

Minireview

Alternative Splicing in the Nuclear Receptor Superfamily Expands Gene Function to Refine Endo-Xenobiotic Metabolism[□]

Andrew J. Annalora, Craig B. Marcus, and Patrick L. Iversen

Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, Oregon (A.J.A., C.B.M., P.L.I.) and United States Army Research Institute for Infectious Disease, Frederick, Maryland (P.L.I.)

Received August 16, 2019; accepted December 31, 2019

ABSTRACT

The human genome encodes 48 nuclear receptor (NR) genes, whose translated products transform chemical signals from endo-xenobiotics into pleiotropic RNA transcriptional profiles that refine drug metabolism. This review describes the remarkable diversification of the 48 human NR genes, which are potentially processed into over 1000 distinct mRNA transcripts by alternative splicing (AS). The average human NR expresses ~21 transcripts per gene and is associated with ~7000 single nucleotide polymorphisms (SNPs). However, the rate of SNP accumulation does not appear to drive the AS process, highlighting the resilience of NR genes to mutation. Here we summarize the altered tissue distribution/function of well characterized NR splice variants associated with human disease. We also describe a cassette exon visualization pictograph methodology for illustrating the location of modular, cassette exons in genes, which can be skipped in-frame, to facilitate the study of their functional relevance to both drug metabolism and NR evolution. We find cassette exons associated with all of the functional domains of NR genes including the DNA and ligand binding domains. The matrix of inclusion or exclusion for functional domain-encoding cassette exons is extensive and capable of significant alterations in cellular phenotypes that modulate endo-xenobiotic metabolism.

Exon inclusion options are differentially distributed across NR subfamilies, suggesting group-specific conservation of resilient functionalities. A deeper understanding of this transcriptional plasticity expands our understanding of how chemical signals are refined and mediated by NR genes. This expanded view of the NR transcriptome informs new models of chemical toxicity, disease diagnostics, and precision-based approaches to personalized medicine.

SIGNIFICANCE STATEMENT

This review explores the impact of alternative splicing (AS) on the human nuclear receptor (NR) superfamily and highlights the dramatic expansion of more than 1000 potential transcript variants from 48 individual genes. Xenobiotics are increasingly recognized for their ability to perturb gene splicing events, and here we explore the differential sensitivity of NR genes to AS and chemical exposure. Using the cassette exon visualization pictograph methodology, we have documented the conservation of splice-sensitive, modular, cassette exon domains among the 48 human NR genes, and we discuss how their differential expression profiles may augment cellular resilience to oxidative stress and fine-tune adaptive, metabolic responses to endo-xenobiotic exposure.

Introduction

The nuclear receptor (NR) concept describes a mediator or sensor that translates chemical changes into physiologic effects (Evans and Mangelsdorf, 2014). These translators bind specific chemical ligands and chromatin leading to regulation of RNA synthesis (Mueller et al., 1958). The NR superfamily of 48 human transcription factors is composed of a modular domain gene structure that includes a DNA binding domain (DBD), a ligand binding domain (LBD), a hinge domain located between the DBD and LBD, and a highly variable amino

(N)-terminal transactivation domain. The functional diversity of NR genes is defined in part by the structural organization of their LBD and the role ligand binding plays in mediating cellular activity. Variations in the LBD's structural scaffold allow NR genes to be subdivided into three functional classes, which include 1) ligand-dependent NRs, which encode a LBD that requires ligand binding to activate both genomic and nongenomic signaling events; 2) ligand-independent or constitutively active NRs, which encode a LBD that is stuck [via mutation or alternative splicing (AS)] in the "on position" in the absence of ligand binding; and 3) inactive or transcriptional repressor NRs, which function as dominant negatives that can no longer interact with coactivators due to the loss of key structural motifs in the LBD (Bridgman et al., 2010). Based on this diversity, ligand binding can trigger a diverse array of

<https://doi.org/10.1124/dmd.119.089102>.

□ This article has supplemental material available at dmd.aspetjournals.org.

ABBREVIATIONS: AhR, aryl hydrocarbon receptor; AR, androgen receptor; AS, alternative splicing; CAR, constitutive androstane receptor; CEVp, cassette exon visualization pictograph; circRNA, circular RNA; DBD, DNA binding domain; DMD, Duchenne muscular dystrophy; ER, estrogen receptor; GWAS, genome-wide association study; HNF4 α , hepatocyte nuclear factor 4 α ; LBD, ligand binding domain; MR, mineralocorticoid receptor; NIH, National Institutes of Health; NR, nuclear receptor; P450, cytochrome P450; PPAR, peroxisome proliferator-activated receptor; PXR, pregnane X receptor; RNAseq, RNA sequencing; RXR, retinoid X receptor; SNP, single nucleotide polymorphism; SSO, splice-switching oligonucleotide; VDR, vitamin D receptor.

cellular responses in NR genes, with ligands being classified according to their role as agonists, antagonists, partial agonists, inverse agonists, superagonists, or selective NR modulator (Burris et al., 2013).

Nearly all human genes are alternately spliced (Lee and Rio, 2015). AS is a cotranscriptional process that generates protein diversity via the inclusion or exclusion of different exons within the mRNA transcripts of a gene. This phenomenon is complex, but it is associated with changes in spliceosome binding preference from weaker to stronger splice sites, alterations in the balance between positive and/or negative splicing factors, dynamic regulation of tissue-specific factors, and transitions in mRNA secondary structure stability (Smith and Valcarcel, 2000; Soller, 2006; Chen and Manley, 2009). Xenobiotics and chemotherapeutic agents (Zaharieva et al., 2012; Lambert et al., 2017), UV-B radiation (Sprung et al., 2011), oxidative stress (Cote et al., 2012; Melangath et al., 2017), and heavy metals (Jiang et al., 2017; Li et al., 2018; Wang et al., 2018) are also known to contribute to AS. These effectors are increasingly being recognized for their potential to alter human drug metabolism and health, via splicing-related mechanisms that expand the scope of their toxic properties.

The premise of this article is to document naturally occurring splice variation patterns in the 48 human NR genes, to clarify the role AS plays in modulating endo-xenobiotic metabolism. While AS is a feature common to most multiexon genes, it appears to play a pivotal role in mediating the diverse set of functions that NRs employ to transduce a discrete, chemical exposure cue into an adaptive, physiologic response. The modular structural domains of NR genes are well conserved across the superfamily, however this review highlights the diversity of “cassette exon” structures operating across the seven human NR gene subfamilies (NR0–NR6) and explores how these features may alter a gene’s susceptibility to environmentally responsive, AS events. In general, a cassette exon is defined as a splicing event whereby an interposing exon, between two other exons, can be either included or skipped during pre-mRNA processing to generate two distinct mRNA transcripts and ultimately, two protein isoforms (Cui et al., 2017). If a cassette exon is expressed at a 100% inclusion rate, it is considered a constitutive exon (Zhang et al., 2016), however, a cassette exon may more accurately be described as any exon that can be conditionally excluded from a mature mRNA transcript without altering the reading frame or protein coding potential. In this regard, many cassette exons are overlooked and considered constitutive exons, simply because they are commonly included in the gene’s reference transcript. In reality, any exon that can be removed without altering the proteins reading frame, can be considered a cassette exon, including those that encode extensions of the amino or carboxy terminus. Modern RNA sequencing advances have greatly improved our ability to study cassette exon utilization and the role AS plays in expanding mRNA transcript diversity, however improved sequencing methodologies are still required to address fundamental questions relating to the role of AS and transcriptome expansion in normal physiology, development, and disease (Wan and Larson, 2018). In 2017, we published a similar review focused on AS in the cytochrome P450 (P450) superfamily that highlighted the tremendous expansion of the human P450 transcriptome over the last decade, which now includes over 1000 unique mRNA transcripts generated from the 57 human P450 genes (Annalora et al., 2017). The goal of the current review is to explore similar trends in the NR superfamily, and to compare the complementary role alternative splicing may play in mediating the expansion of gene function among these highly associated gene superfamilies, as reviewed (Honkakoski and Negishi, 2000).

To explore the role AS plays in expanding NR gene functionality, we annotated all known human NR splice variants listed in the National Institutes of Health (NIH) PubMed, AceView, and Ensembl data bases (Thierry-Mieg and Thierry-Mieg, 2006; Cunningham et al., 2015). From

this computational meta-analysis, we found that the 48 human NR genes encode a transcriptome comprising at least 1004 unique transcript variants (see Table 1). This number represents all currently known reference (or wild-type) and variant NR mRNAs, including those with retained introns, premature termination codons, and those subject to nonsense-mediated decay; we also included some experimentally derived transcript variants listed in PubMed citations that were not represented in the primary data bases; these were identified by searching PubMed for “nuclear receptor-, NR*-, or nuclear hormone receptor-variant,” “splice variant,” “alternative splicing,” or “alternative transcript.”

We discovered that NR transcripts exploit AS options at rates comparable to the P450 superfamily, and that cassette exon utilization in particular, may be a key factor influencing the functional plasticity of NR genes. Numerous splice variants capable of influencing homo- and heterodimer interactions, ligand binding, DNA binding, and subcellular trafficking events were identified or predicted, some of which may help to refine complex physiologic responses to endo-xenobiotic exposures. Endogenous gene splicing programs involving NR genes and their coregulatory proteins, some of which are components of the spliceosome, are also known to exist (Auboeuf et al., 2005). In this regard, NR signaling cascades couple gene transcription with pre-mRNA splicing, allowing a high level of coordination among ligand binding events, gene splicing, and protein translation. This central role allows NR genes to modulate an array of pleiotropic cellular functions, including neurogenesis and steroid hormone signaling, via discrete interactions with tissue-specific, NR coregulators (e.g., **Cofactor of BRCA1 (COBRA1)** or **Arginine and glutamate-rich protein 1 (ARGLU1)**) that direct gene-specific AS events. It is currently thought that this class of splicing regulation is more limited in scope than stochastic splicing cascades triggered by the direct effects of chemical mutagens, UV irradiation, or oxidative stress (Sun et al., 2007; Magomedova et al., 2019).

AS can also be mediated by small molecules (as reviewed, Taladriz-Sender et al., 2019) and nucleic acid drugs, generating new opportunities for modulating aberrant P450 and NR gene activity linked to disease. Antisense oligonucleotides that induce alternate exon splicing by interfering with pre-mRNA processing are called splice-switching oligonucleotides (SSOs) (as reviewed, Croft et al., 2000; Kole et al., 2012; van Roon-Mom and Aartsma-Rus, 2012). SSO technology has primarily been exploited as a therapeutic approach to rare genetic diseases such as Duchenne muscular dystrophy (Fall et al., 2006; McClorey et al., 2006a,b; Adams et al., 2007; Fletcher et al., 2007; Koo and Wood, 2013) as highlighted by the Food and Drug Administration approval of Eteplirsen in 2016 (Syed, 2016). The therapeutic utility of SSOs to modulate the immune response have also been explored (Mourich and Iversen, 2009; Mourich et al., 2014; Panchal et al., 2014), as has ligand-independent signaling in the vitamin D receptor (Annalora et al., 2019). The implications of splice-switching small molecules and SSOs as gene-directed therapeutics targeting the pre-mRNA of NR genes underlie the examination of NR splice variants presented here.

Results

Human Nuclear Receptor Associations with Pathology. The 48 genes in the nuclear receptor superfamily are fundamentally linked to human health and homeostasis, and this review aims to clarify how cellular changes in NR gene expression can promote both disease and health resilience. For clarity, a complete listing of all 48 human NR genes, their common names and abbreviations are listed in the supplemental abbreviations section of the Supplemental Materials. Using public data bases, we identified 31 NR genes already connected to a human disease either from literature citations, links to single

TABLE 1
Summary of human nuclear receptor variants

NR family (no. genes)	Transcript Size (Bases)	Exons	Amino Acids (Residues)	Molecular Mass (kDa)	Transcript Variants	SNPs	SNPs per Variant ^a
NR0 (2)	1011–2021	1 to 2	54–470	6.3–51.7	7	871	124
NR1 (20)	1256–10,998	7–12	374–976	42.7–114.2	474	163,551	345
NR2 (12)	2112–8330	3–15	339–703	36.2–78.4	203	25,260	124
NR3 (9)	2169–13,762	5–11	344–984	38.7–107.1	211	111,634	529
NR4 (3)	2465–4996	7 to 8	609–699	67.9–75.4	75	5409	72
NR5 (2)	1863–3200	7	461–562	51.6–63.4	26	9754	375
NR6 (1)	7052	10	480	54.4	8	12,777	1597
Total (49) ^b	1011–10,998	1–15	54–984	6.3–114.2	1004	342,359	341

^aDerived as total SNPs per total splice variants; equivalent to 0.003 splice variants per SNP. The mean average for total SNPs per total splice variants for all 48 primary human NR genes is 473 SNPs per variant or 0.009 variants per SNP (not shown). The median for the 48 human NR genes is 191 SNPs per variant or 0.005 variants per SNP.
^bThe analysis includes all 48 human NR genes plus the processed human pseudogene *FXRβ* (NR1H5).

nucleotide polymorphisms (SNPs), or genome-wide associative studies (GWAS), as indicated in Table 2. Disease associations were observed for reproductive, nervous, cardiovascular, musculoskeletal, metabolic, sensory, and immune systems, highlighting the central role of NR genes in regulating human health. Numerous associations with a variety of cancers and rare genetic diseases were also identified, and some of these are associated with NR transcript variants listed in the NIH AceView data base (Thierry-Mieg and Thierry-Mieg, 2006). A summary of all known transcript variants identified by searching PubMed, AceView, and the Ensembl data base (Cunningham et al., 2015) are listed in Table 1, as well as all known SNPs listed in the GeneCards data base (Stelzer et al., 2011).

Meta-Analysis of Nuclear Receptor Transcript Variants and SNPs. As summarized in Table 1, we cataloged all known transcript variants for the 48 human NR genes, plus one processed pseudogene, farnesoid X receptor β (*FXRβ* (NR1H5)), and annotated their known association with human disease and any available information regarding their molecular weight, tissue-specific expression profile, and sub-cellular trafficking. The stereo-radar plot shown in Fig. 1 highlights the expansion of NR transcript variants in humans and contrasts these results with the number of known SNPs for each gene. We found the number of mRNA transcript variants expressed for each human NR gene ranged from 1 to 62 variants, with liver X receptor α (*LXRα*) (62), constitutive androstane receptor (*CAR*) (60), and estrogen related receptor γ (*ERRγ*) (47) expressing the highest number of known mRNA transcripts. In contrast, *FXRβ* (1), small heterodimer partner-1 (*SHP-1*) (3), dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1 (*Dax-1*) (4) and *RAR*-related orphan receptor β (*RORβ*) (4) display the fewest number of transcript variants. A comparison of total variant transcripts and total SNPs reveals some divergence between these characteristics as retinoic acid receptor α (*RARα*) (49,231) and *RORα* (39,338) display the highest number of known polymorphisms, but not the highest number of transcript variants [*RARα* (39) and *RORα* (28)]. However, the orphan nuclear receptor *ERRγ* (NR3B3) showed both an extremely high number of transcript variants (47) and SNPs (33,772). If the number of transcript variants is normalized to the total number of SNPs, the xenobiotic sensing NR gene, *CAR* (NR1I3), possesses nearly seven times as many splice variants per SNP (0.06) than any other human NR gene, which on average expresses ~0.009 splice variants per SNP, suggesting that the *CAR* gene, an important regulator of xenobiotic metabolism, may be more sensitive to mutational selection pressures than the average NR gene and/or more functionally dependent on AS mechanisms to expand its xenosensor role (Auerbach et al., 2005; Chen et al., 2010). However, because the evolutionary history of the *CAR* gene is complex compared with other NR genes, having descended into mammals from lobe-finned fishes and tetrapods (Sarcopterygii), but not into reptiles or most birds (Zhao et al., 2015), it is difficult to

understand why this gene may have become more sensitive to AS than other NR genes. *CAR* may simply have experienced higher levels of selection pressure in humans due to its central role mediating adaptation to changes in both diet and environmental exposure.

A Brief Comparison with P450 Genes. Based on our analysis, we found that NR genes express roughly 20.5 transcript variants per gene, compared with 16.7 for the average P450 superfamily gene member (Annalora et al., 2017) (Table 1). This is not a dramatic difference. However, we found 6985 SNPs per gene for NR genes, but only 848 SNPs per gene for P450s. This translates to 341 SNPs per transcript variant for NR genes, as compared with 50 SNPs per transcript variant in the P450s. Given the similar gene size and number of total splice variants per gene, but nearly 7-fold more SNPs per gene, it appears that NR genes are subject to greater mutational selection pressures than P450 genes. However, because the accumulation of single nucleotide variants in NR genes have been inversely correlated to a gene’s age (Popadin et al., 2014; Mackeh et al., 2017), P450 genes are likely more ancient than their cognate NR genes, which due to their younger age, may be more resilient to mutations that might further optimize their expression parameters, function, or interactome.

Nuclear Receptor Splice Variants Associated with Human Disease. While SNPs can undoubtedly have a profound effect on NR gene function via multiple mechanisms (Prakash et al., 2015), there is evidence that the NR gene family is more resilient to mutations than other gene families, including the human leukocyte antigens and histone deacetylases (Mackeh et al., 2017). However, less is known about how SNPs alter the natural splicing patterns of NR genes, and whether these types of variations are more likely to promote disease in humans. Based on this meta-analysis, and a growing appreciation for the role that AS plays in expanding gene function, we speculate that NR transcript variant expression, which can increasingly be monitored by RNA sequencing (RNA-seq) methods, may provide an improved fingerprint for disease diagnosis than genomic studies focused only on SNPs that may or may not alter gene splicing and expression. In this regard, we have listed more than 25 NR transcript variants that have been linked to a human disease (see Table 3). Alternately spliced NR variants can express dominant negative phenotypes (e.g., receptors lacking a complete DBD) or constitutively active phenotypes (e.g., receptors lacking a complete LBD) capable of driving disease, as we have reported for the vitamin D receptor (*VDR*; NR1I1) (Annalora et al., 2019). This phenomenon is perhaps best exemplified by splice variants of the androgen receptor (*AR*; NR3C4) that promote castration-resistant forms of prostate cancer (Dehm et al., 2008; Antonarakis et al., 2014; Kohli et al., 2017). In this meta-analysis we identified 12 additional cancer-related NR gene variants [for *DAX-1*, *SHP*, peroxisome proliferator-activated receptor (*PPAR*) β/δ , *RARα*, *RARγ*, *LXRα*, *LXRβ*, mineralocorticoid receptor (*MR*), estrogen receptor (*ER*) α , nerve

TABLE 2
Representative nuclear receptor association with pathology

NR gene	Group	Disease Associations	SNP/GWAS ^a
DAX-1	NR0B1	X-linked adrenal hyperplasia; azoospermia	Flucloxacillin-induced liver injury X-linked adrenal hypoplasia Autism and schizophrenia BMI, glucose/insulin resistance
SHP	NR0B2	Hepatocellular carcinoma, lipid dystrophy, and obesity	4 kHz hearing threshold
THR α	NR1A1	Resistance to thyroid hormone	Insulin resistance
RAR β	NR1B2	Neural tube defects	Altered blood glucose
RAR γ	NR1B3	Anthracycline cardiotoxicity	T2D
PPAR α	NR1C1	—	Altered white blood cell count
PPAR β/δ	NR1C2	—	Severe acne
PPAR γ	NR1C3	—	Alcoholism, Alzheimer disease, ALS, asthma, diabetic nephropathies
REV-erb α	NR1D1	Disordered circadian clock	Stroke, tobacco use, type II diabetes, macular degeneration
REV-erb β	NR1D2	Chagas cardiomyopathy	Elevated VEGF levels
ROR α	NR1F1	—	Supranuclear palsy; metabolic syndrome
ROR β	NR1F2	Circadian rhythm, retinal degeneration	Inflammatory bowel disease; microbiome
LXR β	NR1H2	Gall bladder cancer risk	Blood metabolite levels
LXR α	NR1H3	Coronary artery disease, T2D	Rheumatoid arthritis, T2D
VDR	NR1I1	Vitamin D-resistant rickets	Macular degeneration, ALS, hypertension
CAR	NR1L3	Zellweger syndrome	Wegener's granulomatosis
HNF4 α	NR2A1	Diabetes mellitus, adenocarcinoma, nephropathies	Age of menarche onset
RXR α	NR2B1	Adenocarcinoma, ADHD, bipolar disorder, autism	Alopecia
RXR β	NR2B2	Alcoholism, cancer (NSCLC)	—
RXR γ	NR2B3	Shared with NR2B1/NR2B2	Increased reaction time
TR2	NR2C1	Prostate neoplasms	Body height, HVA:MHPG ratio
TR4	NR2C2	Alzheimer disease, vitamin A deficiency	ADHD symptoms
TLX	NR2E1	Human lymphoid leukemia	BMI
PNR	NR2E3	Retinitis pigmentosa 37	Breast neoplasms, nicotine dependence, akathisia, schizophrenia
ERR β	NR3B2	Deafness	Elevated LDL cholesterol, alopecia
GR	NR3C1	Rheumatoid arthritis, asthma	Age of menarche onset
MR	NR3C2	Hypertension, pseudohypoadosteronism	Crohn's disease, ulcerative colitis, pancreatic neoplasms
AR	NR3C4	Androgen insensitivity	Macular degeneration, narcolepsy
NURR1	NR4A2	Parkinson's disease, ADHD	
LRH-1	NR5A2	Pancreatic cancer	
GCNF	NR6A1	Prostate cancer marker	

ADHD, attention-deficit/hyperactivity disorder; ALS, amyotrophic lateral sclerosis; BMI, body mass index; HVA:MHPG, 3-methoxy-4-hydroxyphenylglycol:homovanillic acid; LDL, low-density lipoprotein; NSCLC, non-small cell lung cancer; T2D, type II diabetes; VEGF, vascular endothelial growth factor.

^aData obtained in early 2019 from www.GWASCentral.org and GWAS Catalog www.ebi.ac.uk/gwas/search?query.

growth factor IB (NGFIB or nuclear receptor related 77 (Nur77)), nuclear receptor related 1 (NURR1), and steroidogenic factor 1 (SF1)] that lack a complete LBD and may thus be constitutively active. We also identified 14 cancer-related NR variants [for PPAR α , PPAR γ , RAR β , V-erbA-related protein EAR-4 (Rev-erb β), LXR α , LXR β , VDR, pregnane X receptor (PXR), hepatocyte nuclear factor 4 α (HNF4 α), testicular receptor 4 (TR4), AR, ERR α , NGFIB, and NURR1] that lack a complete DBD and may thus represent dominant negatives (based on the NIH AceView data base in early 2018) (Thierry-Mieg and Thierry-Mieg, 2006).

There are currently no crystal structures for any NR splice variants listed here, making the structural consequences of such large-scale modifications to the NR scaffold difficult to predict. However, the elimination of key domains clearly has a profound effect on NR function, tissue distribution, and subcellular trafficking, as highlighted by our current knowledge of where NR splice variants are being expressed and trafficked (see Table 4). Many nuclear receptor ligands are also either substrates or products of P450 enzymes, and there is growing interest in understanding how different classes of NR ligands might alter the splicing patterns of NR target genes, by altering RNA stability directly, or by modulating the activity of the transcription activation complex, spliceosome, or ribosome (Auboeuf et al., 2005, Bhat-Nakshatri et al., 2013, Zhou et al., 2015). It was recently shown that the environmental pollutant 2,3,7,8-tetrachlorodibenzo-p-dioxin can profoundly modulate the AS patterns of aryl hydrocarbon receptor (AhR) target genes (e.g., CYP1A1) in a hepatic mouse model (Villaseñor-Altamirano et al., 2019). The AhR is a ligand-dependent transcription factor, related to NR genes, that is a member

of the basic-helix-loop-helix-Per-ARNT-Sim superfamily. While the mechanisms by which 2,3,7,8-tetrachlorodibenzo-p-dioxin drives global changes in AS via the AhR remain unknown, changes in spliceosome assembly, alternative exon recognition, and tissue-specific, splicing factor expression are thought to be involved. Xenobiotics like phenobarbital and nifedipine are known to promote AS in related, drug-metabolizing enzymes, including CYP2B2 (Desrochers et al., 1996; Zangar et al., 1999), and future studies focused on defining the structure-activity relationships between ligands and receptors that drive the cotranscriptional, AS processes are needed.

Cassette Exon Utilization Links Nuclear Receptor Structure, Function, and Evolution. Despite numerous advances in the structural characterization and modeling of NR genes (Rastinejad et al., 2013), improved understanding of NR protein structure/function is still required to fully understand how these genes evolved substrate specificity, and how they interact as a collection of individual, alternatively spliced genes. Nuclear receptors are a diverse group of transcriptional regulators that play key roles in development, physiology, and reproduction while also acting as sensors of xenobiotic exposure. How they accomplish such diversity in their functional roles while being constrained by a highly conserved structural scaffold, while exposed to a complex and variable cellular interactome remains a mystery. We speculate that the expansive NR gene function observed in humans may be tightly linked to the expression of splice variant products that can fine-tune biologic signaling pathways, through the coordinated expression of modular, cassette exon domains that expand gene function. For example, we have previously shown that exon 8 of the human VDR is a cassette exon that

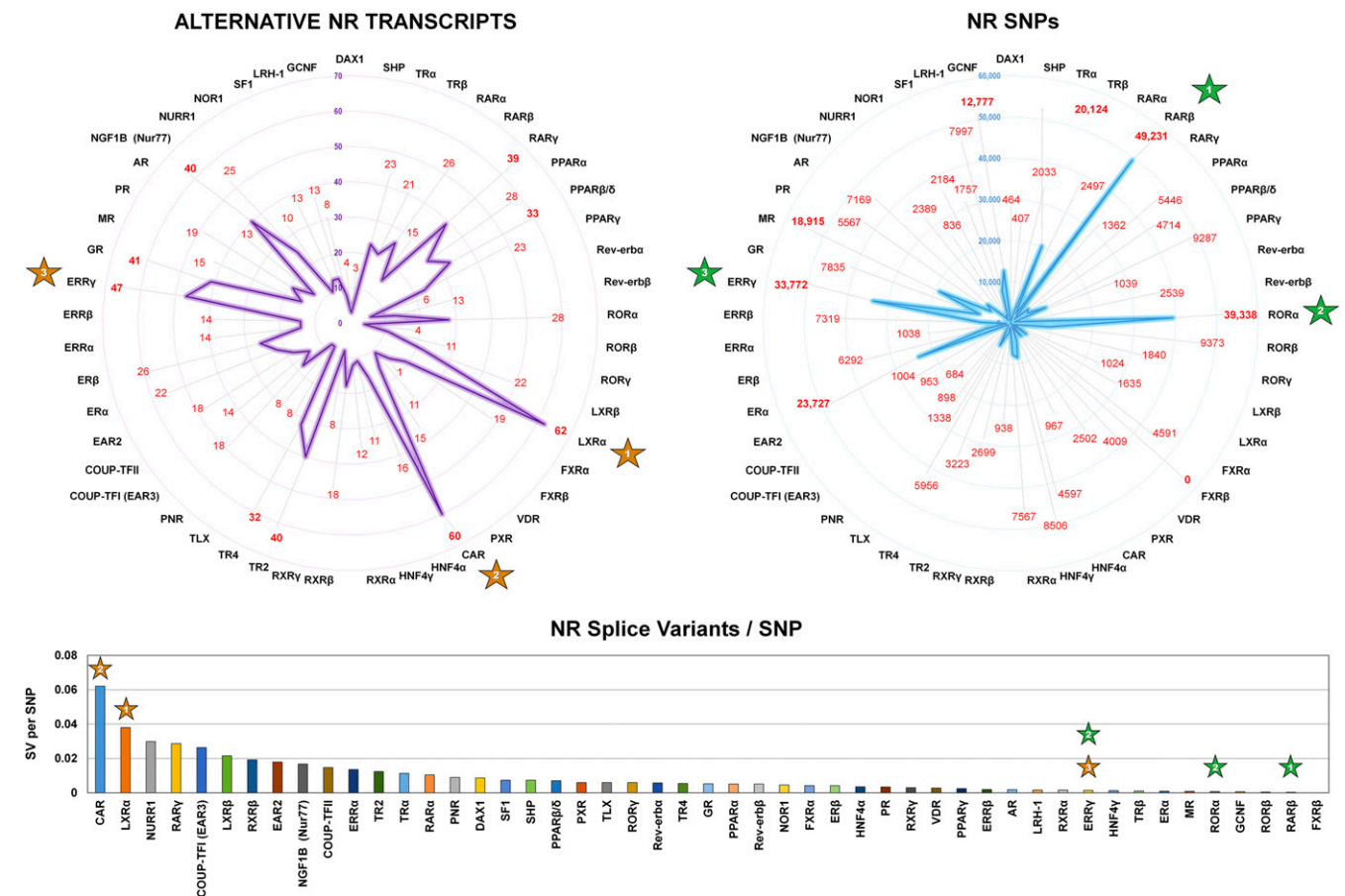


Fig. 1. Stereo meta-analysis of human nuclear receptor transcript variants and SNPs. The total number of mRNA transcript variants and SNPs for each of the 48 human nuclear receptors, and the processed NR pseudogene *FXRβ* (*NR1H5*), were cataloged using the NIH AceView and Ensembl data bases (Thierry-Mieg and Thierry-Mieg, 2006; Cunningham et al., 2015), and GeneCards.org (Stelzer et al., 2011), respectively. Each NR gene expresses, on average, 20.5 unique mRNA transcripts, which is comparable to the 16.7 transcripts per gene documented for the P450 gene superfamily (Annalora et al., 2017). *LXRα* (62), *CAR* (60), and *ERRγ* (47) were found to express the highest number of known mRNA transcripts variants per gene (shown highlighted with gold stars). In contrast, *FXRβ* (1), *SHF-1* (3), *Dax-1* (4), and *RORβ* (4) display the fewest number of transcript variants. Each NR genes is associated with ~6985 SNPs, compared with only 848 SNPs per gene for the related P450 superfamily, with *RARα* (49,231), *RORα* (39,338), and *ERRγ* (33,772) displaying the highest total number of SNPs among human NR genes (shown highlighted with green stars). *FXRβ* (0), *SHF-1* (407), and *Dax-1* (464) also have the lowest number of SNPs, distinguishing them from *RORβ*, which has very few transcript variants (4) but an above average number of SNPs (9373). Interestingly, the xenobiotic sensor *CAR* (*NR1H3*) expresses the most known splice variants per SNP (~0.06 variants per SNP or ~16 SNPs per variant), almost 7-fold more than the average NR gene (~0.009 variants per SNP or ~437 SNPs per variant), which suggests that splice variant expansion may arise from functional pressures rooted in ligand recognition, in addition to more random, mutation-driven events.

can be excluded from mature transcripts, without disrupting protein expression, and that splice variants lacking a portion of the LBD (encoded by exon 8) do not transduce the same ligand-dependent and/or ligand-independent signaling cascades as the wild-type receptor (Annalora et al., 2019). Here, we observed that each NR subfamily has its own unique repertoire of cassette exon structures that may represent an underappreciated level of gene structure/function complexity operating at the transcriptional level.

We speculate that most if not all eukaryotic genes have evolved cassette exons to augment tissue-specific protein function during bouts of elevated environmental stress or chemical exposure. Here we propose a new scheme for visualizing gene structures, defined by cassette exon organization, which emphasizes gene structure over sequence, to facilitate interpretations of how multiple transcript variants can arise from a single gene. As shown in Fig. 2A, our cassette exon visualization pictograph methodology, which we abbreviate as CEViP, stems from the observation that all coding exons exist in one of nine possible reading frames with respect to their intron/exon border, which we refer to as exon junction frame codes: 1–9 [see inset to Fig. 2A; frame code 1 = (0,0); 2 = (0,+1); 3 = (0,+2); 4 = (+1,0); 5 = (+1,+1); 6 = (+1,+2);

7 = (+2,0); 8 = (+2,+1); and 9 = (+2,+2)]. The CEViP modeling convention allows rapid identification of cassette exon structures embedded in transcriptomic sequences, which we define as any modular exonic component organized to be skipped in the mature mRNA transcript without altering the standard reading frame. The CEViP approach also provides new insights into the evolutionary relationships among distinct NR families and subfamilies, which would be more difficult to identify using traditional sequence alignments and hierarchical clustering methods generated by programs like MUSCLE or Clustal X version 2 (Edgar, 2004; Larkin et al., 2007). CEViP modeling provides valuable insight into the splicing potential of an exon and the cumulative sensitivity of a gene to AS mechanisms.

A schematic representation for each of the 48 human NR genes, generated using the CEViP methodology is shown in Supplemental Fig. 1, and a summary of these details is presented in Supplemental Table 1. In Fig. 2B, the four general classes of cassette exons are highlighted in comparison with a prototype transcript that encodes no cassette exons. These four classes include: 1) N- or carboxy (C)-terminal encoding cassette exons; 2) internal cassette exons with

TABLE 3
Nuclear receptor variants associated with human disease

Isoform	Variant	Related Disease	Reference
NR1A1	TR α 2	Tissue-specific processing in kidney and liver TR α 2 expression associated with breast cancer survival	Nakai et al., 1988 Jerzak et al., 2015
NR1B1	RAR α 1 Δ BC RARA fusion	Functional interaction with RAR α 1 PKRARIa:RARA fusion in promyelocytic leukemia	Parrado et al., 2001 Catalano et al., 2007
NR1C1	PPAR α	Truncated variants exert antitumor effects	Thomas et al., 2015
NR1C2	PPAR δ -trunc	3'-truncations (Δ Ex4–9) are dominant negatives	Lundell et al., 2007
NR1C3	PPAR γ Ex3'	Interfere with PPAR γ 1; chemotherapeutic resistance	Kim et al., 2007
NR1F3	ROR- γ t	Drive Th17 cell differentiation	Unutmaz 2009
NR1I1	VDR Ex1f	Underexpression in parathyroid adenoma	Correa et al., 2002
NR1I2	PXR T1–T4	Result in differential expression of UGT isoforms; PXR1 interacts with p53; PXR3 and 4 do not induce target gene expression	Gardner-Stephen et al., 2004 Brewer and Chen 2016
NR1I3	CAR2	CAR2 variant selectively activated by phthalate (DEHP)	DeKeyser et al., 2011
NR2A1	HNF4 α 2	Colon cancer tumor suppressor; HNF4 α 8 is not	Vuong et al., 2015
NR2B2	RXR β 2	Four-residue insertion in LBD; altered ligand binding, but forms heterodimers with TR, VDR, RAR, LXR; performs ligand-independent RXR functions only	Fujita and Mitsuhashi 1999
NR3A1	ER α Δ Ex3, Δ Ex3,7, Δ Ex6, Δ Ex2,4,5; Δ Ex7; Δ Ex5,7	Sixteen variants associated with breast cancer, endometrial adenocarcinoma; dominant negative Δ Ex3, Δ Ex2,4,5, Δ Ex3,7, Δ Ex6 enriched in cancer cells; Δ Ex7 induction regulated by estradiol; Δ Ex5,7 induced by tamoxifen, does not bind DNA promoter element	Poola and Speirs 2001 Horvath et al., 2002
NR3A2	ER β 2, β 5	Differential expression in prostate cancer; use of 5'-most exons (0K, 0N) drive ER β 2, β 5 expression	Lee et al., 2013
NR3C1	GR var	Promotes glucocorticoid resistance in multiple myeloma	Krett et al., 1995
NR3C3	PR Δ 2, Δ 4, PR Δ 5,6, PR Δ 6; delta6/2 ASV	Found in normal and breast cancer tissue; PR Δ 5,6, PR Δ 6 are dominant negatives; PR delta6/2 ASV 52 bp deletion of in exon 6 found in over 60% of breast cancer tissues; altered LBD	Richer et al., 1998 Hisatomi et al., 2003
NR3C4	AR-V7, AR-V9	Linked to castration-resistant prostate cancer; missing LBD but constitutively active; AR-V9 coexpressed with AR-V7, share similar 3' cryptic exon; predictive of drug resistance	Dehm et al., 2008 Antonarakis et al., 2014 Kohli et al., 2017

DEHP, di-2-ethylhexyl phthalate; Th17, type 17 T helper; UGT, UDP-glucuronosyltransferase.

complete or 3) incomplete reading frames, and 4) multiexonic, internal cassettes that must be skipped in combination to retain the standard reading frame. The modeling output of the CEViP methodology can also be used to explore questions relating to NR structure/function as exemplified for HNF4 α (NR2A1) in Fig. 3 and the VDR (NR1I1) in Fig. 4.

In Fig. 3A, the crystal structure of the full-length HNF4 α gene is shown in cylindrical cartoon representation to highlight how structural domains in the NR protein are organized with respect to the gene's nine primary coding exons (Chandra et al., 2013). Modular, cassette exon components of HNF4 α encoded by exons 5, 7, 8, and 9, which allow the protein to be expressed in full length or as multiple transcript variants lacking cassette exon–encoded domains, are also highlighted (Fig. 3B). The four cassette exons found in the human HNF4 α gene encode large segments of the LBD and C terminus, which may allow for the tissue-selective expression of at least seven unique HNF4 α transcript variants under normal conditions, as highlighted in Fig. 3C. Changes in cassette exon utilization patterns, triggered by oxidative stress, chemical exposure, mutation, or disease, may provide the HNF4 α gene with increased flexibility to bind and accommodate different classes of lipid ligands or to induce constitutively active variant forms with ligand-independent functionalities capable of modulating both basal and inducible gene expression patterns.

A second CEViP model is provided for the VDR in Fig. 4A to highlight the positioning of cassette exon 8 with respect to the gene's DBD, LBD, hinge region, and N and C termini. Figure 4B shows the positioning of key secondary structural elements encoded by cassette exon 8 (i.e., α -helices H7 and H8) within the crystal structure of the VDR's LBD (Rochel et al., 2000). When coupled with structural information, CEViP models allows rapid assessment of gene complexity and modularity based on the number and location of cassette exons within a gene's genomic structure; this information

allows predictions of structural plasticity potentially harbored in the alternative transcript pool while providing insight into the sensitivity and responsiveness a gene may have to the environmentally sensitive, AS mechanisms.

By applying the CEViP modeling method to all 48 human NR genes, we revealed that cassette exon utilization patterns are highly conserved within NR families and subfamilies. Indeed, cassette exon patterns offer a simple criterion for NR linkage that is not biased by sequence homology. In Fig. 5 the diversity of cassette exon structures among NR0 and NR1 family genes is highlighted. Primitive NR0 family genes (Dax-1 and SHP) that lack a prototypical DBD and contain no cassette exons are shown in Fig. 5A. These small NR genes may resemble the common ancestor for all modern NR genes, having linked a primitive DNA binding element to an endo-xenobiotic sensor/receptor for the first time (Reitzel et al., 2011). Beyond this early evolutionary advancement, more complex NR gene assemblies emerged, with longer and more complex N- and C-terminal transactivation domains, highly conserved DBD and LBD elements and, ultimately, cassette exon structures. Interestingly, some NR genes, like the NR1C subfamily (PPAR α , PPAR δ , and PPAR γ ; see Fig. 5B), lack any cassette exons in their reference gene coding structure, whereas related genes from the NR1A (TR α and TR β), NR1B (RAR α , RAR β , and RAR γ), NR1D (Rev-erb α and Rev-erb β), and NR1I (VDR, PXR, and CAR) subfamilies all contain at least one cassette exon within the LBD (see Supplemental Fig. 1). Even greater cassette exon complexity is observed in the NR1F (ROR α , ROR β , and ROR γ) and NR1H (LXR α , LXR β , and FXR α) subfamilies, which contain modular domains at both the N and C terminus, and within the DBD, LBD and hinge/dimerization domains (Fig. 5C). Based on only traditional sequence homology criteria, the NR1H family is most highly related to the NR1I subfamily, which includes the VDR (see Supplemental Fig. 2). However, NR1I subfamily genes encode a single cassette exon in the LBD of their transcripts, rather than the multiple

TABLE 4
Human nuclear receptor variant localization and expression

Isoform	Variant Name ^a	Localization/Features ^b	Tissue
NR0B1	DAX-1-var a,b	Cytoplasmic; Δ LBD	Carcinoid lung, adrenals
NR0B2	SHP-var a, b	Mitochondrial; Δ LBD	Liver, kidney
NR1B2	RAR β (4) variant	Cytoplasmic; subnuclear compartments	Mammary epithelial cells
NR1B3	RAR γ -var n	Δ LBD	Esophageal tumor
	-var o	Δ DBD	Embryonic stem cells
NR1C1	PPAR α -var e	Nucleus; Δ DBD	Choriocarcinoma
NR1C2	PPAR β/δ -var g	Cytoplasmic; Δ LBD	Prostate/ovarian cancer
	-var m	Nucleus	Uterus
NR1C3	PPAR γ -var j	Nuclear	Liver
	-var c	Cytoplasmic	Lung, colon macrophage
NR1F1	ROR α -var a, b & d	Cytoplasmic; no membrane associations	Testis, brain
NR1I2	PXR-var c	Cytoplasmic; Δ LBD	Liver, spleen, colon, stomach, tumor tissue
	-var f	Nucleus; Δ DBD	
NR1I3	CAR-var h	Cytoplasmic; Δ LBD/ Δ DBD	Liver
	-var w	Nucleus; Δ DBD	
	-var bb, cc	Mitochondrial; Δ DBD	
NR2B2	RXR β -var e	Nucleus; Δ DBD	B cell
	-var f	Cytoplasmic; Δ LBD/ Δ DBD	Testis, skin
NR2B3	RXR γ -var c	Cytoplasmic; Δ LBD (unspliced)	Eye
NR2F2	COUP-TFII-var g	Endoplasmic reticulum	Kidney, lung, uterus
NR2F6	EAR2-var b	Nuclear	Brain
	-var c	Cytoplasmic; Δ LBD	Lung, kidney, brain
	-var d	Mitochondrial	Eye, kidney
NR3C3	PR-var d	Nucleus	Breast cancer
	PR-var g	Cytoplasmic; Δ LBD	Adipose, aorta
NR3B1	ERR α -var c	Cytoplasmic	Lung, colon, kidney
NR4A2	NURR1-var i	Cytoplasmic; Δ LBD	Lung

EAR2, V-erbA-related gene EAR-2.

^aHuman nuclear receptor splice variants (var) associated with alternative subcellular trafficking or tissue distribution currently listed in the NIH AceView data base: <https://www.ncbi.nlm.nih.gov/IEB/Research/AceView/>.

^bSequence-based predictions of alternative subcellular trafficking for NR variants; Δ LBD refers to loss of a complete LBD coding sequence in the transcript variant; Δ DBD refers to loss of a complete DBD coding sequence in the variant transcript.

cassettes used by NR1H genes in the DBD, LBD, hinge, and N and C termini. This divergence in gene structural assembly would have been difficult to detect using primary sequences alignments alone. A cassette exon analysis suggests that the NR1H subfamily may have emerged from a different common ancestor, one that more closely resembles the NR isoforms found in the NR1B, NR1D, NR1F, or NR6A subfamilies. While these common features may have evolved independently of a common ancestor, based on convergent degeneracy within the original NR scaffold template, the strong conservation of cassette exon structures across NR gene families and subfamilies implies that functional differences imparted by modular gene assembly may have helped drive the dynamic expansion of the NR gene structure into 48 distinct, but overlapping biologic regimes.

In contrast to the NR1 family, all members of the NR2 subfamily have at least one cassette exon, with NR2F (Chicken ovalbumin upstream promoter transcription factor I (COUP-TFI), COUP-TFII, and V-erbA-related gene EAR-2(EAR-2)), NR2A (HNF4 α and HNF4 γ), and NR2E1 (Tailless homologue receptor (TLX)) subfamilies each containing cassette exons that encode portions of the hinge region or LBD. NR2B [retinoid X receptor (RXR) α , RXR β , and RXR γ] genes contain one cassette exon in the DBD/hinge region only, whereas the NR2C subfamily genes (TR2, TR4) use cassette exons in both DBD and LBD (see Supplemental Fig. 1 or Supplemental Table 1). It is notable that the single cassette exon found in the DBD/hinge of NR2B subfamily members (RXRs) is highly conserved across the NR2C subfamily and in all NR3 family members (Supplemental Fig. 3). In addition, all NR3, NR5, and NR6 subfamily members have dual cassette exons encoding portions of the LBD and DBD, with NR3 genes having additional C-terminal cassette exons, in contrast to extra N-terminal cassettes in the NR5 and NR6 families. NR3 family genes are further distinguished from all other NR genes by their N-terminal coding exon, which is not

a cassette exon, but which adds an additional ~400–600 amino acids to the transactivation domain (see Supplemental Fig. 1). In contrast, the NR4 family is largely devoid of any cassette exons, outside of one N-terminal cassette found in NR4A2 (Nurr1), which would make truncations of the transactivation domain and DBD possible. Ultimately, the global NR gene organization provided in Supplemental Fig. 1 reflects the hierarchical complexity of cassette exon utilization that we propose to be operating across the seven NR gene families. It is tempting to speculate that cassette exon evolution has not occurred randomly, and that these cryptic cassette features represent an additional dimension of gene complexity that guide or direct differential responses to cellular stress, on a scale that has yet to be fully comprehended or elucidated. A comparison of possible evolutionary trajectories for the seven human NR gene families based on phylogeny (as depicted in Supplemental Fig. 2) and cassette exon usage (as depicted in Supplemental Fig. 1) is shown in Supplemental Fig. 4.

Utilizing a cassette exon framework, we hypothesize that a single gene can explore the use of multiple, structural scaffolds “on-the-fly” to better detect and respond to chemicals that a tissue may have never been exposed to previously, due to genetic (or epigenetic) changes that alter endogenous metabolic output or via direct exposure to novel, synthetic, or environmental chemicals. In this regard, cassette exon utilization may function to enhance the resilience of a cell’s “chemical immune system” by enabling novel crosstalk networks to emerge among NR gene variants and their coactivators. This process would allow the cell to rapidly reshape the expression patterns of phase 0, I, II, and III metabolic systems (including NRs, P450s, and drug transporters) linked to their signal transduction pathways, as highlighted in Fig. 6. We speculate that this adaptive feedback response helps the cell refine the interplay among cognate, phase I–III metabolic systems, optimizing physiologic responses to endo-xenobiotics through the selection of splice variant

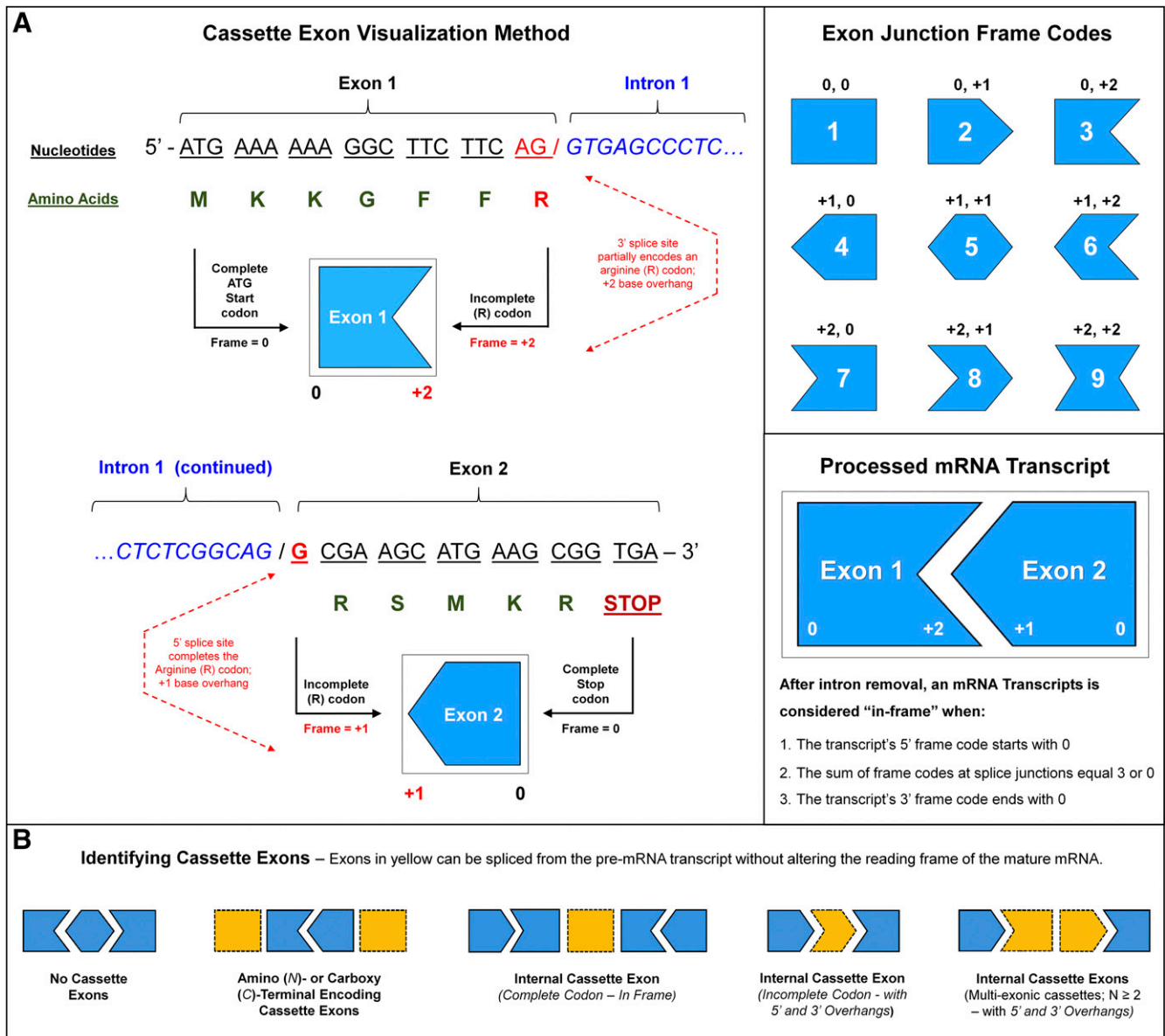


Fig. 2. CEViPs reveal cryptic aspects of NR gene structure. (A) Traditional gene schematics simply highlight the introns and exons encoding an underlying protein, but improved sequencing and annotation of reference genes has now created sufficient data to readily map the reading frame of individual exons at each intron/exon border or splice junction of a given gene. Exons can thus be classified using one of nine exon junction frame codes (see inset) based on the completeness of the codon that defines the 5' and 3' end of each exon in a pre-mRNA subject. The CEViP modeling convention provides insight into the organization of each exon with respect to neighboring exons and allows rapid identification of cassette exons subject to alternative splicing. An example of this methodology implemented upon a prototype two-exon gene is shown. The 5' codon of exon 1 has a complete ATG start codon, which is assigned the value of 0, and the 3' codon partially encodes (two-thirds complete) an arginine (R) codon and is assigned a frame of +2. The complete frame of exon 1 is therefore 0, +2, which matches exon junction code 3 (as shown in the inset). This exon can be spliced to any exon that starts with a +1 frame (i.e., exon junction frame codes 4, 5, or 6) to complete the terminal codon in exon 1. Here, exon 1 is shown joined to exon 2, which is the frame of +1, 0 (or frame code 4). This prototype transcript is considered in-frame because the splice junction components of the transcript sum to a total of 0 or 3, forming a complete three-base codon at each node of the transcript. We used this CEViP methodology to analyze the organization of cassette exons in all 48 human NR genes, and the complete analysis is shown in Supplemental Fig. 1. (B) Using CEViP modeling the location of putative cassette exons can be rapidly assessed, as splice junctions that join to form a complete codon, postsplicing, are easily visualized. Here we highlight the four primary classes of cassette exons that can be visualized with this method (colored orange with a black dotted line border), including: 1) N- or C-terminal encoding cassettes in frame code 1 (0, 0) that can be truncated from either end of the transcript; 2) complete internal cassette in frame code 1 (0, 0) that are flanked by exons with (x, 0) and (0, x) frame codes, where x can be any reading frame; 3) incomplete internal cassettes represented by frame codes 2–9, which provide the proper frame for joining two bordering exons in its absence; and 4) multiexonic cassettes, in any frame code, that must be skipped in combination to allow the remaining transcript to remain in-frame. These features provide valuable insight into the splicing potential of an exon and the cumulative sensitivity of a gene to either natural or environmentally responsive AS mechanisms.

transcripts that promote cellular homeostasis and health resilience after a chemical exposure.

Discussion

AS is a complex cotranscriptional process that is a function of several factors, including but not limited to gene coding structure, and the

expression of tissue-specific splicing factors that vary during aging and development, in genetic diseases, cancer, and infections, and during exposure to a changing environment. AS of pre-mRNA is tightly regulated by multiple cellular factors including a large family of RNA binding proteins known as serine/arginine-rich proteins, that include the heterogeneous nuclear ribonucleoproteins; these splicing factors

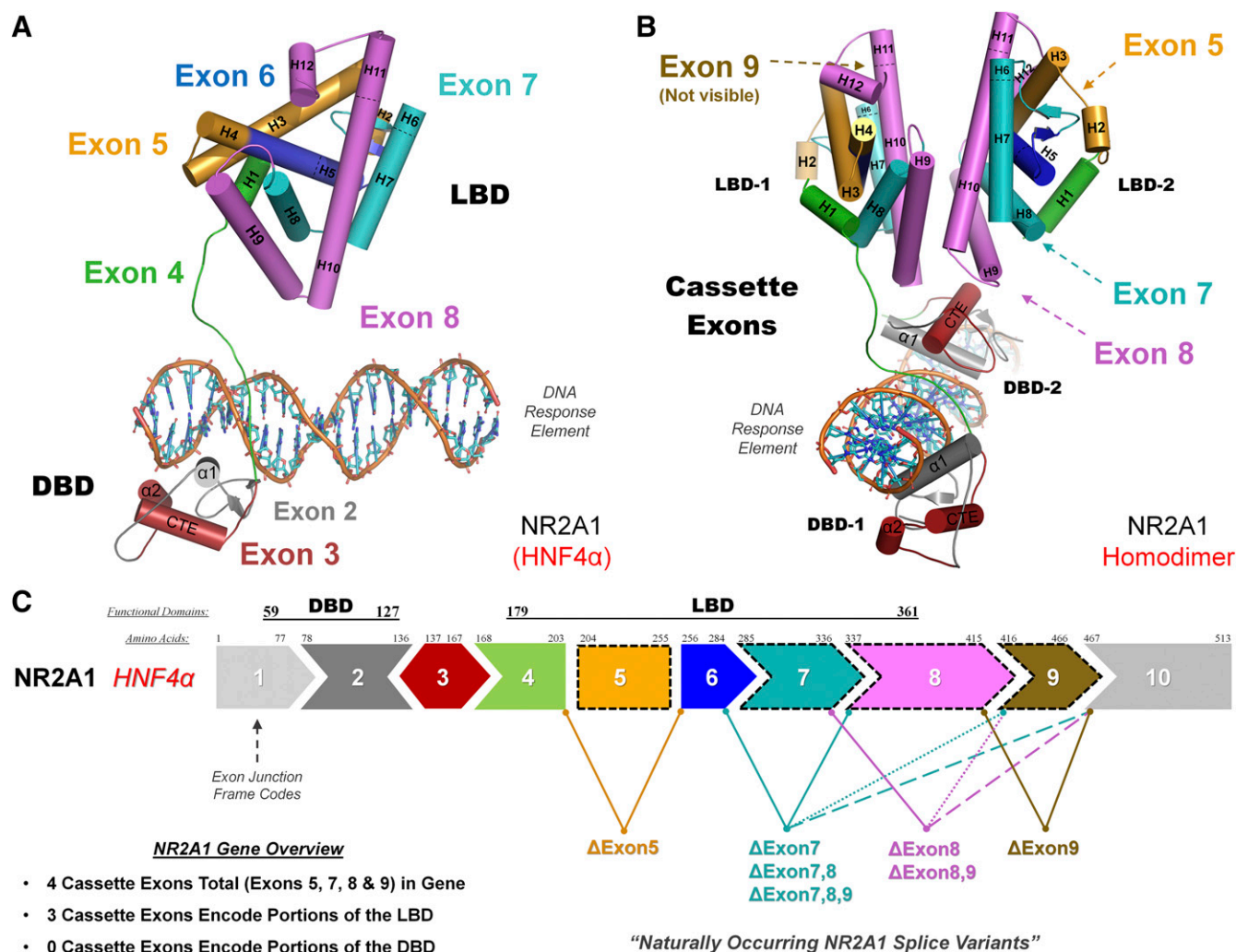


Fig. 3. Cassette exon structure of the prototype nuclear receptor NR2A1. (A) The crystal structure of human NR2A1 (HNF4α; Protein Data Bank identifier: 4IQR), shown here as a cylindrical cartoon diagram shows the structural basis for both the DBD and the LBD of a class I NR bound to a DR1 DNA binding element from the HNF4α promoter. The DBD is composed of three alpha helices [$\alpha 1$, $\alpha 2$, and the C-terminal extension helix (CTE)] partially encoded by exons 2 (gray cylinder) and 3 (red cylinder). The LBD is composed of ten alpha helices (H1–H12). Exon 4 (green cylinder) encodes the hinge region and helix H1. Exon 5 (orange cylinder) encodes helices H2–H4. Exon 6 (blue cylinder) encodes H5 and portions of the H5–H6 loop. Exon 7 (cyan cylinder) encodes portions of the H5–H6 loop and helices H6–H8. Exon 8 (magenta cylinders) encodes helices H9–H12. The human NR2A1 gene also contains two additional, C-terminal encoding, cassette exons (9 and 10; not shown in the crystal structure), which if encoded could extend the C terminus by an additional 76 residues. (B) Class II NRs like HNF4α form homodimers at DR1 sites (as shown), and the positions of putative cassette exons in HNF4α are highlighted. Exon 5 (orange cylinders), exon 7 (cyan cylinders), and exon 8 (magenta cylinders) are organized to be skipped in-frame, creating the potential for splice variants with unique ligand binding and dimerization properties. The position of additional residues encoded by cassette exon 9 are also indicated, highlighting the potential for C-terminal extensions beyond the AF-2 activation function in H12, which may facilitate the fine-tuning of coregulatory interactions under specific cellular conditions. (C) The modular assembly of the HNF4α reference gene is shown using exon junction frame codes based on the CEViP modeling method described in Fig. 2; individual exons are colored using the same scheme used in (A and B). The positions of cassette exons with respect to the DBD and LBD are highlighted, and the seven natural splice variants (Δ Exon5, Δ Exon7, Δ Exon7–8, Δ Exon7–9, Δ Exon8, Δ Exon8–9, and Δ Exon9) encoded within HNF4α's reference gene assembly are shown.

recognize a specific set of splicing regulatory elements that include exonic and intronic splicing enhancers, and exonic and intronic splicing silencers (Busch and Hertel, 2012). RNA polymerase II recruitment of transcription splicing factors in conjunction with the speed of transcription elongation ensure that that each splice site is selected in a competitive environment that promotes condition-specific gene expression and splicing (Lee and Rio, 2015). However, when individual splicing factors are absent or dysfunctional in a cell, dysregulation of highly evolved, splicing regulatory networks can arise and promote cellular stress or disease (as reviewed, Ule and Blencowe, 2019).

For example, spinal muscular atrophy, a rare genetic spinal motor neuron disease, is associated with reduced expression of the survival motor neuron (SMN) protein, a master heterogeneous nuclear ribonucleoprotein assembler, which results in a broad spectrum of splicing

alterations (Zhang et al., 2008). Developmentally regulated DNA methylation events have also been shown to promote the skipping of exon 5 in CD45 expression (Shukla et al., 2011), and epigenetic modes of AS regulation are increasingly recognized (Zhu et al., 2018b). However, nearly half of all age-related AS events (e.g., tissue deterioration, organ dysfunction, telomere attrition, and loss of stem cell renewal) are due to changes in the expression of splicing factors (Mazin et al., 2013). Overexpression of the human oncogenic transcription factor related to an oncogene carried by the avian virus, Myelocytomatosis (v-myc) (MYC) in tumors can also yield multiple oncogenic AS events due to an upregulation of splicing factors (Anczuków and Krainer, 2016; Das et al., 2012). Stem cell-specific AS programs have also been established, and are mediated by RNA binding fox-1 homolog 2 (RBFOX2), a key player in mesenchymal tissue splicing (Venables

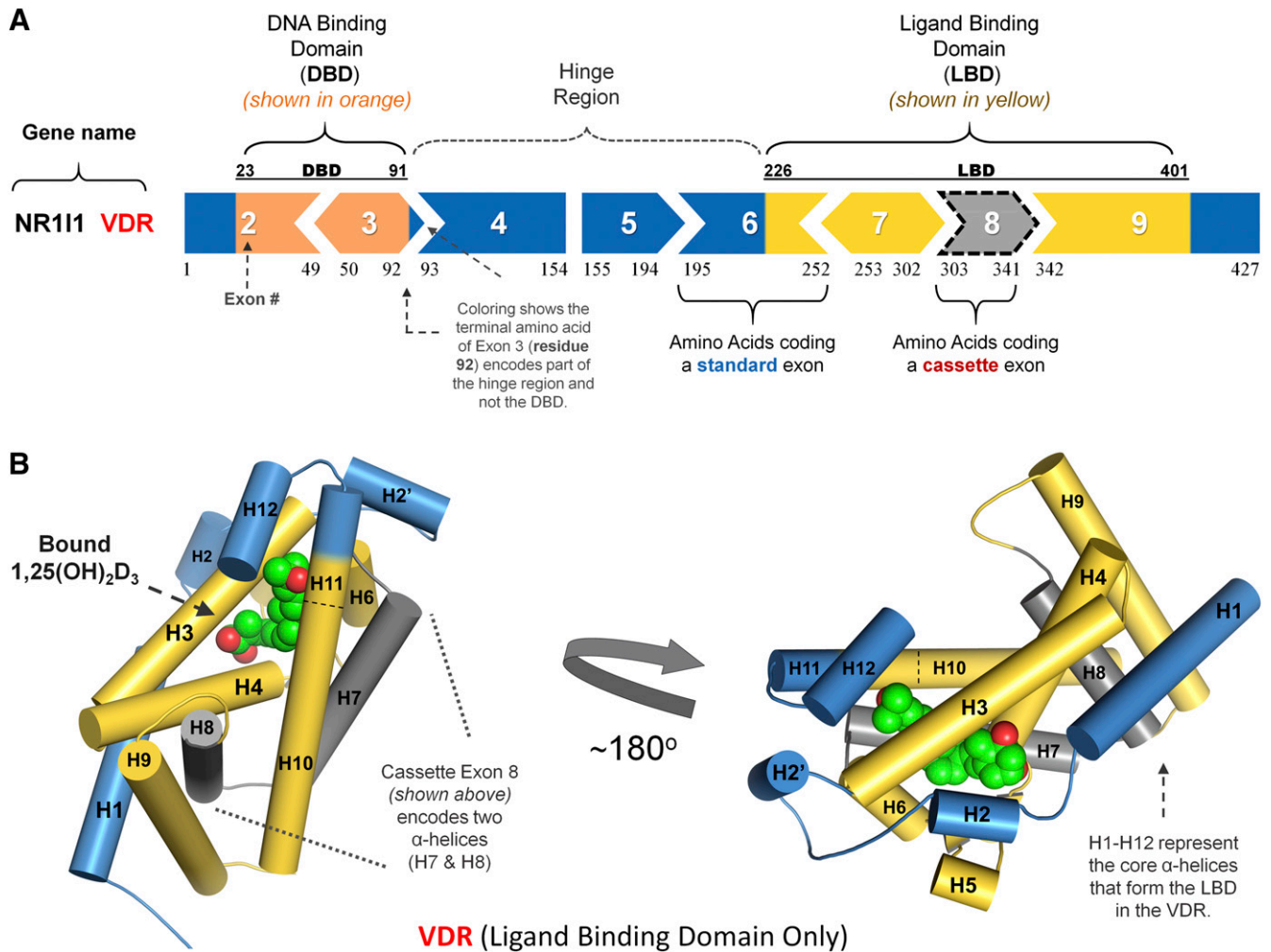


Fig. 4. CEViP modeling of the vitamin D receptor. (A) A schematic representation of the human VDR gene (NR111) is shown using the CEViP modeling method, which has been augmented to highlight key secondary structures in the NR gene family, including the DBD (shown in orange) and the LBD (shown in yellow). The blue region between the DBD and LBD is the hinge region, and the length of individual exon junction frame codes is proportional to the total number of amino acids encoded by each exon, which is also shown numerically. Exon numbers are also noted with respect to the reference gene; in the VDR the reference AUG start site is in exon 2, as exon 1 features are not commonly expressed. Cryptic N-terminal exons were not considered in our analysis because of annotation challenges for some NR genes outside the reference gene sequence. Exon 8 of the VDR is a cassette exon because it contains a (+2, +1) reading frame that can be excluded without disrupting the reading frame of exon 7 (+1, +1) and exon 9 (+2, 0). Exon 8 (in gray) is highlighted with a bold, black dashed line, and this convention is used to distinguish all cassette exons in the NR superfamily, as shown in Supplemental Fig. 1. (B) The canonical LBD of the VDR is composed of 12 α -helical elements (named H1–H12), and the crystal structure (Rochel et al., 2000) is shown here, in two orientations, using a cylindrical cartoon model colored using the same convention as the CEViP model in Fig. 4A. Cassette exon 8 (shown in gray) encodes two α -helices (H7 and H8) that form a significant portion of the 1,25-dihydroxyvitamin D₃ [1,25(OH)₂D₃] binding pocket in the LBD of the VDR. AS of exon 8 has been shown to alter both ligand-dependent and ligand-independent functionalities of the VDR (Annalora et al., 2019), and we have cataloged similar cassette exon structures within the LBD of several additional human NR genes, including those found in the NR 1B, 1H, 1I, 1F, 2A, 2E, 2C, 5A, and 6A subfamilies (see Supplemental Fig. 1).

et al., 2013). Understanding of the scope and importance of tissue-specific AS and tissue-dependent regulation of AS in vertebrates is growing (Blencowe, 2006; Barbosa-Morais et al., 2012; Merkin et al., 2012), as many of these events are lineage-specific, providing new insights into how phenotypic differences arise among vertebrates. The cellular consequences of AS in the NR family are highlighted in Table 4; these events add to the complexity and refinement of expression and phenotype necessary to respond to chemical challenges in the environment.

AS is also associated with evolutionarily significant alterations to the genome (Modrek and Lee, 2003) including, exon duplication events (Letunic et al., 2002), retrotransposon (i.e., Alu element)–mediated exonization (Sorek et al., 2002), exon creation (Pan et al., 2008), and the introduction of premature protein termination codons (Xing and Lee, 2004). AS may also accelerate paths of evolution based on observations that nonsynonymous mutations (*ka*) are favored over synonymous

mutations (*ks*) (Yang and Bielawski, 2000); *ka/ks* is an established selection pressure metric that participates in reading frame preservation by alternate splicing (Xing and Lee, 2005). Our observation that cassette exons may play discrete roles in directing proteome expansion for NR genes, is consistent with the proposed role for AS in fine-tuning elements of gene structure/function and evolution (Bush et al., 2017; Grau-Bové et al., 2018). However, the full extent to which cassette exons are functionally relevant and conserved across other mammalian species remains largely unknown (Sorek et al., 2004).

These insights into the importance of AS events in evolution lead to a sharpened focus on the RNA sequence motifs that may regulate gene splicing in both intronic and exonic locations. Recent studies indicate that there are at least 42 different consensus sequences that are *cis*-elements, which regulate splicing at consensus “GU/AG” splice-site donor/splice-site acceptor exon junctions (Qu et al., 2017b). Mutations that alter these *cis*-elements can profoundly affect gene splicing and

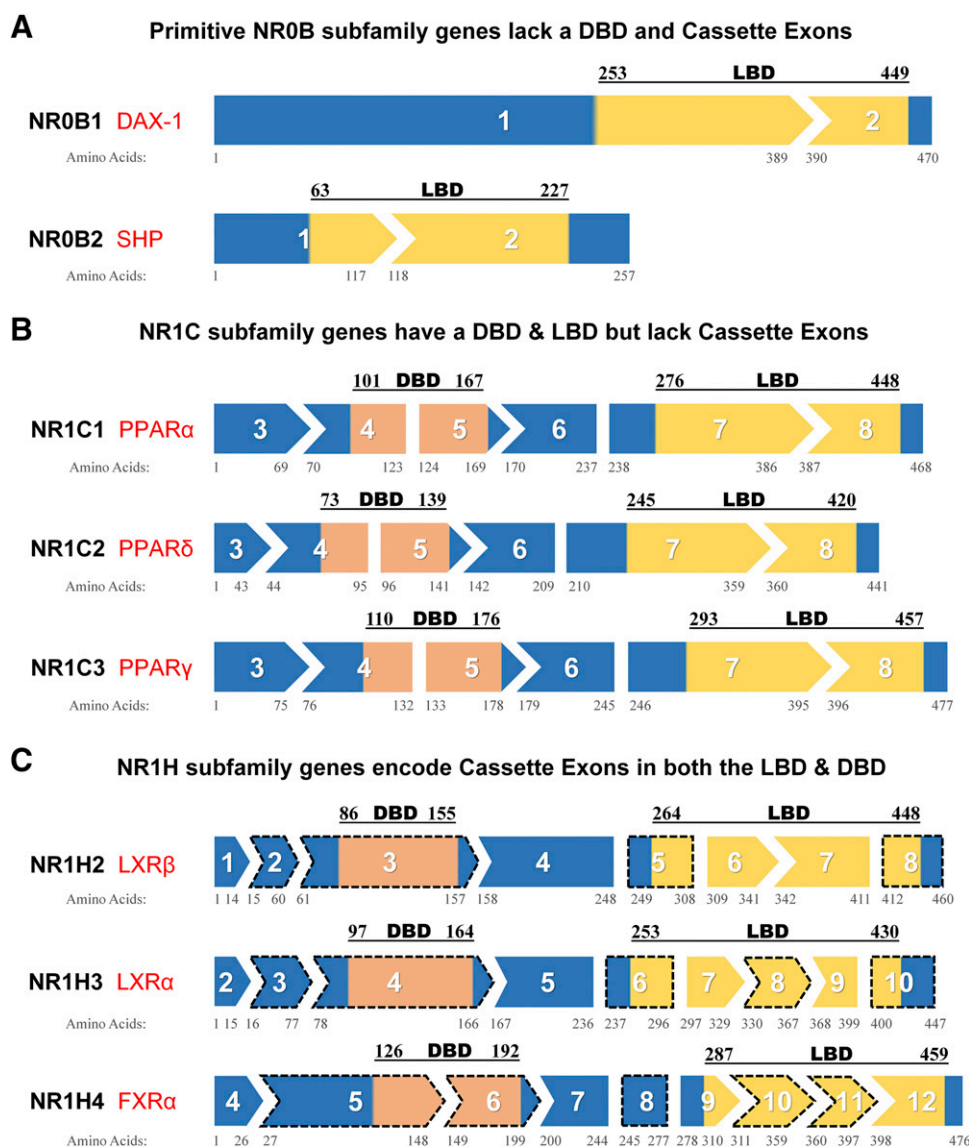


Fig. 5. Exploring cassette exon evolution in the NR0 and NR1 gene families. (A) The NR0B subfamily (i.e., DAX-1 and SHP) is identified as an ancestral NR gene group with unique sequence features compared with all other NR families. Interestingly, these NR genes lack a canonical DBD but encode a classic LBD segmented across two standard, noncassette exons. (B) More highly evolved genes from the NR1C subfamily (PPARs) contain a similar organization of the LBD and C terminus but contain an expanded N terminus containing the canonical DBD (shown in orange) encoded by two standard exons. The NR1C subfamily is distinct from all other NR1 subfamilies, which all contain at least one cassette exon. Two NR4 family members [NR4A1 (NGF1B) and NR4A3 (NOR1)] also contain no cassette exons, but a third member, NR4A2 (NURR1) has one N-terminal cassette exon in the reference gene sequence (see Supplemental Fig. 1). Phylogenetic analysis of NR gene coding sequences completed using a MUSCLE sequence alignment and the PHYLIP Neighbor joining tree drawing program in Unipro UGENE (Okonechnikov et al., 2012), (see Supplemental Fig. 2) does not provide a clear basis for how the NR1C3 subfamily emerged from other NR1 family members, or a common ancestor from the NR2 family, without encoding any cassette exons (see Supplemental Fig. 5). Based on cassette exon usage, the NR1C family may be more closely related to the NR0 and NR4 families than other NR1 or NR2 family genes. (C) The elaborate cassette exon signatures of the NR1H subfamily [NR1H3 (LXRα), NR1H2 (LXRβ), and NR1H4 (FXRα)] highlight the expansion of cassette exon utilization in the NR1 family, with each gene incorporating four to five cassette exons. The functional implications of these differences are currently unknown, but they may subject these genes to higher levels of tissue-specific, alternate splicing.

expression, and gene-directed therapeutics that modulate aberrant splicing are increasingly being explored. For example, protein-truncating mutations define Duchenne muscular dystrophy (DMD), which compromises muscle fiber integrity (Emery, 1989), in contrast to reading frame-preserving mutations, associated with Becker muscular dystrophy, a condition with slower progression and milder symptoms (Heald et al., 1994). Apparent deletion hotspots lead to approximately two-thirds of all DMD cases (Aartsma-Rus et al., 2002), generating new interest in finding antisense therapeutics that induce skipping exons to preserve reading frame. The search for antisense oligonucleotides that can induce exon skipping in the dystrophin pre-mRNA revealed some exons to be refractory to SSO-induced exon skipping (Aartsma-Rus et al., 2005; Wilton et al., 2007). Conversely, multiexon skipping of exons 6 and 8 was accomplished in the canine dystrophin gene to restore reading frame (Yokota et al., 2009; Echigoya et al., 2017). Building from the success of SSO therapeutics such as Eteplirsen in treating DMD in humans (Lim et al., 2017; Charleston et al., 2018), we hypothesized that that single- and multicassette exon skipping approaches targeting nuclear receptor genes might also be feasible, based on the modular nature of the NR protein scaffold and the family-specific conservation of cassette exons within key functional domains.

We have now identified putative cassette exons operating within the LBD and DBD of numerous human NR genes, and their alternative usage can promote the expression of both gain-of-function (i.e., constitutively active) and loss-of-function (i.e., dominant negative) phenotypes that strongly support a significant role for AS in regulating NR gene function. For example, in the AR, aberrant skipping of the exon(s) that encodes the LBD promote the expression of a ligand-independent AR splice variant that drives prostate cancer progression (Dehm et al., 2008; Guo and Qiu, 2011; Daniel and Dehm, 2017; Antonarakis et al., 2014; Kohli et al., 2017). Constitutively active splice variants of the MR (NR3C2) and the AhR, a related basic-helix-loop-helix-Per-ARNT-Sim family nuclear receptor, have also been identified, and these variants lack discrete segments of the LBD (McGuire et al., 2001; Zennaro et al., 2001). Ligand-independent ER (NR3A1) variants have also been reported with implications for the treatment of breast cancer (Heldring et al., 2007; Zhu et al., 2018a; Al-Bader et al., 2011). More directly, our group recently showed that targeted skipping of cassette exon 8 in the human VDR (NR1H1) promotes a splice variant form (Dex8-VDR) that stimulates ligand-independent transcriptional activity capable of suppressing colon cancer cell growth (Annalora et al., 2019). Conversely, AS of exon 3, which encodes a portion of the DBD,

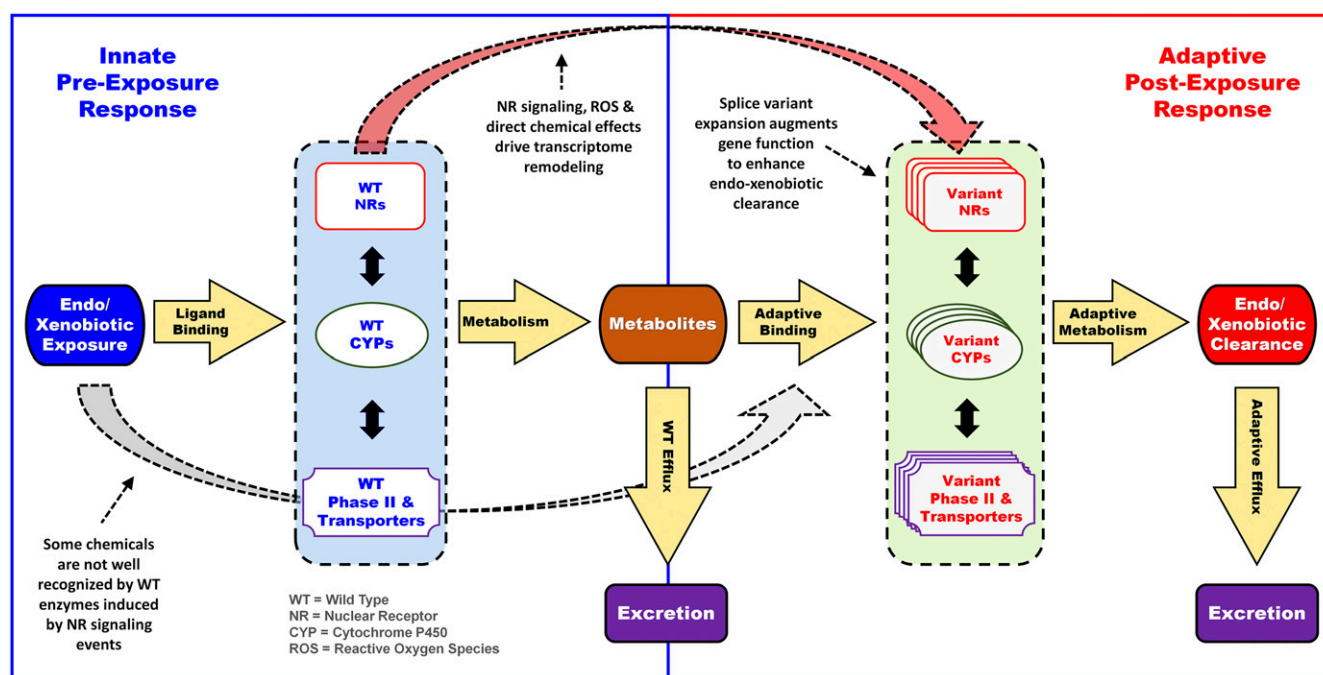


Fig. 6. Alternative splicing of metabolic genes refine adaptive physiologic responses to endo-xenobiotic exposure. NR gene function is modulated, inter alia, by chemical exposures and the spectrum of substrate metabolites generated by phase I metabolic enzymes, exemplified by the P450 superfamily of genes. Access to coordinating NR ligands is also influenced by phase II and phase III drug metabolism, which promote the biotransformation and excretion of soluble metabolites and ligands. Drug transporters also play an important role in this process and appear to have coevolved in concert with related NR and P450 genes, as exemplified by the interplay among PXR, CYP3A4, and P-glycoprotein (also known as multidrug resistance protein 1), a well characterized **ATP-binding cassette (ABC) transporter**, encoded by the **ATP-binding cassette sub-family B member 1 (ABCB1)** gene, also known as multidrug resistance protein 1 (MDR1) (Chen et al., 2012). Ligand-directed NR gene activation, reactive oxygen species (ROS) production, and direct chemical effects on gene transcription, splicing, and translation machinery are among the primary drivers of AS and transcriptome remodeling during periods of elevated cellular stress. We speculate that this adaptive feedback response helps the cell refine the interplay among cognate, phase I–III metabolic systems, optimizing physiologic responses to the environment through splice variant selection that promotes a return to cellular homeostasis after both acute and chronic endo-xenobiotic exposures. WT, wild type.

promoted the expression of a dominant negative splice variant that disrupts the VDR's ligand-independent, gene-repressor function (Annalora et al., 2019). Using the NIH AceView data base (Thierry-Mieg and Thierry-Mieg, 2006), we have already identified 12 NR genes previously linked to cancer that generate splice variants lacking a complete DBD; these include: PPAR β/δ , RAR α , RAR γ , LXR α , LXR β , MR, ER α , NGF1B, NURR1, and Steroidogenic factor 1 (SF1). An additional 14 NR genes previously linked to cancer that express splice variants lacking a complete LBD have also been identified by RNAseq methods; these include PPAR α , PPAR γ , RAR β , Rev-erb β , LXR α , LXR β , VDR, PXR, HNF4 α , TR4, AR, ERR α , NGF1B, and NURR1. Although the mechanisms driving NR splice variant expression in cancer remain poorly understood, we are intrigued by the potential for antisense therapeutics to correct or reprogram aberrant gene expression profiles linked to cancer and other diseases, as reviewed (Martinez-Montiel et al., 2018).

The potential for developing antisense approaches that allow for targeted expression of ligand-independent NR variants capable of signaling in the absence of an endogenous ligand is also intriguing (Annalora et al., 2019). Equally compelling would be the manipulation of the DNA binding domains to create "therapeutic" dominant negative variants capable of reprogramming genomic signaling pathways in the absence of ligands. Our work with the VDR has already demonstrated that splice variants lacking a complete LBD can fold properly and bind antibodies directed toward the reference protein, in addition to forming unique homodimer assemblies (with the reference protein) that migrate differentially on an SDS-PAGE gel (Annalora et al., 2019). However, improved biochemical and structural characterization of NR splice variants, including those lacking well conserved cassette

exon structures, are needed to further improve our understanding of the role AS plays in modulating global NR function. While our current understanding of NR splice variant structure/function remains limited due to the lack of structural information, AS events that remodel the LBD of xenobiotic sensor genes, like PXR or CAR, have the potential to dramatically modify gene function by altering the determinants of ligand binding, cofactor recruitment, and signal transduction (Auerbach et al., 2005; Chen et al., 2010; Wallace and Redinbo, 2013).

It is also important to recognize that cassette exon utilization is also linked to the expression of covalently closed, single-stranded circular RNAs (circRNAs), which are noncoding transcripts consisting of introns and/or exons (Salzman et al., 2012). These transcripts are highly stable, broadly distributed in eukaryotic cells, and increasingly recognized for a role in regulating gene expression and splicing, as they can interact with serine/arginine-rich proteins, microRNA, and long noncoding RNAs. CircRNAs are now implicated for roles in the etiology of a range of human disorders, including cancer, heart disease, diabetes, and aging (Qu et al., 2017a), and they represent an important class of noncoding RNA that are increasingly studied. Connecting functional circRNA expression to their host genes, and linking their assembly to the skipping of discrete cassette exon structures may help to provide deeper insight into the complex structure of eukaryotic gene assembly and greater appreciation for the role that AS events play in guiding host resilience to changes in the genome, microbiome, diet, or environment.

The primary goal of this review is to highlight the dramatic expansion of the NR transcriptome that is observed when all known splice variants (1004 total; Table 1) are recognized. This insight highlights how each NR gene can yield, on average, ~20 splice variants to augment their

gene function in both a species- and tissue-specific manner, based on gene structure, the accumulation of SNPs, variations in splicing factor expression, and changes in cellular stress that may alter DNA and RNA stability. Thermodynamic determinants of gene splicing/expression that operate independently from other posttranscriptional or posttranslational modifications can also further fine-tune NR gene expression and function. Our desire to understand the complexity of cassette exon utilization across the NR superfamily, led us to develop the CEViP methodology shown in Fig. 2A, and it helps to highlight the remarkable level of conservation of modular cassette motifs operating across both NR family and subfamily. This approach provides additional details about the sequences defining each intron/exon border within a gene, facilitating the identification of cassette exons and the visualization of potential alternative transcripts. However, without a comparative analysis of cassette exon structures from related mammalian species, it is currently difficult to contextualize the importance of these conserved features from an evolutionary perspective. We do not currently know if each species has a different cassette exon reservoir for orthologous genes.

If we make predictions of NR evolution based on an analysis of human cassette structures only (see Supplemental Fig. 4), it implies that the NR0 family is the most primitive class of NRs, followed by the NR1C and NR4A subfamilies, which essentially lack cassette exons in the reference gene, followed by the NR2A family which contains up to two cassette exons that are not well conserved in any other subfamily. However, there is strong support in the literature that suggests NR2A1 (HNF α) is the primordial NR gene that emerged from our common metazoan ancestor (demosponges) (Bridgman et al., 2010). In humans, this gene appears structurally complex compared with other NR genes, having four cassette exons; however, we speculate that this may represent organism-specific evolution of the gene, rather than high-level complexity intrinsic to the primordial gene. Moreover, the conservation of cassette exon structures observed in the NR2A subfamily is particularly striking, because although the size and location of NR2A cassette exons appears to have drifted over time, the overall organization of cassette exons has not (Supplemental Fig. 1). Ultimately, our approach for analysis of cassette exon structures imply that the NR2E subfamily of genes (TLX and Photoreceptor-specific nuclear receptor (PNR)) may be more ancestral than the NR2A family, due to the lack of well conserved cassette exon structures in this family compared with all other NR2 family members (Supplemental Fig. 5). While these findings are not definitive, cassette exon visualization provides a new approach to study NR gene evolution across different gene families, species and time. This tool may help to address the remaining controversy regarding the early evolution of NR genes, from an ancestral form participating in endogenous signaling pathways devoid of hormones, to more modern versions that perform a xenobiotic sensor role for the cell (Holzer et al., 2017). Interestingly, there is some evidence that the oldest NR genes may have first arisen in Ctenophores, a more primitive phylum related to Cnidarians, which curiously express NR genes that cluster with the NR2A family, but contain only a consensus LBD, which is strikingly similar to the assembly of NR0B subfamily genes in humans (Reitzel et al., 2011). A phylogenetic analysis of the primary amino acid sequences for all 48 human NR genes (see Supplemental Fig. 2) places the NR0 family at the top of the ancestral tree, but this is likely biased by their lack of a conserved DBD element. Beyond the NR0 family, NR gene expansion appears to have proceeded primarily through the diversification of two gene scaffolds, as exemplified by the NR2E (TLX and PNR) and NR6A (Germ cell nuclear factor (GCNF)) subfamilies. Interestingly, these subgroups possess two of the most unique cassette exon signatures we have observed, particularly for the NR6A (GCNF) gene, which utilizes five cassette

exons that are not fully conserved, or recapitulated, in any other NR subfamilies (Supplemental Fig. 5). One possible explanation is that the utilization of cassette exons was less functionally relevant when the NR2E and NR6A subfamily genes were emerging, and that modern cassette exon utilization patterns, which may have arisen from random events, were standardized later to help maintain the reading frame of duplicated genes. AS mechanisms may have placed different selection pressures on these cassette exon features over time, with some becoming important splice-sensitive drivers of structural adaptation which today enhances gene resilience by allowing for stress-related proteome expansion, and tissue-specific, fine-tuning of NR gene functions. Thus, a deeper analysis of cassette exon utilization patterns across multiple species similar to what has been done for primary coding sequences (Zhao et al., 2015), could help to clarify the organism-specific drivers guiding NR gene diversification in humans as well as other species.

We also recognize that the diversification of internal cassette exons is not the only mechanism by which NR genes adapt their function to changing selection pressures. AS that alters the expression of the activation function 1 (AF-1) and activation function 2 (AF-2) motifs found in either the N or C terminus of the protein can also profoundly alter reference gene function. This is exemplified by the NR3C subfamily (Glucocorticoid receptor (GR), MR, Progesterone receptor (PR), AR), which appears to have evolved a highly extended, N-terminal transactivation domain (~400–600 amino acids in length), encoded by a single, noncassette exon, to augment their signaling functions (Supplemental Fig. 1). An additional set of NR genes (including CAR, VDR, and GR) are known to modulate their function via the expression of naturally occurring, C-terminal variants (van der Vaart and Schaaf, 2009). While the function of these truncated NR variants remains unclear, many are thought to be dominant negatives incapable of efficiently binding endogenous ligands, coregulatory proteins or canonical DNA response elements. However, work involving the role of the AR gene in cancer has demonstrated that some C-terminal splice variants (i.e., AR-V7 and AR-V9) can also promote constitutively active NR phenotypes as well (Dehm et al., 2008; Antonarakis et al., 2014, Kohli et al., 2017). Therefore, internal cassette exons structures that alter the expression of the DBD, LBD, and hinge regions may represent only one strategy NR genes use to augment their cellular function. For example, in the human NR1C subfamily (PPARs) the role of cassette exons may have become superfluous or counterproductive over time and therefore were completely eliminated.

In summary, NR genes are highly prone to AS events and their pattern of cassette exon utilization provides a new lens for interpreting how these biologic sensors have evolved to accommodate different ligands, which are often the products of metabolic or environmental exposure pathways subject to variation over time (Holzer et al., 2017). As highlighted in Fig. 6, the expansion of multiple transcript variants from a single gene provides a cell with the powerful ability to find novel protein-based and noncoding RNA-based solutions to biochemical challenges that disrupt homeostasis, through the expression of novel transcript variants. We previously described how splice variant expression in the P450 superfamily of genes provides an expanding phenotypic cloud of functional plasticity that can augment a P450 gene's ability to promote cell survival and fitness (Annalora et al., 2017). After exploring similar trends in the interrelated NR superfamily, we continue to speculate that splice variant expression, working in concert with reference gene function, provides the basis for sophisticated biologic adaptation to novel chemical insults. Because NR genes coordinate the expression and splicing of virtually all metabolic genes (including P450s and drug transporters) involved in the "chemical immune system," their own sensitivity to AS provides the cell with an additional synergistic layer of complexity for bootstrapping new and increasingly complex

biologic responses to the environment. For example, transcript variants (T1–T3) of PXR (NR1I2), are known to promote differential expression of both CYP3A4 (Lin et al., 2009) and UDP-glucuronosyltransferase 1A isoforms (Gardner-Stephen et al., 2004). The ability of PXR variants to coordinate tissue-specific patterns in the expression of both phase I and II metabolic genes highlight the potential of alternatively spliced NR genes to optimize cellular feedback mechanisms linked to drug influx (phase 0) and all phases (I–III) of drug metabolism.

While there are still some uncertainties regarding the role that AS events play in proteome expansion and the fine-tuning of reference gene functions under normal cellular conditions (Chaudhary et al., 2019), there is increasingly strong biochemical evidence to indicate that the majority of splice variants are translated (Weatheritt et al., 2016) and that they help shape normal human physiology and development by contributing a vast diversity of “functional alloforms” to the human proteome ((Yang et al., 2016)). The role of NR variants in driving idiosyncratic differences in human drug metabolism are currently less well understood, but the ability of PXR splice variants (T1–T3) to fine-tune the expression of both phase I and phase II metabolic systems highlights the potential of NR variants to personalize individual responses to endo-xenobiotic exposure. While NR variants driving human disease are currently more well characterized and understood than those driving interindividual variability in human drug metabolism, there is ample reason to believe that future advances in RNAseq methodologies, coupled with improved computational analysis, will provide improved diagnostics of AS events that may underlie or drive common chemical toxicities, adverse drug events, and individual susceptibility to environmental disease.

Ultimately, the splicing-related mechanisms affecting NR gene function explored in this review correspond well with earlier observations that SNPs and epigenetic modifications that alter the NR gene expression also induce differential expression of drug-metabolizing enzymes and transporters, resulting in variable drug responses that may complicate or limit the success of precision medicine (Prakash et al., 2015). Improved understanding of the role adaptive splicing mechanisms play in fine-tuning drug metabolism is therefore needed to enhance future predictions of chemical toxicity and adverse drug events, and this information would also provide an improved roadmap for implementing precision medicine, which if guided by pharmacogenomics alone, may continue to miss the hidden complexities that render one person's poison another person's cure.

Acknowledgments

This research was supported in part by funds from the Agricultural Research Foundation (ARF) at Oregon State University.

Authorship Contributions

Participated in research design: Annalora, Marcus, Iversen.

Conducted experiments: Annalora, Iversen.

Contributed new reagents or analytic tools: Annalora, Iversen.

Performed data analysis: Annalora, Iversen.

Wrote or contributed to the writing of the manuscript: Annalora, Marcus, Iversen.

References

Aartsma-Rus A, Bremmer-Bout M, Janson AA, den Dunnen JT, van Ommen GJ, and van Deutekom JC (2002) Targeted exon skipping as a potential gene correction therapy for Duchenne muscular dystrophy. *Neuromuscul Disord* **12** (Suppl 1):S71–S77.

Aartsma-Rus A, De Winter CL, Janson AA, Kaman WE, Van Ommen GJ, Den Dunnen JT, and Van Deutekom JC (2005) Functional analysis of 114 exon-internal AONs for targeted DMD exon skipping: indication for steric hindrance of SR protein binding sites. *Oligonucleotides* **15**: 284–297.

Adams AM, Harding PL, Iversen PL, Coleman C, Fletcher S, and Wilton SD (2007) Antisense oligonucleotide induced exon skipping and the dystrophin gene transcript: cocktails and chemistries. *BMC Mol Biol* **8**:57.

Al-Bader M, Ford C, Al-Ayadhy B, and Francis I (2011) Analysis of estrogen receptor isoforms and variants in breast cancer cell lines. *Exp Ther Med* **2**:537–544.

Anczukow O and Krainer AR (2016) Splicing-factor alterations in cancers. *RNA* **22**:1285–1301.

Annalora AJ, Jozic M, Marcus CB, and Iversen PL (2019) Alternative splicing of the vitamin D receptor modulates target gene expression and promotes ligand-independent functions. *Toxicol Appl Pharmacol* **364**:55–67.

Annalora AJ, Marcus CB, and Iversen PL (2017) Alternative splicing in the cytochrome P450 superfamily expands protein diversity to augment gene function and redirect human drug metabolism. *Drug Metab Dispos* **45**:375–389.

Antonarakis ES, Lu C, Wang H, Luber B, Nakazawa M, Roeser JC, Chen Y, Mohammad TA, Chen Y, Fedor HL, et al. (2014) AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N Engl J Med* **371**:1028–1038.

Auboeuf D, Dowhan DH, Duterte M, Martin N, Berget SM, and O'Malley BW (2005) A subset of nuclear receptor coregulators act as coupling proteins during synthesis and maturation of RNA transcripts. *Mol Cell Biol* **25**:5307–5316.

Auerbach SS, Stoner MA, Su S, and Omiecinski CJ (2005) Retinoid X receptor- α -dependent transactivation by a naturally occurring structural variant of human constitutive androstane receptor (NR1I3). *Mol Pharmacol* **68**:1239–1253.

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodenic V, Kutter C, Watt S, Colak R, et al. (2012) The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**:1587–1593.

Bhat-Nakshatri P, Song EK, Collins NR, Uversky VN, Dunker AK, O'Malley BW, Geistlinger TR, Carroll JS, Brown M, and Nakshatri H (2013) Interplay between estrogen receptor and AKT in estradiol-induced alternative splicing. *BMC Med Genomics* **6**:21.

Blencowe BJ (2006) Alternative splicing: new insights from global analyses. *Cell* **126**:37–47.

Brewer CT and Chen T (2016) PXR variants: the impact on drug metabolism and therapeutic responses. *Acta Pharm Sin B* **6**:441–449.

Bridgham JT, Eick GN, Larroux C, Deshpande K, Harms MJ, Gauthier MEA, Ortlund EA, Degnan BM, and Thornton JW (2010) Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. *PLoS Biol* **8**:e1000497.

Burris TP, Solt LA, Wang Y, Crumley C, Banerjee S, Griffett K, Lundasen T, Hughes T, and Kojetin DJ (2013) Nuclear receptors and their selective pharmacologic modulators. *Pharmacol Rev* **65**:710–778.

Busch A and Hertel KJ (2012) Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip Rev RNA* **3**:1–12.

Bush SJ, Chen L, Tovar-Corona JM, and Urrutia AO (2017) Alternative splicing and the evolution of phenotypic novelty. *Philos Trans R Soc Lond B Biol Sci* **372**:20150474.

Catalano A, Dawson MA, Somana K, Opat S, Schwarzer A, Campbell LJ, and Iland H (2007) The PRKAR1A gene is fused to RARA in a new variant acute promyelocytic leukemia. *Blood* **110**: 4073–4076.

Chandra V, Huang P, Potluri N, Wu D, Kim Y, and Rastinejad F (2013) Multidomain integration in the structure of the HNF-4 α nuclear receptor complex. *Nature* **495**:394–398.

Charleston JS, Schnell FJ, Dworzak J, Donoghue C, Lewis S, Chen L, Young GD, Milici AJ, Voss J, DeAlwis U, et al. (2018) Eteplirsen treatment for Duchenne muscular dystrophy: exon skipping and dystrophin production. *Neurology* **90**:e2146–e2154.

Chaudhary S, Khokhar W, Jabre I, Reddy ASN, Byrne LJ, Wilson CM, and Syed NH (2019) Alternative splicing and protein diversity: plants versus animals. *Front Plant Sci* **10**:708.

Chen M and Manley JL (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10**:741–754.

Chen T, Tompkins LM, Li L, Li H, Kim G, Zheng Y, and Wang H (2010) A single amino acid controls the functional switch of human constitutive androstane receptor (CAR) 1 to the xenobiotic-sensitive splicing variant CAR3. *J Pharmacol Exp Ther* **332**:106–115.

Chen Y, Tang Y, Guo C, Wang J, Boral D, and Nie D (2012) Nuclear receptors in the multidrug resistance through the regulation of drug-metabolizing enzymes and drug transporters. *Biochem Pharmacol* **83**:1112–1126.

Correa P, Akerstrom G, and Westin G (2002) Exclusive underexpression of vitamin D receptor exon 1f transcripts in tumors of primary hyperparathyroidism. *Eur J Endocrinol* **147**:671–675.

Cote GJ, Zhu W, Thomas A, Martin E, Murad F, and Sharina IG (2012) Hydrogen peroxide alters splicing of soluble guanylyl cyclase and selectively modulates expression of splicing regulators in human cancer cells. *PLoS One* **7**:e41099.

Croft L, Schandorff S, Clark F, Burrage K, Arctander P, and Mattick JS (2000) ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. *Nat Genet* **24**:340–341.

Cui Y, Cai M, and Stanley HE (2017) Comparative analysis and classification of cassette exons and constitutive exons. *BioMed Res Int* **2017**:7323508.

Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. (2015) Ensembl 2015. *Nucleic Acids Res* **43**: D662–D669.

Daniel M and Dehm SM (2017) Lessons from tissue compartment-specific analysis of androgen receptor alterations in prostate cancer. *J Steroid Biochem Mol Biol* **166**:28–37.

Das S, Anczukow O, Akerman M, and Krainer AR (2012) Oncogenic splicing factor SRSF1 is a critical transcriptional target of MYC. *Cell Rep* **1**:110–117.

Dehm SM, Schmidt LJ, Heemers HV, Vessella RL, and Tindall DJ (2008) Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Res* **68**:5469–5477.

DeKeyser JG, Laurenzana EM, Peterson EC, Chen T, and Omiecinski CJ (2011) Selective phthalate activation of naturally occurring human constitutive androstane receptor splice variants and the pregnane X receptor. *Toxicol Sci* **120**:381–391.

Desrochers M, Christou M, Jefcoate C, Belzil A, and Anderson A (1996) New proteins in the rat CYP2B subfamily: presence in liver microsomes of the constitutive CYP2B3 protein and the phenobarbital-inducible protein product of alternatively spliced CYP2B2 mRNA. *Biochem Pharmacol* **52**:1311–1319.

Echigoaya Y, Nakamura A, Nagata T, Urasawa N, Lim KRQ, Trieu N, Panesar D, Kuraoka M, Moulton HM, Saito T, et al. (2017) Effects of systemic multiexon skipping with peptide-conjugated morpholinos in the heart of a dog model of Duchenne muscular dystrophy. *Proc Natl Acad Sci USA* **114**:4213–4218.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**:1792–1797.

Emery AE (1989) Clinical and molecular studies in Duchenne muscular dystrophy. *Prog Clin Biol Res* **306**:15–28.

- Evans RM and Mangelsdorf DJ (2014) Nuclear receptors, RXR, and the big bang. *Cell* **157**: 255–266.
- Fall AM, Johnsen R, Honeyman K, Iversen P, Fletcher S, and Wilton SD (2006) Induction of revertant fibres in the mdx mouse using antisense oligonucleotides. *Genet Vaccines Ther* **4**:3–15.
- Fletcher S, Honeyman K, Fall AM, Harding PL, Johnsen RD, Steinhaus JP, Moulton HM, Iversen PL, and Wilton SD (2007) Morpholino oligomer-mediated exon skipping averts the onset of dystrophic pathology in the mdx mouse. *Mol Ther* **15**:1587–1592.
- Fujita A and Mitsuhashi T (1999) Differential regulation of ligand-dependent and ligand-independent functions of the mouse retinoid X receptor beta by alternative splicing. *Biochem Biophys Res Commun* **255**:625–630.
- Gardner-Stephen D, Heydel JM, Goyal A, Lu Y, Xie W, Lindblom T, Mackenzie P, and Radominska-Pandya A (2004) Human PXR variants and their differential effects on the regulation of human UDP-glucuronosyltransferase gene expression. *Drug Metab Dispos* **32**: 340–347.
- Grau-Bové X, Ruiz-Trillo I, and Irimia M (2018) Origin of exon skipping-rich transcriptomes in animals driven by evolution of gene architecture. *Genome Biol* **19**:135.
- Guo Z and Qiu Y (2011) A new trick of an old molecule: androgen receptor splice variants taking the stage? *Int J Biol Sci* **7**:815–822.
- Heald A, Anderson LV, Bushby KM, and Shaw PJ (1994) Becker muscular dystrophy with onset after 60 years. *Neurology* **44**:2388–2390.
- Heldring N, Pike A, Andersson S, Matthews J, Cheng G, Hartman J, Tjugue M, Ström A, Treuter E, Warner M, et al. (2007) Estrogen receptors: how do they signal and what are their targets. *Physiol Rev* **87**:905–931.
- Hisatomi H, Kohno N, Wakita K, Nagao K, Hirata H, Hikiji K, and Harada S (2003) Novel alternatively spliced variant with a deletion of 52 BP in exon 6 of the progesterone receptor gene is observed frequently in breast cancer tissues. *Int J Cancer* **105**:182–185.
- Holzer G, Markov GV, and Laudet V (2017) Evolution of nuclear receptors and ligand signaling: toward a soft key-lock model? *Curr Top Dev Biol* **125**:1–38.
- Honkakoski P and Negishi M (2000) Regulation of cytochrome P450 (CYP) genes by nuclear receptors. *Biochem J* **347**:321–337.
- Horvath G, Leser G, Helou K, and Henriksson M (2002) Function of the exon 7 deletion variant estrogen receptor alpha protein in an estradiol-resistant, tamoxifen-sensitive human endometrial adenocarcinoma grown in nude mice. *Gynecol Oncol* **84**:271–279.
- Jerzak KJ, Cockburn J, Pond GR, Pritchard KJ, Narod SA, Dhesy-Thind SK, and Bane A (2015) Thyroid hormone receptor α in breast cancer: prognostic and therapeutic implications. *Breast Cancer Res Treat* **149**:293–301.
- Jiang P, Hou Z, Bolin JM, Thomson JA, and Stewart R (2017) RNA-seq of human neural progenitor cells exposed to lead (Pb) reveals transcriptome dynamics, splicing alterations and disease risk associations. *Toxicol Sci* **159**:251–265.
- Kim HJ, Hwang JY, Kim HJ, Choi WS, Lee JH, Kim HJ, Chang KC, Nishinaka T, Yabe-Nishimura C, and Seo HG (2007) Expression of a peroxisome proliferator-activated receptor gamma 1 splice variant that was identified in human lung cancers suppresses cell death induced by cisplatin and oxidative stress. *Clin Cancer Res* **13**:2577–2583.
- Kohli M, Ho Y, Hillman DW, Van Etten JL, Henzler C, Yang R, Sperger JM, Li Y, Tseng E, Hon T, et al. (2017) Androgen receptor variant AR-V9 is coexpressed with AR-V7 in prostate cancer metastases and predicts abiraterone resistance. *Clin Cancer Res* **23**:4704–4715.
- Kole R, Krainer AR, and Altman S (2012) RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nat Rev Drug Discov* **11**:125–140.
- Koo T and Wood MJ (2013) Clinical trials using antisense oligonucleotides in duchenne muscular dystrophy. *Hum Gene Ther* **24**:479–488.
- Krett NL, Pillay S, Moalli PA, Greipp PR, and Rosen ST (1995) A variant glucocorticoid receptor messenger RNA is expressed in multiple myeloma patients. *Cancer Res* **55**:2727–2729.
- Lambert CA, Garbacki N, and Colige AC (2017) Chemotherapy induces alternative transcription and splicing: facts and hopes for cancer treatment. *Int J Biochem Cell Biol* **91** (Pt B):84–97.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace JM, Wilm A, Lopez R, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947–2948.
- Lee MT, Ouyang B, Ho SM, and Leung YK (2013) Differential expression of estrogen receptor beta isoforms in prostate cancer through interplay between transcriptional and translational regulation. *Mol Cell Endocrinol* **376**:125–135.
- Lee Y and Rio DC (2015) Mechanisms and regulation of alternative pre-mRNA splicing. *Annu Rev Biochem* **84**:291–323.
- Letunic I, Copley RR, and Bork P (2002) Common exon duplication in animals and its role in alternative splicing. *Hum Mol Genet* **11**:1561–1567.
- Li Y, He B, Gao J, Liu QS, Liu R, Qu G, Shi J, Hu L, and Jiang G (2018) Methylmercury exposure alters RNA splicing in human neuroblastoma SK-N-SH cells: implications from proteomic and post-transcriptional responses. *Environ Pollut* **238**:213–221.
- Lim KR, Maruyama R, and Yokota T (2017) Eteplirsen in the treatment of Duchenne muscular dystrophy. *Drug Des Devel Ther* **11**:533–545.
- Lin YS, Yasuda K, Assem M, Cline C, Barber J, Li CW, Kholodovych V, Ai N, Chen JD, Welsh WJ, et al. (2009) The major human pregnane X receptor (PXR) splice variant, PXR.2, exhibits significantly diminished ligand-activated transcriptional regulation. *Drug Metab Dispos* **37**: 1295–1304.
- Lundell K, Thulin P, Hamsten A, and Ehrenborg E (2007) Alternative splicing of human peroxisome proliferator-activated receptor delta (PPAR delta): effects on translation efficiency and trans-activation ability. *BMC Mol Biol* **8**:70.
- Mackeh R, Marr AK, Dargham SR, Syed N, Fakhro KA, and Kino T (2017) Single-nucleotide variations of the human nuclear hormone receptor genes in 60,000 individuals. *J Endocr Soc* **2**: 77–90.
- Magomedova L, Tiefenbach J, Zilberman E, Le Billan F, Voisin V, Saikali M, Boivin V, Robitaille M, Gueroussou S, Irimia M, et al. (2019) ARGLU1 is a transcriptional coactivator and splicing regulator important for stress hormone signaling and development. *Nucleic Acids Res* **47**: 2856–2870.
- Martinez-Montiel N, Rosas-Murrieta NH, Anaya Ruiz M, Monjaraz-Guzman E, and Martinez-Contreras R (2018) Alternative splicing as a target for cancer treatment. *Int J Mol Sci* **19**:E545.
- Mazin P, Xiong J, Liu X, Yan Z, Zhang X, Li M, He L, Somel M, Yuan Y, Phoebe Chen YP, et al. (2013) Widespread splicing changes in human brain development and aging. *Mol Syst Biol* **9**: 633.
- McCloy G, Fall AM, Moulton HM, Iversen PL, Rasko JE, Ryan M, Fletcher S, and Wilton SD (2006a) Induced dystrophin exon skipping in human muscle explants. *Neuromuscul Disord* **16**: 583–590.
- McCloy G, Moulton HM, Iversen PL, Fletcher S, and Wilton SD (2006b) Antisense oligonucleotide-induced exon skipping restores dystrophin expression in vitro in a canine model of DMD. *Gene Ther* **13**:1373–1381.
- McGuire J, Okamoto K, Whitelaw ML, Tanaka H, and Poellinger L (2001) Definition of a dioxin receptor mutant that is a constitutive activator of transcription: delineation of overlapping repression and ligand binding functions within the PAS domain. *J Biol Chem* **276**:41841–41849.
- Melangath G, Sen T, Kumar R, Bawa P, Srinivasan S, and Vijayraghavan U (2017) Functions for fission yeast splicing factors SpSlu7 and SpPrp18 in alternative splice-site choice and stress-specific regulated splicing. *PLoS One* **12**:e0188159.
- Merkin J, Russell C, Chen P, and Burge CB (2012) Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338**:1593–1599.
- Modrek B and Lee CJ (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat Genet* **34**:177–180.
- Mourich DV and Iversen PL (2009) Splicing in the immune system: potential targets for therapeutic intervention by antisense-mediated alternative splicing. *Curr Opin Mol Ther* **11**:124–132.
- Mourich DV, Oda SK, Schnell FJ, Crumley SL, Hauck LL, Moentenich CA, Marshall NB, Hinrichs DJ, and Iversen PL (2014) Alternative splice forms of CTLA-4 induced by antisense mediated splice-switching influences autoimmune diabetes susceptibility in NOD mice. *Nucleic Acid Ther* **24**:114–126.
- Mueller GC, Herranen AM, and Jervell KF (1958) Studies on the mechanism of action of estrogens. *Recent Prog Horm Res* **14**:95–129, NaN–139.
- Nakai A, Seino S, Sakurai A, Szilak I, Bell GI, and DeGroot LJ (1988) Characterization of a thyroid hormone receptor expressed in human kidney and other tissues. *Proc Natl Acad Sci USA* **85**:2781–2785.
- Okonechnikov K, Golosova O, and Fursov M; UGENE team (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**:1166–1167.
- Pan Q, Shai O, Lee LJ, Frey BJ, and Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**: 1413–1415.
- Panchal RG, Mourich DV, Bradfute S, Hauck LL, Warfield KL, Iversen PL, and Bavari S (2014) Induced IL-10 splice altering approach to antiviral drug discovery. *Nucleic Acid Ther* **24**: 179–185.
- Parrado A, Despouy G, Kraïba R, Le Pogam C, Dupas S, Choquette M, Robledo M, Larghero J, Bui H, Le Gall I, et al. (2001) Retinoic acid receptor alpha1 variants, RARalphaDeltaB and RARalphaDeltaBC, define a new class of nuclear receptor isoforms. *Nucleic Acids Res* **29**: 4901–4908.
- Poola I and Speirs V (2001) Expression of alternatively spliced estrogen receptor alpha mRNAs is increased in breast cancer tissues. *J Steroid Biochem Mol Biol* **78**:459–469.
- Popadin KY, Gutierrez-Arcelus M, Lappalainen T, Buil A, Steinberg J, Nikolaev SI, Lukowski SW, Bazykin GA, Seplyarskiy VB, Ioannidis P, et al. (2014) Gene age predicts the strength of purifying selection acting on gene expression variation in humans. *Am J Hum Genet* **95**: 660–674.
- Prakash C, Zuniga B, Song CS, Jiang S, Cropper J, Park S, and Chatterjee B (2015) Nuclear receptors in drug metabolism, drug response and drug interactions. *Nucl Receptor Res* **2**:101178.
- Qu S, Zhong Y, Shang R, Zhang X, Song W, Kjems J, and Li H (2017a) The emerging landscape of circular RNA in life processes. *RNA Biol* **14**:992–999.
- Qu W, Cingolani P, Zeeberg BR, and Ruden DM (2017b) A bioinformatics-based alternative mRNA splicing code that may explain some disease mutations is conserved in animals. *Front Genet* **8**:38.
- Rastinejad F, Huang P, Chandra V, and Khorasanzadeh S (2013) Understanding nuclear receptor form and function using structural biology. *J Mol Endocrinol* **51**:T1–T21.
- Reitzel AM, Pang K, Ryan JF, Mullikin JC, Martindale MQ, Baxevas AD, and Tarrant AM (2011) Nuclear receptors from the ctenophore Mnemiopsis leidyi lack a zinc-finger DNA-binding domain: lineage-specific loss or ancestral condition in the emergence of the nuclear receptor superfamily? *EvoDevo* **2**:3.
- Richer JK, Lange CA, Wierman AM, Brooks KM, Tung L, Takimoto GS, and Horwitz KB (1998) Progesterone receptor variants found in breast cells repress transcription by wild-type receptors. *Breast Cancer Res Treat* **48**:231–241.
- Rochel N, Wurtz JM, Mitschler A, Klaholz B, and Moras D (2000) The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand. *Mol Cell* **5**:173–179.
- Salzman J, Gawad C, Wang PL, Lacayo N, and Brown PO (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**: e30733.
- Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, and Oberdoerffer S (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**:74–79.
- Smith CW and Valcárcel J (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* **25**:381–388.
- Soller M (2006) Pre-messenger RNA processing and its regulation: a genomic perspective. *Cell Mol Life Sci* **63**:796–819.
- Sorek R, Ast G, and Graur D (2002) Alu-containing exons are alternatively spliced. *Genome Res* **12**:1060–1067.
- Sorek R, Shamir R, and Ast G (2004) How prevalent is functional alternative splicing in the human genome? *Trends Genet* **20**:68–71.
- Sprung CN, Li J, Hovan D, McKay MJ, and Forrester HB (2011) Alternative transcript initiation and splicing as a response to DNA damage. *PLoS One* **6**:e25758.
- Stelzer G, Dalah I, Stein TI, Satanower Y, Rosen N, Nativ N, Oz-Levi D, Olender T, Belinky F, Bahir I, et al. (2011) In-silico human genomics with GeneCards. *Hum Genomics* **5**:709–717.
- Sun J, Blair AL, Ayar SE, and Li R (2007) Cofactor of BRCA1 modulates androgen-dependent transcription and alternative splicing. *J Steroid Biochem Mol Biol* **107**:131–139.
- Syed YY (2016) Eteplirsen: first global approval. *Drugs* **76**:1699–1704.
- Taladriz-Sender A, Campbell E, and Burley GA (2019) Splice-switching small molecules: a new therapeutic approach to modulate gene expression. *Methods* **167**:134–142.
- Thierry-Mieg D and Thierry-Mieg J (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* **1**:S12.1–14.
- Thomas M, Bayha C, Klein K, Müller S, Weiss TS, Schwab M, and Zanger UM (2015) The truncated splice variant of peroxisome proliferator-activated receptor alpha, PPAR α -tr, autonomously regulates proliferative and pro-inflammatory genes. *BMC Cancer* **15**:488.
- Ule J and Blencowe BJ (2019) Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol Cell* **76**:329–345.

- Unutmaz D (2009) RORC2: the master of human Th17 cell programming. *Eur J Immunol* **39**: 1452–1455.
- van der Vaart M and Schaaf MJ (2009) Naturally occurring C-terminal splice variants of nuclear receptors. *Nucl Recept Signal* **7**:e007.
- van Roon-Mom WM and Aartsma-Rus A (2012) Overview on applications of antisense-mediated exon skipping. *Methods Mol Biol* **867**:79–96.
- Venables JP, Lapasset L, Gadea G, Fort P, Klinck R, Irimia M, Vignal E, Thibault P, Prinos P, Chabot B, et al. (2013) MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. *Nat Commun* **4**:2480.
- Villaseñor-Altamirano AB, Watson JD, Prokopec SD, Yao CQ, Boutros PC, Pohjanvirta R, Valdés-Flores J, and Elizondo G (2019) 2,3,7,8-Tetrachlorodibenzo-p-dioxin modifies alternative splicing in mouse liver. *PLoS One* **14**:e0219747.
- Vuong LM, Chellappa K, Dhahbi JM, Deans JR, Fang B, Bolotin E, Titova NV, Hoverter NP, Spindler SR, Waterman ML, et al. (2015) Differential effects of hepatocyte nuclear factor 4 α isoforms on tumor growth and T-cell factor 4/AP-1 interactions in human colorectal cancer cells. *Mol Cell Biol* **35**:3471–3490.
- Wallace BD and Redinbo MR (2013) Xenobiotic-sensing nuclear receptors involved in drug metabolism: a structural perspective. *Drug Metab Rev* **45**:79–100.
- Wan Y and Larson DR (2018) Splicing heterogeneity: separating signal from noise. *Genome Biol* **19**:86.
- Wang W, Chen J, Luo L, Li Y, Liu J, and Zhang W (2018) Effect of cadmium on kitl pre-mRNA alternative splicing in murine ovarian granulosa cells and its associated regulation by miRNAs. *J Appl Toxicol* **38**:227–239.
- Weatheritt RJ, Sterne-Weiler T, and Blencowe BJ (2016) The ribosome-engaged landscape of alternative splicing. *Nat Struct Mol Biol* **23**:1117–1123.
- Wilton SD, Fall AM, Harding PL, McClorey G, Coleman C, and Fletcher S (2007) Antisense oligonucleotide-induced exon skipping across the human dystrophin gene transcript. *Mol Ther* **15**:1288–1296.
- Xing Y and Lee CJ (2004) Negative selection pressure against premature protein truncation is reduced by alternative splicing and diploidy. *Trends Genet* **20**:472–475.
- Xing Y and Lee C (2005) Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. *Proc Natl Acad Sci USA* **102**:13526–13531.
- Yang X, Coulombe-Huntington J, Kang S, Sheynkman GM, Hao T, Richardson A, Sun S, Yang F, Shen YA, Murray RR, et al. (2016) Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* **164** (4):805–817.
- Yang Z and Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* **15** (12):496–503.
- Yokota T, Lu QL, Partridge T, Kobayashi M, Nakamura A, Takeda S, and Hoffman E (2009) Efficacy of systemic morpholino exon-skipping in Duchenne dystrophy dogs. *Ann Neurol* **65** (6): 667–676.
- Zaharieva E, Chipman JK, and Soller M (2012) Alternative splicing interference by xenobiotics. *Toxicology* **296**:1–12.
- Zangar RC, Okita JR, Kim H, Thomas PE, Anderson A, Edwards RJ, Springer DL, and Okita RT (1999) Effect of calcium channel antagonists nifedipine and nicardipine on rat cytochrome P-450 2B and 3A forms. *J Pharmacol Exp Ther* **290**:1436–1441.
- Zennaro MC, Souque A, Viengchareun S, Poisson E, and Lombès M (2001) A new human MR splice variant is a ligand-independent transactivator modulating corticosteroid action. *Mol Endocrinol* **15**:1586–1598.
- Zhang X, Peng Q, Li L, and Li X (2016) Recognition of alternatively spliced cassette exons based on a hybrid model. *Biochem Biophys Res Commun* **471**:368–372.
- Zhang Z, Lotti F, Dittmar K, Younis I, Wan L, Kasim M, and Dreyfuss G (2008) SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* **133** (4):585–600.
- Zhao Y, Zhang K, Giesy JP, and Hu J (2015) Families of nuclear receptors in vertebrate models: characteristic and comparative toxicological perspective. *Sci Rep* **5**:8554.
- Zhou R, Chun RF, Lisse TS, Garcia AJ, Xu J, Adams JS, and Hewison M (2015) Vitamin D and alternative splicing of RNA. *J Steroid Biochem Mol Biol* **148**:310–317.
- Zhu H, Huang Y, Su H, Ma Y, Tao Y, Liao DJ, Liu Y, and Feng Z (2018a) Identification of a novel human estrogen receptor- α splice variant able to enhance malignant biological behaviors of breast cancer cells. *Oncol Lett* **15**:5339–5344.
- Zhu LY, Zhu YR, Dai DJ, Wang X, and Jin HC (2018b) Epigenetic regulation of alternative splicing. *Am J Cancer Res* **8**:2346–2358.

Address correspondence to: Dr. Andrew J. Annalora, Department of Environmental and Molecular Toxicology, Oregon State University, 1007 Agriculture and Life Sciences Bldg., Corvallis, OR 97331. E-mail: Andrew.Annalora@oregonstate.edu
