Global analysis reveals SRp20- and SRp75-specific mRNPs in cycling and neural cells

Minna-Liisa Änkö[1], Lucia Morales[1], Ian Henry[1], Andreas Beyer[2] & Karla M Neugebauer[1]

Members of the SR protein family of RNA-binding proteins have numerous roles in mRNA metabolism, from transcription to translation. To understand how SR proteins coordinate gene regulation, comprehensive knowledge of endogenous mRNA targets is needed. Here we establish physiological expression of GFP-tagged SR proteins from stable transgenes. Using the GFP tag for immunopurification of mRNPs, mRNA targets of SRp20 and SRp75 were identified in cycling and neurally induced P19 cells. Genome-wide analysis showed that SRp20 and SRp75 associate with hundreds of distinct, functionally related groups of transcripts that change in response to neural differentiation. Knockdown of either SRp20 or SRp75 led to up- or downregulation of specific transcripts, including identified targets, and rescue by the GFP-tagged SR proteins proved their functionality. Thus, SR proteins contribute to the execution of gene-expression programs through their association with distinct endogenous mRNAs.

SR proteins are a family of at least seven canonical RNA binding proteins (SF2, SC35, 9G8, SRp20, SRp40, SRp55 and SRp75). They share a common domain structure comprising one or two RNA recognition motifs (RRMs) at their N termini and an arginine/serine-rich (RS) domain at their C termini[1–3]. Both domains can directly contact RNA[4], although the RRM appears to determine the binding specificity of SR proteins[5–7]. The RS domain can also mediate protein-protein interactions[8] and is a target of kinases that modulate SR protein activities in response to signaling[9,10]. SR proteins are expressed at varying levels from tissue to tissue[11,12], suggesting important roles in the development and maintenance of cell lineages. Because depletion of individual SR proteins leads to the death of cells or organisms[3], each SR protein appears to be essential. Consistent with this, SR proteins are conserved among metazoans, yet the number and identities of transcripts associated with each SR protein have so far remained mysterious.

The best-understood function of SR proteins is in regulating constitutive and alternative pre-mRNA splicing[1,3,13–15]. It is currently thought that SR proteins primarily stimulate splice-site selection through site-specific binding on pre-mRNA, which leads to the stable association of spliceosome components with nearby 5′ and 3′ splice site. However, recent studies have implicated SR proteins in additional functions, including transcription, genomic stability, mRNA export and translation[2]. Indeed, SR proteins are present in RNPs after splicing is completed[5,16,17]. Moreover, all SR proteins except SC35 undergo nucleocytoplasmic shuttling[5,18], suggesting widespread roles for SR proteins in gene expression.

To broadly understand how SR proteins govern gene expression in cells, it is important to determine which genes are regulated by individual family members. However, only few endogenous genes regulated by individual SR proteins have been identified[16,19–21], making

it difficult to dissect the specific steps in gene expression that require any given SR protein. Binding sequences for some SR proteins have been determined using SELEX and related *in vitro* methods[14,22]. The shortness (four to ten nucleotides) and degeneracy of the binding sites implies that a single factor can regulate a large number of genes, yet these properties make *in silico* prediction of transcripts regulated by any SR protein extremely difficult[23]. Recently, *in vivo* cross-linking was used to define mRNA targets of SF2; this study identified thousands of SF2 target sites, which resembled the SELEX sequences[24]. However, it is currently unclear whether binding target mRNAs depends only on short RNA sequences or whether other *trans*-acting factors or RNA secondary structures contribute to SR protein association with mRNAs. Moreover, SR protein target mRNAs may vary with cell type or in response to cell signaling.

Here we study SR protein expression and function in the context of neural differentiation using mouse P19 cells, which are diploid and multipotent. Upon treatment with retinoic acid (RA), P19 cells differentiate into neural cells via two major switches in gene expression, first from 16 h to 4 d and later at 6 d after the start of treatment[25,26]. We focused on SRp20 and SRp75 because their protein levels changed during differentiation. SRp75 is the largest member of the SR protein family, with two RRMs and a very long, highly phosphorylated RS domain[27]. Little is known about the functions or targets of SRp75 (refs. 28–31). SRp20 is the smallest member of the SR protein family, having only one RRM and an RS domain almost four times shorter than that of of SRp75. A few genes regulated by SRp20 have been previously identified *in vitro*[28,32–35]. Both SRp20 and SRp75 shuttle between the nucleus and cytoplasm[5,18]. We established P19 cell lines expressing GFP-tagged versions of SRp20 and SRp75 under physiological control by bacterial artificial chromosome (BAC)

recombineering[5,36] and systematically identify mRNPs specifically bound by SRp20 and SRp75. This global analysis reveals that SRp20 and SRp75 are present in distinct and largely nonoverlapping mRNPs in both cycling and differentiated cells. The detection of SRp20 and SRp75 in mature mRNPs indicates widespread roles for SR proteins in gene expression.

RESULTS

### Establishment of P19-SRp20-BAC and P19-SRp75-BAC cell lines

To study *in vivo* functions of individual SR proteins, we have previously established expression of GFP-tagged SR proteins under physiological control using the BAC recombineering[5,36]. Here we introduced the GFP-tagged BACs encoding human SRp20 and SRp75 (**Fig. 1a**) into mouse P19 cells. We chose the human versions of the SR proteins to allow for the distinction between the endogenous genes and the transgenes at the transcript level and to permit functional analysis of the transgenes (see below). The human and mouse homologs are expected to be functionally indistinguishable, as amino acid identity is 100% and 96% for SRp20 and SRp75, respectively (**Supplementary Fig. 1**). We isolated stable P19 clones (named P19-SRp20-BAC and P19-SRp75-BAC) expressing each SR protein from BACs and showed that they express each protein in a predominantly nucleoplasmic pattern, as expected (**Fig. 1b**). Upon RA treatment, parental and BAC-transgenic cells expressed neural differentiation markers, such as beta-III-tubulin (**Fig. 1b**) and NeuN (data not shown). This indicates that the expression of the BAC-encoded transgenes does not interfere with neural differentiation.

### Regulation of SR protein expression during neural induction

We investigated the temporal expression patterns of SR proteins during neural differentiation in all three cell lines (P19 WT, P19-SRp20-BAC and P19-SRp75-BAC). We performed western blotting using an antibody (mAb104) specific for common phosphoepitopes on all SR protein family members as well as anti-SRp20 (mAb7B4) and anti-GFP antibodies[5,37,38]. Characteristic patterns of all SR proteins were expressed in both undifferentiated and neural cells (**Fig. 1c**). During a time course of RA treatment, the expression of most SR proteins remained relatively constant. Notably, SRp20 and SRp75 reactivity

increased during neural induction, peaking at 4–6 d after the addition of RA. These results were validated with anti-SRp20, which does not depend on phosphorylation for detection (**Fig. 1c**). Although no antibody specific for SRp75 is available, reactivity of SRp75-GFP with phosphodependent mAb104 indicates that the tagged SR proteins are substrates for regulation by endogenous SR protein kinases. Notably, the expression of the GFP-tagged SR proteins was regulated in parallel with the endogenous proteins (**Fig. 1c**).

The recapitulation of endogenous expression patterns suggested that introduction of the SR protein–encoding transgenes may lead to homeostatic expression of each SR protein, owing to previously reported autoregulation at the level of splicing, stability and/or translation (refs. 39–41 and references therein). To investigate this, we used western blotting, quantitative RT-PCR (RT-qPCR) and microarray analysis to compare SRp20 and SRp75 protein and mRNA levels among the cell lines. **Figure 2a** shows that the endogenous SRp20 or SRp75 protein was reduced when the human SR protein homolog was expressed from the BAC. Total mRNA levels (endogenous plus transgene) of each SR protein measured by RT-qPCR in the transgenic cell lines were statistically identical to the parental cell line (**Fig. 2b**). These data indicate that homeostatic expression of SR proteins takes place when transgenes are introduced under the control of their own promoters and intron regulatory elements. Chromatin immuno-precipitation (ChIP) analysis showed the recruitment of SRp20 and SRp75 to their respective genes (**Supplementary Fig. 2**), suggesting a splicing-mediated mechanism of autoregulation. It was important for subsequent analysis that SR proteins were not overexpressed in the transgenic cell lines, as even small changes in cellular SR protein concentrations can produce changes in alternative splicing[19]. To address overall differences in gene expression among the three cell lines, we performed gene-expression microarray analysis (**Fig. 2c**). The correlation coefficients between parental (WT) and BAC transgenic cells were high and similar to the levels of correlation between individual replicates (**Supplementary Table 1**). These data imply that tagged SR
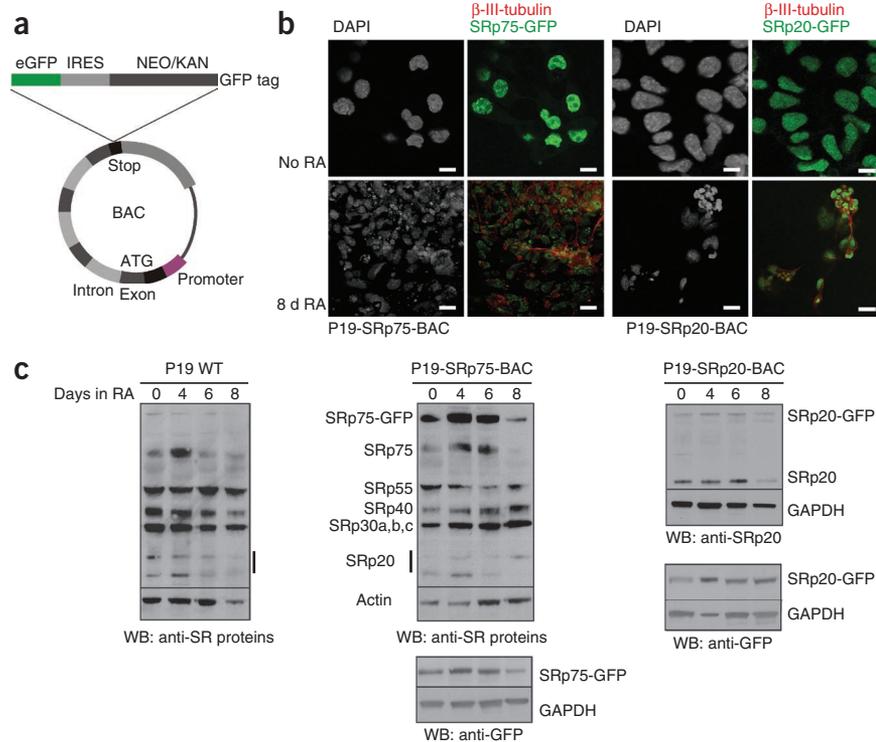
Figure 1 Expression of endogenous and GFP-tagged SRp20 and SRp75 in P19 cells during neural differentiation. (**a**) Schematic representation of the SR protein BAC-constructs used to establish stable P19 cell lines. IRES, internal ribosome entry site; NEO/KAN, neomycin/kanamycin resistance. (**b**) Fluorescent images showing the expression of the SRp75-GFP (left) and SRp20-GFP (right) in undifferentiated cells (No RA) and 8 d after retinoic acid (RA) induction. β-III-tubulin was used in immunostaining as a marker for neurites (red) and DAPI to mark the nuclei. Scale bar, 10 μm. (**c**) Western blot of P19 cell extracts at different stages of neural differentiation. mAb104 was used to detect all members of the SR protein family in parental P19 (WT) and P19-SRp75-BAC–expressing cells. mAb7B4 (anti-SRp20) antibody was used to detect SRp20 in P19-SRp20-BAC cells, and anti-GFP antibody was used to detect the GFP-tagged SR proteins in P19-SRp20-BAC and P19-SRp75-BAC cells. Anti-actin or anti-GAPDH antibody served as loading controls.
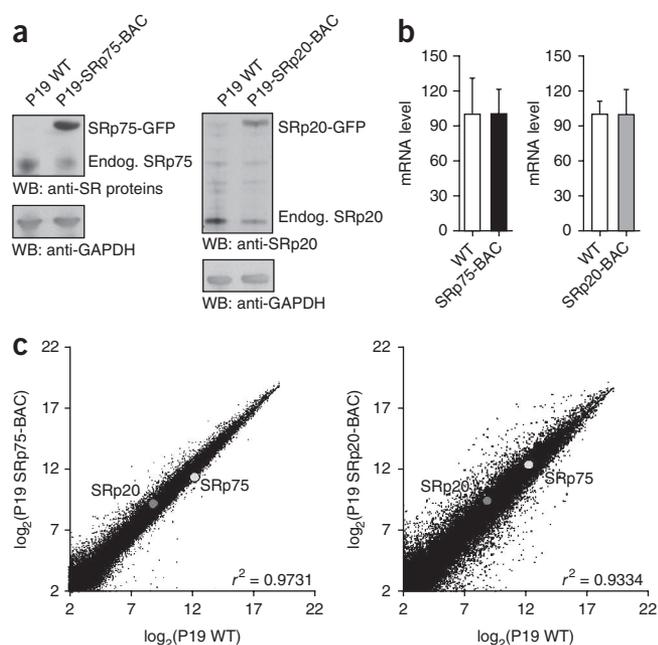
Figure 2 Total SR protein expression is not increased in BAC transgenic cell lines. (**a**) Western blot analysis of the SR protein levels in the transgenic P19 cell lines (P19-SRp20-BAC and P19-SRp75-BAC). mAb104 was used to detect SRp75 and mAb7B4 (anti-SRp20) antibody to detect SRp20; anti-GAPDH antibody served as loading control. Endog., endogenous. (**b**) Quantification of the total SR protein transcript levels with RT-qPCR in P19 parental (WT) and BAC transgenic cells. Primers detecting mRNAs transcribed from both the endogenous SRp20 (*SFRS3*) or SRp75 (*SFRS4*) and the corresponding transgenes were used. The signals were normalized to RT-qPCR signals obtained for *Pol2R* mRNA ($n = 6$, error is s.e.m.). The differences between WT and BAC cell lines were not statistically significant (unpaired Student's *t*-test). (**c**) Global analysis of gene expression in the parental P19 cells (WT) and transgenic cell lines using whole mouse genome microarray ($n = 3$). The larger points show the expression of the SRp20 and SRp75. The $\log_2$ values of quantile-normalized intensities are plotted.

proteins are functional, because they participate in autoregulatory feedback; we further address functionality below. Taken together, SR protein expression by BAC transgenesis in P19 cells provides an excellent model system for studying SR protein function.

The expression levels of many genes are altered during neural differentiation. In P19 cells induced with RA, the first set of genes are induced from 16 h onwards up to 4 d from the start of treatment, and a second larger induction peak occurs around 6 d (ref. 26). As expected, gene-expression microarray analysis showed decreased expression of stem-cell markers and increased expression of neural markers 8 d after RA treatment in both P19-SRp20-BAC and P19-SRp75-BAC cell lines (**Fig. 3a**). To validate the microarray analysis, we analyzed the expression of the neurally induced *FOS* gene in detail (**Fig. 3b**). The temporal pattern of *FOS* induction in the transgenic BAC cell lines was identical to that of the parental cells. In addition, we analyzed the functional homogeneity of the genes that increased in expression more than two-fold upon neural induction in the BAC cell lines by searching for enriched gene ontology terms within the Panther classification system[42] and found terms related to differentiation and development among the top categories (**Supplementary Fig. 3**).
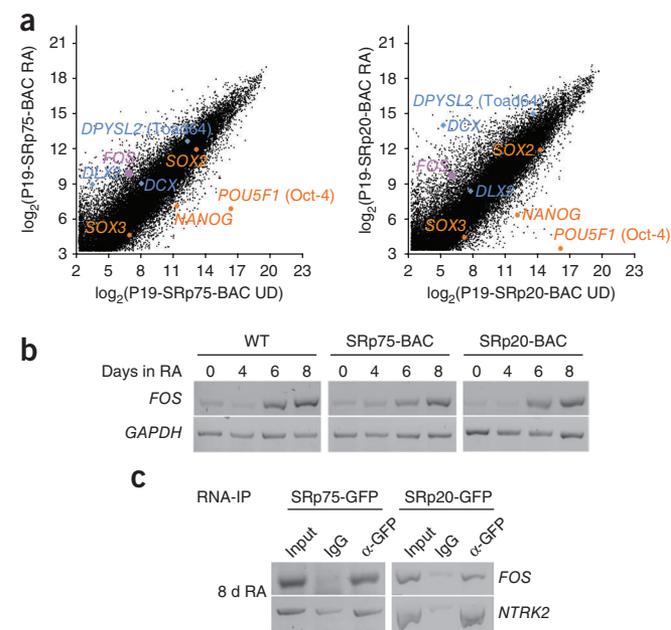
SRp20 and SRp75 protein levels increased within the first phase of differentiation, before many of the neural-specific genes are induced (see **Fig. 1c**). Therefore, we asked whether these SR proteins might be involved in the regulation of the expression of neural transcripts. RNA immunoprecipitation (RIP) followed by RT-PCR analysis
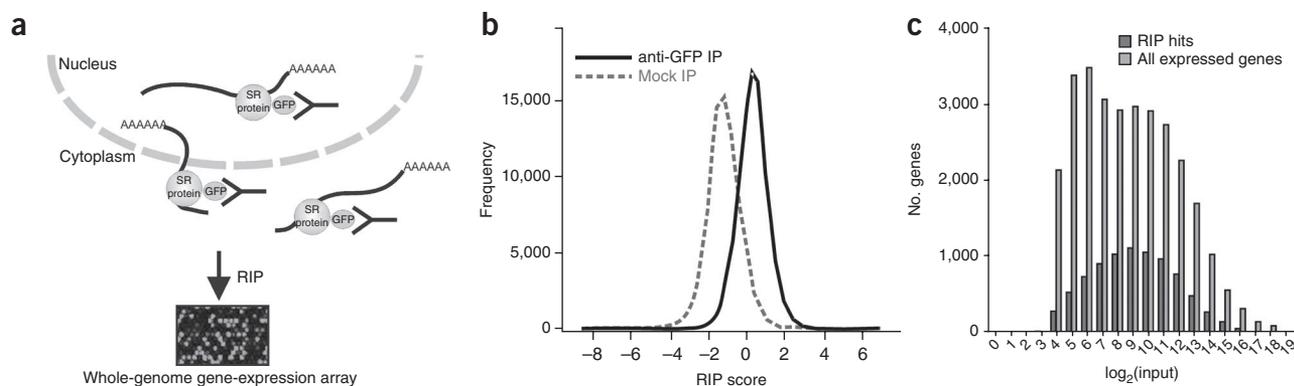
revealed that SRp20 and SRp75 were associated with mRNAs encoding c-fos and the neural-specific form of the TrkB receptor (**Fig. 3c**). *FOS* and *NTRK2* mRNAs were nearly undetectable by RT-PCR in the immunoprecipitates from cycling cells and are apparently not expressed in the absence of RA (**Fig. 3b** and data not shown). The association in neural cells was confirmed by ChIP, which showed the co-transcriptional recruitment of SRp20 and SRp75 to *FOS* and *NTRKB* genes at different stages of the neural differentiation process (**Supplementary Fig. 4**). Thus, both SR proteins were upregulated during differentiation and specifically associated with two neural-specific transcripts.

RIP-chip identifies mRNAs associated with SRp20 and SRp75
To date, only few *in vivo* targets of SRp20 or SRp75 have been reported. To identify mRNAs associated with SRp20 and SRp75 at any stage of gene expression following polyadenylation, we performed a global RIP analysis (RIP-chip) of SRp20-GFP and SRp75-GFP in the transgenic cell lines (**Fig. 4a**). We used mRNA immunoprecipitated via the GFP tag from undifferentiated and neural whole-cell extracts to prepare probes that were hybridized onto a whole mouse genome microarray. Along with the specific anti-GFP immunoprecipitations, we hybridized nonspecific IgG

Figure 3 Changes in gene expression upon differentiation are linked to SR protein function. (**a**) Changes in global gene expression upon neural differentiation in P19-SRp20-BAC cells (left) and P19-SRp75-BAC cells (right) as measured by whole mouse genome microarray. The expression of stem-cell markers was decreased (*NANOG, POU5F1, SOX2, SOX3* in orange) and that of neural markers was increased (*DLX2, DCX, FOS, DPYSL2* in blue/purple) upon neural induction. (**b**) RT-PCR analysis of *FOS* expression in the course of neural differentiation in P19 WT, P19-SRp20-BAC and P19-SRp75-BAC cells. *GAPDH* served as a control. (**c**) RNA immunoprecipitation of neurally expressed *FOS* and *NTRK2* mRNAs with anti-GFP antibody from P19-SRp20-BAC and P19-SRp75-BAC cells 8 d after retinoic acid induction. IgG was used as nonimmune control; 10% (v/v) of the starting material was used for the input.

Figure 4 RIP-chip analysis of SRp20- and SRp75-associated mRNAs in undifferentiated and neural P19 cells. (**a**) Schematic representation of the experimental setup for genome-wide detection of polyadenylated mRNAs that co-purify with GFP-tagged SR proteins. (**b**) Frequency distribution of the RIP scores ($\log_2$(RIP/input)) for specific immunoprecipitations (anti-GFP) and mock immunoprecipitations (IgG). All RIP scores from four experiments are plotted together. (**c**) RIP-chip analysis identified hits independent of the gene expression level. Frequency histogram shows the expression levels of RIP hits and all expressed genes in undifferentiated and neural cells relative to their gene expression levels. (**d**) Validation of RIP hits of SRp20 and SRp75 (RIP score > 1, FDR < 0.05) and nonhits by RT-PCR. Three independent RIPs were performed and analyzed; representative images are shown.

immunoprecipitations and input samples (total gene expression). Following quantile normalization of raw signals extracted from at least three replicates each, we filtered out signals representing unexpressed genes, and we calculated a RIP score ($\log_2$(RIP / input)) representing the enrichment in the immunoprecipitation over the input samples for each expressed transcript. A detailed description of the data analysis appears in Online Methods, and a flow chart of the analysis is shown in **Supplementary Figure 5**.
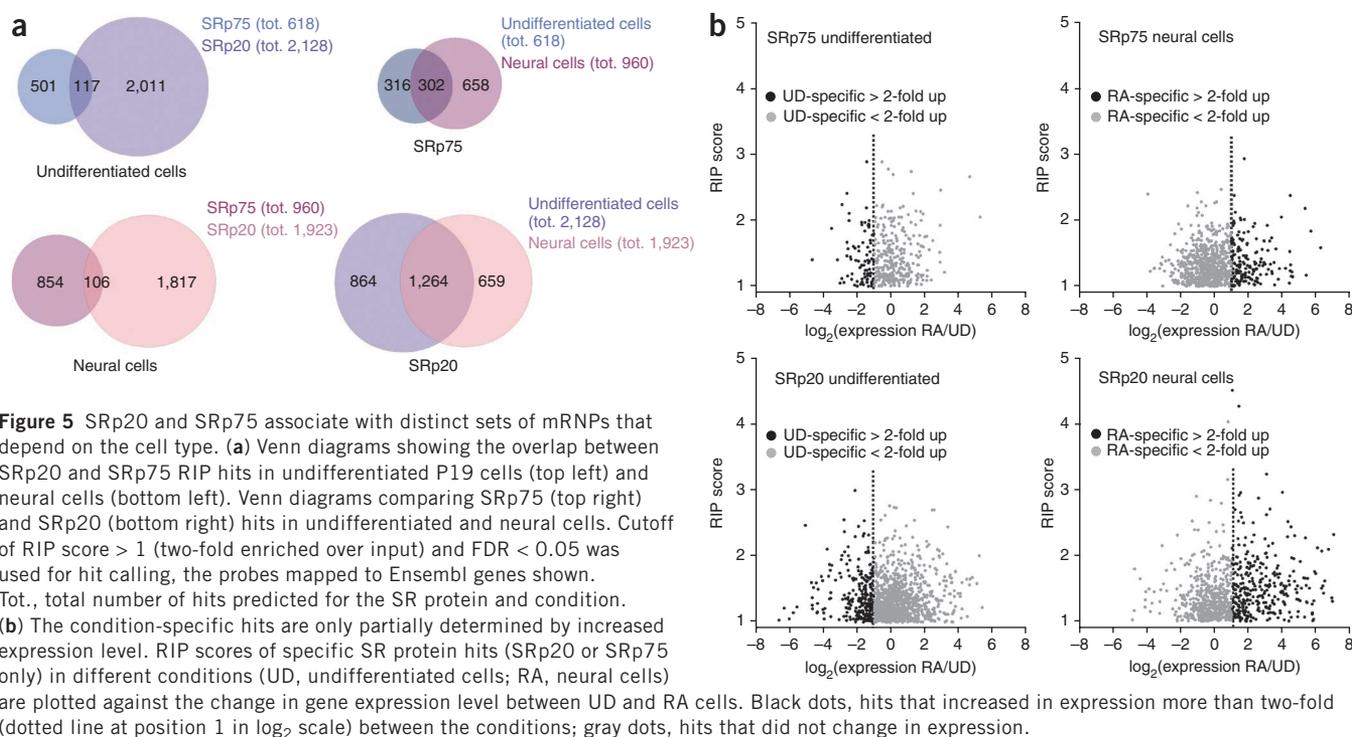
The distributions of the RIP scores for SRp20 and SRp75 were distinctly different from the RIP score distribution for the mock RIP (**Fig. 4b**), showing the specificity of the immunoprecipitations. The data analysis was not biased to low or highly expressed genes, because hits were identified over the whole range of gene expression (**Fig. 4c**). Based on the RIP score ranking and ANOVA analysis with correction for multiple testing (false discovery rate, FDR), we identified groups of target transcripts for SRp20 and SRp75 as 'RIP hits' and validated them by RT-PCR (**Fig. 4d**). The RIP score and FDR cutoffs used for the 'hit calling' were RIP score > 1 (≥ 2-fold enrichment over input) and FDR < 0.05. All RIP hits based on these criteria are listed in **Supplementary Table 2**. Transcripts with high RIP scores and low FDR-corrected P values in SRp75 RIP-chip but not in SRp20 RIP-chip could be specifically co-immunoprecipitated with SRp75 (**Fig. 4d**: *CLK4*, *CNPY1*, *PSMD4* and *VEGFA*). Conversely, we could show the specificity of SRp20 to its RIP hits (*FRAG1, NT5DC2, CLN6* and *SLAIN2*). *SFPQ* scored very low in both SRp20 and SRp75 RIP-chips, and the RT-PCR results confirmed this observation. Based on the RT-PCR validation, *PTBP1* and *FN1* were immunoprecipitated by both SRp20 and SRp75, although they were high-confidence hits only in the SRp20 dataset. We detected very few false positives (*CAR11* for SRp75, and *CCDC50* and *VEGFA* for SRp20). To further validate the specificity of the RIP, we performed RIPs from formaldehyde–cross-linked samples with stringent washing conditions (**Supplementary Fig. 6**). The results of the cross-linked RIPs corresponded well with the non–cross-linked RIP data.

Cell type–specific association of mRNAs with SRp20 and SRp75

For the further analysis of mRNAs immunoprecipitated by SRp20 or SRp75, we used the cutoff of RIP score > 1 and FDR 0.05. Note that, irrespective of the cutoff used for hit calling, the overlaps and relationships between the four datasets (SRp20 and SRp75, undifferentiated and neural cells) showed a similar pattern. SRp20 associated with a greater number of mRNAs than SRp75: 13.1% (undifferentiated cells) and 12.4% (neural cells) of all expressed genes associated with SRp20-GFP, whereas 3.9% and 5.8% associated with SRp75-GFP. Notably, SRp20- and SRp75-associated mRNAs showed very little overlap (**Fig. 5a**). We detected only 19% of SRp75 and 6% of SRp20 RIP hits in undifferentiated cells in both groups. Similarly, SRp20 and SRp75 RIP hits were very distinct from one another in neural cells; the overlap between SR proteins was 11% (SRp75) and 6% (SRp20) (**Fig. 5a**). We also compared the RIP hits to genes previously identified as targets of SRp20 or SRp75 (**Supplementary Table 3**); we detected some but not all. These discrepancies suggest either differences between cell lines used or differences in the assays used to detect target mRNAs.

Does neural differentiation lead to changes in the mRNAs associated with a given SR protein? Comparison of the RIP hits in undifferentiated cells with the neural cells revealed an overlap of ~50% for each SR protein, indicating that the other ~50% of hits is condition specific (**Fig. 5a**). The simplest explanation is that these condition-specific hits reflect changes in global gene expression. Unexpectedly, only 5–10% of neural-specific RIP hits represented genes turned on by the RA treatment (data not shown). In other words, 90–95% of RIP hits identified in neural cells were detectably expressed in undifferentiated cells, where they were not detected as hits. Although 22% of

Figure 5 SRp20 and SRp75 associate with distinct sets of mRNPs that depend on the cell type. (**a**) Venn diagrams showing the overlap between SRp20 and SRp75 RIP hits in undifferentiated P19 cells (top left) and neural cells (bottom left). Venn diagrams comparing SRp75 (top right) and SRp20 (bottom right) hits in undifferentiated and neural cells. Cutoff of RIP score > 1 (two-fold enriched over input) and FDR < 0.05 was used for hit calling, the probes mapped to Ensembl genes shown. Tot., total number of hits predicted for the SR protein and condition. (**b**) The condition-specific hits are only partially determined by increased expression level. RIP scores of specific SR protein hits (SRp20 or SRp75 only) in different conditions (UD, undifferentiated cells; RA, neural cells) are plotted against the change in gene expression level between UD and RA cells. Black dots, hits that increased in expression more than two-fold (dotted line at position 1 in $\log_2$ scale) between the conditions; gray dots, hits that did not change in expression.
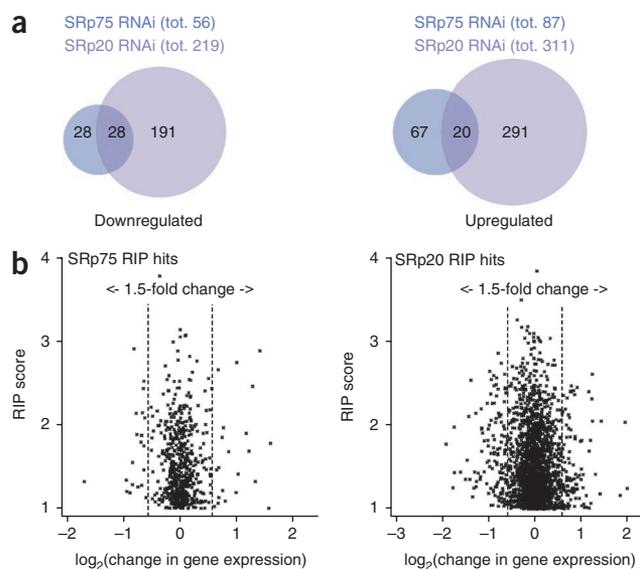
expressed genes were >2-fold up- or downregulated between undifferentiated and neural P19 cells, only 61% (SRp20) and 26% (SRp75) of neural-specific hits increased in gene expression by >2-fold upon RA treatment (**Fig. 5b**). Similarly, only 27% (SRp20) and 28% (SRp75) of hits specific for undifferentiated cells were more highly expressed in this condition. Thus, differential gene expression is not the only determining factor for detection of cell type–specific hits.

To determine whether functionally related sets of genes might be coordinately regulated by SRp20 or SRp75, we analyzed the functional homogeneity of the RIP hits by searching for enriched gene ontology terms within the Panther classification system[42]. Notably, in both cell lines and conditions, terms related to transcriptional regulation were significantly enriched (**Supplementary Fig. 7**). In addition, we identified specific functional groups—for example, 'SNARE proteins' for SRp75, and 'cell proliferation and differentiation' for SRp20 in the undifferentiated cells. In parallel, we asked whether SRp20 and SRp75 RIP hits were distinguished by other features, such as gene architecture. In accordance with our previous study[5], both SRp20 and SRp75 had intronless genes as hits. Otherwise, we detected no preference for chromosomal position, exon number, total transcript length or 3′ and 5′ untranslated region (UTR) length within the RIP hits when compared to random datasets of equal size (**Supplementary Fig. 8a** and data not shown).

In vitro binding motifs for SRp20 have been previously identified using SELEX[43,44], and a single binding motif found in bovine papillomavirus type 1 (BPV-1) pre-mRNA for SRp75 has been proposed[45]. We searched for these motifs within the RIP hits. All SRp20 hits contained at least one of the three SRp20 motifs queried, and 79% of SRp75 hits had the proposed SRp75 motif within the transcript. We further analyzed the positions of the motifs within 5′ UTRs, coding sequences and 3′ UTRs. No enrichment of the motifs at particular positions along the transcripts could be detected (**Supplementary Fig. 8b**), consistent with previous reports[23]. Neither motif was enriched in any set of RIP hits as compared to the entire transcriptome

(**Supplementary Table 4**). Thus, the mere presence of binding motifs determined *in vitro* could not predict the *in vivo* RIP hits obtained here, because their occurrence is so commonplace. Taken together, SRp20 and SRp75 RIP hits differed in cellular function, not gene architecture or sequence, suggesting specific cellular roles of SR proteins dependent on the differentiation state.


Figure 6 Expression of SRp20 and SRp75 RIP hits is compromised by SR protein depletion. (**a**) Venn diagrams presenting the transcripts affected more than 1.5-fold upon SRp20 or SRp75 knockdown (Ensembl genes; *P* value < 0.05, one-way ANOVA). Tot., total number of affected genes upon the SR protein depletion. (**b**) Change of gene expression of SRp75 RIP hits upon SRp75 knockdown (left) and SRp20 RIP hits upon SRp20 knockdown (right). Dashed line, 1.5-fold change in gene expression (knockdown relative to control).

Figure 7 Misexpression of RIP hits upon depletion of SR proteins and rescue by GFP-tagged SR proteins expressed from BACs. (**a**) Outline of the experiment. SRp20 or SRp75 RNAi was performed in the parental P19 WT cells and in the BAC transgenic cells. The RNAi was designed to target only the endogenous mouse SR protein. (**b**) RT-PCR analysis of SRp20 and SRp75 RIP hit expression in undifferentiated P19 cells 24 h after the start of RNAi. *VEGFA, PTBP1* and *SLAIN2* expression was analyzed in P19 parental cells (WT) and P19-SRp20-BAC or P19-SRp75-BAC cell lines after RNAi. Ctrl, control; n.a., not applicable.
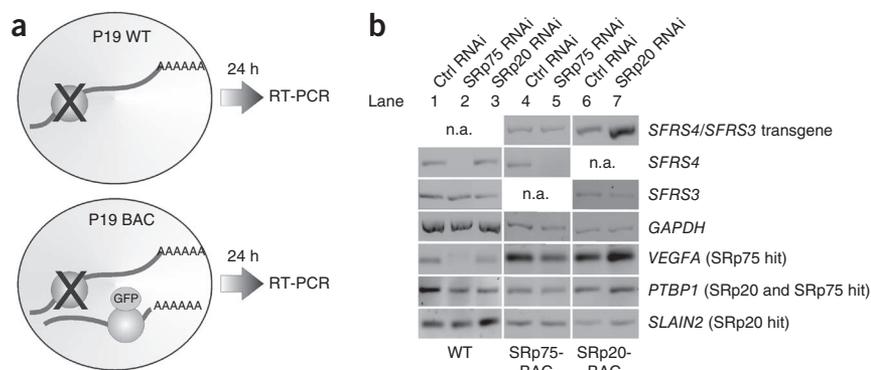

Knockdown of SRp20 and SRp75 changes expression of RIP hits

To determine whether SRp20 and SRp75 RIP hits depend on the respective SR protein for expression, we performed whole mouse genome microarray analysis after SRp20 or SRp75 depletion in the parental P19 cell line. We achieved RNA interference (RNAi) by the endoribonuclease-prepared small interfering RNA (esiRNA) method to minimize off-target effects[46]. The data was processed similar to the RIP-chip data. We compared normalized probe intensities from the knockdown samples to control samples; we used a cutoff based on fold change (>1.5) and *P* value ($P < 0.05$) to identify affected genes. All affected genes based on this cutoff are listed in **Supplementary Table 5**. Notably, distinct sets of genes were both up- and downregulated upon SRp20 and SRp75 knockdown (**Fig. 6a**). Similar to RIP-chip hits, we observed little overlap of affected genes upon SRp20 and SRp75 depletion. Furthermore, the knockdown of SRp20, which had a substantially greater number of RIP hits than

SRp75, also led to the misregulation of a greater number of transcripts than the knockdown of SRp75. When we performed the knockdowns in the BAC transgenic cells, where the transgene is resistant to the RNAi, expression of the affected genes was restored to control levels (**Supplementary Fig. 9**), showing that GFP-tagged SR proteins expressed from the transgenes rescue the knockdown phenotype. Next, we specifically analyzed the expression of the identified RIP hits upon the knockdowns. Note that SRp75 depletion affected some SRp20 RIP hits and vice versa (**Supplementary Fig. 10**); this underscores the possibility that some indirect effects may occur and suggests a complex network of gene regulation by SR proteins. Notably, we did indeed detect changes in the expression of some RIP-hit mRNAs upon depletion of the relevant SR protein (**Fig. 6b**), indicating that SRp20 and SRp75 are required for the expression of at least a subset of identified RIP hits. RIP hits were both up- and downregulated upon SR protein depletion.
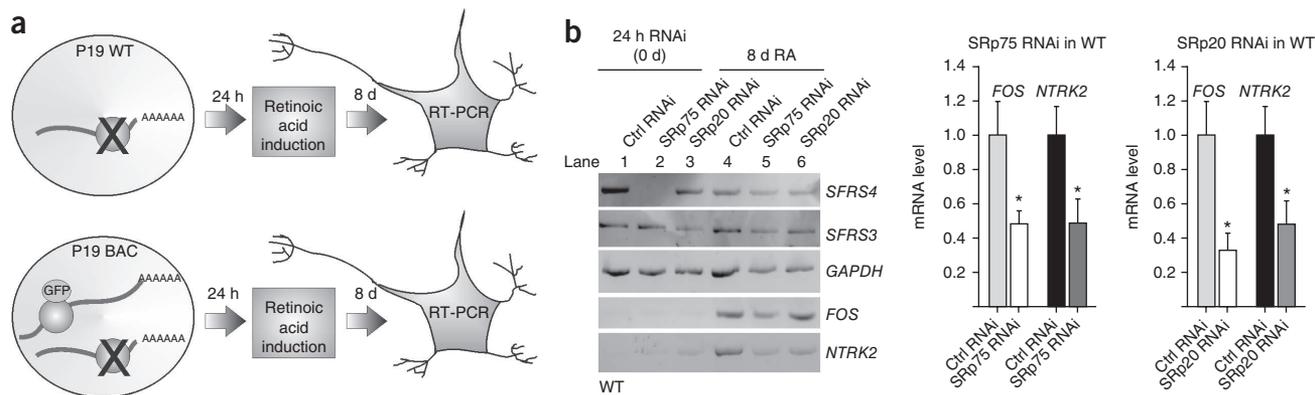
Figure 8 SR protein depletion leads to misregulation of neural gene expression. (**a**) Outline of the experiment. SRp20 or SRp75 RNAi was performed in the parental P19 WT cells and in the BAC transgenic cells. RA treatment began 24 h after transfection, and cells were harvested 8 d later. The RNAi was designed to target only the endogenous mouse SR protein. (**b**) RT-PCR analysis of neural SRp20 and SRp75 RIP hits *FOS* and *NTRK2* upon RNAi followed by neural differentiation in the parental P19 cells (WT). Bar graphs, quantification of the expression of *FOS* and *NTRK2* by RT-qPCR (normalized to *ACTB*; *n* = 6, unpaired Student's *t*-test; error is s.e.m.; *, $P < 0.05$ for control versus specific RNAi). Lighter set of bars, *FOS*; darker set of bars, *NTRK2*. Ctrl, control. (**c**) RT-PCR analysis of *FOS* and *NTRK2* expression in the P19-SRp20-BAC and P19-SRp75-BAC cell lines after SRp20 and SRp75 RNAi, respectively. Bar graphs, quantification of the expression of *FOS* and *NTRK2* by RT-qPCR (normalized to *ACTB*, *n* = 3, unpaired Student's *t*-test; error is s.e.m.). Lighter set of bars, *FOS*; darker set of bars, *NTRK2*.
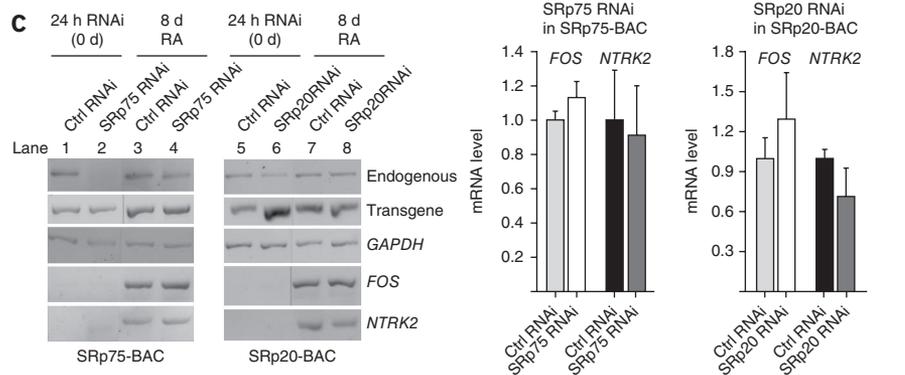
To validate and extend these results, we analyzed mRNA levels of specific RIP hits by RT-PCR after SR protein knockdown (**Fig. 7a**). In undifferentiated cells, the expression of an SRp75-specific RIP hit *VEGFA* (see **Fig. 4d**) was downregulated upon SRp75 depletion (**Fig. 7b**, lane 2). We identified *PTBP1* as both an SRp20- and SRp75-specific RIP hit; accordingly, *PTBP1* expression decreased upon either SRp20 or SRp75 depletion (**Fig. 7b**, lanes 2 and 3). *SLAIN2*, an SRp20-specific RIP hit, was upregulated by SRp20 depletion (**Fig. 7b**, lane 3). When we performed the knockdowns in the respective BAC cell lines, where the human transgenes were resistant to RNAi, we observed no change in *VEGFA*, *PTBP1* or *SLAIN2* expression (**Fig. 7b**, lanes 5 and 7). This further confirms that the GFP-tagged SR proteins expressed from BACs are functional and can replace their mouse homologs.

Because SRp20 and SRp75 expression increases early in the pathway of neural differentiation (see **Fig. 1**), we wondered whether depletion of either protein could produce long-lasting effects in neurally differentiated cells (**Fig. 8a**). Note that, although mRNAs encoding SRp20 (*SFRS3*) and SRp75 (*SFRS4*) are reduced 24 h after esiRNA transfection, both mRNAs return to normal levels by 8 d (**Fig. 8b**). Notably, induction of *FOS* was significantly reduced (**Fig. 8b**, lanes 5 and 6, and lighter bars in the graph) in P19 cells depleted of SRp20 and SRp75 before neural differentiation. Similarly, the expression of the neural-specific form of *NTRK2* was significantly reduced in the SRp20- or SRp75-RNAi–treated P19 cells (**Fig. 8b**, lanes 5 and 6, and darker bars in the graph). The expression of the reference gene *ACTB* was not affected by the knockdowns. When we depleted endogenous SRp20 or SRp75 in P19-SRp20-BAC or P19-SRp75-BAC cells, we observed no significant change in the *FOS* induction or *NTRK2* expression (**Fig. 8c**, lanes 4 and 8, and bar graphs), consistent with the functionality of the tagged SR proteins and showing the specificity of the knockdown effects even 8 d after start of RNAi. Taken together, SRp20 and SRp75 were required for the correct expression of the RIP hits analyzed in both undifferentiated and neural cells.

DISCUSSION

We have identified endogenous mRNAs associated with two SR proteins in cycling and neurally differentiated P19 cells using RNA immunopurification (RIP) followed by gene-expression microarray analysis. To achieve this aim, we established physiological expression of GFP-tagged SRp20 and SRp75 encoded on BACs. Expression of both endogenous and tagged forms of each protein was upregulated during neural differentiation. Notably, the tagged SR proteins could rescue specific knockdown of the endogenous SR proteins, demonstrating their functionality. We can draw four major conclusions. First, each SR protein associated with hundreds of distinct mRNAs in cycling or neurally differentiated P19 cells. A relatively small fraction of the SRp20 and SRp75 RIP hits identified were shared in either condition, indicating a striking specificity in mRNP composition with respect to individual SR proteins. Second, the detection of unique RIP hits upon neural differentiation was only partially due to global changes in gene expression, indicating cell type–specific composition of mRNPs. Third, mRNAs associated with each SR protein fell into functionally related groups, suggesting that each SR protein may coregulate genes involved in distinct cellular processes. Fourth, depletion of SRp20 or SRp75 compromised the expression of some RIP hits; we detected both up- and downregulated genes upon knockdown. Taken together, we propose that SR proteins can regulate distinct, functionally related sets of genes through their association with particular mRNAs in mRNPs after splicing and can thereby influence the deployment of cellular differentiation programs.

To develop a uniform affinity tag for SR proteins, we previously established GFP tagging of SR proteins by BAC recombineering, which preserves endogenous genetic control elements and facilitates the generation of stable cell lines of choice[5]. The GFP-tagged SR proteins appear to be fully functional by several criteria. The tagged proteins are localized properly to the nucleus and undergo nucleocytoplasmic shuttling, as expected[5]. Furthermore, both in this study and in the previous one in HeLa cells, the GFP-tagged SR proteins associate co-transcriptionally with nascent RNA and post-transcriptionally with endogenous mRNAs[5]. Most importantly, we show here that SRp20-GFP and SRp75-GFP can rescue the function of the endogenous SR protein upon knockdown (**Figs. 7** and **8** and **Supplementary Fig. 9**).

Our conclusion that the GFP-tagged SR proteins are functional is reinforced by the observation that SRp20-GFP and SRp75-GFP expressed from BACs participate in an autoregulatory feedback mechanism, leading to homeostasis of SR protein levels in cells (**Fig. 2**). Among others, SRp20 regulates its own expression through alternative splicing[34]. Recently, it was shown that all SR protein genes harbor ultraconserved regions that code alternative cassette exons[40,41]. In most cases, the inclusion of these cassette exons by alternative splicing leads to nonsense-mediated decay. Furthermore, SF2 appears to autoregulate its expression through several mechanisms at the level of splicing, transcription and translation[39]. Consistent with the predicted autoregulatory mechanisms, the total expression of SRp20 or SRp75 in the transgenic cell lines was similar to the level of each protein in the parental line. Moreover, global gene expression in the transgenic lines was unchanged from the parental cells, and the SRp20 and SRp75 transgenes were coregulated with the endogenous genes upon neural induction. Therefore, our transgenic cell lines provide a model system for studying the functional, tagged SR proteins within the endogenous context.

Using the GFP tag as a uniform handle to access SRp20- and SRp75-containing mRNPs, we performed RIP-chip on cycling and neurally differentiated P19 cells. RIP-chip identifies polyadenylated mRNAs associated with each SR protein in the context of any cellular process in either the nucleus or the cytoplasm and has been previously used to identify mRNA targets for other RNA-binding proteins[47–50]. We performed the RIP in the absence of cross-linking, and it yielded specific mRNAs, because none of the statistically defined RIP hits were recovered in mock RIPs (**Fig. 4**). The RIP hits do not appear to be the result of SR proteins unbinding and rebinding mRNAs in the extract for two reasons. First, if mixing were random, we would expect little correlation among the mRNAs isolated by RIP; on the contrary, individual RIPs were highly reproducible (**Supplementary Table 1**). Second, the target mRNAs identified could be validated by immunoprecipitation from both cross-linked and non–cross-linked extracts (**Fig. 4** and **Supplementary Fig. 6**).

Until now, the number of transcripts regulated by one SR protein has been unknown; our results imply that single SR proteins may have very different numbers of target mRNAs. Notably, more transcripts depended on SRp20 than SRp75 for correct expression (**Fig. 6**). Moreover, a greater number of mRNA transcripts was associated with SRp20 than with SRp75 (**Fig. 5**). These differences may reflect the more robust shuttling behavior of SRp20 compared to SRp75 (ref. 5), yielding more mRNAs dependent on SRp20 in the cytoplasm. Furthermore, we found almost no overlap between the mRNAs associated with SRp20 and those associated with SRp75 in either condition analyzed. This is in agreement with a RIP-chip study of two *Drosophila melanogaster* SR proteins, SF2 and SRp55, in which a similarly low level of overlap was found[19]. Therefore, comparison of SR proteins to

one another indicates little general redundancy between SR protein targets. The specificity of SR proteins for mRNA targets is further supported by studies that examined SR protein association with nascent RNA; specific combinations of SR proteins associate with the Balbiani ring genes of *Chironomus tentans*[51] and with a limited number of human genes tested[5]. The global analysis conducted here facilitates a comprehensive analysis of the distinct sets of mRNAs associated with SRp20 and SRp75. Analysis of annotated gene functions associated with the two sets of mRNAs revealed functional relationships. For example, SRp20 and SRp75 associated with transcripts encoding proteins involved in transcription and nucleic acid binding. This was also observed for SF2 in a recent study[16] and coincides with a higher rate of alternative splicing among transcription-factor genes[52]. Moreover, contrast, SRp20 preferentially associated with transcripts encoding proteins involved in developmental processes. This suggests that SR proteins might act as 'master switches' in the expression of genes essential for the execution of related cellular processes.

The global identification of SR protein–specific mRNPs in cycling and neural cells provides a framework for a comprehensive understanding of how SR proteins control gene-expression programs. SR proteins can bind co-transcriptionally to pre-mRNAs as soon as transcription begins (see also ref. 5). They can regulate splice-site selection co-transcriptionally or post-transcriptionally and may remain associated with mature mRNPs[17]. Alternatively, SR proteins may be lost from some targets after splicing and may newly associate with others to fulfill additional functions. Both SRp20 and SRp75 shuttle to the cytoplasm, and SR proteins are involved in the export of some mRNAs[5,18,51,53,54], suggesting that the mRNPs containing SRp20 or SRp75 may be dependent on SR proteins for nucleocytoplasmic transport. Whether these SR proteins are broadly required for mRNA nucleocytoplasmic transport is an open question. Notably, intronless mRNAs were observed to be associated with SRp20 and SRp75. Similarly, the splicing factor U2AF65 was found to associate with intronless mRNAs by RIP-chip, consistent with its role in mRNA export[49,55]. The gene-expression analysis of SR protein knockdowns performed in parallel provides evidence for a complex regulatory network of gene expression by SR proteins. Many, but not all, RIP targets were up- or downregulated upon SR protein depletion. The observed gene-expression phenotypes upon knockdown could be due to defects in any of the above processes, and some changes in gene expression could be indirect. For instance, *PTBP1* was among the affected genes. *PTBP1* encodes PTB, a well-studied regulator of alternative splicing that has been shown to regulate splicing during neuronal differentiation[56]. However, the specific cases of up- and downregulation examined could be rescued by the corresponding BAC-encoded SR protein, indicating that these phenotypes were specific for the SR protein targeted by RNAi. Possible other roles for SRp20 and SRp75 in gene expression include the regulation of transcriptional elongation and protein translation, though neither SRp20 nor SRp75 has been implicated in these processes to date[2,16,57].

The investigation of SR protein targets in the context of neural induction offers insights into the relationship between changes in gene expression and recognition by SR proteins. The simplistic expectation was that newly expressed genes would add to the potential pool of mRNPs containing SR proteins. Upon global analysis, however, a more complex picture emerged: half of the mRNAs associated with each SR protein in neural cells were also expressed in cycling cells but were not RIP hits. Thus, the differentiation state of the cell affects the sets of mRNAs associated with SRp20 and SRp75. One noteworthy possibility is that additional cofactors become expressed or activated in the neural cells, changing the association

of SR proteins with mRNAs. Indeed, global gene-expression analysis revealed that a number of RNA-binding proteins, kinases and phosphatases were increased in expression upon neural induction in P19 cells (M.-L.Ä. and K.M.N., unpublished data). Further experiments will be needed to determine how SR protein association with the same mRNA can be modulated by neural differentiation. Whether other SR proteins are involved remains an open question. The demonstrated potential of individual SR proteins to regulate functionally related groups of genes invokes the idea of an 'mRNP code', in which gene-expression programs can be coordinated by RNA-binding proteins with multiple functions.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

AUTHOR CONTRIBUTIONS
M.-L.Ä. and K.M.N. designed the experiments; M.-L.Ä. conducted the experiments; A.B. and L.M. designed the RIP-chip data analysis; I.H., L.M. and M.-L.Ä. performed the data analysis; M.-L.Ä. and K.M.N. wrote the manuscript.

1. Shepard, P.J. & Hertel, K. The SR protein family. *Genome Biol.* **10**, 242 (2009).
2. Zhong, X.-Y., Wang, P., Han, J., Rosenfeld, M.G. & Fu, X.-D. SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol. Cell* **35**, 1–10 (2009).
3. Lin, S. & Fu, X.D. SR proteins and related factors in alternative splicing. *Adv. Exp. Med. Biol.* **623**, 107–122 (2007).
4. Shen, H., Kan, J.L.C. & Green, M.R. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol. Cell* **13**, 367–376 (2004).
5. Sapra, A.K. *et al.* SR protein family members display diverse activities in the formation of nascent and mature mRNPs *in vivo*. *Mol. Cell* **34**, 179–190 (2009).
6. Mayeda, A., Screaton, G.R., Chandler, S.D., Fu, X.-D. & Krainer, A.R. Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. *Mol. Cell. Biol.* **19**, 1853–1863 (1999).
7. Caceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L. & Krainer, A.R. Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J. Cell Biol.* **138**, 225–238 (1997).
8. Wu, J.Y. & Maniatis, T. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061–1070 (1993).
9. Blaustein, M. *et al.* Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat. Struct. Mol. Biol.* **12**, 1037–1044 (2005).
10. Sanford, J.R., Ellis, J.D., Cazalla, D. & Caceres, J.F. Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor 2/alternative splicing factor. *Proc. Natl. Acad. Sci. USA* **102**, 15042–15047 (2005).
11. Hanamura, A., Caceres, J.F., Mayeda, A., Franza, B.R. Jr. & Krainer, A.R. Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA* **4**, 430–444 (1998).
12. Zahler, A.M., Neugebauer, K.M., Lane, W.S. & Roth, M.B. Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science* **260**, 219–222 (1993).

13. Hertel, K.J. Combinatorial control of exon recognition. *J. Biol. Chem.* **283**, 1211–1215 (2008).
14. Blencowe, B.J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37–47 (2006).
15. Smith, C.W.J. & Valcarcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* **25**, 381–388 (2000).
16. Sanford, J.R. *et al.* Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PLoS ONE* **3**, e3369 (2008).
17. Merz, C., Urlaub, H., Will, C.L. & Luhrmann, R. Protein composition of human mRNPs spliced *in vitro* and differential requirements for mRNP protein recruitment. *RNA* **13**, 116–128 (2007).
18. Caceres, J.F., Screaton, G.R. & Krainer, A.R. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev.* **12**, 55–66 (1998).
19. Gabut, M., Dejardin, J., Tazi, J. & Soret, J. The SR family proteins B52 and dASF/SF2 modulate development of the *Drosophila* visual system by regulating specific RNA targets. *Mol. Cell. Biol.* **27**, 3087–3097 (2007).
20. Ding, J.-H. *et al.* Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *EMBO J.* **23**, 885–896 (2004).
21. Karni, R. *et al.* The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* **14**, 185–193 (2007).
22. Ray, D. *et al.* Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat. Biotechnol.* **27**, 667–670 (2009).
23. Wang, J., Smith, P.J., Krainer, A.R. & Zhang, M.Q. Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. *Nucleic Acids Res.* **33**, 5053–5062 (2005).
24. Sanford, J.R. *et al.* Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* **19**, 381–394 (2009).
25. McBurney, M.W., Jones-Villeneuve, E.M., Edwards, M.K. & Anderson, P.J. Control of muscle and neuronal differentiation in a cultured embryonal carcinoma cell line. *Nature* **299**, 165–167 (1982).
26. Wei, Y., Harris, T. & Childs, G. Global gene expression patterns during neural differentiation of P19 embryonic carcinoma cells. *Differentiation* **70**, 204–219 (2002).
27. Zahler, A.M., Neugebauer, K.M., Stolk, J.A. & Roth, M.B. Human SR proteins and isolation of a cDNA encoding SRp75. *Mol. Cell. Biol.* **13**, 4023–4028 (1993).
28. Liang, H., Tuan, R.S. & Norton, P.A. Overexpression of SR proteins and splice variants modulates chondrogenesis. *Exp. Cell Res.* **313**, 1509–1517 (2007).
29. Faa, V. *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.* **284**, 30024–30031 (2009).
30. Ramchatesingh, J., Zahler, A.M., Neugebauer, K.M., Roth, M.B. & Cooper, T.A. A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. *Mol. Cell. Biol.* **15**, 4898–4907 (1995).
31. Exline, C.M., Feng, Z. & Stoltzfus, C.M. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 Vif gene expression. *J. Virol.* **82**, 3921–3931 (2008).
32. Galiana-Arnoux, D. *et al.* The CD44 Alternative v9 exon contains a splicing enhancer responsive to the SR proteins 9G8, ASF/SF2, and SRp20. *J. Biol. Chem.* **278**, 32943–32953 (2003).
33. Lim, L.P. & Sharp, P.A. Alternative splicing of the fibronectin EIIIB exon depends on specific TGCATG repeats. *Mol. Cell. Biol.* **18**, 3900–3906 (1998).
34. Jumaa, H. & Nielsen, P. The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.* **16**, 5077–5085 (1997).
35. Lou, H., Neugebauer, K.M., Gagel, R.F. & Berget, S.M. Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Mol. Cell. Biol.* **18**, 4977–4985 (1998).
36. Poser, I. *et al.* BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods* **5**, 409–415 (2008).
37. Roth, M.B., Zahler, A.M. & Stolk, J.A. A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *J. Cell Biol.* **115**, 587–596 (1991).
38. Neugebauer, K.M. & Roth, M.B. Distribution of pre-mRNA splicing factors at sites of RNA polymerase II transcription. *Genes Dev.* **11**, 1148–1159 (1997).
39. Sun, S., Zhang, Z., Sinha, R., Karni, R. & Krainer, A.R. SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nat. Struct. Mol. Biol.* **17**, 306–312 (2010).
40. Ni, J.Z. *et al.* Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**, 708–718 (2007).
41. Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C. & Brenner, S.E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**, 926–929 (2007).
42. Thomas, P.D. *et al.* Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res.* **34**, W645–650 (2006).
43. Schaal, T.D. & Maniatis, T. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.* **19**, 1705–1719 (1999).
44. Cavaloc, Y., Bourgeois, C.F., Kister, L. & Stevenin, J. The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA* **5**, 468–483 (1999).
45. Zheng, Z.M., He, P.J. & Baker, C.C. Structural, functional, and protein binding analyses of bovine papillomavirus type 1 exonic splicing enhancers. *J. Virol.* **71**, 9096–9107 (1997).
46. Kittler, R. *et al.* Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nat. Methods* **4**, 337–344 (2007).
47. Penalva, L.O., Tenenbaum, S.A. & Keene, J.D. Gene expression analysis of messenger RNP complexes. *Methods Mol. Biol.* **257**, 125–134 (2004).
48. Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D. & Brown, P.O. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol.* **6**, e255 (2008).
49. Gama-Carvalho, M., Barbosa-Morais, N., Brodsky, A., Silver, P. & Carmo-Fonseca, M. Genome-wide identification of functionally distinct subsets of cellular mRNAs associated with two nucleocytoplasmic-shuttling mammalian splicing factors. *Genome Biol.* **7**, R113 (2006).
50. Hieronymus, H. & Silver, P.A. Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. *Nat. Genet.* **33**, 155–161 (2003).
51. Bjork, P. *et al.* Specific combinations of SR proteins associate with single pre-messenger RNAs *in vivo* and contribute different functions. *J. Cell Biol.* **184**, 555–568 (2009).
52. Talavera, D., Orozco, M. & de la Cruz, X. Alternative splicing of transcription factors' genes: beyond the increase of proteome diversity. *Comp. Funct. Genomics* doi:10.1155/2009/905894 (published online 12 July 2009).
53. Huang, Y. & Steitz, J.A. Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol. Cell* **7**, 899–905 (2001).
54. Huang, Y., Gattoni, R., Stévenin, J. & Steitz, J.A. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol. Cell* **11**, 837–843 (2003).
55. Blanchette, M., Labourier, E., Green, R.E., Brenner, S.E. & Rio, D.C. Genome-wide analysis reveals an unexpected function for the *Drosophila* splicing factor U2AF50 in the nuclear export of intronless mRNAs. *Mol. Cell* **14**, 775–786 (2004).
56. Boutz, P.L. *et al.* A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* **21**, 1636–1652 (2007).
57. Sanford, J.R., Gray, N.K., Beckmann, K. & Caceres, J.F. A novel role for shuttling SR proteins in mRNA translation. *Genes Dev.* **18**, 755–768 (2004).

ONLINE METHODS

Cell culture conditions and treatments. P19 cells were cultured in DMEM (4.5 g l^{-1} glucose), supplemented with 10% (v/v) FCS and penicillin and streptomycin under humified 5% CO_2 at 37 °C. The neural differentiation was induced by allowing the cells to aggregate in the presence of 1 μM retinoic acid for 4 d, after which the aggregates were dissociated and cultured on cell-culture dishes first for 24 h and then for 36–48 h in the presence of cytosine arabinoside (Ara-C). After the removal of Ara-C, the cells were cultured until they were harvested on the 8th day after neural induction.

BAC constructs and cell lines. BACs harboring the genes encoding SRp20 (gene name *SFRS3*) and SRp75 (*SFRS4*) were ordered from BACPAC Resources Center (http://bacpac.chori.org). The EGFP-IRES-Neo cassette was PCR-amplified with primers carrying 50 nucleotides of homology to the targeting sequence. Recombineering of the BACs was performed as described[5,36]. The tagged BACs were transfected into P19 mouse teratocarcinoma cells using Effectene (Qiagen). After selection with geneticin, single-cell clones were selected using fluorescence-associated cell sorting. The expression level of the tagged protein was determined with western blot analysis, RT-PCR and RT-qPCR.

Immunocytochemistry. The undifferentiated or differentiated cells were fixed with 4% (w/v) paraformaldehyde and stained for neural markers. The antibodies used were TU-20 (Abcam) against neural β-III-tubulin and the secondary antibody anti–mouse Alexa 568 (Invitrogen). The samples were counterstained for 4′-6-diamidino-2-phenylindole (DAPI) to mark the nucleus.

SDS-PAGE and western blotting. Whole-cell extracts were prepared using NET-2 buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% (v/v) Nonidet P-40 and Complete protease inhibitor cocktail, Roche). We separated 50–200 μg of total protein on 4–12% NuPage gradient precast gel (Invitrogen) or 10% (w/v) SDS-PAGE gel. The proteins were transferred onto nitrocellulose membranes and probed with the antibodies indicated. The antibodies used were mAb104 (ref. 37) against the SR protein family and mAb7B4 against SRp20 (ref. 38), goat anti-GFP (a kind gift of D. Drechsel), mouse anti-GAPDH (Novus) and mouse anti-actin (Sigma).

RNA extraction, cDNA synthesis and (q)PCR. Total RNA was extracted with acidic phenol:chloroform (Ambion). RNA was pretreated with DNaseI before reverse transcription. RNA (2.5–5.0 μg) was used in a 20-μl RT-PCR reaction with Superscript III reverse transcriptase (Invitrogen) using oligo-d(T)$_{18}$ reverse primers. One-twentieth of the reaction was used as a template for conventional PCR and 1/100 for quantitative PCR. The primer sequences used in PCR analysis are available upon request.

RNA immunoprecipitation. All steps for RIPs were performed at 4 °C. The cell pellet was suspended into RIP lysis buffer (20 mM Tris-HCl, pH 7.5, 0.2 M KCl, 2 mM MgCl$_2$, 1% (v/v) Nonidet P-40, complete protease inhibitor cocktail and RNAseOUT (Invitrogen)). The cells were forced through a 27G needle 20 times, and the extract was cleared by centrifugation. We stored 10% (v/v) of the extract as the input, and the remaining extract was used for the immunoprecipitation with goat anti-GFP antibody or goat nonimmune IgG. The extracts were diluted with the lysis buffer (Nonidet P-40 concentration decreased to 0.5% (v/v)). After incubation with antibodies, Gamma Bind protein G beads (GE Healthcare) were added and the immunoprecipitated complexes were collected. The beads with bound immunocomplexes were washed five times with the lysis buffer

(0.5% (v/v) Nonidet P-40). The bound RNA was eluted by adding NET-2 buffer, 1% (w/v) SDS, 20 μg glycogen (Roche) and acidic phenol:chloroform (Ambion) at 37 °C for 1 h. After the elution, RNA was precipitated with ethanol and suspended into RNase-free water. RNA from the input samples was extracted as from the immunoprecipitated samples. The RNA was used in a 20-μl RT-PCR reaction as above.

RIP microarray analysis. Total RNA or RIP RNA was extracted as described. The integrity of the samples was verified with 2100 Bioanalyzer (Agilent Technologies). The contaminating rRNA traces were used to control the integrity of the RIP RNA. From experience, samples lacking rRNA traces are of poor quality. The RNA was amplified and labeled with Low RNA Input Linear Amplification Kit Plus and hybridized onto mouse Whole Genome Microarrays (Agilent Technologies). The raw data was obtained with Agilent Feature Extraction Software. To analyze the RIP-chip data, we established an analysis protocol to identify the enriched RNAs (see also **Supplementary Fig. 5**). Signal intensities of the same sample type were quantile-normalized (specific immunoprecipitation, input or mock immunoprecipitation). The first inflection point of the intensity histogram of the input was used as a cutoff to remove low values (nonexpressed genes). The median was taken across the replicates and used to calculate a RIP score (log$_2$(IP/input)) for each probe. The RIP scores were calculated for both specific and mock immunoprecipitations. The statistical significance of the RIP scores was determined with a one-way ANOVA between the input, mock immunoprecipitation and specific immunoprecipitation samples followed by a correction for multiple hypotheses testing (FDR). For the hit calling, a RIP score and FDR thresholds were set, requiring the specific immunoprecipitation signals to be significantly greater than the mock immunoprecipitation values. The raw and quantile-normalized microarray data are deposited in the ArrayExpress at EPI.

Production and transfection of endoribonuclease-prepared siRNA. The template sequences for esiRNA production were chosen using the Deqor software[58]. The *SFRS4* esiRNA template sequence (nucleotides 1030–1452 of NM_020587) and the *SFRS3* template sequence (nucleotides 545–1054 of NM_013663) contain no perfect cross-silencers to known genes, as predicted by Deqor. The template sequences are given in **Supplementary Methods**. The esiRNAs targeting mouse *SFRS3* or *SFRS4*, or *GFP* as a negative control, were produced according to previous work[59] and transfected into P19 cells with Lipofectamine 2000 (Invitrogen). The cells were collected 24 h after transfection.

Gene expression microarray analysis after RNAi. RNAi was performed as described above. The samples were processed as for the RIP-chip. The raw data was processed with Agilent Feature Extraction Software. The data was further processed with Partek Genomics Suite 6.5 (Partek, Inc.). The arrays were quantile-normalized across the samples, and the median was taken from the three replicate samples. The significantly changed genes were determined by one-way ANOVA comparing the three samples (control, SRp75 and SRp20 RNAi), and a fold-change threshold was set. The raw and quantile-normalized microarray data are deposited in the ArrayExpress at EPI.

58. Henschel, A., Buchholz, F. & Habermann, B. DEQOR: a web-based tool for the design and quality control of siRNAs. *Nucleic Acids Res.* **32**, W113–W120 (2004).
59. Kittler, R., Heninger, A.-K., Franke, K., Habermann, B. & Buchholz, F. Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. *Nat. Methods* **2**, 779–784 (2005).

SUPPLEMENTARY MATERIAL

Global Analysis Reveals SRp20- and SRp75-specific mRNPs in Cycling and Neural Cells

Minna-Liisa Änkö, Lucia Morales, Ian Henry, Andreas Beyer and Karla M. Neugebauer

Sequence comparison of mouse and human SRp20 (SFRS3 gene)

```
NPS@ Network Protein Sequence Analysis.
http://npsa-pbil.ibcp.fr

Alignment data :
Alignment length : 164
Identity (*) : 164 is 100.00 %
Strongly similar (:) : 0 is 0.00 %
Weakly similar (.) : 0 is 0.00 %
Different : 0 is 0.00 %
Sequence 0001 : mousex0 ( 164 residues).
Sequence 0002 : humanx1 ( 164 residues).


                    10        20        30        40        50        60
                     |         |         |         |         |         |
mouse         MHRDSCPLDCKVYVGNLGNNGNKTELERAFGYYGPLRSVWVARNPPGFAFVEFEDPRDAA
human         MHRDSCPLDCKVYVGNLGNNGNKTELERAFGYYGPLRSVWVARNPPGFAFVEFEDPRDAA
Prim.cons.    MHRDSCPLDCKVYVGNLGNNGNKTELERAFGYYGPLRSVWVARNPPGFAFVEFEDPRDAA
Homology      ************************************************************

                    70        80        90        100       110       120
                     |         |         |         |         |         |
mouse         DAVRELDGRTLCGCRVRVELSNGEKRSRNRGPPPSWGRRPRDDYRRRSPPPRRRSPRRRS
human         DAVRELDGRTLCGCRVRVELSNGEKRSRNRGPPPSWGRRPRDDYRRRSPPPRRRSPRRRS
Prim.cons.    DAVRELDGRTLCGCRVRVELSNGEKRSRNRGPPPSWGRRPRDDYRRRSPPPRRRSPRRRS
Homology      ************************************************************

                    130       140       150       160
                     |         |         |         |
mouse         FSRSRSRSLSRDRRRERSLSRERNHKPSRSFSRSRSRSRSNERK
human         FSRSRSRSLSRDRRRERSLSRERNHKPSRSFSRSRSRSRSNERK
Prim.cons.    FSRSRSRSLSRDRRRERSLSRERNHKPSRSFSRSRSRSRSNERK
Homology      *******************************************
```

Supplementary Figure 1. Sequence comparison of human and mouse SRp20 and SRp75 proteins. The sequence comparison shows that human and mouse SRp20 proteins are identical and SRp75 proteins 96% identical in sequence. The differences between human and mouse SRp75 lie within the repetitive RS domain.

Sequence comparison of mouse and human SRp75 (SFRS4 gene)
NPS@ Network Protein Sequence Analysis.
http://npsa-pbil.ibcp.fr

Alignment data :SFRS4
Alignment length : 497
Identity (*) : 434 is 87.32 %
Strongly similar (:) : 21 is 4.23 %
Weakly similar (.) : 24 is 4.83 %
Different : 18 is 3.62 %
Sequence 0001 : mouse (491 residues).
Sequence 0002 : human (494 residues).

```
                    10        20        30        40        50        60
                     |         |         |         |         |         |
mouse       MPRVYIGRLSYQARERDVERFFKGYGKILEVDLKNGYGFVEFDDLRDADDAVYELNGKDL
human       MPRVYIGRLSYQARERDVERFFKGYGKILEVDLKNGYGFVEFDDLRDADDAVYELNGKDL
Prim.cons.  MPRVYIGRLSYQARERDVERFFKGYGKILEVDLKNGYGFVEFDDLRDADDAVYELNGKDL
Homology    ************************************************************

                    70        80        90       100       110       120
                     |         |         |         |         |         |
mouse       CGERVIVEHARGPRRDGSYGSGRSGYGYRRSGRDKYGPPTRTEYRLIVENLSSRCSWQDL
human       CGERVIVEHARGPRRDGSYGSGRSGYGYRRSGRDKYGPPTRTEYRLIVENLSSRCSWQDL
Prim.cons.  CGERVIVEHARGPRRDGSYGSGRSGYGYRRSGRDKYGPPTRTEYRLIVENLSSRCSWQDL
Homology    ************************************************************

                   130       140       150       160       170       180
                     |         |         |         |         |         |
mouse       KDYMRQAGEVTYADAHKGRKNEGVIEFVSYSDMKRALEKLDGTEVNGRKIRLVEDKPGSR
human       KDYMRQAGEVTYADAHKGRKNEGVIEFVSYSDMKRALEKLDGTEVNGRKIRLVEDKPGSR
Prim.cons.  KDYMRQAGEVTYADAHKGRKNEGVIEFVSYSDMKRALEKLDGTEVNGRKIRLVEDKPGSR
Homology    ************************************************************

                   190       200       210       220       230       240
                     |         |         |         |         |         |
mouse       RRRSYSRSRSHSRSRSRSRHSRKSRSRSGSSKSSHSKSRSRSRSGSHSRSKSRSRSQSRS
human       RRRSYSRSRSHSRSRSRSRHSRKSRSRSGSSKSSHSKSRSRSRSGSRSRSKSRSRSQSRS
Prim.cons.  RRRSYSRSRSHSRSRSRSRHSRKSRSRSGSSKSSHSKSRSRSRSGS2SRSKSRSRSQSRS
Homology    *********************************************:**************

                   250       260       270       280       290       300
                     |         |         |         |         |         |
mouse       RSKKEKSRSPSKDNKSRSRSRSPDKSRSKSKDHAEDKLQNNDSAGKAKSHSPSRHDSKS-
human       RSKKEKSRSPSKE-KSRSRSHSAGKSRSKSKDQAEEKIQNNDNVGKPKSRSPSRHKSKSK
Prim.cons.  RSKKEKSRSPSK2NKSRSRS2S22KSRSKSKD2AE2K2QNND22GK2KS2SPSRH2SKSK
Homology    ************: ******:*..********:**:*:****..**.**:*****.***

                   310       320       330       340       350       360
                     |         |         |         |         |         |
mouse       -RSRSQERRAEEEERRRSVSRARSQEKSRSQEKSLLKSRSRSRSKVGSRSRSRSKDKRKGR
human       SRSRSQERRVEEEKRGSVSRGRSQEKSLRQSRS--RSRSKGGSRSRSRSRSKSKDKRKGR
Prim.cons.  SRSRSQERR2EEE2R2SVSR2RSQEKS22Q22SLL2SRS222S222SRSRS2SKDKRKGR
Homology     ********.***:* ****.******  *.:*   :***:. *:  *****:********

                   370       380       390       400       410       420
                     |         |         |         |         |         |
mouse       KRSRDESRSRSRS--KSERSRKHSSKRDSKVSSSSSKKKKDTDHS--RSPSRSVSKEREH
human       KRSREESRSRSRSKSERSRKRGSKRDSKAGSSKKKKEDTDRSQSRSPSRSVSKEREH
Prim.cons.  KRSR2ESRSRSRSKSKSERSRK22SKRDSK22SS22KKK2DTD2SQSRSPSRSVSKEREH
Homology    ****:********* *******:.******..**..***:***:*   ************

                   430       440       450       460       470       480
                     |         |         |         |         |         |
mouse       AKAESGQRESRAEGESEAPNPEPRARSRSTSKSKPNVPAESRSRSKSASKTRSRSKSPSR
human       AKSESSQREGRGESENAGTNQETRSRSRSNSKSKPNLPSESRSRSKSASKTRSRSKSRSR
Prim.cons.  AK2ES2QRE2R2E2E2222N2E2R2RSRS2SKSKPN2P2ESRSRSKSASKTRSRSKS2SR
Homology    **:**.***.*.*.*. ..* *.*:****.******:*:***************** **

                   490
                     |
mouse       SASRSPSRSRSRSHSRS
human       SASRSPSRSRSRSHSRS
Prim.cons.  SASRSPSRSRSRSHSRS
Homology    ****************
```

Supplementary Figure 1 continued.
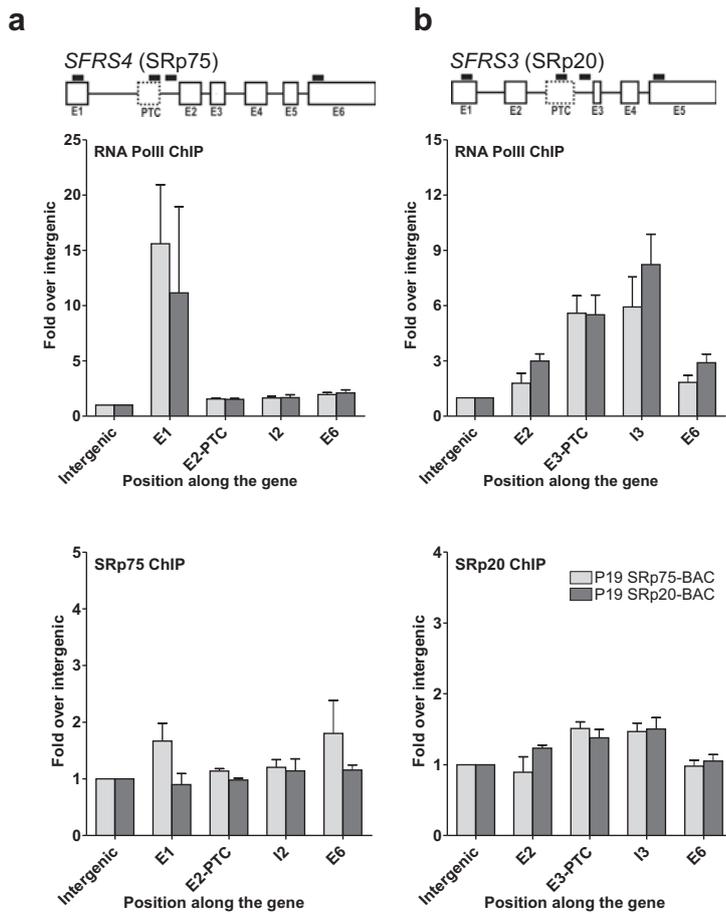
Supplementary Figure 2. ChIP analysis showing co-transcriptional recruitement of SRp20 and SRp75 to their genes. (a) Recruitment of SRp20 and SRp75 to SRp75 gene (*SFRS4*), and **(b)** to SRp20 gene (*SFRS3*). Error bars as s.e.m. and *n*=4-6. Positions of the primers used for the qPCR analysis are highlighted in the schematic representation of the genes analyzed. The exons containing the premature termination codons are marked with PTC.
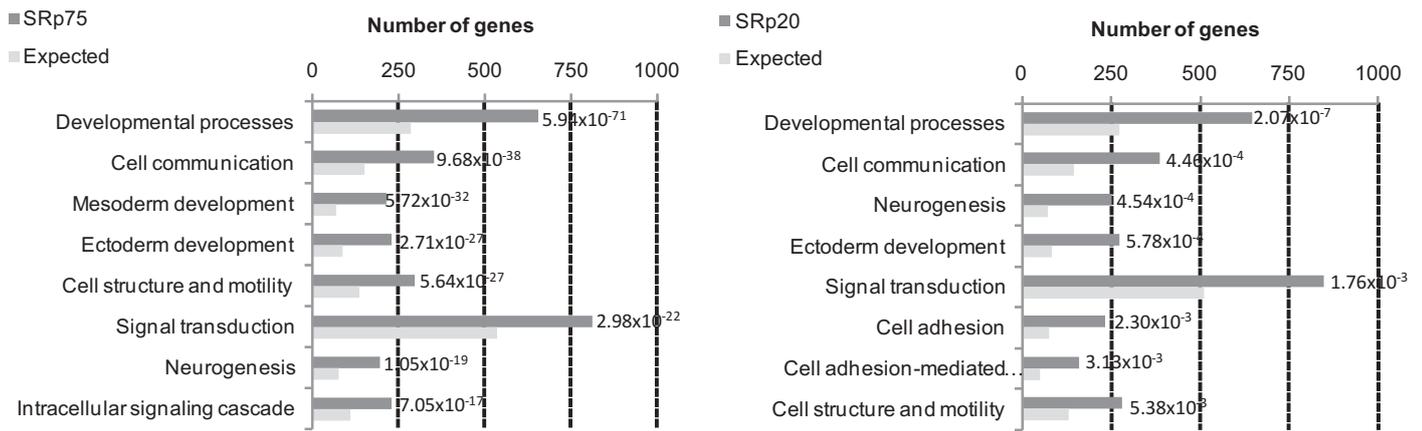
Supplementary Figure 3. Expression of genes involved in differentiation and development is increased in P19 BAC cell lines upon neural induction. Panther ontologies of mRNAs that increased in expression more than 2-fold upon neural differentiation. The top eight categories of biological process are shown for P19-SRp20-BAC and P19-SRp75-BAC cells. The dark grey bars show the observed number of genes in the GO category among the genes that increased in expression, and the light grey bars show the expected number of genes based on the size of the data set; p-value with Bonferroni correction for multiple testing is shown.
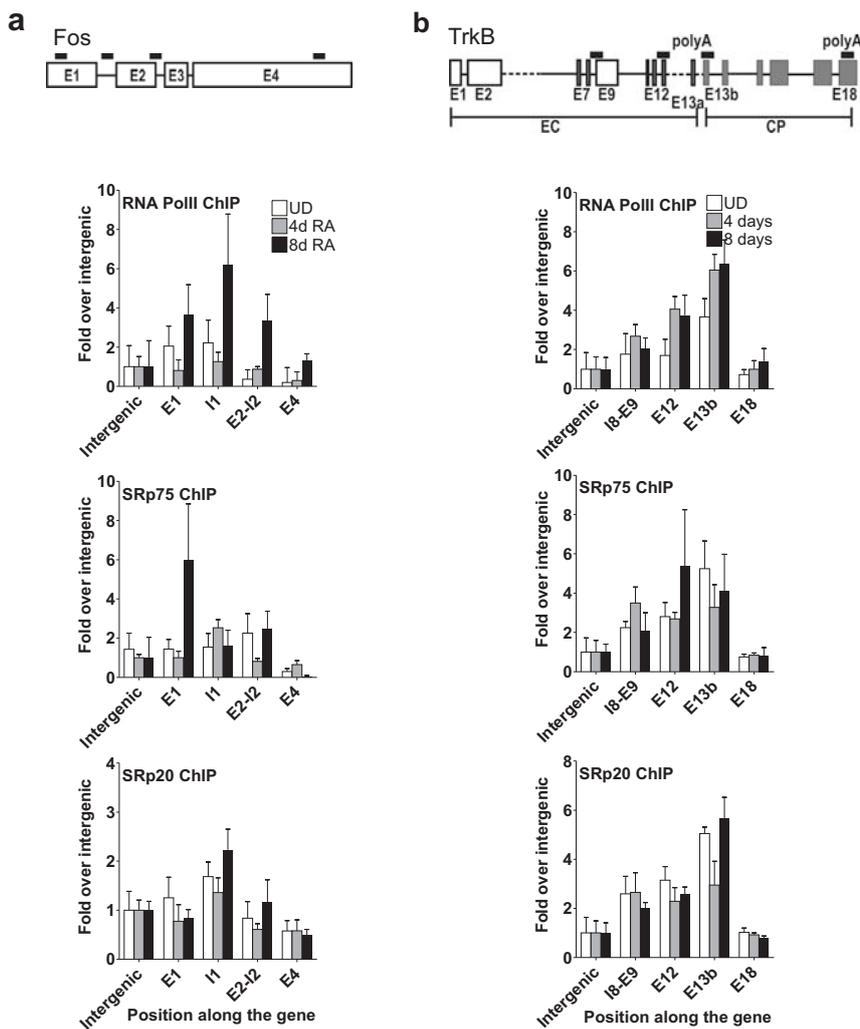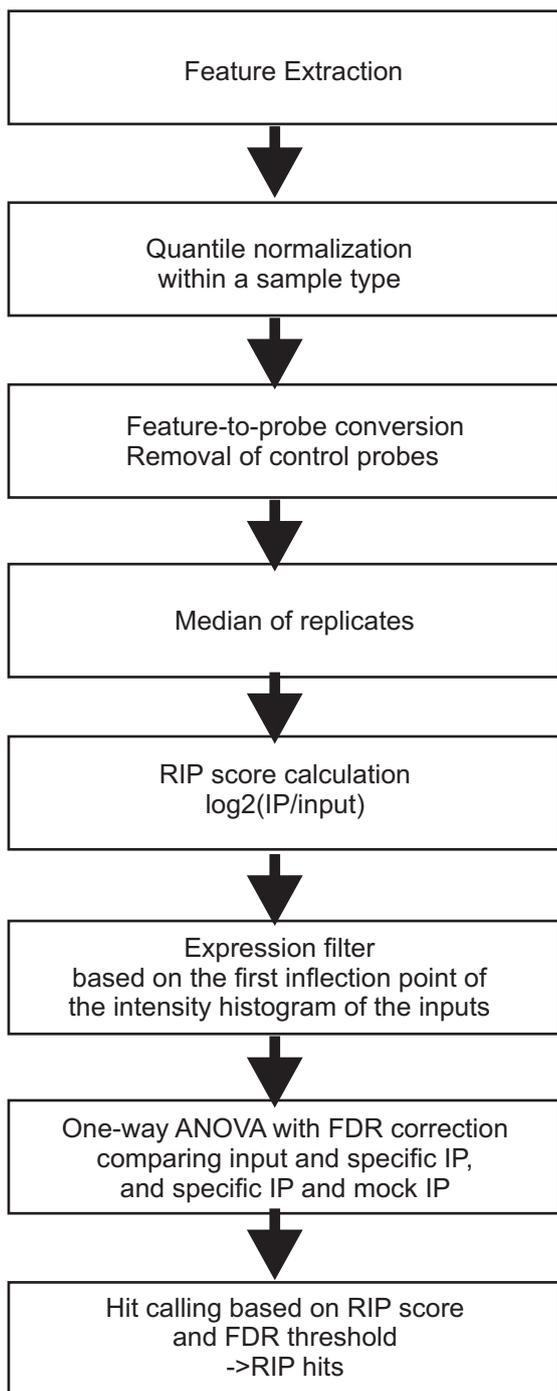
Supplementary Figure 4. ChIP analysis showing co-transcriptional recruitment of SRp20 and SRp75 to target genes. ChIP analysis of **(a)** *FOS*, and **(b)** *NTRK2* at different stages of neural differentiation in P19 cells. Error bars are s.e.m. and *n*=4-6. Positions of the primers used for the qPCR analysis are highlighted in the schematic representation of the genes analyzed. EC marks the part of *NTRK2* encoding the extracellular domain expressed in all cells, and CP the part of *NTRK2* encoding the cytoplasmic domain expressed only in neural cells.

```
┌─────────────────────────────────────┐
│                                     │
│         Feature Extraction          │
│                                     │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│       Quantile normalization        │
│        within a sample type         │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│     Feature-to-probe conversion     │
│      Removal of control probes      │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│         Median of replicates        │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│        RIP score calculation        │
│           log2(IP/input)            │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│          Expression filter          │
│   based on the first inflection point of │
│   the intensity histogram of the inputs │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│  One-way ANOVA with FDR correction  │
│     comparing input and specific IP,│
│      and specific IP and mock IP    │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│    Hit calling based on RIP score   │
│          and FDR threshold          │
│             ->RIP hits              │
└─────────────────────────────────────┘
```

**Supplementary Figure 5. A flow chart of the RIP-chip data analysis.**
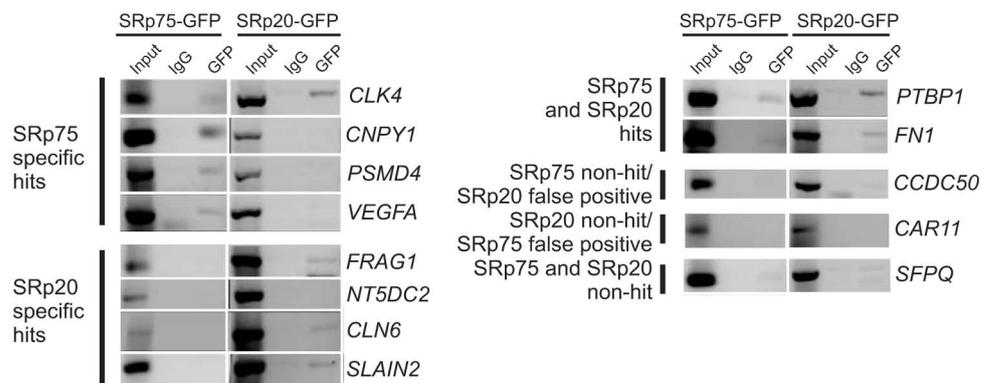
Supplementary Figure 6. Validation of RIP hits using formaldehyde cross-linking prior to RNA immunoprecipitation. The mRNAs shown in Figure 4d were analyzed with RT-PCR after crosslinked RIP. Stringent washing conditions were applied to the crosslinked RIP. The details of the experiment can be found in the Supplementary Methods.
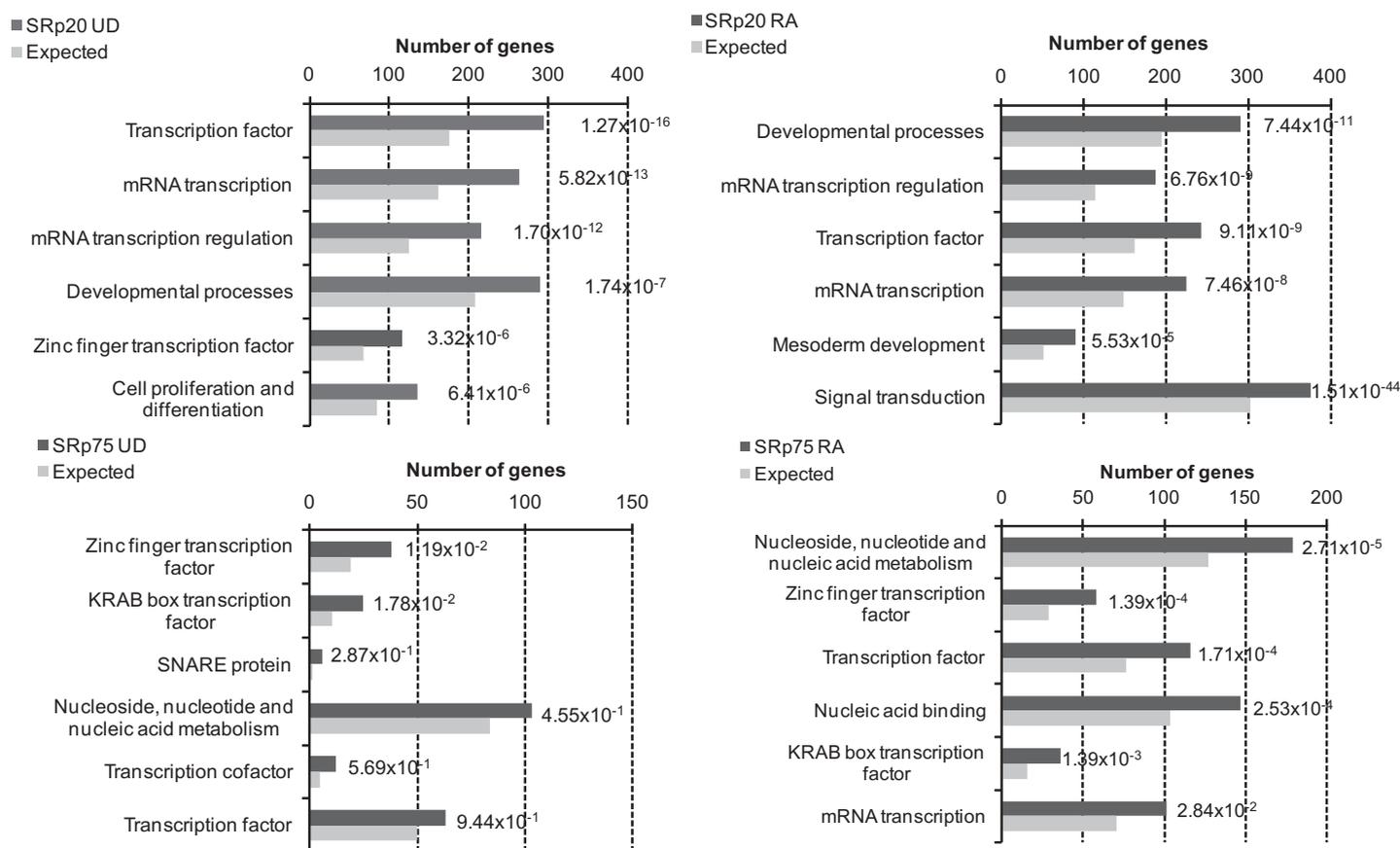
Supplementary Figure 7. SRp20 and SRp75 RIP hits are functionally related.

Panther ontologies of the hit sets presented in Figure 5 and Supplementary Table 2.

The top six categories detected are shown for each cell line and condition. The dark

grey bars show the number of genes in the GO category among the hit gene sets and

the light grey bars the expected number of genes based on the size of the data set in

each cell line and condition. UD=undifferentiated cells or RA=neural cells; p-value

with Bonferroni correction for multiple testing is shown.
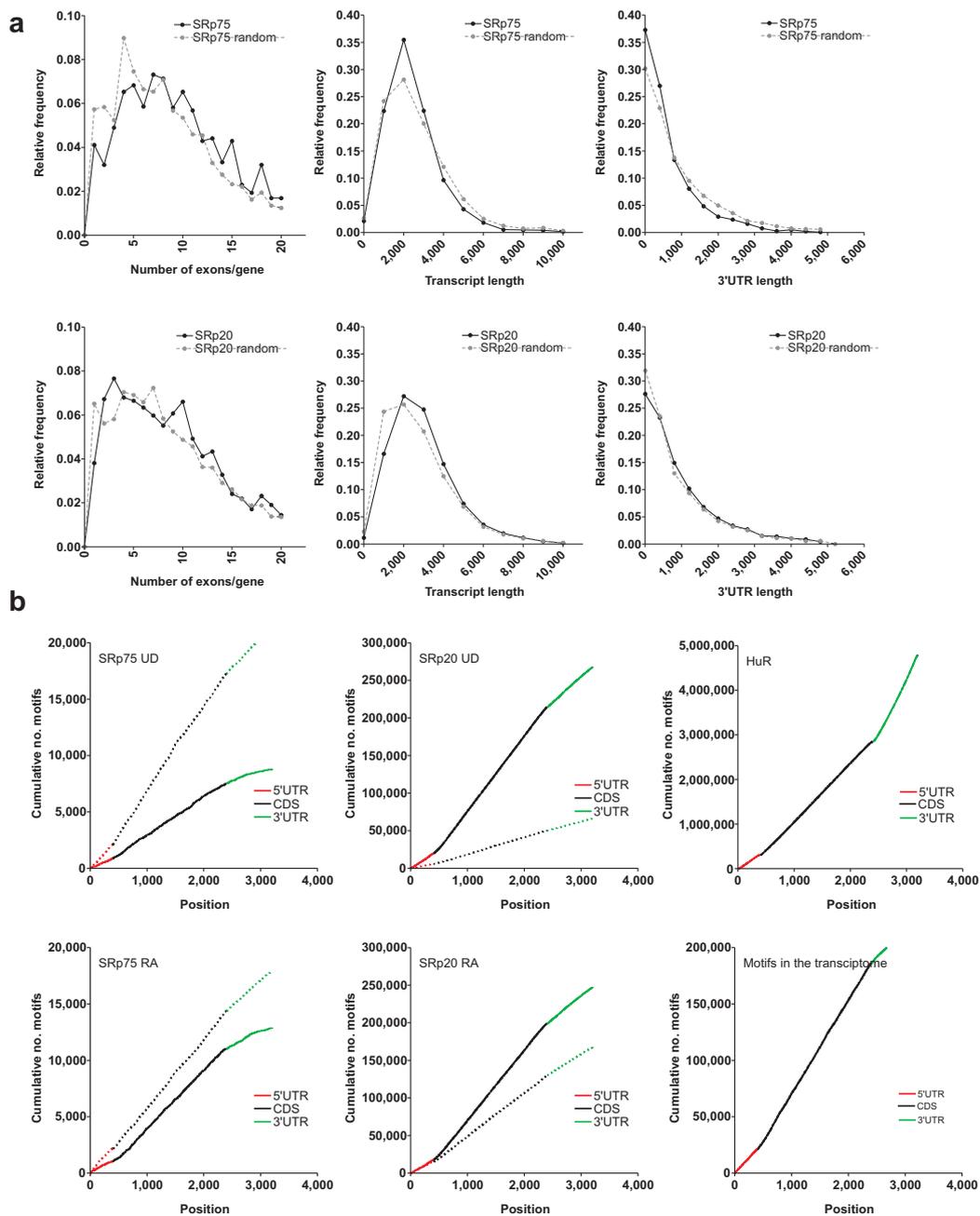
Supplementary Figure 8. Analysis of gene architecture and sequence features of the SRp20 and SRp75 RIP hits. (a) Frequency histograms representing the exon number, transcript length and 3'UTR length among the hit sets. The black solid line shows the specific hits and the gray dotted line the random data sets of equal size that were analyzed in parallel. **(b)** Distribution of the previously identified binding motifs of SRp20 and SRp75 along the 5' untranslated regions (5'UTR), coding sequence (CDS) and 3'UTRs within the hit mRNAs. The dashed lines in the graphs represent the analysis of SRp75 motifs within the SRp20 hits and *vice versa*. UD=undifferentiated cells or RA=neural cells. The distribution of the HuR binding motifs (expected to be biased to 3'UTRs) within all mouse genes was analyzed as a control. The distribution of the SRp20 and SRp75 motifs within the mouse transcriptome is also shown (the lines for SRp20 and SRp75 motifs completely overlapping).
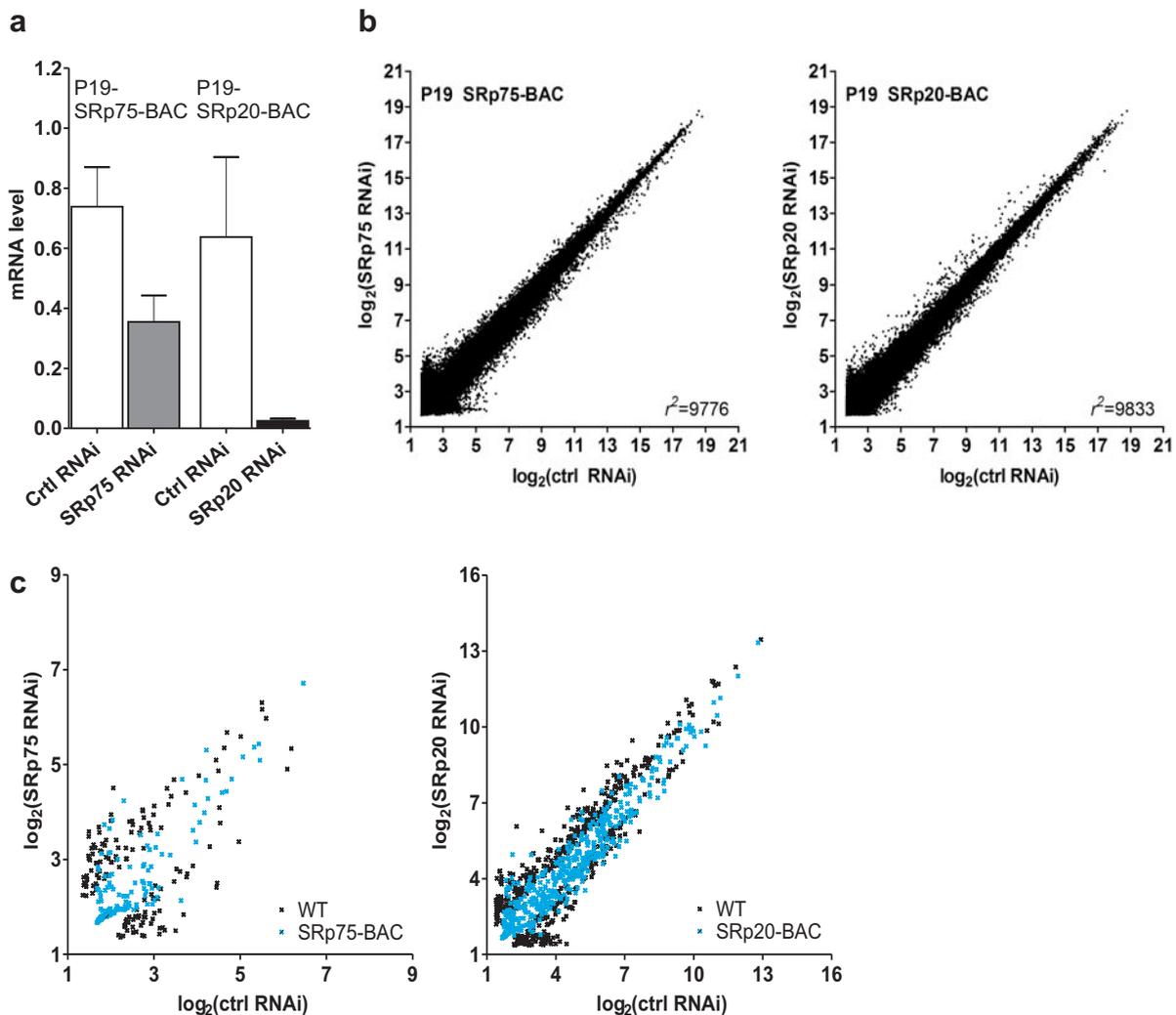
Supplementary Figure 9. Genome-wide phenotypic rescue of SR protein depletion by the BAC transgenes. SRp20 or SRp75 RNAi similar to done in Fig. 6 was performed in the P19-SRp20-BAC or P19-SRp75-BAC cells, respectively. **(a)** Quantification of knockdown efficiency by RT-qPCR. The primers used were specific for the endogenous SR protein targeted by the RNAi. **(b)** No significant changes in global gene expression were detected in BAC transgenic cells upon endogenous SRp20 or SRp75 depletion. Normalized signal intensities of the SR protein knockdowns are compared to control knockdowns. **(c)** The gene expression (normalized signal intensities) of the genes that changed in expression more than 1.5-fold with p-value<0.05 (One-way ANOVA) in the parental cells upon SRp20 or SRp75 depletion (see Fig. 6 and text) are presented in black crosses. The expression of the corresponding genes in the BAC cells upon knockdown are presented in blue crosses.
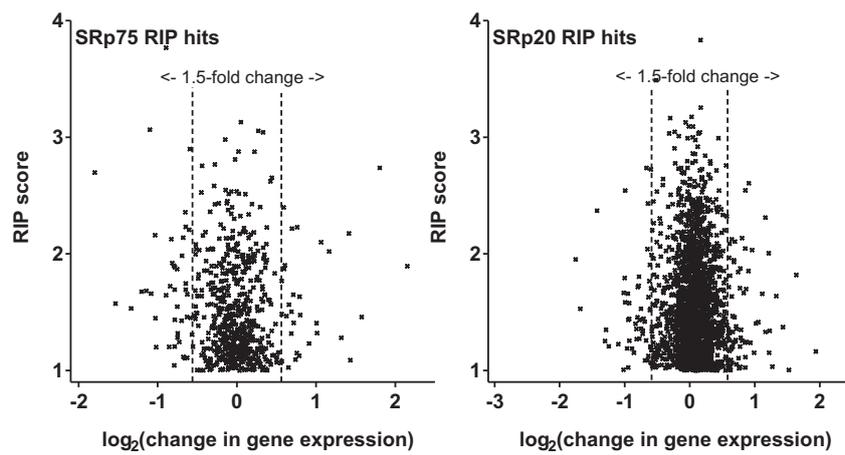
Supplementary Figure 10. SRp75 knockdown affects the expression of SRp20 RIP hits and SRp20 knockdown SRp75 RIP hits. Gene expression change of SRp75 RIP hits upon SRp20 knockdown (left) and SRp20 RIP hits upon SRp75 knockdown (right), data presented as in Fig. 6b.

Supplementary Table 1. Correlation of individual replicates of gene expression

microarray for undifferentiated P19 SRp20-BAC and P19 SRp75-BAC cells.

Comparison	r^2
SRp75-BAC R1 vs R2	0.9356
SRp75-BAC R1 vs R3	0.9280
SRp75-BAC R2 vs R3	0.9395
SRp20-BAC R1 vs R2	0.9879
SRp20-BAC R1 vs R3	0.9778
SRp20-BAC R2 vs R3	0.9735
Average r^2	0.9571

Supplementary Table 2. RIP hits of SRp20 and SRp75 in undifferentiated P19 cells and cells after 8 days of retinoic acid induction. Values as $\log_2$(intensity) or RIP score. The complete table can be found as a separate file.

Supplementary Table 3. Comparison of identified *in vivo* RIP-chip targets of SRp75 and SRp20 to known target genes. UD=undifferentiated cells, RA=neural cells and ND=not detected. The ones marked in bold have been shown to be regulated by SRp20 or SRp75 previously (references given).

Gene name	RIP-chip SRp75	RIP-chip SRp20	Expressed	References
CD44	ND	**RA**	UD, RA	[1]
FN1	**UD**	**UD, RA**	UD, RA	[2, 3, 4]
RAC1	ND	**ND**	UD, RA	[5]
SFRS3	ND	**ND**	UD, RA	[6]
FOS	**RA**	**RA**	(UD), RA	[7]
MAPT	ND	**ND**	UD, RA	[8]
CALCA	ND	**UD**	UD, RA	[9]

Supplementary Table 4. Relative density of SRp20 and SRp75 binding motifs among identified RIP hits and over the entire mouse transcriptome. The sequence motifs analyzed: SRp75 motif GAAGGA, SRp20 motif1 [A/T]C[A/T][A/T]C, SRp20 motif2 CTC[T/G]TC[C/T] and SRp20 motif3 GCTCCTCTTC. Relative density is the ratio of the average density of motifs in the given transcript region per unit length to the average density of motifs across all considered transcripts per unit length. First, it can be seen that the relative density of motifs in any gene region among the "hits" is very similar to the relative density of motifs for the same gene region across the transcriptome, indicating that motif enrichment cannot be used to predict experimentally determined hits. Second, the position of the detected motifs within transcripts reveals enrichment of motifs within the coding region (CDS) among the hits; however, this enrichment is also seen across the transcriptome, indicating that filtering of motif searches based on selected mRNA regions would not improve hit prediction.

RELATIVE DENSITY			
"Hits"	5'UTR	CDS	3'UTR
SRp75 UD	0.818	1.284	0.596
SRp75 RA	0.818	1.294	0.578
SRp20 UD m1	0.716	1.195	0.823
SRp20 UD m2	0.980	0.991	1.017
SRp20 UD m3	1.093	1.100	0.847
SRp20 RA m1	0.713	1.203	0.804
SRp20 RA m2	0.943	1.006	1.008
SRp20 RA m3	1.455	1.069	0.785

RELATIVE DENSITY			
Transcriptome	5'UTR	CDS	3'UTR
SRp75	0.933	1.244	0.666
SRp20 m1	0.786	1.131	0.866
SRp20 m2	1.127	0.944	1.049
SRp20 m3	1.130	1.120	0.795

Supplementary Table 5. List of genes that changed more than 1.5-fold in expression upon SRp75 or SRp20 knockdown compared to the control (p-value<0.05, one-way ANOVA). The complete table can be found as a separate file.

SUPPLEMENTARY METHODS

Chromatin immunoprecipitation and quantitative PCR

ChIP and real-time PCR protocols were according to[10]. Equal amount of extract from undifferentiated or differentiated cells was used per ChIP. Antibodies used for the ChIP were 4H8 against RNA polymerase II (Abcam or Santa Cruz Biotechnology) and anti-GFP (kind gift from Dr. David Drechsel). Mouse nonimmune IgG was used for the mock ChIPs. ChIP data is represented as %input after correction to the nonimmune control and as fold enrichment above an intergenic region on chromosome 13 where no annotated genes are found. Primer sequences used in the qPCR analysis can be obtained upon request.

Motif search for SRp20 and SRp75 binding sites

The probes from each of the four RIP target data sets (SRp20 UD, SRp20 RA, SRp75 UD and SRp75 RA) were blasted against the Ensembl mouse transcriptome (Ensembl v38). Those probes that uniquely matched individual genes were taken and all matching transcripts of that gene with annotated 5' and 3'UTR regions were considered further. Transcripts were scaled for easy comparison with the coding sequence (CDS) to a length of 2000 positions, the 5' UTR scaled to 400 positions and the 3'UTR scaled to 800 positions. The distribution of the individual motifs[11-13] was calculated for all 5' UTR, 3'UTR and CDS and were plotted onto the scaled transcripts. In addition, the number of motifs occurring at each of the scaled positions was calculated to identify regions where motifs were most commonly found. Cumulative distribution curves were plotted to show how the frequency of motifs altered at each scaled position along the transcript.

Further analysis was carried out to investigate whether motifs occurred more frequently in the UTRs or the CDS. To ensure that an unbiased comparison was made, the lengths of the UTRs and coding sequences were taken into account along with the total

1

number of motifs in the transcript. This was done by calculating the ratio of the average density of motifs in a transcript region (i.e. 5'UTR, CDS, 3'UTR) to the average density of motifs across the transcriptome.

Corrected motif frequency (f) for a region $f_{region} = \dfrac{N_{region} \times l_{region}}{N_{transcript} \times l_{transcript}}$

N = number of motifs, l = length of sequence

RNA immunoprecipitation with formaldehyde crosslinking

The cells on culture dishes were crosslinked with 0.1% (v/v) formaldehyde in phosphate buffered saline for 10min at room temperature. All subsequent steps of RNA immunoprecipitations were performed at +4°C. The cell pellet was suspended into RIP lysis buffer (20mM Tris-HCL pH 7.5, 0.2M KCl, 2mM MgCl$_2$, 1% (v/v) Nonidet P-40, Complete protease inhibitor cocktail and RNAseOUT (Invitrogen)). The crosslinked immunocomplexes were solubilized using sonication and the extract was cleared by centrifugation. Ten percent of the extract was stored as the input and the remaining extract was used for the immunoprecipitation with goat anti-GFP antibody or goat nonimmune IgG. The extracts were diluted with the lysis buffer (Nonidet P-40 concentration decreased to 0.5%, v/v). After incubation with antibodies, Gamma Bind protein G-beads (GE Healthcare) were added and the immunoprecipitated complexes were collected. The beads with bound immunocomplexes were washed using buffers with increasing stringency as in ChIP (see above). The samples were uncrosslinked for 45min at 70°C in NET-2 buffer with 1% (w/v) SDS. The RNA was extracted with acidic phenol:chloroform. RNA was precipited with ethanol and suspended into RNase-free water. RNA from the input samples was uncrosslinked and extracted as from the immunoprecipitated samples. The RNA was used in a 20μl RT-PCR reaction as in uncrosslinked RIP.

2

Template sequences for esiRNA production

Sequence used as template for the production of *SFRS3* esiRNA.

AAATCACAAGCCGTCTCGATCCTTCTCTAGGTCTCGTAGC
CGATCTAGGTCAAATGAAAGGAAATAGAAGACCAGTTTGC
AAAAGTGGTGTACAGGAAATAACTTCATCTGACAGGAGTA
TGTACAGGAAATTAAAGTTTTGTTTGAGACTTCATAAGCT
TGGTGCATTTTTAAGATGGTTTAGCTGTTTAAATTTGTTT
TGTCTCTTGGAACAGTGACACACAAAACAATGTAATTCTC
TATGGTTTTCAGATGGATCATAAGAGGCACGTGATATCAA
GAATTGTTACTTTACAATGTTCCCTTAAGCAAGATTTAAT
TTTCTTTGAATTTTAGTTTTTCATAGACTGAAATAAACCT
TAGGTCCTGCCCAGTTTTAAGTGTGATGTACTAATGATAT
AAAGCAACTGGCGGAAATTGAAAGAAGCTATAGTCCTCTA
GTAGCTGAGACACTGTGGCACTGTGGGTGGAATGATAAAG
CGGTGTTTAA<u>GAGCTGCTGTGAACACAAGC</u>

Sequence used as template for the production of *SFRS4* esiRNA.

<u>AAGGTCCTAACTGGCTCTGC</u>CACGCTGGAACTGCCG
AGAAGTGTTTTGTACATGTTGGTAGCCGTAGCACAAGAGT
GAAGTAGAACACCCGTCACTGCTGTACATTAACTCCCTAA
AGGTGTGTCTCAGTTGTTCAATCTCAGTGCTTCCTCGGTC
AGCCTCCAGGCGCCAGGCCTTCCCGCTCTACTGAAAGCAG
CTCCTCAGACCTCCCCTTACTCACAGTAGGACACCCAGAC
GCCTGCCTTTCAGGCCTGGCCACGGCTATAGGGAGCTCGG
CACCCAGACGGCTGGCTTCTCAGGCTGGAGTGATGGCGTA
GGTAGGTGTGCTGAGCTCAGCCGTCTGCCCTTGAATCGAT
GCCCTTTGATGTGTGCCACGTAGTGAAAGTGCAAGTCTTC
AGTCTCC<u>CACCACTTCGGTTTCCTGTT</u>

The underlined sequences were used for PCR priming.

SUPPLEMENTARY REFERENCES

1.   Galiana-Arnoux, D. et al. The CD44 alternative v9 exon contains a splicing enhancer responsive to the SR proteins 9G8, ASF/SF2, and SRp20. *J. Biol. Chem.* **278**, 32943-32953 (2003).

2.   Cramer, P. et al. Coupling of transcription with alternative splicing: RNA Pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Molecular Cell* **4**, 251-258 (1999).

3.   Lim, L.P. & Sharp, P.A. Alternative splicing of the fibronectin EIIIB exon depends on specific TGCATG repeats. *Mol. Cell. Biol.* **18**, 3900-3906 (1998).

4.   Liang, H., Tuan, R.S. & Norton, P.A. Overexpression of SR proteins and splice variants modulates chondrogenesis. *Experimental Cell Research* **313**, 1509-1517 (2007).

5.   Goncalves, V., Matos, P. & Jordan, P. Antagonistic SR proteins regulate alternative splicing of tumor-related Rac1b downstream of the PI3-kinase and Wnt pathways. *Hum. Mol. Genet.*, ddp317 (2009).

6.   Jumaa, H. & Nielsen, P. The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO Journal* **16**, 5077-85 (1997).

7.   Sapra, A.K. et al. SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo. *Mol Cell* **34**, 179-190 (2009).

8.   Mabon, S.A. & Misteli, T. Differential recruitment of pre-mRNA splicing factors to alternatively spliced transcripts in vivo. *PLoS Biology* **3**, e374 (2005).

9.   Lou, H., Neugebauer, K.M., Gagel, R.F. & Berget, S.M. Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Mol. Cell. Biol.* **18**, 4977-4985 (1998).

10.   Listerman, I., Sapra, A.K. & Neugebauer, K.M. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat Struct Mol Biol* **13**, 815-822 (2006).

11.   Schaal, T.D. & Maniatis, T. Selection and characterization of pre-mRNA splicing enhancers: Identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.* **19**, 1705-1719 (1999).

12.   Cavaloc, Y., Bourgeois, C.F., Kister, L. & Stevenin, J. The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA* **5**, 468-483 (1999).

13.   Zheng, Z.M., He, P.J. & Baker, C.C. Structural, functional, and protein binding analyses of bovine papillomavirus type 1 exonic splicing enhancers. *J. Virol.* **71**, 9096-9107 (1997).