integrative Genomics Viewer (IGV) [88, 89]. Coverage profiles of iCLIP tags and peaks were created using ngs.plot [90].

RNA affinity purification

RNA affinity purification was performed as previously described [8]. For each purification 100 pmol of RNA oligonucleotides (Additional file 16: Materials and methods) was coupled to 50 μ l of streptavidin-coupled magnetic beads (Life) and incubated with HeLa nuclear extract (CilBiotech s.a.). After washing, bound proteins were investigated by western blotting. The experiments were performed two or three independent times.

Knockdown of hnRNP A1

HeLa cells were cultured in RPMI 1640 with L-glutamine (Lonza) supplemented with streptomycin, penicillin, and glutamine. Cells were seeded in 6-well plates and transfected in duplicate with siRNA smartpool targeting hnRNP A1 (L-008221-00) or non-targeting siRNA (D-001810-10-20) (both from Dharmacon) using Lipofectamine RNAiMAX (Life Technologies) following the manufacturer's instructions. After 24 h, the cells were transfected again with siRNAs using Lipofectamine RNAiMAX and the forward transfection protocol. After 96 h in total, the cells were harvested for RNA and protein analysis. Knockdown was performed three independent times. RNA was isolated using Isol-RNA Lysis Reagent (5 Prime). The RNA was treated with RQ1 RNase-Free DNase (Promega) and 1 μ g of RNA was reverse-transcribed using the High-Capacity Reverse Transcription Kit (Life Technologies). cDNA was used for qPCR using FastStart Essential DNA Green Master (Roche) and the LightCycler 480 instrument II (Roche), or PCR using TEMPase Hotstart Mastermix (Ampliqon), and analyzed by agarose gel electrophoresis (using an Advanced Analytical Technologies system). Agarose gels were semi-quantified using the ImageJ software package [91]. Primer sequences can be found in Additional file 16: Materials and methods.

Western blotting

Protein was extracted using M-PER Mammalian Protein Extraction Reagent (Thermo Scientific). Proteins were separated on a 4–12 % Bis-Tris gel using the NuPAGE SDS Gel System (Life Technologies) and membranes were probed with anti-T7-tag antibody (AK42, Cold Spring Harbor Laboratory), anti-hnRNP A1 (R9778, Sigma-Aldrich), anti-hnRNP H (sc-10042, Santa Cruz Biotechnology), anti-TDP-43 (10782-2-AP, Proteintech Group), or anti-HPRT (HPA006360, Sigma-Aldrich).

RNA sequencing

Three biological replicates of hnRNP A1 knockdown in HeLa cells were prepared for RNA sequencing using the TruSeq Stranded Total RNA library Prep Kit, Human/mouse/rat (Illumina). Samples were paired-end sequenced using Illumina's HiSeq 1500. Reads were subsequently mapped with Spliced Transcripts Alignment to a Reference (STAR) [92], and gene expression analysis was carried out using HTSeq [93] and DESeq2 [94] with conditional quantile normalization [95]. We used SAJR [96] to detect alternative splicing and select significant events as those with FDR <0.1. FDR values for all analyses were calculated by adjusting *p* values using the Benjamini-Hochberg method. To compare significant splicing changes of cassette exons, we extracted cassette exons that overlapped the intronic regions with cassette exon missplicing indicated in the Huelga et al. data made public at <http://rnabind.ucsd.edu> and with identical cassette exon lengths. Log2 fold-change estimates from SAJR analysis were computed by taking the log2 of the ratio between conditions of the average inclusion ratios with 1 % pseudo-inclusion added to each condition.

Transfection of splice-switching oligonucleotides

A549, HEK293, or HeLa cells were cultured in RPMI 1640 with L-glutamine (Lonza) supplemented with streptomycin, penicillin, and glutamine. Cells were reverse-transfected in duplicate using Lipofectamine RNAiMAX (Life Technologies) following the manufacturer's instructions using either 40 nM splice-switching oligonucleotides (SSOs) or, for the dose-response curve, 1–16 nM. The 2'-OMe phosphorothioate SSOs (Additional file 16: Materials and methods) were synthesized by DNA Technology. The cells were harvested 48 h after transfection, and RNA was isolated using Isol-RNA Lysis Reagent (5 Prime) and reverse-transcribed using the High-Capacity Reverse Transcription Kit (Life Technologies). PCR was performed using TEMPase Hotstart Mastermix (Ampliqon) and analyzed by agarose gel electrophoresis. Bands were quantified using either ImageJ [91] or the Fragment Analyzer (Advanced Analytical Technologies).

SKA2 minigenes

A fragment of the human *SKA2* gene encompassing exon 2, the first 200 bp of intron 2, an *Xho*I cleavage site, the last 1105 bp of intron 2, exon 3, the first 1000 bp of intron 3, a *Bam*HI cleavage site, the last 200 bp of intron 3, and the first 200 bp of exon 4 was cloned into pcDNA3.1+ (GenScript). The mutant minigene (*Mut*) contains A > C substitutions at five potential hnRNP A1 binding sites adjacent to the exon 3 5' splice site. Minigenes were transfected in duplicate into HeLa cells with X-tremeGENE 9 (Roche) according to the manufacturer's instructions.

Additional files

Additional file 1: Figure S1. hnRNP A1 iCLIP. **A.** The inducible expression of T7-hnRNP A1 analyzed by western blotting using either anti-hnRNP A1 antibody or anti-T7 antibody. **B.** iCLIP was performed three independent times. Autoradiography from one representative iCLIP is shown. Western blotting of input samples confirms antibody-specific immunoprecipitation of hnRNP A1. **C.** Scoring matrix based on the frequencies of bases in the hnRNP A1 binding motif (Fig. 1). (TIF 1098 kb)

Additional file 2: Table S1. iCLIP raw reads and alignment. **Table S2.** hnRNP A1-regulated cassette exons in disease-associated genes. **Table S3.** Comparison with Huelga et al. hnRNP A1 HITS-CLIP. **Table S4.** Comparison with Huelga et al. hnRNP A1 knockdown. (DOCX 52 kb)

Additional file 3: Figure S2. High-resolution read overlap heat map of hnRNP A1 iCLIP reads. Density of read start (*top*), center (*middle*), and end (*bottom*) indicated by colors and according to read length (*y-axis*) (20–35 nt) compared to reads of 41 bases (group R). The distributions are normalized to the total number of reads of that length. The hnRNP A1 iCLIP reads cluster around the read start sites, indicating that the hnRNP A1 crosslinking sites are located close to the read start site. (TIF 4883 kb)

Additional file 4: Lists of hnRNP A1 iCLIP binding peaks, hnRNP A1 iCLIP binding peaks in cassette exons including 50 bp in flanking introns, and gene expression and exon usage changes after hnRNP A1 knockdown in HeLa cells (FDR <0.1). In addition, a list of the genes with hnRNP A1 binding peaks in cassette exons including 50 bp in flanking introns present in the Orphanet, ClinVar (Pathogenic LoF), or Developmental Disorders (DDG2P) databases (NCBI). The complete output from the RNA sequencing and iCLIP analysis can be found in the ArrayExpress database (<https://www.ebi.ac.uk/arrayexpress/>) under accession number E-MTAB-3612. (XLSX 2197 kb)

Additional file 5: Figure S3. hnRNP A1-repressed exons have lower GC content than unaffected exons. The GC content in the 3' splice site (3'ss), 5' splice site (5'ss), and exon region of exons activated or repressed by hnRNP A1 or neutral exons is shown. **p* value < 0.05. *ns* non-significant. Error bars are standard error of mean. (TIF 2004 kb)

Additional file 6: Figure S4. hnRNP A1 iCLIP identifies well-known hnRNP A1 binding sites in *PKM* and the N1 silencer of *SMN1/SMN2*. **A.** hnRNP A1 iCLIP reads in *PKM* exons 9 and 10. hnRNP A1 reads are depicted as horizontal *gray bars*. The read density is depicted as vertical *gray bars*, and significant binding peaks are marked with *blue bars*. The gene is on the antisense strand. An hnRNP A1 binding peak across the reported splicing silencer (*ESS*) in *PKM* exon 10 (*red bar*). **B.** Non-uniquely aligned reads (*open bars*) located at the N1 hnRNP A1-dependent silencer (*N1 ISS*) in *SMN1/SMN2* intron 7. (TIF 2535 kb)

Additional file 7: Figure S5. hnRNP A1 binds *BRCA1* exon 11 and *BRCA2* exon 11 and SNPs disrupt hnRNP A1 binding. **A.** IGV screenshot of the distribution of hnRNP A1 iCLIP reads in *BRCA1*. **B.** The same in *BRCA2*. hnRNP A1 iCLIP reads are enriched in the long exons 11. **C.** RNA-affinity chromatography of biotin-conjugated RNA oligonucleotides carrying wild-type or SNP-containing *BRCA1* or *BRCA2* sequences, and subsequent western blotting with hnRNP A1 or as control TDP43 antibody. The sequences of the biotin-conjugated RNA oligonucleotides are listed. hnRNP A1 motifs are *underscored*, and the SNP variant disrupting the hnRNP A1 binding motif is *red*. (TIF 2004 kb)

Additional file 8: Figure S6. hnRNP A1 binds downstream of the pseudoexon in *SKA2* exon 3. **A.** Western blot of proteins purified by RNA-affinity chromatography of biotin-conjugated RNA oligonucleotides covering all the putative hnRNP A1 binding sites near the 5' splice site of *SKA2* exon 3. hnRNP A1 binding motifs are *underscored*, and mutations disrupting the hnRNP A1 binding motifs are shown in *red*. Introduction of the mutations reduces hnRNP A1 binding. Representative of two experiments. **B.** Surface plasmon resonance imaging (*SPRi*) using RNA oligonucleotides containing the *SKA2* sequence downstream of the 5' splice site or mutant sequence (*MUT*) where the hnRNP A1 binding motifs were mutated (Fig. 4). RNA oligonucleotides were immobilized in array format on a hydrogel-coated gold surface. Binding of hnRNP A1 was measured in real time by following changes of the SPR angles at all printed positions of the array during 3-min injections of hnRNP A1 protein over the entire surface. Five injections of a

twofold titration series from 6.25–100 nM hnRNP A1 were injected in sequence from the lowest concentration to the highest. A continuous flow of SPR buffer (10 mM HEPES/KOH pH 7.9, 150 mM KCl, 10 mM MgCl₂, 0.5 mM DTT) flowed over the surface before, between, and after the hnRNP A1 injections to measure baseline and dissociation kinetics. Responses for a calibration curve were created after the concentration series by measuring SPR responses from defined dilutions of glycerol in running buffer (ranging from 5–0 % glycerol) and of pure water as defined by the automated calibration routine of IBIS MX-96. *Left:* Increasing concentrations of recombinant hnRNP A1 were injected over the surface to monitor concentration-dependent association. *Right:* Binding of hnRNP A1 at equilibrium. No binding was observed to the *SKA2* MUT oligonucleotides, suggesting that hnRNP A1 binds in a sequence-dependent way. *k_a*, *k_d*: kinetic association and dissociation rate constants, *K_D*: (*k_d/k_a*): equilibrium dissociation constant calculated from the ratio *k_d/k_a*, *K_D* (eq.): equilibrium dissociation constant calculated from steady-state binding responses, +/-: average and +/- standard deviation of *n* = 1. (TIF 1425 kb)

Additional file 9: Figure S7. *SKA2* minigene. **A.** *SKA2* minigenes containing the wild-type sequence or mutated hnRNP A1 binding sites downstream of the exon 3 5' splice site were transfected into HeLa cells. **B.** Inclusion levels were quantified using the fragment analyzer. *PSI* percent spliced in. *n* = 2. **C.** *SKA2* SSO transfections improve splicing of the *SKA2* exon 3 wild-type minigene. Quantified using the fragment analyzer. *PSI* percent spliced in. *n* = 2. **D.** Increasing concentrations of *SKA2* SSO improves *SKA2* exon 3 inclusion. Inclusion ratios were quantified using the fragment analyzer. For comparison the effect of 40 nM SSO is shown (as in Fig. 4). (TIF 1295 kb)

Additional file 10: Figure S8. Mutations in the hnRNP A1 binding motifs downstream of the *MTRR* pseudoexon reduce hnRNP A1 binding affinity. The first UAG motif is responsible for hnRNP A1 binding. Surface plasmon resonance imaging using RNA oligonucleotides containing the *MTRR* wild-type or mutant sequences downstream of the 5' splice site. *Left:* Increasing concentrations of recombinant hnRNP A1 were injected over the surface to monitor concentration-dependent association. *Right:* Binding of hnRNP A1 at equilibrium. No binding was observed to the *MTRR* MUT C and MUT D oligonucleotides, suggesting that the proximal UAG motif is necessary for hnRNP A1 binding. *k_a*, *k_d*: kinetic association and dissociation rate constants, *K_D*: (*k_d/k_a*): equilibrium dissociation constant calculated from the ratio *k_d/k_a*, *K_D* (eq.): equilibrium dissociation constant calculated from steady-state binding responses, +/-: average and +/- standard deviation of *n* = 2–4. (TIF 3125 kb)

Additional file 11: Figure S9. hnRNP A1 binds downstream of the pseudoexon in *COL4A5*. **A.** Western blot with hnRNP A1 or as control hnRNP H antibody of proteins purified by RNA-affinity chromatography of biotin-conjugated RNA oligonucleotides covering the downstream region of the *COL4A5* pseudoexon 5' splice site. Disruption of the hnRNP A1 binding motif reduces hnRNP A1 binding. The motifs are scored using our generated scoring matrix (Additional file 1: Figure S1). Representative of two experiments. **B.** Surface plasmon resonance imaging using RNA oligonucleotides containing the *COL4A5* sequence downstream of the 5' splice site or mutant sequences (*MUT*) where the hnRNP A1 binding motifs were mutated. *Left:* Increasing concentrations of recombinant hnRNP A1 were injected over the surface to monitor concentration-dependent association. *Right:* Binding of hnRNP A1 at equilibrium. No binding was observed to the *COL4A5* MUT oligonucleotides, suggesting that hnRNP A1 binds here in a sequence-dependent way. *k_a*, *k_d*: kinetic association and dissociation rate constants, *K_D*: (*k_d/k_a*): equilibrium dissociation constant calculated from the ratio *k_d/k_a*, *K_D* (eq.): equilibrium dissociation constant calculated from steady-state binding responses, +/-: average and +/- standard deviation of *n* = 2. (TIF 1448 kb)

Additional file 12: Figure S10. hnRNP A1 binding peak near the *COL4A5* pseudoexon. **A.** RT-PCR analysis shows that knockdown of hnRNP A1 does not increase *COL4A5* pseudoexon inclusion. **B.** hnRNP A1 binding peak (*blue bar*) near the 5' splice site of the *COL4A5* pseudoexon. The SSO binding site is shown. **C.** *COL4A5* SSO transfections in HeLa and HEK293 cells decrease pseudoexon inclusion. (TIF 1448 kb)

Additional file 13: Figure S11. Increasing the concentration of the *MTRR* SSO improves *MTRR* pseudoexon inclusion. Transfection of increasing concentrations of *MTRR* SSO into HeLa cells improves *MTRR*

pseudoexon inclusion. Quantification of the *MTRR* pseudoexon inclusion level after MTRR SSO treatment was based on data from the fragment analyzer. For comparison, the effect of 40 nM SSO is shown (same as Fig. 5). (TIF 672 kb)

Additional file 14: Figure S12. Increased affinity of hnRNP A1 to the *ETFDH* mutant RNA oligonucleotide. Surface plasmon resonance imaging: RNA oligonucleotides containing the *ETFDH* wild-type or mutant sequences where the important UAG is mutated to UCG were immobilized in array format on a hydrogel-coated gold surface. *Left:* Increasing concentrations of hnRNP A1 were injected over the surface to monitor concentration-dependent association. *Right:* Binding of hnRNP A1 at equilibrium. k_a , k_d : kinetic association and dissociation rate constants, K_D : (k_d/k_a): equilibrium dissociation constant calculated from the ratio k_d/k_a , K_D (eq.): equilibrium dissociation constant calculated from steady-state binding responses, +/-: average and +/- standard deviation of $n = 4$. (TIF 1621 kb)

Additional file 15: Figure S13. hnRNP A1 binds to hnRNP A2/B1 and hnRNP A1. hnRNP A1 iCLIP binding peaks were abundant in the pre-mRNAs encoding hnRNP proteins, in particular in the 3' UTR region. **A.** hnRNP A1 iCLIP reads across the *HNRNP A2/B1* gene. **B.** hnRNP A1 iCLIP reads around the alternative exon 7B in *HNRNP A1*. (TIF 3350 kb)

Additional file 16: Materials and methods. Lists of RNA oligonucleotide sequences, splice-switching oligonucleotide sequences, and primer sequences. Description of the surface plasmon resonance imaging. (DOCX 22 kb)

Acknowledgements

We thank Ronni Nielsen, Tine Christiansen, Aleksandra Kulus, Tanja Bruun, Kristian Traantoft Rasmussen, and Jane Serup Pedersen for expert technical assistance. The authors have no conflict of interest to declare.

Funding

Funding was provided by Sundhed og Sygdom, Det Frie Forskningsråd (11-107174); Natur og Univers, Det Frie Forskningsråd (4181-00515); and Novo Nordisk Fonden (DK) (61310-0128 and 15430) to Professor Brage S. Andresen. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

Supporting data can be found in Additional file 4. All sequencing data have been deposited in the ArrayExpress database (<https://www.ebi.ac.uk/arrayexpress/>) under accession number E-MTAB-3612.

Authors' contributions

GHB and BSA designed the study. GHB performed most of the experiments and analyzed the results. TKD analyzed all the RNA sequencing and iCLIP data. JB-J helped perform and analyze the SPRi experiments. AM and KO helped with the iCLIP experiment, and ARK provided the HeLa cells with inducible expression of T7-tagged hnRNP A1. GHB and BSA wrote the manuscript implementing comments and edits from all authors. All authors read and approved the final manuscript.

Competing interest

The authors declare that they have no competing interests.

Author details

¹Department of Biochemistry and Molecular Biology and The Villum Center for Bioanalytical Sciences, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. ²Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan. ³Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA.

Received: 18 January 2016 Accepted: 23 June 2016

Published online: 05 July 2016

References

1. Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. Using positional distribution to identify splicing elements and predict pre-mRNA processing

- defects in human genes. *Proc Natl Acad Sci U S A*. 2011;108(27):11093–8. doi:10.1073/pnas.1101135108. PubMed PMID: ISI:000292376700038.
- Sterne-Weiler T, Howard J, Mort M, Cooper DN, Sanford JR. Loss of exon identity is a common mechanism of human inherited disease. *Genome Res*. 2011;21(10):1563–71. doi:10.1101/gr.118638. PubMed PMID: ISI:000295407800001.
- Lorson CL, Hahnen E, Androphy EJ, Wirth B. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc Natl Acad Sci U S A*. 1999;96(11):6307–11. PubMed PMID: 10339583, PubMed Central PMCID: PMC26877, Epub 1999/05/26.
- Cartegni L, Hastings ML, Calarco JA, de Stanchina E, Krainer AR. Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. *Am J Hum Genet*. 2006;78(1):63–77. doi:10.1086/498853. PubMed PMID: 16385450, PubMed Central PMCID: PMC1380224, Epub 2005/12/31.
- Kashima T, Rao N, David CJ, Manley JL. hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. *Hum Mol Genet*. 2007;16(24):3149–59. doi:10.1093/hmg/ddm276. Epub 2007/09/22.
- Singh NN, Androphy EJ, Singh RN. An extended inhibitory context causes skipping of exon 7 of SMN2 in spinal muscular atrophy. *Biochem Biophys Res Commun*. 2004;315(2):381–8. doi:10.1016/j.bbrc.2004.01.067. Epub 2004/02/10.
- Kashima T, Manley JL. A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat Genet*. 2003;34(4):460–3. doi:10.1038/ng1207. Epub 2003/07/02.
- Nielsen KB, Sorensen S, Cartegni L, Corydon TJ, Doktor TK, Schroeder LD, et al. Seemingly neutral polymorphic variants may confer immunity to splicing-inactivating mutations: a synonymous SNP in exon 5 of MCAD protects from deleterious mutations in a flanking exonic splicing enhancer. *Am J Hum Genet*. 2007;80(3):416–32. doi:10.1086/511992. PubMed PMID: ISI:000244403300003.
- Du L, Gatti RA. Progress toward therapy with antisense-mediated splicing modulation. *Curr Opin Mol Ther*. 2009;11(2):116–23. PubMed PMID: 19330717, PubMed Central PMCID: PMC2753608, Epub 2009/03/31.
- Veltrop M, Aartsma-Rus A. Antisense-mediated exon skipping: taking advantage of a trick from Mother Nature to treat rare genetic diseases. *Exp Cell Res*. 2014;325(1):50–5. doi:10.1016/j.yexcr.2014.01.026. Epub 2014/02/04.
- Hua Y, Vickers TA, Okunola HL, Bennett CF, Krainer AR. Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *Am J Hum Genet*. 2008;82(4):834–48. doi:10.1016/j.ajhg.2008.01.014. PubMed PMID: 18371932, PubMed Central PMCID: PMC2427210, Epub 2008/03/29.
- Singh NK, Singh NN, Androphy EJ, Singh RN. Splicing of a critical exon of human Survival Motor Neuron is regulated by a unique silencer element located in the last intron. *Mol Cell Biol*. 2006;26(4):1333–46. doi:10.1128/MCB.26.4.1333-1346.2006. PubMed PMID: 16449646, PubMed Central PMCID: PMC1367187, Epub 2006/02/02.
- Singh NN, Shishimorova M, Cao LC, Gangwani L, Singh RN. A short antisense oligonucleotide masking a unique intronic motif prevents skipping of a critical exon in spinal muscular atrophy. *RNA Biol*. 2009;6(3):341–50. PubMed PMID: 19430205, PubMed Central PMCID: PMC2734876, Epub 2009/05/12.
- Rigo F, Hua Y, Krainer AR, Bennett CF. Antisense-based therapy for the treatment of spinal muscular atrophy. *J Cell Biol*. 2012;199(1):21–5. doi:10.1083/jcb.201207087. PubMed PMID: 23027901, PubMed Central PMCID: PMC3461520, Epub 2012/10/03.
- Fu XD, Ares Jr M. Context-dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews Genet*. 2014. doi:10.1038/nrg3778. Epub 2014/08/13.
- Anko ML, Muller-McNicoll M, Brandl H, Curk T, Gorup C, Henry I, et al. The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol*. 2012;13(3):R17. doi:10.1186/gb-2012-13-3-r17. PubMed PMID: 22436691, PubMed Central PMCID: PMC3439968, Epub 2012/03/23.
- Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, et al. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep*. 2012;1(2):167–78. doi:10.1016/j.celrep.2012.02.001. PubMed PMID: 22574288, PubMed Central PMCID: PMC3345519, Epub 2012/05/11.
- Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, et al. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol*. 2010;17(7):909–15. doi:10.1038/nsmb.1838. PubMed PMID: 20601959, PubMed Central PMCID: PMC3000544, Epub 2010/07/06.

19. Masuda A, Andersen HS, Doktor TK, Okamoto T, Ito M, Andresen BS, et al. CUGBP1 and MBNL1 preferentially bind to 3' UTRs and facilitate mRNA decay. *Sci Rep.* 2012;2:209. doi:10.1038/srep00209. PubMed PMID: 22355723, PubMed Central PMCID: PMC3250574, Epub 2012/02/23.
20. Pandit S, Zhou Y, Shiue L, Coutinho-Mansfield G, Li H, Qiu J, et al. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol Cell.* 2013;50(2):223–35. doi:10.1016/j.molcel.2013.03.001. PubMed PMID: 23562324, PubMed Central PMCID: PMC3640356, Epub 2013/04/09.
21. Roszbach O, Hung LH, Khrameeva E, Schreiner S, Konig J, Curk T, et al. Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. *RNA Biol.* 2014;11(2):146–55. doi:10.4161/ma.27991. PubMed PMID: 24526010, PubMed Central PMCID: PMC2661799, Epub 2014/02/15.
22. Sanford JR, Wang X, Mort M, Vanduy N, Cooper DN, Mooney SD, et al. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* 2009;19(3):381–94. doi:10.1101/gr.082503.108. PubMed PMID: 19116412, PubMed Central PMCID: PMC2661799, Epub 2009/01/01.
23. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP identifies Nova-regulated RNA networks in the brain. *Science.* 2003;302(5648):1212–5. doi:10.1126/science.1090095. Epub 2003/11/15.
24. Weyn-Vanhenyck SM, Mele A, Yan Q, Sun S, Farny N, Zhang Z, et al. HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep.* 2014;6(6):1139–52. doi:10.1016/j.celrep.2014.02.005. PubMed PMID: 24613350, PubMed Central PMCID: PMC3992522, Epub 2014/03/13.
25. Burd CG, Dreyfuss G. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* 1994;13(5):1197–204. PubMed PMID: 7510636, PubMed Central PMCID: PMC394929, Epub 1994/03/01.
26. Jean-Philippe J, Paz S, Caputi M. hnRNP A1: the Swiss army knife of gene expression. *Int J Mol Sci.* 2013;14(9):18999–9024. doi:10.3390/ijms140918999. PubMed PMID: 24065100, PubMed Central PMCID: PMC3794818, Epub 2013/09/26.
27. Clery A, Sinha R, Anczukow O, Corriero A, Moursy A, Daubner GM, et al. Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc Natl Acad Sci U S A.* 2013;110(30):E2802–11. doi:10.1073/pnas.1303445110. PubMed PMID: 23836656, PubMed Central PMCID: PMC3725064, Epub 2013/07/10.
28. Eperon IC, Makarova OV, Mayeda A, Munroe SH, Caceres JF, Hayward DG, et al. Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol.* 2000;20(22):8303–18. PubMed PMID: 11046128, PubMed Central PMCID: PMC102138, Epub 2000/10/25.
29. Bonomi S, di Matteo A, Buratti E, Cabianca DS, Baralle FE, Ghigna C, et al. HnRNP A1 controls a splicing regulatory circuit promoting mesenchymal-to-epithelial transition. *Nucleic Acids Res.* 2013;41(18):8665–79. doi:10.1093/nar/gkt579. PubMed PMID: 23863836, PubMed Central PMCID: PMC3794575, Epub 2013/07/19.
30. Tange TO, Damgaard CK, Guth S, Valcarcel J, Kjems J. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J.* 2001;20(20):5748–58. doi:10.1093/emboj/20.20.5748. PubMed PMID: 11598017, PubMed Central PMCID: PMC125679, Epub 2001/10/13.
31. Okunola HL, Krainer AR. Cooperative-binding and splicing-repressive properties of hnRNP A1. *Mol Cell Biol.* 2009;29(20):5620–31. doi:10.1128/MCB.01678-08. PubMed PMID: 19667073, PubMed Central PMCID: PMC2756886, Epub 2009/08/12.
32. Nasim FU, Hutchison S, Cordeau M, Chabot B. High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. *RNA.* 2002;8(8):1078–89. PubMed PMID: 12212851, PubMed Central PMCID: PMC1370318, Epub 2002/09/06.
33. Blanchette M, Chabot B. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* 1999;18(7):1939–52. doi:10.1093/emboj/18.7.1939. PubMed PMID: 10202157, PubMed Central PMCID: PMC1171279, Epub 1999/04/15.
34. Martinez-Contreras R, Fiset JF, Nasim FU, Madden R, Cordeau M, Chabot B. Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol.* 2006;4(2):e21. doi:10.1371/journal.pbio.0040021. PubMed PMID: 16396608, PubMed Central PMCID: PMC1326234, Epub 2006/01/07.
35. Boukakis G, Patrinoiu-Georgoula M, Lekaraku M, Valavanis C, Gualias A. Deregulated expression of hnRNP A/B proteins in human non-small cell lung cancer: parallel assessment of protein and mRNA levels in paired tumour/non-tumour tissues. *BMC Cancer.* 2010;10:434. doi:10.1186/1471-2407-10-434. PubMed PMID: 20716340, PubMed Central PMCID: PMC2933625, Epub 2010/08/19.
36. Ushigome M, Ubagai T, Fukuda H, Tsuchiya N, Sugimura T, Takatsuka J, et al. Up-regulation of hnRNP A1 gene in sporadic human colorectal cancers. *Int J Oncol.* 2005;26(3):635–40. Epub 2005/02/11.
37. Pino I, Pio R, Toledo G, Zabalegui N, Vicent S, Rey N, et al. Altered patterns of expression of members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family in lung cancer. *Lung Cancer.* 2003;41(2):131–43. Epub 2003/07/23.
38. Patry C, Bouchard L, Labrecque P, Gendron D, Lemieux B, Toutant J, et al. Small interfering RNA-mediated reduction in heterogeneous nuclear ribonucleoprotein A1/A2 proteins induces apoptosis in human cancer cells but not in normal mortal cell lines. *Cancer Res.* 2003;63(22):7679–88. Epub 2003/11/25.
39. Berson A, Barbash S, Shaltiel G, Goll Y, Hanin G, Greenberg DS, et al. Cholinergic-associated loss of hnRNP-A/B in Alzheimer's disease impairs cortical splicing and cognitive function in mice. *EMBO Mol Med.* 2012;4(8):730–42. doi:10.1002/emmm.201100995. PubMed PMID: 22628224, PubMed Central PMCID: PMC3494073, Epub 2012/05/26.
40. Olsen RK, Broner S, Sabaratnam R, Doktor TK, Andersen HS, Bruun GH, et al. The ETFDH c.158A > G variation disrupts the balanced interplay of ESE- and ESS-binding proteins thereby causing missplicing and multiple Acyl-CoA dehydrogenation deficiency. *Hum Mutat.* 2014;35(1):86–95. doi:10.1002/humu.22455. Epub 2013/10/15.
41. Dobrowolski SF, Andersen HS, Doktor TK, Andresen BS. The phenylalanine hydroxylase c.30C > G synonymous variation (p.G10G) creates a common exonic splicing silencer. *Mol Genet Metab.* 2010;100(4):316–23. doi:10.1016/j.ymgme.2010.04.002. Epub 2010/05/12.
42. Yu CY, Theusch E, Lo K, Mangravite LM, Naidoo D, Kutilova M, et al. HNRNPA1 regulates HMGCR alternative splicing and modulates cellular cholesterol metabolism. *Hum Mol Genet.* 2014;23(2):319–32. doi:10.1093/hmg/ddt422. PubMed PMID: 24001602, PubMed Central PMCID: PMC3869353, Epub 2013/09/05.
43. Bruun GH, Doktor TK, Andresen BS. A synonymous polymorphic variation in ACAD10 exon 11 affects splicing efficiency and may affect fatty acid oxidation. *Mol Genet Metab.* 2013;110(1–2):122–8. doi:10.1016/j.ymgme.2013.06.005. Epub 2013/07/03.
44. Huppertz I, Attig J, D'Ambrogio A, Easton LE, Sibley CR, Sugimoto Y, et al. iCLIP: protein-RNA interactions at nucleotide resolution. *Methods.* 2014; 65(3):274–87. doi:10.1016/j.jymeth.2013.10.011. PubMed PMID: 24184352, PubMed Central PMCID: PMC3988997, Epub 2013/11/05.
45. Caceres JF, Misteli T, Sreter GR, Spector DL, Krainer AR. Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol.* 1997;138(2):225–38. PubMed PMID: 9230067, PubMed Central PMCID: PMC2138183, Epub 1997/07/28.
46. Allemand E, Guil S, Myers M, Moscat J, Caceres JF, Krainer AR. Regulation of heterogeneous nuclear ribonucleoprotein A1 transport by phosphorylation in cells stressed by osmotic shock. *Proc Natl Acad Sci U S A.* 2005;102(10):3605–10. doi:10.1073/pnas.0409889102. PubMed PMID: 15738418, PubMed Central PMCID: PMC553333, Epub 2005/03/02.
47. Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, et al. iCLIP—transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J Vis Exp.* 2011;(50). Epub 2011/05/12. doi: 10.3791/2638. PubMed PMID: 21559008; PubMed Central PMCID: PMC3169244
48. Zhang C, Darnell RB. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol.* 2011;29(7):607–14. doi:10.1038/nbt.1873. PubMed PMID: 21633356, PubMed Central PMCID: PMC3400429, Epub 2011/06/03.
49. Granneman S, Kudla G, Petfalski E, Tollervey D. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci U S A.* 2009;106(24):9613–8. doi:10.1073/pnas.0901997106. PubMed PMID: 19482942, PubMed Central PMCID: PMC2688437, Epub 2009/06/02.
50. Ule J, Jensen K, Mele A, Darnell RB. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods.* 2005;37(4):376–86. doi:10.1016/j.jymeth.2005.07.018. Epub 2005/11/30.
51. Hauer C, Curk T, Anders S, Schwarzl T, Alleaume AM, Sieber J, et al. Improved binding site assignment by high-resolution mapping of RNA-

- protein interactions using iCLIP. *Nat Commun.* 2015;6:7921. doi:10.1038/ncomms8921. Epub 2015/08/12.
52. Huang M, Rech JE, Northington SJ, Flicker PF, Mayeda A, Krainer AR, et al. The C-protein tetramer binds 230 to 240 nucleotides of pre-mRNA and nucleates the assembly of 40S heterogeneous nuclear ribonucleoprotein particles. *Mol Cell Biol.* 1994;14(1):518–33. PubMed PMID: 8264621, PubMed Central PMCID: PMC358402, Epub 1994/01/01.
 53. Palhais B, Praestegaard VS, Sabaratnam R, Doktor TK, Lutz S, Burda P, et al. Splice-shifting oligonucleotide (SSO) mediated blocking of an exonic splicing enhancer (ESE) created by the prevalent c.903+469T>C MTRR mutation corrects splicing and restores enzyme activity in patient cells. *Nucleic Acids Res.* 2015;43:4627–39. doi:10.1093/nar/gkv275. Epub 2015/04/17.
 54. Sorek R, Ast G. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* 2003;13(7):1631–7. doi:10.1101/gr.1208803. PubMed PMID: 12840041, PubMed Central PMCID: PMC403736, Epub 2003/07/04.
 55. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol.* 2004;11(2-3):377–94. doi:10.1089/1066527041410418.
 56. Chen M, David CJ, Manley JL. Concentration-dependent control of pyruvate kinase M mutually exclusive splicing by hnRNP proteins. *Nat Struct Mol Biol.* 2012;19(3):346–54. doi:10.1038/nsmb.2219. PubMed PMID: 22307054, PubMed Central PMCID: PMC3698866, Epub 2012/02/07.
 57. David CJ, Chen M, Assanah M, Canoll P, Manley JL. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature.* 2010;463(7279):364–8. doi:10.1038/nature08697. PubMed PMID: 20010808, PubMed Central PMCID: PMC2950088, Epub 2009/12/17.
 58. Clower CV, Chatterjee D, Wang Z, Cantley LC, Vander Heiden MG, Krainer AR. The alternative splicing repressors hnRNP A1/A2 and PTB influence pyruvate kinase isoform expression and cell metabolism. *Proc Natl Acad Sci U S A.* 2010;107(5):1894–9. doi:10.1073/pnas.0914845107. PubMed PMID: 20133837, PubMed Central PMCID: PMC2838216, Epub 2010/02/06.
 59. Wang Z, Chatterjee D, Jeon HY, Akerman M, Vander Heiden MG, Cantley LC, et al. Exon-centric regulation of pyruvate kinase M alternative splicing via mutually exclusive exons. *J Mol Cell Biol.* 2012;4(2):79–87. doi:10.1093/jmcb/mjr030. PubMed PMID: 22044881, PubMed Central PMCID: PMC3493165, Epub 2011/11/03.
 60. Wang Z, Jeon HY, Rigo F, Bennett CF, Krainer AR. Manipulation of PK-M mutually exclusive alternative splicing by antisense oligonucleotides. *Open Biol.* 2012;2(10):120133. doi:10.1098/rsob.120133. PubMed PMID: 23155487, PubMed Central PMCID: PMC3498831, Epub 2012/11/17.
 61. Kashima T, Rao N, Manley JL. An intronic element contributes to splicing repression in spinal muscular atrophy. *Proc Natl Acad Sci U S A.* 2007;104(9):3426–31. doi:10.1073/pnas.0700343104. PubMed PMID: 17307868, PubMed Central PMCID: PMC1805620, Epub 2007/02/20.
 62. Cartegni L, Krainer AR. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat Genet.* 2002;30(4):377–84. doi:10.1038/ng854. Epub 2002/04/02.
 63. Doktor TK, Schroeder LD, Vested A, Palmfeldt J, Andersen HS, Gregersen N, et al. SMN2 exon 7 splicing is inhibited by binding of hnRNP A1 to a common ESS motif that spans the 3' splice site. *Hum Mutat.* 2011;32(2):220–30. doi:10.1002/humu.21419. Epub 2010/12/02.
 64. Jayaprakash AA, Santamaria A, Jayachandran U, Chan YW, Benda C, Nigg EA, et al. Structural and functional organization of the Ska complex, a key component of the kinetochore-microtubule interface. *Mol Cell.* 2012;46(3):274–86. doi:10.1016/j.molcel.2012.03.005. Epub 2012/04/10.
 65. Wang Y, Zhang Y, Zhang C, Weng H, Li Y, Cai W, et al. The gene pair PRR11 and SKA2 shares a NF-Y-regulated bidirectional promoter and contributes to lung cancer development. *Biochim Biophys Acta.* 2015;1849(9):1133–44. doi:10.1016/j.bbagg.2015.07.002. Epub 2015/07/15.
 66. Ishii S, Nakao S, Minamikawa-Tachino R, Desnick RJ, Fan JQ. Alternative splicing in the alpha-galactosidase A gene: increased exon inclusion results in the Fabry cardiac phenotype. *Am J Hum Genet.* 2002;70(4):994–1002. doi:10.1086/339431. PubMed PMID: 11828341, PubMed Central PMCID: PMC379133, Epub 2002/02/06.
 67. Pagani F, Buratti E, Stuani C, Bendix R, Dork T, Baralle FE. A new type of mutation causes a splicing defect in ATM. *Nat Genet.* 2002;30(4):426–9. doi:10.1038/ng858. Epub 2002/03/13.
 68. King K, Flinter FA, Nihalani V, Green PM. Unusual deep intronic mutations in the COL4A5 gene cause X linked Alport syndrome. *Hum Genet.* 2002;111(6):548–54. doi:10.1007/s00439-002-0830-3. Epub 2002/11/19.
 69. Rincon A, Aguado C, Desviat LR, Sanchez-Alcudia R, Ugarte M, Perez B. Propionic and methylmalonic acidemia: antisense therapeutics for intronic variations causing aberrantly spliced messenger RNA. *Am J Hum Genet.* 2007;81(6):1262–70. doi:10.1086/522376. PubMed PMID: 17966092, PubMed Central PMCID: PMC2276355, Epub 2007/10/30.
 70. Faa V, Incani F, Meloni A, Corda D, Masala M, Baffico AM, et al. Characterization of a disease-associated mutation affecting a putative splicing regulatory element in intron 6b of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. *J Biol Chem.* 2009;284(44):30024–31. doi:10.1074/jbc.M109.032623. PubMed PMID: 19759008, PubMed Central PMCID: PMC2781556, Epub 2009/09/18.
 71. Davis RL, Homer VM, George PM, Brennan SO. A deep intronic mutation in FGB creates a consensus exonic splicing enhancer motif that results in afibrinogenemia caused by aberrant mRNA splicing, which can be corrected in vitro with antisense oligonucleotide treatment. *Hum Mutat.* 2009;30(2):221–7. doi:10.1002/humu.20839. Epub 2008/10/15.
 72. Homolova K, Zavadakova P, Doktor TK, Schroeder LD, Kozich V, Andresen BS. The deep intronic c.903 + 469 T > C mutation in the MTRR gene creates an SF2/ASF binding exonic splicing enhancer, which leads to pseudoexon activation and causes the cblE type of homocystinuria. *Hum Mutat.* 2010;31(4):437–44. doi:10.1002/humu.21206. PubMed PMID: 20120036, PubMed Central PMCID: PMC3429857, Epub 2010/02/02.
 73. Sterne-Weiler T, Sanford JR. Exon identity crisis: disease-causing mutations that disrupt the splicing code. *Genome Biol.* 2014;15(1):201. doi:10.1186/gb4150. PubMed PMID: 24456648, PubMed Central PMCID: PMC4053859, Epub 2014/01/25.
 74. Guo R, Li Y, Ning J, Sun D, Lin L, Liu X. HnRNP A1/A2 and SF2/ASF regulate alternative splicing of interferon regulatory factor-3 and affect immunomodulatory functions in human non-small cell lung cancer cells. *PLoS One.* 2013;8(4):e62729. doi:10.1371/journal.pone.0062729. PubMed PMID: 23658645, PubMed Central PMCID: PMC3639176, Epub 2013/05/10.
 75. Staffa A, Cochrane A. Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. *Mol Cell Biol.* 1995;15(8):4597–605. PubMed PMID: 7623851, PubMed Central PMCID: PMC230700, Epub 1995/08/01.
 76. Zhu J, Mayeda A, Krainer AR. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell.* 2001;8(6):1351–61. Epub 2002/01/10.
 77. Mayeda A, Krainer AR. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell.* 1992;68(2):365–75. Epub 1992/01/24.
 78. Sun H, Chasin LA. Multiple splicing defects in an intronic false exon. *Mol Cell Biol.* 2000;20(17):6414–25. PubMed PMID: 10938119, PubMed Central PMCID: PMC86117, Epub 2000/08/11.
 79. Mayeda A, Munroe SH, Caceres JF, Krainer AR. Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J.* 1994;13(22):5483–95. PubMed PMID: 7957114, PubMed Central PMCID: PMC395506, Epub 1994/11/15.
 80. Chabot B, Blanchette M, Lapierre I, La Branche H. An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol Cell Biol.* 1997;17(4):1776–86. PubMed PMID: 9121425, PubMed Central PMCID: PMC232024, Epub 1997/04/01.
 81. Hutchison S, LeBel C, Blanchette M, Chabot B. Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5' splice site selection in the hnRNP A1 mRNA precursor. *J Biol Chem.* 2002;277(33):29745–52. doi:10.1074/jbc.M203633200. Epub 2002/06/13.
 82. Luan W, Wang Y, Chen X, Shi Y, Wang J, Zhang J, et al. PKM2 promotes glucose metabolism and cell growth in gliomas through a mechanism involving a let-7a/c-Myc/hnRNP1 feedback loop. *Oncotarget.* 2015;6(15):13006–18. doi:10.18632/oncotarget.3514. Epub 2015/05/08. PubMed PMID: 25948776; PubMed Central PMCID: PMC4536995.
 83. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60. doi:10.1093/bioinformatics/btp324. PubMed PMID: 19451168, PubMed Central PMCID: PMC2705234, Epub 2009/05/20.
 84. Lovci MT, Ghanem D, Marr H, Arnold J, Gee S, Parra M, et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat Struct Mol Biol.* 2013;20(12):1434–42. doi:10.1038/nsmb.2699. PubMed PMID: 24213538, PubMed Central PMCID: PMC3918504, Epub 2013/11/12.
 85. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory

- elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576–89. doi:10.1016/j.molcel.2010.05.004. PubMed PMID: 20513432, PubMed Central PMCID: PMC2898526, Epub 2010/06/02.
86. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*. 2012;28(16):2184–5. doi:10.1093/bioinformatics/bts356. Epub 2012/06/30.
87. Shen L, Shao N, Liu X, Nestler E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics*. 2014;15:284. doi:10.1186/1471-2164-15-284. PubMed PMID: 24735413, PubMed Central PMCID: PMC4028082, Epub 2014/04/17.
88. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24–6. doi:10.1038/nbt.1754. PubMed PMID: 21221095, PubMed Central PMCID: PMC3346182, Epub 2011/01/12.
89. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178–92. doi:10.1093/bib/bbs017. PubMed PMID: 22517427, PubMed Central PMCID: PMC3603213, Epub 2012/04/21.
90. Loh YH, Shen L. Analysis and visualization of ChIP-Seq and RNA-Seq sequence alignments using ngs.plot. *Methods Mol Biol*. 2016;1415:371–83. doi:10.1007/978-1-4939-3572-7_18.
91. Rasband WS. ImageJ U.S. National Institutes of Health, Bethesda, Maryland, USA 1997–2015. <http://imagej.nih.gov/ij/>. Accessed 2 Jul 2016.
92. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21. doi:10.1093/bioinformatics/bts635. PubMed PMID: 23104886, PubMed Central PMCID: PMC3530905.
93. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166–9. doi:10.1093/bioinformatics/btu638. PubMed PMID: 25260700, PubMed Central PMCID: PMC4287950.
94. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi:10.1186/PREACCEPT-8897612761307401. Epub 2014/12/18.
95. Hansen KD, Irizarry RA, Wu Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics*. 2012;13(2):204–16. doi:10.1093/biostatistics/kxr054. PubMed PMID: 22285995, PubMed Central PMCID: PMC3297825.
96. Mazin P, Xiong J, Liu X, Yan Z, Zhang X, Li M, et al. Widespread splicing changes in human brain development and aging. *Mol Syst Biol*. 2013;9:633. doi:10.1038/msb.2012.67. PubMed PMID: 23340839; PubMed Central PMCID: PMC3564255.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

