3-1-2010

An Exonic Splicing Enhancer within a Bidirectional Coding Sequence Regulates Alternative Splicing of an Antisense mRNA

Valerie K. Salato
Marquette University

Nathaniel W. Rediske
Marquette University

Chao Zhang
Marquette University

Michelle Laura Hastings
Chicago Medical School

Stephen Munroe
Marquette University, stephen.munroe@marquette.edu

An Exonic Splicing Enhancer within a Bidirectional Coding Sequence Regulates Alternative Splicing Of an Antisense mRNA

Valerie K. Salato

Department of Biological Sciences, Marquette University
Milwaukee, WI

Nathaniel W. Rediske

Department of Biological Sciences, Marquette University
Milwaukee, WI

Chao Zhang

Department of Biological Sciences, Marquette University
Milwaukee, WI

Michelle L. Hastings

Department of Biological Sciences, Marquette University
Milwaukee, WI

Stephen H. Munroe

Department of Biological Sciences, Marquette University
Milwaukee, WI

Abstract: The discovery of increasing numbers of genes with overlapping sequences highlights the problem of expression in the context of constraining regulatory elements from more than one gene. This study identifies regulatory sequences encompassed within two genes that overlap in an antisense orientation at their 3′ ends. The genes encode the α-thyroid hormone receptor gene (TRα or NR1A1) and Rev-erbα (NR1D1). In mammals TRα pre-mRNAs are alternatively spliced to yield mRNAs encoding functionally antagonistic proteins: TRα1, an authentic thyroid hormone receptor; and TRα2, a non-hormone-binding variant that acts as a repressor. TRα2-specific splicing requires two regulatory elements that overlap with Rev-erbα sequences. Functional mapping of these elements reveals minimal splicing enhancer elements that have evolved within the constraints of the overlapping Rev-erbα sequence. These results provide insight into the evolution of regulatory elements within the context of bidirectional coding sequences. They also demonstrate the ability of the genetic code to accommodate multiple layers of information within a given sequence, an important property of the code recently suggested on theoretical grounds.

Keywords: alternative splicing, exonic splicing enhancer, intronic splicing enhancer, antisense RNA, genetic code, exon evolution, nuclear receptor proteins

Introduction

A single genomic locus can give rise to more than one distinct protein product through alternative RNA splicing, alternative promoters and differential 3′ end processing.[1] Because transcription can occur on either strand of the DNA, a single genomic locus can also produce RNAs from the opposite DNA strand, resulting in potential formation of complementary transcripts termed cis-encoded natural antisense transcripts (cis-NATs or cis-antisense RNAs).[2, 3] Regulatory mechanisms associated with antisense transcription and alternative splicing are at present poorly understood. However, it is clear that the combination of alternative splicing and overlapping transcription greatly increases the complexity and information density of the human transcriptome.[4–8] It has been estimated that upwards of 95% of protein-coding transcription units give rise to alternatively spliced mRNAs[7, 9–11] and more than 70% of transcription units in the human genome are transcribed in both directions.[5, 12, 13]

In this study we examine requirements for alternative processing within the mammalian α-thyroid hormone (T3) receptor gene (TRα). This gene yields two alternatively spliced mRNAs that encode two functionally antagonistic proteins, TRα1 and TRα2.[14] TRα1 (NR1A1a) is an authentic nuclear receptor that together with the related receptors, TRβ1 (NR1A2a) and TRβ2 (NR1A2b), encoded by the TRβ gene, mediates the cellular response to T3.[15] Each of these nuclear receptors has an N-terminal DNA-binding domain that recognizes T3 response elements (TREs) and a C-terminal ligand-binding domain that binds T3 and activates T3-responsive genes. In contrast, TRα2 (NR1A1b) is a non-hormone-binding variant that differs from TRα1 in its C-terminal domain and neither binds T3 nor activates transcription. TRα2 is thought to modulate the response to T3 by dominant-negative interference with the hormone-responsive isoforms.[14, 16] Consistent with this hypothesis, expression of TRα2 is developmentally regulated in many tissues, including brain, where it is the predominant TRα isoform.[14, 17, 18]

The structure of TRα2 reflects its unique 3' exon, which is located downstream of the polyadenylation site for TRα1 mRNA. This exon overlaps the 3' exon of another mRNA on the opposite strand of the chromosome that encodes a third nuclear receptor, Rev-erbα (NR1D1).[19, 20] Although long regarded as a ligandless orphan receptor, Rev-erbα has recently been shown to bind heme specifically in its canonical ligand-binding domain.[21, 22] In the presence of this ligand, the interaction of Rev-erbα with co-repressors is responsive to physiological regulators such as nitric oxide. These results and the well-documented role of Rev-erbα as a component of the mammalian molecular clock demonstrate that Rev-erbα, like TRα1, serves as a critical physiological regulator.[21–24] The overlap between TRα2 and Rev-erbα may have important consequences for the balance between TRα1 and TRα2, as indicated by *in vivo* and *in vitro* analysis.[25–28]

Several recent studies have demonstrated that requirements for splicing regulatory elements constrain the coding sequence of alternatively spliced mRNAs.[29–32] Since the overlap between the Rev-erbα and TRα genes includes a portion of the C-terminal coding sequence of both genes, it seemed likely that requirements for coding might represent an overriding constraint that restricts regulatory elements for TRα2 splicing to sites outside of the overlap. Contrary to

this expectation, experiments reported here define two splicing enhancers, TR-ISE3 and ESX10, that are located within the overlap between TRα2 and Rev-erbα where they flank the TRα2-specific 3' splice site (ss). Most notably, ESX10 maps directly to the bidirectional coding sequence (BCS) associated with the overlap between coding sequences for both TRα2 and Rev-erbα. These results provide a striking demonstration of the ability of the genetic code to accommodate multiple layers of information and raise interesting questions concerning the evolution of this gene overlap.[33]

Results

Sequence and structure of the TRα and Rev-erbα overlap

As shown in Figure 1, TRα1 mRNA is formed by splicing exon 1 through exon 9 whereas TRα2 mRNA is formed by splicing the 5' end of exon 9 (exon 9A) to exon 10. Exon 10 overlaps the final exon of Rev-erbα such that the 3' end of Rev-erbα mRNA overlaps the 3'ss of TRα2 mRNA and vice versa. Together the two transcription units occupy 34 kb on rat chromosome 10 with a 1.1 kb overlap. The 263 bp exon-exon overlap between TRα2 and Rev-erbα includes the 200 bp BCS comprising 67 codons at the C-terminus of each protein that share an in-phase overlap. Furthermore the poly(A) site of TRα2 mRNA maps within 50 bp of the final 5' ss of Rev-erbα mRNA, and that of Rev-erbα mRNA maps between the poly(A) site of TRα1 and the 3'ss of exon 10 imposing additional constraints on this region.
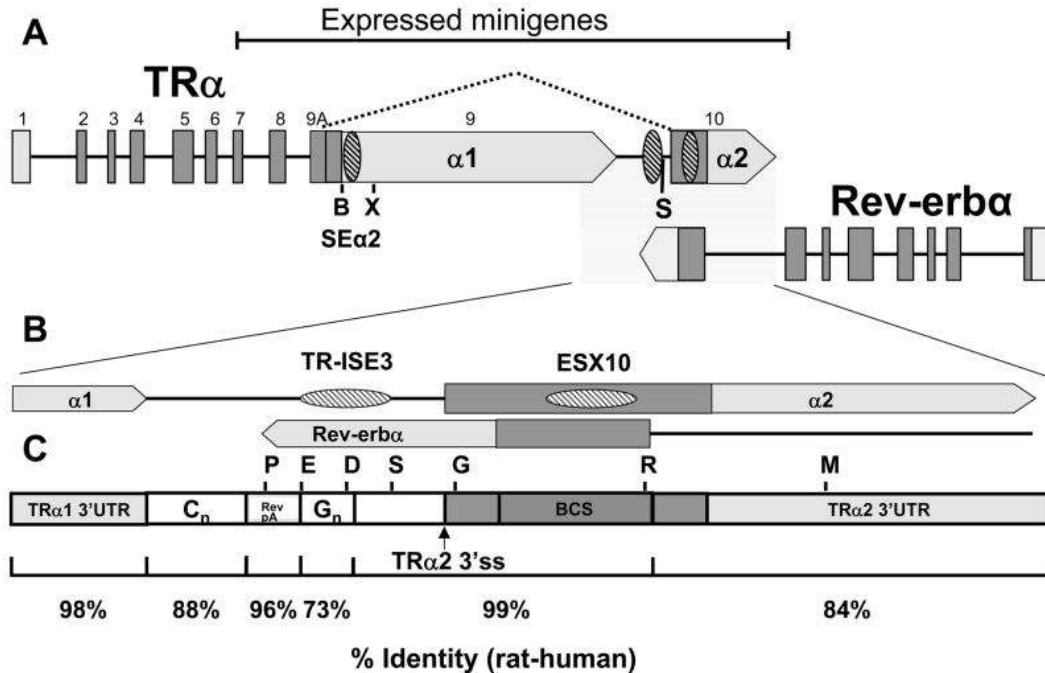
Figure 1 Structure and processing of TRα2 mRNA. **A.** Schematic representation of the structure of the TRα and Rev-erbα genes which overlap by 1.1 kb on rat chromosome 10 (position 316154–338490; reference assembly RGSC v3.4). Boxes represent exons which are numbered for TRα, pointed boxes are 3' terminal exons and horizontal lines are introns. The positions of three splicing enhancers including SEα2, previously described [34], TR-ISE3 and ESX10, are represented by hatched ellipses. Splicing of exons 9A and 10 in TRα2 mRNA is indicated by a dotted angled line. Dark gray shading represents coding sequences, light gray shading untranslated regions. The position of the minigenes cloned in pErbA is indicated by a bracket at top. Bold letters (B,S, X) indicate endpoints of deletions. The TRα/Rev-erbα overlap region is highlighted by shading. **B.** To-scale representation of the 1.4 kb region shaded in Panel A. Splicing enhancers TR-ISE3 and ESX10 are indicated by hatched ellipses. **C.** Segmented bar indicates regions defined by processing sites, base composition and conservation. Segments are labeled as follows: pA is poly(A) site; Rev is Rev-erbα; C_n and G_n denote C-rich and G-rich regions, respectively; BCS is bidirectional coding sequence; and 3' UTR is the 3' untranslated region. Single letters above bar indicate endpoints of deletions within minigenes. Numbers at bottom indicate sequence identity between rat and human.

As a first step towards understanding constraints introduced by the overlap between these genes we compared the genomic sequences of six mammals (human, chimp, mouse, rat, dog and cow) across the region including the overlap between TRα2 and Rev-erbα mRNAs and the adjacent 3' sequence of TRα1 mRNA (Supplementary data, Table S1). Although sequences throughout the region where TRα1, TRα2 and Rev-erbα mRNAs converge are highly conserved, local variations in

sequence composition and conservation reflect diverse constraints. Within the region highlighted in [Figure 1C](#) there are three areas of very high conservation, illustrated here for the rat-human alignment. These include a stretch of 200 bp at the 3' end of exon 9 of TRα1 mRNA, 100 bp upstream of the 3'ss for exon 10 in TRα2 mRNA and 300 bp at the 5' end of exon 10 ([Table S1](#)). Between the TRα1 poly(A) site and the 3' exon of TRα2 70 bp of sequence antisense the Rev-erbα poly(A) site are also highly conserved (96% identity). This region is flanked upstream and downstream by C-rich and G-rich regions (C_n and G_n, respectively; [Figure 1C](#)) that are less tightly conserved with numerous runs of three or more consecutive C or G residues. These features reflect multiple, overlapping requirements for transcription, processing and translation on one or both strands.

The TRα2-specific 3'splice site is suboptimal

To understand functional constraints introduced by the overlap with Rev-erbα and how they may relate to regulation of TRα expression, we examined sequences directly involved in TRα2 splicing. Many alternative splice sites are suboptimal in their match to consensus sequences and their efficiency of splicing. For example, a previous study showed that replacement of a single non-consensus nucleotide at the 5'ss of TRα2 (+5C/G) greatly increases TRα2 splicing [28]. Since the 3'ss of exon 10 is also a poor match to the consensus sequence, with AAG (vs. CAG) immediately preceding the splice site ([Figure 2A](#)), the non-consensus adenosine was replaced with cytidine (−3A/C) in a minigene construct, pErbAΔBS, that lacks the upstream SEα2 enhancer required for efficient splicing.[28, 34] This single-nucleotide substitution results in a substantial increase in splicing, although it is less than that seen when the +5C/G mutation is incorporated into the 5'ss of this construct ([Figure 2B](#)). These results demonstrate that both 3'ss and 5'ss are suboptimal and contribute to the regulation of TRα1/TRα2 mRNA levels. The lower splicing efficiency of the 3'ss substitution may reflect other suboptimal features of this site such as a non-consensus branch point (Hastings, unpublished result).
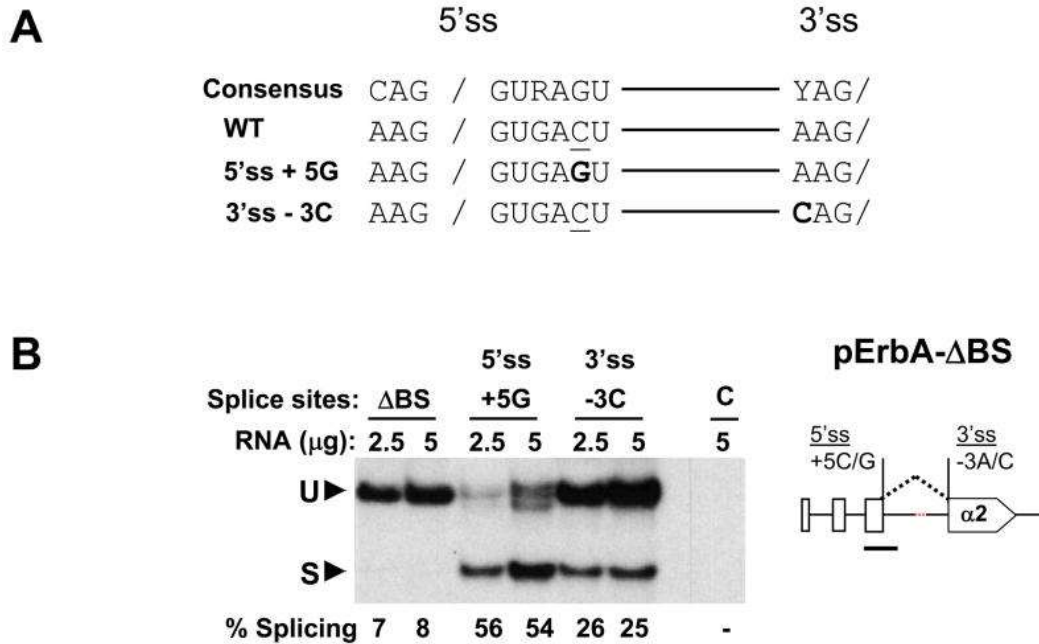
Figure 2 The TRα2-specific splice sites are suboptimal. **A.** Sequences of 5' and 3' splice site sequences and corresponding point mutations. **B.** RNase protection assays of TRα2-splicing of wildtype and mutant ErbAΔBS pre-mRNAs in HEK293 transfected cells. The control lane (C) shows cells transfected with vector plasmid express no detectable endogenous TRα1 or TRα2 mRNA. Structure of the ErbAΔBS minigene is shown at right with the position of the probe indicated by the bold line. Arrows labeled "U" and "S" indicate bands corresponding to unspliced and spliced RNAs, respectively, and % splicing is shown below each lane. Double bands in some lanes represent partial protection of ends of RNase trimmed probes hybridized to unspliced RNA. Results are representative of two independent experiments.

Sequences at the 3' end of the regulated intron enhance TRα2 mRNA splicing

Because suboptimal splice sites often are regulated by intronic or exonic splicing enhancer elements (ISEs or ESEs), requirements for sequences adjacent to the TRα2 3'ss in splicing regulation were evaluated through a series of deletions upstream of the 3' ss. Previous studies showed that one such splicing enhancer element, SEα2, lies within an 80 nt region directly downstream of the TRα1 coding sequence.[28, 34] Here, two series of deletions in the 3'ss region were examined that differ in the presence or absence of the SEα2 splicing enhancer[34], one derived from pErbAΔXS and the other from pErbAΔBS (Figure 3A). RNase protection assays show a dramatic stimulation of TRα2 splicing in RNA from the ΔBD, ΔBE and ΔBP minigenes. These

results demonstrate that sequences upstream of the polypyrimidine also contribute to TRα2 splicing efficiency (Figure 3B). A parallel increase was seen in splicing of transcripts ErbAΔXD, ErbAΔXE and ErbAΔXP compared with that of the parent construct, ErbAΔXS (Figure 3B). As the difference in splicing observed with ErbAΔBE and ErbAΔBP is small relative to that between ErbAΔBS and ErbAΔBD or ErbAΔBE (with parallel results for the series ErbAΔXS through ErbAΔXP), we conclude that the sequences between site "E" and "P" in ΔBE and ΔBP contribute marginally to TRα2 splicing. Consequently, the ES fragment (present in ΔBE and ΔXE but absent in ΔBS and ΔXS) includes an intronic splicing enhancer which we designated TR-ISE3.
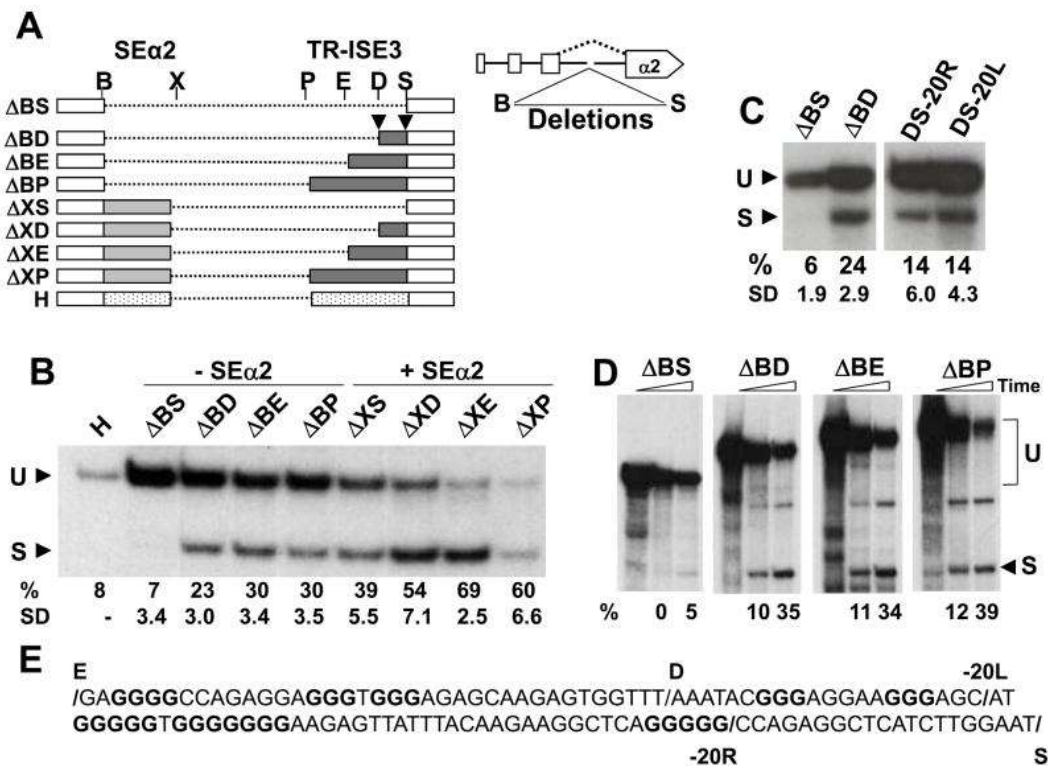

Figure 3 Intronic sequences adjacent to the TRα2 3'ss are required for efficient splicing of TRα2 mRNA. **A.** Structure of ErbA minigene pre-mRNA is shown on top and the structures of 9 deletion constructs within the final exon of TRα2 mRNA are shown below. Splicing of ErbAΔBS and ErbAΔXS pre-mRNAs, has been described previously [28]. Light shading of BX fragment indicates SEα2 enhancer, dark shading sequences overlapping TR-ISE3. Endpoints of deletions are indicated by letters B, X, P, E, D,S and by inverted triangles for deletions DS-20L and DS-20R shown in panel E. Speckled shading indicates 300 nts of heterologous sequence in ErbAΔBS-H derived from lambda phage. **B.** RNase protection assays showing splicing of deleted pre-mRNAs expressed in HEK293 as in Fig. 2. Splicing of ΔBS series, lacking 5'ss proximal SEα2

enhancer, and ΔXS series including the SEα2 enhancer is shown [34]. % Splicing (average of 3 experiments except for ErbA-H) and standard deviation (SD) is shown at bottom of figure. **C.** RNase protection assays showing effect of deleting 20 nts from either the 5′ end (-20L) and 3′ end (-20R) of the DS fragment present in ErbAΔBD as indicated in Panel A (inverted triangles) and Panel E. % splicing and standard deviation shown for 6–12 experiments. **D.** *In vitro* splicing of ^{32}P-labeled transcripts following 0, 45 and 90 min incubation with HeLa cell nuclear extract. % splicing is representative of three experiments. **E.** Sequence of TR-ISE3 with downstream endpoints of deletions indicated with (/) and labels. Runs of 3 or more G residues are shown in bold. "U" and "S" indicate unspliced and spliced RNAs.

Interestingly, the effects of SEα2 and TR-ISE3 appear additive; that is, the increase in splicing of transcripts in the series ErbAΔXS through ErbAΔXP closely matches that observed with the series ErbAΔBS through ErbAΔBP. When the percent splicing of ErbAΔBS is subtracted from the percent splicing observed with ErbAΔBD, ErbAΔBE and ErbAΔBP, the results are similar to those obtained by subtracting the percent splicing of ErbAΔXS from that for ErbAΔXD, ErbAΔXE and ErbAΔXP (e.g. 16%, 23% and 23% vs. 15%, 20% and 21%, respectively, as shown in Figure 3B). The additive effects of SEα2 and TR-ISE3 suggest that these two enhancers function independently of one another.

Mapping of TR-ISE3 was confirmed by additional experiments. Deletion of 20 nts from either end of the DS fragment in ErbαΔBD (DS-20L or DS-20R) containing the TR-ISE3 sequence in ErbAΔBD, decreased splicing by 40% relative to ErbAΔBD (Fig 3C). To determine if TR-ISE3 is active *in vitro* as well as *in vivo* a parallel series of transcripts were tested for their ability to undergo splicing when incubated *in vitro* with HeLa cell nuclear extract.[26, 34] Transcripts containing sequences corresponding to ΔBD, ΔBE and ΔBP showed a large increase in splicing when compared to the minigene transcript α2-ΔBS (Figure 3D). In contrast to their *in vivo* expression, however, similar levels of splicing were observed *in vitro* for all three RNAs, α2-ΔBE, α2-ΔBD and α2-ΔBP, suggesting that either only the 3′end of TR-ISE3 is active *in vitro* or that the longer introns in α2-ΔBE and α2-ΔBP counteract the stimulatory effects of the full-length TR-ISE3 enhancer.

Efficient TRα2 mRNA splicing requires exon 10 sequences

The presence of enhancer elements at either end of the TRα2 intron raises the question whether intronic sequences alone are sufficient to regulate splicing at the flanking suboptimal 5' and 3' splice sites. To assess the possible contribution of sequences within exon 10 to TRα2 splicing, we constructed a series of deletions within exon 10. First, 400 nts were deleted from exon 10 of the plasmid pErbAΔBE starting 31 nt downstream of the 3'ss. This deletion (ΔGM) removes almost half the length of exon 10, while leaving sequences adjacent to the splice site and poly(A) sites intact (Figure 4A). When expressed in HEK 293 cells the transcript ErbAΔGM was very poorly spliced in comparison with the parent plasmid, pErbAΔBE (Figure 4B). In fact, the splicing of pErbAΔGM was only slightly greater than that of pErbAΔBS lacking both TR-ISE3 and SEα2 (Figures 4B, 4C). Next, two other transcripts each containing 200 nt deletions that represented either half of the ΔGM deletion were similarly tested. ErbAΔRM, the transcript with the 3' deletion, was spliced at almost the same level as a transcript with the intact exon sequence. In contrast, ErbAΔGR, the transcript with the 5' deletion, showed a substantial decrease in splicing as shown in Figures 4B and 4C. These data suggest the presence of a splicing enhancer element within the fragment deleted in ErbAΔGR.
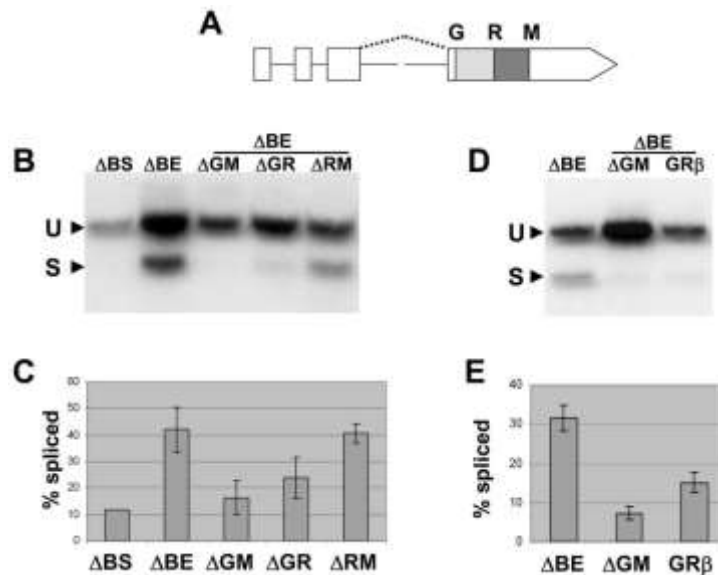

Figure 4 Sequences within exon 10 are required for efficient TRα2 splicing. **A.** Diagram of ErbAΔBE minigene transcript showing endpoints of exon 10 deletions. **B.** RNase protection assays showing splicing of TRα2 minigene transcripts with exon 10 deletions transfected in HEK 293 cells as described in Fig. 2. **C.** Quantitation of splicing

(average of 6 experiments except for ΔBS where n=2). **D.** RNase protection assays on transfected minigenes with sequences deleted in ErbAΔGR replaced by homologous sequence from antisense strand of Rev-erbβ gene (GRβ). **E**. Quantitation of splicing (average of 5–8 experiments).

Because a variety of adventitious effects may be associated with deletions due to the unintended introduction of new sequence elements or alteration of transcript stability, the activity of the region encompassed by the ΔGR deletion was examined by two alternate approaches. First, the GR fragment (deleted in pErbAΔGR) was replaced in the parental construct with a size-matched sequence (GRβ) derived from the antisense strand of the homologous Rev-erbβ gene (NR1D2). Although this fragment is 63% homologous to GR from TRα2, the Rev-erbβ gene lacks an overlapping antisense mRNA comparable to TRα2 and thus is also unlikely to have a functional splicing enhancer on its antisense strand. The resulting construct, pErbAGRβ, showed low levels of splicing, only 50% that of wildtype, as shown in Figure 4D and 4E. This result taken together with the deletion analysis is consistent with the presence of an exonic splicing enhancer.

To examine the role of exonic sequence elements in enhancing splicing of TRα2 *in vitro*, a series of constructs were prepared in which the sequence deleted in pErbAΔGR was fused to the 3' end of pre-mRNA derived from exons 3 and 4 from the Drosophila dsx gene but lacking a functional splicing enhancer.[34–36] First, a 200 bp fragment corresponding to the GR region of exon 10 and the homologous GRβ antisense sequence from Rev-erbβ were compared for their ability to promote splicing of the truncated dsx pre-mRNA in HeLa cell nuclear extract. The transcript with the GR fragment derived from TRα2 (Dsx-GR) increased splicing of Dsx pre-mRNA to a level much higher than the parent Dsx minigene transcript, as shown in Figures 5B and 5C. In contrast, dsx pre-mRNA with the GRβ sequence added was spliced at a very low level, similar to that of the control Dsx vector transcript (Dsx-GRβ, Figure 5D). As a positive control, the dsx pre-mRNA with a very strong ASLV enhancer added[35] was tested in parallel and found to splice at a level similar to that of Dsx-GR (Figures 5C, 5D).
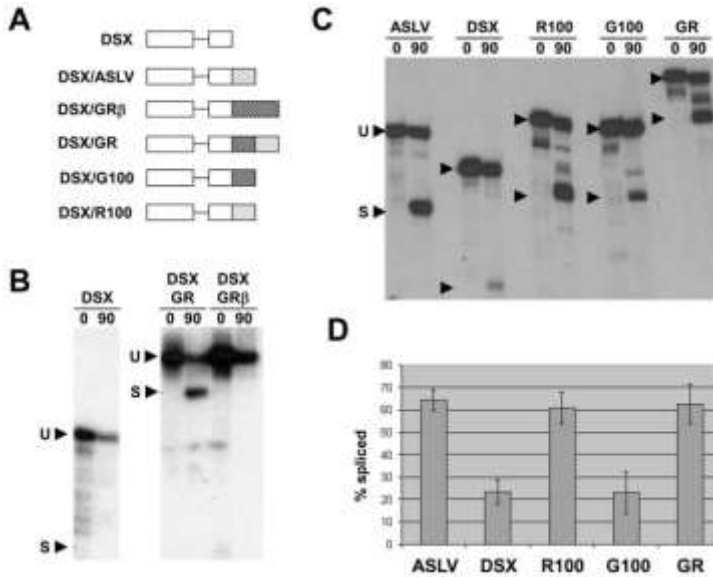
Figure 5 Sequences within the bidirectional coding region of exon 10 promote splicing of dsx pre-mRNA. **A.** Structure of chimeric dsx transcripts. Open boxes represent exons 3 and 4 of Drosophila dsx pre-mRNA, shaded boxes represent added sequences from TRα2 exon 10 or ASLV. G100, R100 are 5' and 3' half of GR fragment, respectively. **B** and **C.** Splicing of ^{32}P-labeled dsx pre-mRNAs *in vitro*. Arrows next to each lane indicate unspliced (U, top) and spliced (S, bottom) RNAs. **D.** Quantitation of splicing of dsx pre-mRNAs, average of 3–9 experiments.

To further map sequences required for enhancer activity, the GR fragment was split into two fragments of 113 nts each, designated G100 and R100, from the 5' and 3' half, respectively, and each fragment was tested separately. As shown in Figure 5C, fragment R100 enhanced splicing of dsx pre-mRNA almost to the same level as the GR fragment. In contrast, the G100 fragment was much less efficient, as seen by the relatively high ratio of unspliced to spliced RNA after incubation in HeLa cell nuclear extract. Quantitation of splicing of Dsx-G100 in multiple splicing assays showed that only about 20% of this transcript was spliced in contrast to about 60% of Dsx-R100 (Figs. 5C and 5D). The results shown in Figures 4 and and5,5, taken together, demonstrate that sequences located within the 3' half of the GR fragment constitute an exonic enhancer, designated ESX10 (**e**xonic **s**plicing enhancer in e**x**on **10**).

Multiple sequences within ESX10 enhance splicing

RNA Biology, Vol. 7, No. 2 (March 2010): pg. 179-190. DOI. This article is © Taylor & Francis and permission has been granted for this version to appear in e-Publications@Marquette. Taylor & Francis does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Taylor & Francis.

12

Since exonic splicing enhancers are often comprised of short sequence elements, typically 6–9 nt in length, that correspond to binding sites for individual proteins,[37–39] we wished to identify shorter elements within the R100 fragment. The length of R100 was thus subdivided into 6 overlapping segments, each of which was 42 nucleotides long (with the exception of fragment"A" that was 43 nts). These subfragments of R100 were designated R42αA-R42αF as shown in Figure 6A. The ends of each fragment were staggered 14 nt apart such that each fragment overlapped the next by 28 nt, to minimize the chance of truncating active elements. Since the 3' end of the pre-mRNA affects the stability of the extended dsx pre-mRNA, each fragment was terminated with the same 5 nt sequence, AGATC, based on the XbaI runoff site used for transcripts derived from the plasmid vector, pDsx. Each of the sequences (R42αA-R42αF) was matched with a homologous sequence derived from the antisense strand of the Rev-erbβ gene (R42βA-R42βF) to provide a parallel set of six controls.
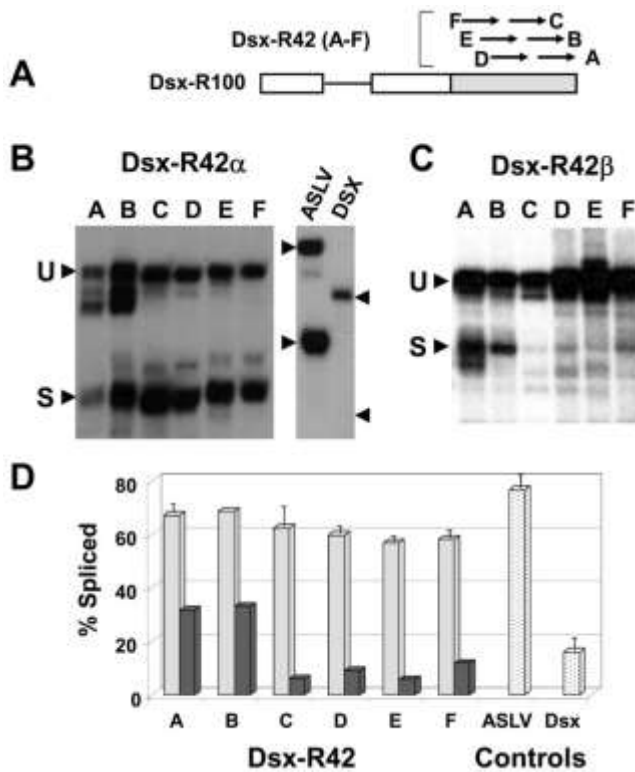

Figure 6 Mapping of enhancer sequences within R100 fragment. **A.** R100 fragment was subdivided into 6 overlapping fragments 42–43 nts in length, R42A R42F, as shown in diagram. **B.** and **C**. Splicing of ^{32}P-labeled chimeric dsx transcripts for 90 min *in vitro* with HeLa cell nuclear extract. R42αA to R42αF fragments with positive (Dsx-

13

ASLV) and negative (Dsx, vector-only) controls (B) or R42βA to R42βF fragments using homologous sequences from the antisense strand of Rev-erbβ (C). **D.** Quantitation of results in experiments shown in Panels B and C. Light shaded bars: Dsx-R42α pre-mRNAs; dark shaded bars: Dsx-R42β pre-mRNAs; speckled shading: controls.

As seen in Figure 6B, each of the 42 (or 43) nt sequences derived from the TRα2 R100 sequence promoted splicing of dsx pre-mRNA to approximately the same extent. Notably, the strong ASLV enhancer, which is almost the same length as the 42 nt ESX10 fragments, was only slightly more active than R42αA-R42αF. In contrast, each of four sequences derived from R100β (R42βC through R42βF) showed negligible activity (Figures 6C and 6D), while the remaining two Rev-erbβ antisense fragments, R42βA and R42βB, showed much lower activity than that associated with the corresponding R42αA and R42αB fragments from ESX10. To test whether the 5 nt extension present on all the fragments might contribute to the observed splicing enhancer activity, R42αF was prepared with and without the extension. Both fragments were found to splice with identical efficiency (result not shown). These results suggest that multiple, functionally redundant enhancer elements are present within the R100 sequence.

Given the configuration of the overlapping test fragments, as few as two enhancer elements located within the 14 nt overlaps between R42αA, R42αB and R42αC and between R42αD, R42αE and R42αF might account for the consistent enhancement observed with these fragments, It is more probable, however, that 3 or 4 elements are dispersed within the R100 region. To further evaluate the role of such elements a 70 nt region at the 5' end of R100 was examined. This region includes the full length of fragments R42αD, R42αE and R42αF but entirely excludes R42αA. All possible combinations of either R42αF or R42βF sequence with 28 nts from αD or βD (R28αD' and R28βD') were tested for enhancer activity when fused to dsx pre-mRNA as shown in Figure 7B. The two chimeric 70 nt fragments displayed intermediate activity relative to the 70 nt sequences from TRα or the antisense of Rev-erbα (that is R70αFαD' > R70αFβD'> R70βFαD' ≫ R70βFβD') in proportion to the relative content of TRα sequence (Figure 7A). These sequences reflect the presence of 28 substitutions (60% identity) within the 70 nt region.
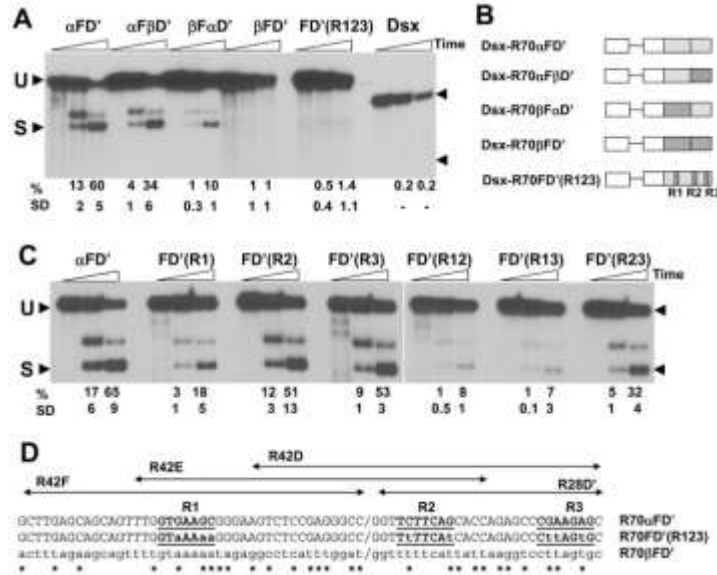
[**Figure 7**](#) Enhancer activity of sub-fragments of the exonic splicing enhancer, ESX10. **A.** Splicing of dsx pre-mRNA *in vitro* in HeLa cell nuclear extract showing enhancement with 70 nt fragments derived from the 5' end of R100. R42αF or R42βF was fused to 28 nts from the 3' end of αD or βD fragment to create R70αFαD' R70αFβD', R70βFαD' and R70βFαD'. **B.** Schematic representation of transcripts shown in panels A and C. Open boxes are Dsx exons, light shaded boxes are TRα2 exon 10 sequence, and hatched boxes homologous Rev-erbβ antisense sequences. **C.** Splicing of dsx-pre-mRNAs *in vitro* after 0, 1 and 2 h in which one or two of the three RESCUE elements (R1, R2 or R3) present in Dsx-R70FD' (R123) contain the substitutions shown below. In Panels A and C 0,1,2 h splicing reactions are shown together with average percent splicing (%) and standard deviation (SD) for four experiments. **D.** Sequence of R70αFD', R70FD' (R123) and R70βFβD' fragments. Positions differing in R70αFD' and R70βFD' are marked with asterisks. The sequence of R70βFD' and substitutions in R70FD' (R123) derived from R70βFD' sequences are shown in lowercase. Arrows indicate the overlapping sequences R42F, R42E and R42D as well as the boundary (marked with "/") between R42F and R28D'. Underlined sequences are three 7 nt elements, R1, R2 and R3, each comprised of two overlapping ESE-associated hexamers according to the analysis of Fairbrother et al. [40].

To further define the sequences requirements for enhancer activity, the bioinformatic approach of Fairbrother et al. [38, 40] was employed to identify three heptamers (R1, R2, R3; underlined in [Figure 7D](#)) comprised of overlapping hexamers with exonic enhancer activity. Interestingly, one or two of these elements are present in five of the six R42 fragments. Each of the heptamers is altered by two or three nucleotide substitutions in the antisense Rev-erbβ sequence ([Figure 7D](#)). When the eight substitutions present in these elements were introduced together into R70αFD', splicing of the resulting

chimera, R70FD(R123), was inhibited more than 95% ([Figure 7A](#)). The effect of these substitutions introduced in R70FD(R123) is greater than that of a much larger number of substitutions in either R70αFβF or R70βFαD', demonstrating that at least two of the elements are essential for the activity of ESX10. Individual and pairwise introduction of mutations in R1, R2 and R3 elements into the wildtype R70αFD' were also analyzed. Individually, the R1 substitution showed the strongest effect and while R2 and R3 alone each had little effect ([Figure 7C](#)). However, the combination of the mutations in R1 and either R2 (R70FD' (R12)) or R3 (R70FD' (R13)) was very effective in blocking R70 enhancer activity and the effect on splicing of five substitutions in R2 and R3 was comparable to that of 11 substitutions in R70 αFβD' ([Figure 7C](#)). Thus, the enhancer element represented by R1 acts synergistically with that represented by R2 and R3 to promote splicing of the heterologous Dsx transcript.

Evolution of ESX 10

To explore the evolution of the overlap, analysis of the region was extended to non-mammalian vertebrates, which lack TRα2. The convergent orientation and proximity of the TRα and Rev-erbα genes in non-mammalian vertebrates is apparent in the frog, *Xenopus tropicalis,* where neither the TRα2 coding sequences nor the 3' ss specific for TRα2 mRNA are present (not shown). Within the sequence corresponding to the BCS, the mammalian Rev-erbα amino acid consensus sequence differs from that of *X. tropicalis* in only ten of 67 codons ([Figure 8A and 8B](#)). Nine of these changes are clustered within the R100 fragment although it accounts for only a little more than half the length of the BCS. A total of 19 nucleotide substitutions are associated with these nine replacements. This pattern of clustered substitutions maximizes the number of base changes and the range of amino acid selection relative to the number of amino acid replacements ([Figure 8B](#)). Thus, a small number of largely conservative amino acid changes in the sequence of Rev-erbα is sufficient to accommodate both coding and splicing of TRα2.

Figure 8 Clustering of non-synonymous nucleotide substitutions within ESX10 in a non-mammalian vertebrate, lacking TRα2. **A.** Distribution of nucleotides substitutions within the amphibian sequence homologous to the bidirectional coding sequence of the mammalian TRα2. The sequencesα from the TRα2 strand of six mammals ([Table S1](#)), and one non-mammalian vertebrate (the frog, *Xenopus tropicalis*; Joint Genome Institute Assembly 4.1) were aligned starting with the codon antisense to the stop codon for Rev-erbα. (boxed) and extend to the Rev-erbα 3'ss (3'ss arrow). Synonymous substitutions in frog relative to the mammalian consensus are indicated with an "S", non-synonymous substitutions with an asterisk. Double-headed arrows shown positions of G100, R100 and R42F and R28D' sequences similar to those in [Figure 7D](#). **B.** Distribution of amino acid substitutions in final exon of Rev-erbα. The sequence shown in Panel A is translated for the Rev-erbα antisense strand. Six mammals are represented: the sequence shown for rat Rev-erbα is identical to that of mouse, cow and dog; human is identical to chimpanzee. The G100 fragment extends beyond the sequence shown. **C.** *In vitro* splicing for 0,1 and 2 h chimeric Dsx-R70 transcripts derived from the rat TRα2 sequence (αFD', [Fig. 7C](#)) and the homologous antisense strand of frog Rev-erbα (xFD') as indicated schematically at right. Light shading indicates rat sequence, dark shading indicates frog sequences and open boxes Dsx exons as in [Fig. 7B](#). Percent splicing (%) and standard deviation (SD) are based on averages of three independent experiments.

Interestingly, both the R1 and R2 elements are perfectly conserved in *X. tropicalis* while R3 is altered by a single substitution.

Comparison of the splicing enhancer activity of amphibian sequences in the Dsx minigene confirmed that the 70 nucleotide frog sequence, R70-xFD', had only 60% activity of the homologous rat R70-αFD' ([Figure 8C]). However, the frog antisense sequence displayed significant activity relative to the ratantisense Rev-erbβ shown in [Figure 7A]. For example, activity of the chimeric fragment xFαD' indicates that the frog xF sequence and the rat αD' sequences complement one another to a significant extent (47% splicing of xFαD' vs. 56% for αFD'; [Figure 8C]). These results suggest that ancient, conserved sequences with the 3' end of Rev-erbα mRNA provided a foundation for the evolution of an exonic splicing enhancer on the antisense strand.

Discussion

In this study we have identified two splicing enhancers in the TRα gene, TR-ISE3, an intronic splicing enhancer, and ESX10, an exonic splicing enhancer. Both promote use of the alternative 3'ss for the terminal exon of TRα2 mRNA and map within the antisense overlap between the 3' ends TRα2 and Rev-erbα mRNAs. Interestingly, ESX10 maps entirely within a bidirectional coding sequence that includes sequences coding for both TRα2 and Rev-erbα ([Figure 1]). Thus the sequence of this exonic splicing enhancer (ESE) is almost completely constrained by the in-phase overlap of codons for the two mRNAs. The C-terminal sequence of Rev-erbα includes part of the heme-binding pocket which is homologous to the canonical ligand-binding domain present in other nuclear receptor proteins. The C-terminus of TRα2 confers on this receptor an array of distinctive properties including site-specific phosphorylation and wide-ranging effects on DNA binding, heterodimerization and co-repressor interactions.[14, 41, 42] The presence of ESX10 within the Rev-erbα-TRα coding overlap demonstrates the ability of RNA sequences to encode three layers of information, two complementary coding sequences and a splicing regulatory element, within a single sequence.

Identification of TR-ISE3

The TR-ISE3 intronic splicing enhancer was mapped to the 3' end of the final intron of TRα2 where it appears to act independently of

SEα2, another previously described intronic splicing enhancer at the 5' end of the same intron.[34] A trivial explanation for these results is that the activity of TR-ISE3 is based simply on a non-specific increase in intron length. This possibility can be ruled out for several reasons: first, the intron in the pErbAΔBS is over 200 nts, much longer than that required for efficient splicing;[43] second, the ErbAΔBS is efficiently spliced when a single nucleotide substitution is made within either the 5' or 3'ss (Figure 2B); and third, replacement of the BP fragment of pErbAΔBP with a size-matched heterologous fragment in ErbA-H (Figure 3B) of the same length yielded very low splicing, comparable to that of ErbAΔBS.

The sequence of TR-ISE3 includes most of the G-rich region (G_n, Figure 1B) and a portion of the highly conserved region just upstream of the 3'ss polypyrimidine tract. Although the G-rich region (Figures 1C and and3E)3E) is less well conserved than the splice-site proximal portion of TR-ISE3 (Table S1), similar runs of G residues are present in many other introns and have been shown to promote alternative and constitutive splicing.[34, 44] Such sequences are often specifically recognized by hnRNP H/F proteins.[34, 45–47] Additionally, G-rich MAZ elements (consensus sequence G_5AG_5) affect alternative splicing and polyadenylation through transcriptional pausing.[48–50] Since the G-rich region in TR-ISE3 is also downstream of the TRα1 poly(A) site, it may be part of a multifunctional locus that impacts transcription as well as processing.

Characterization of ESX10

Results presented here demonstrate that the coding sequence of TRα2 overlapping that of Rev-erbα includes ESX10, an exonic splicing enhancer (ESE) located 140–250 nt downstream of the 3' ss. This region accounts for a substantial portion of the splicing activity that is abolished by the deletion of the splice site proximal portion of the exon in ErbAΔGR (Figure 4). RESCUE analysis[38, 40] revealed a provocative distribution of three enhancer-related heptanucleotide sequences (R1, R2 and R3) across a 70 nucleotide region within R100 (Figure 7D). Substitution of three nucleotides in R1 alone results in a loss of almost 75 % of splicing activity. While substitutions within R2 and R3 individually result in only modest decrease in splicing, the combination

of five substitutions within these two closely spaced heptamers decreases splicing activity by about 40% suggesting that R2 and R3 together constitute a single functional subelement, R2/R3, of the ESX10 enhancer. Substitutions within all three elements decrease splicing activity by >95% (Figure 7A). These results suggest a hierarchy of interacting enhancer elements within the R100 fragment that act synergistically to promote TRα2 splicing. The non-overlapping fragment R42αA is also active in heterologous splicing assays (Fig. 6). Thus ESX10 is a complex ESE comprised of at least three active subelements mapping to sites at R1, R2/R3 and R42A.

Another bioinformatic approach for identifying exonic splicing enhancers involves analysis of consensus binding sites for SR proteins, often associated with ESEs, as implement in the program ESEfinder.[37, 39, 51] ESEfinder revealed a large number of candidate binding sites for SR proteins SF2/ASF, SC35, SRp55 and SRp75 within the R100 region of ESX10 (results not shown) with SC35, SRp 40 and SF2/ASF (SFRS1) associated with sites R1, R2 and R3, respectively. In contrast, far fewer sites are present within the homologous region of the antisense Rev-erbβ strand. However, substitutions in R1 that reduced splicing activity had little effect on the ESEfinder score for SC35 binding at this site. Thus further biochemical studies are required to determine the array of proteins and other factors interacting with ESX10.

Co-evolution of TRα2 splicing and ESX 10

Most ESEs map within coding exons of mRNAs[52, 53] where the dual constraints introduced by requirements for coding and splicing are discernible in the statistical bias of synonymous mutations near the ends of exons adjacent to splice sites.[7, 31, 54, 55] The location of ESX10 within the coding sequences for both TRα2 and Rev-erbα shows that this region simultaneously accommodates at least three layers of specific information and illustrates the surprising capacity of the genetic code to accommodate additional information as recently described by Itzkovitz and Alon.[56, 57] The high level of sequence conservation and the presence of numerous functional constraints make it possible to map the evolutionary history of this locus in detail to reconstruct the evolutionary process by which information required for TRα2 splicing and coding has been superimposed on Rev-erbα

coding sequence. Although Rev-erbα but not TRα is present in model organisms such as flies and nematodes, detailed phylogenetic analysis indicate that homologs of both nuclear receptor proteins were present in the urbilaterian ancestor more than 540 MYA.[58, 59] In addition, the convergent orientation of TRα and Rev-erbα is conserved among many non-mammalian vertebrates (e.g. fish and amphibians), which lack TRα2, as well as in the more distantly related and non-overlapping TRβ and Rev-erbβ genes. These observations strongly suggest that the proximity and orientation of these genes have ancient functions independent of TRα2 expression.

In the genome of the frog *X. tropicalis* coding sequences of TRα1 and Rev-erbα are situated 15 kb apart (vs. only 3.6 kb in mammals) and no sequences corresponding to TRα2 are present, consistent with TRα2 being a recent evolutionary innovation specific to mammals.[33] Since the C-termini of Rev-erbα and Rev-erbβ are tightly conserved in vertebrates, only a limited number of changes are associated with the appearance of TRα2 in the mammalian lineage. Within the R100 fragment the frequency of non-synonymous substitution is 10-fold higher (0.17 substitutions/nt) than in the rest of the BCS, while the occurrence of synonymous substitutions is relatively constant throughout. The clustering of non-synonymous substitutions in R100 maximizes the number of base changes for a given number of amino acid replacements. These changes are sufficient to overlay information for splicing enhancement on the coding sequences for both Rev-erbα and TRα2. The evolution of TRα2 is likely to have been facilitated by the presence of such sequences compatible with ESX10 activity as evident in the conservation of both R1 and R2 on the antisense strand of Rev-erbα in *X. tropicalis* (Fig. 8A) but not that of the more distantly related Rev-erbβ (Fig. 7D). Consistent with this hypothesis, the *X. tropicalis* R1 sequence is able to activate splicing of a Dsx chimera that includes R2 and R3 from rat (compare αFD' and xFαD' in Figure 8C). This analysis suggests that initial steps in the evolution of TRα2 may require only the evolution of a marginally efficient 3'ss sequence on the antisense strand of the Rev-erbα 3' UTR. This scenario is consistent with general requirements for the evolution of alternatively spliced exons in many other genes.[32, 60] According to this model, the amino acid sequence of TRα2 and an increasingly efficient splicing enhancer (ESX10) may then have

coevolved after establishment of an inefficiently spliced TRα2 terminal exon, antisense to Rev-erbα.

Role of antisense overlap in regulating TRα expression

Antisense transcription is widespread in mammalian genomes[4, 12, 13, 42, 61–63] The convergent overlap orientation observed for TRα2 and Rev-erbα mRNAs is the most frequently observed arrangement for overlapping mRNAs[12, 62] although a coding-coding overlap is less common, representing only 10% of overlapping genes.[64] Several lines of evidence support a role for Rev-erbα expression in repressing TRα2 splicing: antisense RNA from the 3' end of Rev-erbα efficiently blocks splicing *in vitro*, and increase in Rev-erbα mRNA expression is associated with a decrease in the expression of TRα2 mRNA relative to TRα1 *in vivo* in various cells types.[25–28]

The presence of multiple splicing enhancer elements within the overlap region ([Figure 1](#)) suggests several specific possible mechanisms for antisense regulation. For example, base-pairing within Rev-erbα and TRα2 mRNA may block binding of splicing factors as suggested by previous in vitro experiments[26] and the proximity of TR-ISE3 the 3' end of the complementary Rev-erbα mRNA may be important for initiating base-pairing as shown in prokaryotes.[65] However, mechanisms associated with antisense RNA are very diverse, involving both transcriptional and post-transcriptional regulation and including positive as well as negative effects on target expression.[5, 66–68]

Since both TRα and Rev-erbα regulate many aspects of mammalian development and physiology, it is not surprising that their expression and function is regulated at many levels.[42, 69- 71] Over the past several decades it has become evident that gene expression is regulated by many different processes that are often closely linked.[4, 72, 73] The evolution of regulatory elements within regions containing multiple layers of genetic information, as demonstrated here for ESX10, may prove a common theme as our ability to interpret the complex information encoded within the mammalian genome continues to advance.

Materials and Methods

Plasmid constructs

Construction of pErbA, pErbAΔBS and pErbAΔXS has been previously described.[28] pErbAΔBS(5'ss+5G) was constructed by swapping the 5' portion of pCMV-erbAm(5'ss+5G) with wildtype sequence of pErbAΔBS. pErbAΔBS(3'ss-3A) was constructed using PCR to replace a single nucleotide at the 3' ss as shown in Figure 2A. pErbAΔBD and pErbAΔBP are similar to pErbAΔBS except for the addition of fragments defined by Dra I and Pvu II sites, 81 and 160 bp, respectively, upstream of the SspI site that marks the 3' end of the ΔBS deletion. pErbAΔBE includes a similar fragment 118 bp in length that was constructed via PCR with an upstream end halfway between the sites defining pErbAΔBD and pErbAΔBP. ErbAΔXS, pErbAΔXD, pErbAΔXE and pErbΔXP include deletions identical to those in pErbAΔBS through pErbAΔBP except that each includes a 250 bp Bsu36I/XbaI fragment at the 5' end of the deletion marked by the Bsu36I site. Deletions were constructed in exon 10 of pErbAΔBE as follows: pErbAΔGM was created by digesting pErbAΔBE with BsrGI and PflMI, Klenow endfilling and ligating; pErbAΔGR was made from pErbAΔBE by introducing a deletion between the unique BsrGI site and the EcoRI site in exon 10; finally, pErbAΔRM was constructed in a similar way by deleting sequences between the same EcoRI site and a PflMI site 220 bp downstream in exon 10. The replacement of the BsrGI-EcoRI fragment with the corresponding antisense Rev-erbβ fragment was achieved by cloning a cDNA copy of the human Rev-erbβ gene into these sites in pErbAΔBE.

pDsx constructs were created from pDsx/KX as previously described.[34, 35, 74] pDsxGR has a 200 nt PCR fragment copied from the BsrGI/EcorI fragment in exon 10 of TRα2 inserted between the KpnI and Xba I site of pDsx/KX. This fragment was created with pErbA as a template using a reverse primer that included EcoRI and XbaI sites matching those in the vector at its 5' end and a forward primer with BsrGI and KpnI sites at its 5' end. The PCR fragment and Dsx vector were cut with KpnI and XbaI and ligated. pDsxR100 was created using the EcoRI/XbaI primer described above and a second primer with a KpnI site adjacent to the site at the center of the GR fragment.

pDsxG100 was created in a similar manner, using the forward primer with BsrGI and KpnI sites added and a reverse primer with XbaI site complement to sequences adjacent to the start of R100. Dsx constructs containing extensions shorter than those associated with R100 or G100 were prepared as PCR products using pDsx-derived plasmids as templates with a forward primer upstream of the T7 promoter and a reverse primer with the desired wildtype or mutant sequences as shown in Table S2. All plasmids generated by PCR were sequenced. Further information on the sequence of all plasmids is available on request.

Plasmid transfection, transcription and in vitro splicing

pErbA plasmids were transfected into HEK 293 cells via calcium phosphate-mediated transfection as described.[27] ^{32}P-labeled RNA transcripts were synthesized from either linearized plasmids or PCR templates containing a bacteriophage promoter as previously described.[34, 75] Templates for the pDsx/KX-ESX10 transcripts were linearized with XbaI, pDsx-ASLV was linearized with MluI and pDsx with BamHI. *In vitro* splicing was carried out as previously described.[26, 34, 36] To start each reaction HeLa cell nuclear extract was added to a solution containing the ^{32}P-UTP labeled transcripts. The reaction was stopped at the times indicated and the products were run on a 5.5% polyacrylamide gel.

To determine the ratio of TRα1/TRα2 mRNA, a ^{32}P-labeled probe that overlaps the TRα2 5' splice site was used in the RNase protection assay as described.[27, 34] The probe covers 162 nt at the 3' most portion of TRα1 or 135 nt complementary to exon 9A of TRα2 mRNA. Following electrophoresis, gels were scanned with an Ambis radioanalytic scanner or Storm phosphoimager. Counts from each band were corrected for uridine content in calculating the percent splicing.[26, 34]

Supplementary Material

S1 Table S1: Conservation of the overlap between TRα2 and Rev-erbα mRNAs. Crosstables for pairwise comparison regions of overlap shown in Figure 1. Lengths of regions refer to the genomic sequence of the rat (NW_047339) and vary slightly between genomes. Two crosstables are shown

for the 3' end of TRα1 mRNA (TRα1 3'UTR in Fig. 1C) since the chimpanzee sequence (NW_001226923) includes only the final 116 bp. The bidirectional coding sequence is 200 bp within the exon-exon overlap, while intron 7 of Rev-erbα mRNA overlaps with 46 bp of TRα2 coding sequence and 470 bp of the TRα2 3' UTR. Other sequences in the alignment are NT_010755 (human), NT_165773 (mouse), NW_876332 (dog) and NW_001493677 [11]. Positions of the TRα1, TRα2 and Rev-erbα poly(A) sites are based on RNase protection mapping and cDNA sequences.[19, 28]

Table S1: Identity within the region of overlap between TRα2 and Rev-erbα mRNAs					
3'end TRa1(205 BP)					
	DOG	RAT	MOUSE	HUMAN	
COW	96	96	96	97	
DOG		96	96	98	
RAT			99	98	
MOUSE				98	
3' end of TRa1 (116 bp including Chimp)					
	DOG	RAT	MOUSE	HUMAN	CHIMP
COW	100	100	100	100	100
DOG		100	100	100	100
RAT			100	100	100
MOUSE				100	100
HUMAN					100
C-rich (140 BP)					
	DOG	RAT	MOUSE	HUMAN	CHIMP
COW	96	89	87	93	94
DOG		90	89	97	96
RAT			96	94	93
MOUSE				92	91
HUMAN					100
Rev-erba poly(A) proximal region (70 BP)					
	DOG	RAT	MOUSE	HUMAN	CHIMP
COW	97	91	94	96	96
DOG		94	97	98	98
RAT			97	96	96
MOUSE				98	98
HUMAN					100
G-rich Region (78 BP)					
	DOG	RAT	MOUSE	HUMAN	CHIMP
COW	85	73	74	76	78
DOG		80	80	80	80

RAT			96	77	77
MOUSE				76	76
HUMAN					99

Conserved region upstream of TRa2 3'ss (99 BP)					
	DOG	RAT	MOUSE	HUMAN	CHIMP
COW	99	99	99	99	99
DOG		98	98	98	98
RAT			100	100	100
MOUSE				100	100
HUMAN					100

Exon-Exon Overlap of TRa2 and Rev-erba mRNA (163 BP)					
	DOG	RAT	MOUSE	HUMAN	CHIMP
COW	98	99	99	98	98
DOG		99	99	99	99
RAT			100	99	99
MOUSE				99	99
HUMAN					100

Intron 7 of Rev-erba mRNA (620 BP)					
	DOG	RAT	MOUSE	HUMAN	CHIMP
COW	86	83	81	86	86
DOG		83	83	90	90
RAT			94	85	86
MOUSE				84	85
HUMAN					99

Exon 7 of Rev-erba (211 bp)					
	DOG	RAT	MOUSE	HUMAN	CHIMP
COW	95	93	95	96	96
DOG		90	93	94	94
RAT			94	93	93
MOUSE				95	95
HUMAN					100

S2 Table S2: Primers used for construction of plasmids and transcription templates. Forward (F) and reverse (R) orientations of the primers are indicated. The primers for the Dsx-R42 templates include an overlap with the exon 4 of dsx mRNA at their 3' end. The reverse primers for R100 and R42A Dsx constructs include at their 5' end an EcoRI site that is not conserved in Rev-erba.

Table S2: Primers used for plasmids constructions.		
Description of plasmid construct	Direction	Sequence (5' to 3')
pErbADXE	F	AGCTCTAGAGGGGCCAGAGGAGGGTG
pErbADXE	R	CATGGACGTAGTGGCCGGACT
pErbADBS(3'ss-3C)	F	GCTTTTCCTTTCGTCTCGTACAGGAG
pErbADBS(3'ss-3C)	R	ACGAAAAGGAAAAGCAAAGGAAAAGGTTAGTAATATTGTTATAA
pDsx (upstream primer)	F	CGGCCAGTGAATTGTAATACG
pDsx/KX-R100	F	CGGGTACCGCTTGAGCAGCAGTTTGGTG
pDsx/KX-G100	R	GGCTCTAGATGCCGGACCTGCGGACCCTG
pDsx/KX-GR and pDsx/KX-R100	R	GGCTCTAGAGAATTCCGCTTCGGTGGAGC
pDsx/KX-GR and pDsx/KX-G100	F	CGGGTACCTGTACAAGGGGGCAGCGGCAG
Dsx/KX-R42aA	R	CTAGA GAA TTCCGCTTCG GTGGAGCAGC TCCAGGAGAC GCTGCTGCGG TACCGCGATCCAAGCTTATC
Dsx/KX-R42aB	R	CTAGA GGAGCAGC TCCAGGAGAC GCTGCTGCGG GCTCTTCGGG CTCT TACCGCGATCCAAGCTTATC
Dsx/KX-R42aC	R	CTAGA AGAC GCTGCTGCGG GCTCTTCGGG CTCTGGTGCT GAAGAACC TACCGCGATCCAAGCTTATC
Dsx/KX-R42aD	R	CTAGA GCTCTTCGGG CTCTGGTGCT GAAGAACCGG CCCTCGGAGA CT TACCGCGATCCAAGCTTATC
Dsx/KX-R42aE	R	CTAGA GGTGCT GAAGAACCGG CCCTCGGAGA CTTCCCGCTT CACCAA TACCGCGATCCAAGCTTATC
Dsx/KX-R42aF	R	CTAGA GG CCCTCGGAGA CTTCCCGCTT CACCAAACTG CTGCTCAAGC TACCGCGATCCAAGCTTATC
Dsx/KX-R42bA	R	CTAGA GAA TTC CAACTCT GTGGAGGCTT TGCAGGAAAC TCTCATTCGT TACCGCGATCCAAGCTTAT
Dsx/KX-R42bB	R	CTAGA GGAGCAGC TCCAGGAGAC GCTGCTGCGG GCTCTTCGGG CTCT TACCGCGATCCAAGCTTATC
Dsx/KX-R42bC	R	CTAGA AAAC TCTCATTCGT GCACTAAGGA CCTTAATAAT GAAAAACC TACCGCGATCCAAGCTTATC
Dsx/KX-R42bD	R	CTAGA GCACTAAGGA CCTTAATAAT GAAAAACCAT CCAAATGAGG CC TACCGCGATCCAAGCTTATC
Dsx/KX-R42bE	R	CTAGA AATAAT GAAAAACCAT CCAAATGAGG CCTCTATTTT TACAAA TACCGCGATCCAAGCTTATC
Dsx/KX-R42bF	R	CTAGA AT CCAAATGAGG CCTCTATTTT TACAAAACTG CTTCTAAAGT TACCGCGATCCAAGCTTATC
Dsx/KX-R70aFD'	R	CTAGAGCTCTTCGGGCTCTGGTGCTG

Dsx/KX-R70aFbD'	R	CTAGA GCA CTA AGG ACC TTA ATA ATG AAA AAC CGG CCC TCG GAG ACT TCC CG
Dsx/KX-R70bFaD'	R	CTAGA GCT CTT CGG GCT CTG GTG CTG AAG AAC CAT CCA AAT GAG GCC TCT AT
Dsx/KX-R70bFD'	R	CTAGA GCA CTA AGG ACC TTA ATA ATG
Dsx/KX-R42aFD5(3'XbaI)	R	GGCCCTCGGAGACTTCCCGCTTCACCAAACTGCTGCTC AAGCTACCGCGATCCAAGCTTATC
Dsx/KX-R42aFD12(3')	R	CTAGACCTTCCCGCTTCACCAAACTGCTGCTCAAGCTAC CGCGATCCAAGCTTATC
Dsx/KX-R42aFD12(5')	R	CTAGAGGCCCTCGGAGACTTCCCGCTTCACCAAACTAC CGCGATCCAAGCTTATC
Dsx/KX-R70aFD'(R123)	R	CTA GA GCAC TAA GGG CTC TGG TGA TGA AAA ACC GGC CCT CGG AGA CTT CCC t
Dsx/KX-R70aFD'(R12)	R	CTAGA GCTCTTC GGGCTCT GGTGaTG AAaAACC GGCCCTC GGAGACT TCCCt
Dsx/KX-R70aFD'(R13)	R	CTAGA GCACTAA GGGCTC TGG TGC TGA AGA ACC GGC CCT CGG AGA CTT CCC t
Dsx/KX-R70aFD'(R23)	R	CTAGA GCaCTaa GGGCTCT GGTGaTG AAaAACC GGCCCTC GGAGACT TC
R70 aFD(R1)	R	CTAGA GCTCTTCGGGCTCTGGTGCTGAAGAACCGGCCCTCGA GACTTCCC T
R70 aFD(R2)	R	CTAGA GCTCTTC GGGCTCT GGTGaTG AAaAACC GGCCCTC GGAGACT TC
R70 aFD(R3)	R	CTA GAG CAC TAA GGG CTC TGG TGC TGA AGA AC
R70xFD'	R	CTAGA GCc CTT CGtt CcC T ca TcC TGA AGA Aca Gt Cca aac Gat A Cc TCC CGC
R70xFaD'	R	CTAGA GC TCT TCG GGC TCT GG TGC TGA AGA ACC Gt Cca aac GAt ACc TCC CGC
R70aFxD'	R	CTAGA GC cCT TCG ttCcC T ca TcC TGA AGA Aca GG CCC TC GGA GAC TTC CCGCT

Abbreviations used

T3 thyroid hormone
TRα α-thyroid hormone receptor
3'ss 3' splice site
5'ss 5' splice site
nt nucleotides
bp base pairs
kb kilo base pairs

BCS bi-directional coding sequence
ESE exonic splicing enhancer
ISE intronic splicing enhancer

References

1. Black DL. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem. 2003;72:291–336.
2. Munroe SH. Diversity of antisense regulation in eukaryotes: multiple mechanisms, emerging patterns. J Cell Biochem. 2004;93:664–71.
3. Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casari G. In search of antisense. Trends Biochem Sci. 2004;29:88–94.
4. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. Nat Rev Genet. 2007;8:413–23.
5. Munroe SH, Zhu J. Overlapping transcripts, double-stranded RNA and antisense regulation: a genomic perspective. Cell Mol Life Sci. 2006;63:2102–18.
6. Hastings ML, Krainer AR. Pre-mRNA splicing in the new millennium. Curr Opin Cell Biol. 2001;13:302–9.
7. Matlin AJ, Clark F, Smith CW. Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol. 2005;6:386–98.
8. Gustincich S, Sandelin A, Plessy C, Katayama S, Simone R, Lazarevic D, et al. The complexity of the mammalian transcriptome. J Physiol. 2006;575:321–32.
9. Modrek B, Lee C. A genomic view of alternative splicing. Nat Genet. 2002;30:13–9.
10. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456:470–6.
11. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008;40:1413–5.
12. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense transcription in the mammalian transcriptome. Science. 2005;309:1564–6.
13. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW. The antisense transcriptomes of human cells. Science. 2008;322:1855–7.

14. Lazar MA. Thyroid hormone receptors: multiple forms, multiple possibilities. Endocr Rev. 1993;14:184–93.

15. Committee on Nuclear Receptor Nomenclature. A unified nomenclature system for the nuclear receptor superfamily. Cell. 1999;97:161–3.

16. Tagami T, Kopp P, Johnson W, Arseven OK, Jameson JL. The thyroid hormone receptor variant alpha2 is a weak antagonist because it is deficient in interactions with nuclear receptor corepressors. Endocrinology. 1998;139:2535–44.

17. Strait KA, Schwartz HL, Perez-Castillo A, Oppenheimer JH. Relationship of c-erbA mRNA content to tissue triiodothyronine nuclear binding capacity and function in developing and adult rats. J Biol Chem. 1990;265:10514–21.

18. Keijzer R, Blommaart PJ, Labruyere WT, Vermeulen JL, Doulabi BZ, Bakker O, et al. Expression of thyroid hormone receptors A and B in developing rat tissues; evidence for extensive posttranscriptional regulation. J Mol Endocrinol. 2007;38:523–35.

19. Lazar MA, Hodin RA, Darling DS, Chin WW. A novel member of the thyroid/steroid hormone receptor family is encoded by the opposite strand of the rat c-erbA alpha transcriptional unit. Mol Cell Biol. 1989;9:1128–36.

20. Miyajima N, Horiuchi R, Shibuya Y, Fukushige S, Matsubara K, Toyoshima K, et al. Two erbA homologs encoding proteins with different T3 binding capacities are transcribed from opposite DNA strands of the same genetic locus. Cell. 1989;57:31–9.

21. Yin L, Wu N, Curtin JC, Qatanani M, Szwergold NR, Reid RA, et al. Rev-erbalpha, a heme sensor that coordinates metabolic and circadian pathways. Science. 2007;318:1786–9.

22. Raghuram S, Stayrook KR, Huang P, Rogers PM, Nosie AK, McClure DB, et al. Identification of heme as the ligand for the orphan nuclear receptors REV-ERBalpha and REV-ERBbeta. Nat Struct Mol Biol. 2007;14:1207–13.

23. Preitner N, Damiola F, Lopez-Molina L, Zakany J, Duboule D, Albrecht U, et al. The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator. Cell. 2002;110:251–60.

24. Pardee KI, Xu X, Reinking J, Schuetz A, Dong A, Liu S, et al. The structural basis of gas-responsive transcription by the human nuclear hormone receptor REV-ERBbeta. PLoS Biol. 2009;7:e43.

25. Lazar MA, Hodin RA, Cardona G, Chin WW. Gene expression from the c-erbA alpha/Rev-ErbA alpha genomic locus. Potential regulation of alternative splicing by opposite strand transcription. J Biol Chem. 1990;265:12859–63.

26. Munroe SH, Lazar MA. Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA. J Biol Chem. 1991;266:22083–6.

27. Hastings ML, Milcarek C, Martincic K, Peterson ML, Munroe SH. Expression of the thyroid hormone receptor gene, erbAalpha, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels. Nucleic Acids Res. 1997;25:4296–300.

28. Hastings ML, Ingle HA, Lazar MA, Munroe SH. Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring antisense RNA. J Biol Chem. 2000;275:11507–13.

29. Chamary JV, Parmley JL, Hurst LD. Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat Rev Genet. 2006;7:98–108.

30. Warnecke T, Parmley JL, Hurst LD. Finding exonic islands in a sea of non-coding sequence: splicing related constraints on protein composition and evolution are common in intron-rich genomes. Genome Biol. 2008;9:R29.

31. Willie E, Majewski J. Evidence for codon bias selection at the pre-mRNA level in eukaryotes. Trends Genet. 2004;20:534–8.

32. Xing Y, Lee C. Alternative splicing and RNA selection pressure--evolutionary consequences for eukaryotic genomes. Nat Rev Genet. 2006;7:499–509.

33. Forrest D, Sjoberg M, Vennstrom B. Contrasting developmental and tissue-specific expression of alpha and beta thyroid hormone receptor genes. EMBO J. 1990;9:1519–28.

34. Hastings ML, Wilson CM, Munroe SH. A purine-rich intronic element enhances alternative splicing of thyroid hormone receptor mRNA. RNA. 2001;7:859–74.

35. Tanaka K, Watakabe A, Shimura Y. Polypurine sequences within a downstream exon function as a splicing enhancer. Mol Cell Biol. 1994;14:1347–54.

36. Tian M, Maniatis T. A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. Cell. 1993;74:105–14.

37. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res. 2003;31:3568–71.

38. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. Predictive identification of exonic splicing enhancers in human genes. Science. 2002;297:1007–13.

39. Liu HX, Zhang M, Krainer AR. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes Dev. 1998;12:1998–2012.

40. Fairbrother WG, Yeo GW, Yeh R, Goldstein P, Mawson M, Sharp PA, et al. RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. Nucleic Acids Res. 2004;32:W187–90.

41. Katz D, Lazar MA. Dominant negative activity of an endogenous thyroid hormone receptor variant (alpha 2) is due to competition for binding sites on target genes. J Biol Chem. 1993;268:20904–10.

42. Engstrom PG, Suzuki H, Ninomiya N, Akalin A, Sessa L, Lavorgna G, et al. Complex Loci in human and mouse genomes. PLoS Genet. 2006;2:e47.

43. Green MR. Pre-mRNA splicing. Annu Rev Genet. 1986;20:671–708.

44. McCullough AJ, Berget SM. G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Mol Cell Biol. 1997;17:4562–71.

45. Caputi M, Zahler AM. Determination of the RNA binding specificity of the heter ogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. J Biol Chem. 2001;276:43850–9.

46. Han K, Yeo G, An P, Burge CB, Grabowski PJ. A combinatorial code for splicing silencing: UAGG and GGGG motifs. PLoS Biol. 2005;3:e158.

47. Martinez-Contreras R, Fisette 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:e21.

48. Ashfield R, Patel AJ, Bossone SA, Brown H, Campbell RD, Marcu KB, et al. MAZ-dependent termination between closely spaced human complement genes. EMBO J. 1994;13:5656–67.

49. Robson-Dixon ND, Garcia-Blanco MA. MAZ elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon IIIb. J Biol Chem. 2004;279:29075–84.

50. Roberts GC, Gooding C, Mak HY, Proudfoot NJ, Smith CW. Co-transcriptional commitment to alternative splice site selection. Nucleic Acids Res. 1998;26:5568–72.

51. Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Hum Mol Genet. 2006;15:2490–508.

52. Mayeda A, Screaton GR, Chandler SD, Fu XD, Krainer AR. Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. Mol Cell Biol. 1999;19:1853–63.

53. Schaal TD, Maniatis T. Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol Cell Biol. 1999;19:261–73.

54. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet. 2002;3:285–98.

55. Parmley JL, Chamary JV, Hurst LD. Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. Mol Biol Evol. 2006;23:301–9.

56. Itzkovitz S, Alon U. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Res. 2007;17:405–12.

57. Bollenbach T, Vetsigian K, Kishony R. Evolution and multilevel optimization of the genetic code. Genome Res. 2007;17:401–4.

58. Escriva H, Bertrand S, Laudet V. The evolution of the nuclear receptor superfamily. Essays Biochem. 2004;40:11–26.

59. Segraves WA, Hogness DS. The E75 ecdysone-inducible gene responsible for the 75B early puff in Drosophila encodes two new members of the steroid receptor superfamily. Genes Dev. 1990;4:204–19.

60. Xiao X, Wang Z, Jang M, Burge CB. Coevolutionary networks of splicing cis-regulatory elements. Proc Natl Acad Sci U S A. 2007;104:18583–8.

61. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, et al. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell. 2004;116:499–509.

62. Chen J, Sun M, Kent WJ, Huang X, Xie H, Wang W, et al. Over 20% of human transcripts might form sense-antisense pairs. Nucleic Acids Res. 2004;32:4812–20.

63. Kiyosawa H, Mise N, Iwase S, Hayashizaki Y, Abe K. Disclosing hidden transcripts: mouse natural sense-antisense transcripts tend to be poly(A) negative and nuclear localized. Genome Res. 2005;15:463–74.

64. Veeramachaneni V, Makalowski W, Galdzicki M, Sood R, Makalowska I. Mammalian overlapping genes: the comparative perspective. Genome Res. 2004;14:280–6.

65. Nordstrom K, Wagner EG. Kinetic aspects of control of plasmid replication by antisense RNA. Trends Biochem Sci. 1994;19:294–300.

66. Shearwin KE, Callen BP, Egan JB. Transcriptional interference--a crash course. Trends Genet. 2005;21:339–45.

67. Beiter T, Reich E, Williams RW, Simon P. Antisense transcription: a critical look in both directions. Cell Mol Life Sci. 2009;66:94–112.

68. Mahmoudi S, Henriksson S, Corcoran M, Mendez-Vidal C, Wiman KG, Farnebo M. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. Mol Cell. 2009;33:462–71.

69. Bookout AL, Jeong Y, Downes M, Yu RT, Evans RM, Mangelsdorf DJ. Anatomical profiling of nuclear receptor expression reveals a hierarchical transcriptional network. Cell. 2006;126:789–99.

70. Burris TP. Nuclear hormone receptors for heme: REV-ERBalpha and REV-ERBbeta are ligand-regulated components of the mammalian clock. Mol Endocrinol. 2008;22:1509–20.

71. Flamant F, Samarut J. Thyroid hormone receptors: lessons from knockout and knock-in mutant mice. Trends Endocrinol Metab. 2003;14:85–90.

72. Darnell JE., Jr Variety in the level of gene control in eukaryotic cells. Nature. 1982;297:365–71.

73. Bartel DP, Chen CZ. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet. 2004;5:396–400.

74. McNally LM, McNally MT. An RNA splicing enhancer-like sequence is a component of a splicing inhibitor element from Rous sarcoma virus. Mol Cell Biol. 1998;18:3103–11.

75. Munroe SH. Antisense RNA inhibits splicing of pre-mRNA in vitro. EMBO J. 1988;7:2523–32.

About the Authors:

Address correspondence to: Stephen H. Munroe, Department of Biological Sciences, Wehr Life Sciences Building 109, Marquette University, PO Box 1881, Milwaukee, WI 53201, Tel: 414-288-1485, FAX: 414-288-7357, Email: munores@marquette.edu

Valerie K. Salato: Current address: Department of Pathology, Medical College of Wisconsin, Milwaukee, WI 53226.

Nathaniel W. Rediske: Current address: School of Medicine, Oregon Health and Science University, Portland, OR 97239.

Michelle L. Hastings: Current address: Department of Cell Biology and Anatomy, Chicago Medical School, Rosalind Franklin University of Medicine and Science, North Chicago, IL 60064