DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

Minireview

Title: Alternative Splicing in the Nuclear Receptor Superfamily Expands Gene Function to Refine Endo-Xenobiotic Metabolism.

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

Department of Environmental and Molecular Toxicology, Oregon State University, 1007 Agriculture & Life Sciences Building, Corvallis, OR 97331; USA - *AJA, CBM, PLI*

United States Army Research Institute for Infectious Disease, 1425 Porter Street, Frederick, MD 21702; USA - *PLI*

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

Running Title: Alternative Splicing of Human Nuclear Receptors.

Corresponding author and person to whom reprint requests should be addressed:

Andrew J. Annalora, PhD

Department of Environmental and Molecular Toxicology

Oregon State University

1007 Agriculture & Life Sciences Building

Corvallis, OR 97331; USA

Phone: 541-737-2706

Fax: 541-737-0497

E-mail: Andrew.Annalora@oregonstate.edu

Text Pages – 24

Tables – 4

Figures – 6

References - 146

Word Count

Abstract – 250

Introduction – 1332

Discussion - 3298

Abbreviations – AR - androgen receptor; AS – alternative splicing; CAR - Constitutive androstane receptor; CEViP – Cassettte Exon Visualization Pictograph; CYP – Cytochrome P450; DMD - Duchenne muscular dystrophy; DBD – DNA binding domain; Endo-xenobiotic – endobiotic and xenobiotic; GWAS – genome wide association study; HNF4α - Hepatocyte nuclear factor 4α; LBD – ligand binding domain; NR – nuclear receptor; P450 - Cytochrome P450; PPAR - peroxisome proliferator-activated receptor; PXR

- Pregnane X receptor; RXR - retinoid X receptor; SNP – single nucleotide polymorphism; SSO - splice-switching oligonucleotides; VDR - vitamin D receptor.

ABSTRACT

The human genome encodes 48 nuclear receptor (NR) genes, whose translated products transform chemical signals from endo-xenobiotics into pleotropic 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 (SNP). 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 (CEViP) 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 AS on the human NR superfamily and highlights the dramatic expansion of more than 1,000 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 CEViP 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 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 are 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 N-terminal transactivation domain. The functional diversity of NR genes are 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: i.) ligand-dependent NRs, which encode a LBD that requires ligand binding to activate both genomic and non-genomic signaling events; ii) ligand-independent or constitutively-active NRs, which encode a LBD that is stuck (via mutation or AS) in the 'on position' in the absence of ligand binding; and iii) 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. (Bridgham et al., 2010). Based on this diversity, ligand binding can trigger a diverse array of cellular responses in NR genes, with ligands being classified according to their role as agonists, antagonists, partial agonists, inverse agonists, super agonists or selective NR modulator (Burris et al., 2013).

Nearly all human genes are alternately spliced (Lee and Rio, 2015). AS is a co-transcriptional 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 (Soller, 2006; Smith and Valcarcel, 2000; 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 multi-exon genes, it appears to plays a pivotal role in mediating the diverse set of functions that NRs employ to transduce a discrete, chemical exposure cue into an adaptive, physiological 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 7 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 (CYP) superfamily that highlighted the tremendous expansion of the human CYP transcriptome over the

last decade, which now includes over 1,000 unique mRNA transcripts generated from the 57 human CYP 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 NIH PubMed, AceView and Ensembl databases (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 comprised of at least 1,004 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 databases; 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 CYP 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 physiological responses to endo-xenobiotic exposures. Endogenous gene splicing programs involving NR genes and their co-regulatory 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 co-regulators (e.g. COBRA1 or 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 CYP 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 (SSO) (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; McClorey et al., 2006b; Adams et al., 2007; Fletcher et al., 2007; Koo and Wood, 2013) as highlighted by the FDA approval of Eteplirsen in 2016 (Syed, 2016). The therapeutic utility of SSO to modulate the immune response have also been explored (Mourich et al., 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 databases, we identified 31 NR genes already connected to a human disease either from literature citations, links to single 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 database (Thierry-Mieg and Thierry-Mieg, 2006). A summary of all known transcript variants identified by searching Pubmed, Aceview and the Ensembl database (Cunningham et al., 2015) are listed in **Table 1**, as well as all known single nucleotide polymorphisms (SNPs) listed in the Genecards database (Stelzer et al, 2011).

Meta-Analysis of Nuclear Receptor Transcript Variants and Single Nucleotide Polymorphisms (SNPs).

As summarized in **Table 1**, we catalogued all known transcript variants for the 48 human nuclear receptor (NR) genes, plus one processed pseudogene (FXRβ (NR1H5)), and annotated their known association with human disease and any available information regarding their molecular weight, tissue-specific expression profile, and subcellular trafficking. The stereo-radar plot shown in **Figure 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 1-62 variants, with LXRα (62), CAR (60) and ERRγ (47) expressing the highest number of known mRNA transcripts. In contrast, FXRβ (1), SHP-1 (3), Dax-1 (4) and 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 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 7 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 to 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 Cytochrome P450 (CYP) Genes.

Based on our analysis, we found that NR genes express roughly 20.5 transcript variants per gene, compared to 16.7 for the average CYP 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 CYPs. This translates to 341 SNPs per transcript variant for NR genes, as compared to 50 SNPs per transcript variant in the CYPs. 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 CYP genes. However, because the accumulation of single nucleotide variants (SNV) in NR genes have been inversely correlated to a gene's age (Podapin et al., 2014; Mackeh et al., 2017), CYP 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 (HLA) and histone deacetylases (HDAC) (Mackeh

et al., 2017). However, less is known about how SNPs alter the natural splicing patterns of NR genes, and whether or not 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 RNAseq 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 DAX1, SHP, PPARβ/δ, RARα, RARγ, LXRα, LXRβ, MR, ERα, Nur77, NURR1 and SF1) that lack a complete LBD and may thus be constitutively active. We also identified 14 cancer-related NR variants (for PPARα, PPARγ, RARB, Rev-erbβ, LXRα, LXRβ, VDR, PXR, HNF4α, TR4, AR, ERRα, Nur77, and NURR1) that lack a complete DBD and may thus represent dominant negatives (based on the NIH Aceview database – 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 CYP 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 (TCDD) 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 bHLH-PAS (basic-helix-loop-helix-Per-ARNT-Sim) superfamily. While the mechanisms by which TCDD 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 co-transcriptional, 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 biological 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 vitamin D receptor (VDR) is a cassette exon that can be excluded from mature transcripts, without disrupting protein expression, and that splice variants lacking a

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

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 under-appreciated 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 **Figure 2A**, our cassette exon visualization pictograph methodology, which we abbreviate as (CEViP), stems from the observation that all coding exons exist in one of 9 possible reading frames with respect to their intron/exon border, which we refer to as exon junction frame codes: 1-9 (see inset to **Figure 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 ClustalX2 (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 Figure 1**, and a summary of these details is presented in **Supplemental Table 1**. In **Figure 2B**, the four general classes of cassette exons are highlighted in comparison to a prototype transcript that encodes no cassette exons. These 4 classes include 1.) amino (N)- or carboxy (C)-terminal encoding cassette exons; 2.) internal cassette exons with complete or 3.) incomplete

reading frames, and 4.) multi-exonic, 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 HNF4α (NR2A1) in **Figure 3** and the VDR (NR1I1) in **Figure 4.**

In **Figure 3A**, the crystal structure of the full-length HNF4α gene is shown in cylindrical cartoon representation to highlight how structural domains in the NR protein are organized with respect to the gene's 9 primary coding exons (Chandra et al., 2013). Modular, cassette exon components of HNF4α 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 (**Figure 3B**). The 4 cassette exons found in the human HNF4α gene encode large segmetns of the LBD and C-terminus, which may allow for the tissue-selective expression of at least 7 unique HNF4α transcript variants under normal conditions, as highlighted in **Figure 3C.** Changes in cassette exon utilization patterns, triggered by oxidative stress, chemical exposure, mutation or disease, may provide the HNF4α 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 **Figure 4A** to highlight the positioning of cassette exon 8 with respect to the gene's DBD, LBD, hinge region and N- and C-terminus. **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.

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

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 **Figure 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 **Figure 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 (PPARα, PPARδ, and PPARγ; see **Figure 5B**), lack any cassette exons in their reference gene coding structure, while related genes from the NR1A (TRα and TRβ), NR1B (RARα, RARβ, and RARγ), NR1D (Rev-erbα and Rev-erbβ), and NR1I (VDR, PXR, and CAR) subfamilies all contain at least one cassette exon within the LBD (see **Supplemental Figure 1**). Even greater cassette exon complexity is observed in the NR1F (RORα, RORβ, and RORγ) and NR1H (LXRα, LXRβ, and FXRα) subfamilies, which contain modular domains at both the N- and C-terminus, and within the DBD, LBD and hinge/dimerization domains (**Figure 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 Figure 2**). However, NR1I subfamily genes encode a single cassette exon in the LBD of their transcripts, rather than the multiple cassettes utilized by NR1H genes in the DBD, LBD, hinge and N- and C-terminus. 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 independent 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 biological regimes.

In contrast to the NR1 family, all members of the NR2 subfamily have at least one cassette exon, with NR2F (COUP-TFI, COUP-TFII, and EAR-2), NR2A (HNF4α and HNF4γ), and NR2E1 (TLX) subfamilies each containing cassette exons that encode portions of the hinge region or LBD. NR2B (RXRα, RXRβ and RXRγ) genes contain one cassette exon in the DBD/hinge region only, while the NR2C subfamily genes (TR2, TR4) utilize cassette exons in both DBD and LBD (see **Table 5** or **Supplemental Figure 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 Figure 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 Figure 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 Figure 1** reflects the hierarchical complexity of cassette exon utilization that we propose to be operating across the 7 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 7 human NR gene families based on phylogeny (as depicted in **Supplemental Figure 2**) and cassette exon usage (as depicted in **Supplemental Figure 1**) is shown in **Supplemental Figure 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, CYPs and drug transporters) linked to their signal transduction pathways, as highlighted in **Figure 6**. We speculate that this adaptive feedback response helps the cell refine the interplay among cognate, phase I-III metabolic systems, optimizing physiological responses to endo-xenobiotics through the selection of splice variant transcripts that promote cellular homeostasis and health resilience after a chemical exposure.

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

DISCUSSION

AS is a complex co-transcriptional 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 SR proteins, or serine/arginine (SR)-rich proteins, that include the heterogeneous nuclear ribonucleoproteins (hnRNPs); these splicing factors recognize a specific set of splicing regulatory elements (SREs) that include exonic and intronic splicing enhancers (ESEs and ISEs), and exonic and intronic splicing silencers (ESSs and ISSs) (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 (SMA), a rare genetic spinal motor neuron disease, is associated with reduced expression of the SMN protein, a master hnRNP 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 MYC in tumors can also yield multiple oncogenic AS events due to an upregulation of splicing factors (Anczukow and Krainer, 2016; Das et al., 2012). Stem cell-specific AS programs have also been established, and are mediated by RBFOX2, a key player in mesenchymal tissue splicing (Venables 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 non-synonymous 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 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 (Aarstma-Rus et al., 2005; Wilton et al., 2007). Conversely, multi-exon skipping of exons 6 and 8 was accomplished in the canine dystrophin gene to restore reading frame (Yokota T et al. 2009; Echigoya et al., 2018). 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 multi-cassette 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 androgen receptor (AR; NR3C4) 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; Daniel and Dehm, 2017; Guo and Qiu, 2011; Antonarakis et al., 2014; Kohli et al., 2017). Constitutively active splice variants of the mineralocorticoid receptor (MR; NR3C2) and the aryl-hydrocarbon receptor (AhR), a related bHLH/PAS family nuclear receptor, have also been identified, and these variants lack discrete segments of the LBD (Zennaro et al., 2001; McGuire et al., 2001). Ligand-independent estrogen receptor (ER; NR3A1) variants have also been reported with implications for the treatment of breast cancer (Heldring et al., 2007; Al-Bader et al., 2011; Zhu et al., 2018a). More directly, our group recently showed that targeted skipping of cassette exon 8 in the human vitamin D receptor (VDR; NR1I1) 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, 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 database (Thierry-Mieg D. and Thierry-Mieg J., 2006), we have already identified 12 NR genes previously linked to cancer that generate splice variants lacking a complete DBD; these include: PPARβ/δ, RARα, RARγ, LXRα, LXRβ, MR, ERα, NGF1B (Nur77), NURR1, and 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: PPARα, PPARγ, RARβ, Rev-erbβ, LXRα, LXRβ, VDR, PXR, HNF4α, TR4, AR, ERRα, NGF1B (Nur77) 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 re-programming 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 towards 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 non-coding 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 SR proteins, microRNA (miRNA) and long noncoding RNAs (lncRNA). 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 post-transcriptional or post-translational 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 **Figure 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 Figure 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 2 cassette exons that are not well conserved in any other subfamily. However, there is strong support in the literature that suggests NR2A1 (HNFα) is the primordial NR gene that emerged from our common metazoan ancestor (demosponges) (Bridgham et al., 2010). In humans, this gene appears structurally-complex compared to other NR genes, having 4 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 while the size and location of NR2A cassette exons appears to have drifted over time, the overall organization of cassette exons has not (**Supplemental Figure 1**). Ultimately, our approach for analysis of cassette exon structures imply that the NR2E subfamily of genes (TLX and PNR) may be more ancestral than the NR2A family, due to the lack of well conserved cassette exon structures in this family compared to all other NR2 family members (**Supplemental Figure 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 Figure 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 (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 5 cassette exons that are not fully conserved, or re-capitulated, in any other NR subfamilies (**Supplemental Figure 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 AF-1 and 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 (GR, MR, PR, AR), which appears to have evolved a highly extended, N-terminal transactivation domain (~400-600 amino acids in length), encoded by a single, non-cassette exon, to augment their signaling functions (**Supplemental Figure 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, co-regulatory 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 utilize 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 biological 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 **Figure 6**, the expansion of multiple transcript variants from a single gene provides a cell with the powerful ability to find novel protein- and non-coding 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 CYP superfamily of genes provides an expanding phenotypic cloud of functional plasticity that can augment a CYP 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 biological adaptation to novel chemical insults. Because NR genes coordinate the expression and splicing of virtually all metabolic genes (including CYPs 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 biological responses to the environment. For example, transcript variants (T1-T3) of the pregnane X receptor (PXR; NR1I2), are known to promote differential expression of both CYP3A4 (Lin et al., 2009) and UDP-glucuronosyltransferase (UGT) 1A isoforms (Garnder-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.

ACKNOWLEDGEMENTS

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, and Iversen

Conducted experiments: Annalora, and Iversen

Contributed new reagents or analytic tools: Annalora, and Iversen

Performed data analysis: Annalora, and Iversen

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

REFERENCES

Aartsma-Rus, A, Bremmer-Bout, M, Janson, AA, Den Dunnen, JT, Van Ommen, GJ, 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, 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(4)**:284-97.

Adams AM, Harding PL, Iversen PL, Coleman C, Fletcher S, 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, Francis I (2011) Analysis of estrogen receptor isoforms and variants in breast cancer cell lines. *Exp Ther Med* **2(3)**:537-544.

Annalora AJ, Marcus CB, 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(4)**:375-389. doi: 10.1124/dmd.116.073254.

Annalora AJ, Marcus CB, and Ivrsen, PL (2019) Alternative Splicing of the Vitamin D Receptor Modulates Target Gene Expression and Promotes Ligand-Independent Functions. *Toxicology and Applied Pharmacology* **364**:55-67. doi: 10.1016/j.taap.2018.12.009.

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(11)**:1028-38. doi: 10.1056/NEJMoa1315815.

Anczuków O, Krainer AR (2016) Splicing-factor alterations in cancers. *RNA* **22(9)**:1285-301. doi: 10.1261/rna.057919.116. Review.

Auboeuf D, Dowhan DH, Dutertre 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, Omiecinski CJ (2005) Retinoid X receptor-alpha-dependent transactivation by a naturally occurring structural variant of human constitutive androstane receptor (NR1I3). *Mol Pharmacol* **68(5)**:1239-53.

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, et al. (2012) The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338(6114)**:1587-93. doi: 10.1126/science.1230612.

Bhat-Nakshatri P, Song EK, Collins NR, Uversky VN, Dunker AK, O'Malley BW, Geistlinger TR, Carroll JS, Brown M, Nakshatri H. (2013) Interplay between estrogen receptor and AKT in estradiol-induced alternative splicing. *BMC Med Genomics*. **6**:21. doi: 10.1186/1755-8794-6-21.

Blencowe BJ (2006) Alternative splicing: new insights from global analyses. *Cell* **126(1)**:37-47. Review.

Brewer CT, Chen T (2016) PXR variants: the impact on drug metabolism and therapeutic responses. *Acta Pharm Sin B* **6(5)**:441-449. Review.

Bridgham JT, Eick GN, Larroux C, Deshpande K, Harms MJ, Gauthier MEA, Ortlund EA, Degnan BM, Thornton JW (2010) Protein evolution by molecular tinkering: Diversification of the Nuclear Receptor Superfamily from a Ligand-Dependent Ancestor. *PLoS Biol* **8(10)**: e1000497.

Burris TP, Solt LA, Wang Y, Crumbley C, Banerjee S, Griffett K, Lundasen T, Hughes T, Kojetin DJ (2013) Nuclear Receptors and their selective pharmacologic modulators. *Pharmacologic Reviews* **65**:710-778.

Busch A, Hertel KJ (2012) Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip Rev RNA* **3(1)**:1-12. doi: 10.1002/wrna.100.

Bush SJ, Chen L, Tovar-Corona JM, Urrutia AO (2017) Alternative splicing and the evolution of phenotypic novelty. *Philos Trans R Soc Lond B Biol Sci.* **372(1713). pii**: 20150474. Review.

Catalano A, Dawson MA, Somana K, Opat S, Schwarer A, Campbell LJ, Iland H (2007) The PRKAR1A gene is fused to RARA in a new variant acute promyelocytic leukemia. *Blood* **110(12):**4073-6.

Chen M, Manley JL (2009) Mechanisms of alternate 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, 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(1)**:106-15.

Chandra V, Huang P, Potluri N, Wu D, Kim Y, Rastinejad F (2013) Multidomain integration in the structure of the HNF-4α nuclear receptor complex. *Nature* **495(7441)**:394-8. doi: 10.1038/nature11966.

Charleston JS, Schnell FJ, Dworzak J, Donoghue C, Lewis S, Chen L, Young GD, Milici AJ, Voss J, DeAlwis U, Wentworth B, Rodino-Klapac LR et al. (2018) Eteplirsen treatment for Duchenne muscular dystrophy: Exon skipping and dystrophin production. *Neurology* **90(24)**:e2146-e2154. Erratum in: *Neurology*. (2018) **91(13)**:637.

Chaudhary S, Khokhar W, Jabre I, Reddy ASN, Byrne LJ, Wilson CM, Syed NH (2019) Alternative Splicing and Protein Diversity: Plants Versus Animals. *Front PlantSci* **10**:708. doi: 10.3389/fpls.2019.00708. Review.

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

Chong HK, Biesinger J, Seo YK, Xie X, Osborne TF (2012) Genome-wide analysis of hepatic LRH-1 reveals a promoter binding preference and suggests a role in regulating genes of lipid metabolism in concert with FXR. BMC *Genomics* **13:**51.

Correa P, Akerstrom G, Westin G (2002) Exclusive underexpression of vitamin D receptor exon 1f transcripts in tumors of primary hyperparathyroidism. *Eur J Endocrinol* **147(5)**:671-5.

Cote GJ, Zhu W, Thomas A, Martin E, Murad F, Sharina IG (2017) Hydrogen peroxide alters splicing of soluble guanylyl cyclase and selectively modulates expression of splicing regulators in human cancer cells. *PLoS One* **7(7):**e41099.

Croft L, Schandorff S, Clark F, Burragge K, Arctander P, 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, Stanley HE (2017) Comparative Analysis and Classification of Cassette Exons and Constitutive Exons. *Biomed Res Int*. **2017**:7323508. doi:10.1155/2017/7323508.

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(Database issue)**:D662-9. doi: 10.1093/nar/gku1010.

Daniel M, Dehm SM (2017) Lessons from tissue compartment-specific analysis of androgen receptor alterations in prostate cancer. *J Steroid Biochem Mol Biol* **166**:28-37. doi: 10.1016/j.jsbmb.2016.04.016.

Das S, Anczuków O, Akerman M, Krainer AR (2012) Oncogenic splicing factor SRSF1 is a critical transcriptional target of MYC. *Cell Rep* **1(2)**:110-7.

Dehm SM, Schmidt LJ, Heemers HV, Vessella RL, 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(13)**:5469-77.

DeKeyser JG, Laurenzana EM, Peterson EC, Chen T, Omiecinski CJ (2011) Selective phthalate activation of naturally occurring human constitutive androstane receptor splice variants and the pregnane X receptor. *Toxicol Sci* **120(2)**:381-91. doi: 10.1093/toxsci/kfq394.

Desrochers M, Christou M, Jefcoate C, Belzil A, 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(8)**:1311-9.

Echigoya 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 U S A* **114(16)**:4213-4218. doi: 10.1073/pnas.1613203114.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32(5)**:1792-7.

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

Evans R, Mangelsdorf DJ (2014) Nuclear receptors, RXR, and the Big Bang. *Cell* **157:**255-266.

Fall AM, Johnsen R, Honeyman K, Iversen P, Fletcher S, Wilton SD (2006) Induction of revertant fibres in the mdx mouse using antisense oligonucletoides. *Genet Vaccines Ther* **4**:3-15.

Feng XJ, Gao H, Gao S, Li Z, Li H, Liu J, Wang JJ, Huang XY, Liu M, Zou J, et al. (2015) The orphan receptor NOR1 participates in isoprenaline-induced cardiac hypertrophy by regulating PARP-1. *Br J Pharmacol* **172:**2852-2863.

Fletcher S, Honeyman K, Fall AM, Harding PL, Johnson RD, Steinhaus JP, Moulton HM, Iversen PL, Wilton SD (2007) Morpholino Oligomer-Mediated Exon Skipping Averts the Onset of Dystrophic Pathology in the mdx Mouse. *Mol Ther* **15(9):**1587-1592.

Fujita A, 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(3)**:625-30.

Gardner-Stephen D, Heydel JM, Goyal A, Lu Y, Xie W, Lindblom T, Mackenzie P, Radominska-Pandya A (2004) Human PXR variants and their differential effects on the regulation of human UDP-glucuronosyltransferase gene expression. *Drug Metab Dispos* **32(3)**:340-7.

Grau-Bové X, Ruiz-Trillo I, Irimia M (2018) Origin of exon skipping-rich transcriptomes in animals driven by evolution of gene architecture. *Genome Biol*. **19(1)** :135. doi: 10.1186/s13059-018-1499-9.

Guo Z, Qiu Y (2011) A new trick of an old molecule: androgen receptor splice variants taking the stage?! *Int J Biol Sci* **7(6)**:815-22. Review.

Heald A, Anderson LV, Bushby KM, Shaw PI (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, Tujague M, Ström A, Treuter E, Warner M, Gustafsson JA (2007) Estrogen receptors: how do they signal and what are their targets. *Physiol Rev* **87(3)**:905-31. Review.

Hisatomi H, Kohno N, Wakita K, Nagao K, Hirata H, Hikiji K, 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(2)**:182-5.

Holzer G, Markov GV, Laudet V (2017) Evolution of Nuclear Receptors and Ligand Signaling: Toward a Soft Key-Lock Model? *Curr Top Dev Biol* **125**:1-38. doi: 10.1016/bs.ctdb.2017.02.003.

Honkakoski P, Negishi M (2000) Regulation of cytochrome P450 (CYP) genes by nuclear receptors. *Biochem J* **347(Pt 2)**:321-37. Review.

Horvath G, Leser G, Helou K, 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(2)**:271-9.

Jerzak KJ, Cockburn J, Pond GR, Pritchard KI, Narod SA, Dhesy-Thind SK, Bane A (2015) Thyroid hormone receptor α in breast cancer: prognostic and therapeutic implications. *Breast Cancer Res Treat* **149(1)**:293-301. doi:10.1007/s10549-014-3235-9.

Jiang P, Hou Z, Bolin JM, Thomson JA, 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(1)**:251-265.

Kane CD, Means AR (2000) Activation of orphan receptor-mediated transcription by Ca(2+)/calmodulin-dependent protein kinase IV. *EMBO J* **19(4):**691-701.

Kim HJ, Hwang JY, Kim HJ, Choi WS, Lee JH, Kim HJ, Chang KC, Nishinaka T, Yabe-Nishimura C, 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(9)**:2577-83.

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(16)**:4704-4715. doi: 10.1158/1078-0432.CCR-17-0017.

Kole R, Krainer AR, Altman S (2012) RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nat Rev Drug Discov* **11(2)**:125-40.

Koo T, Wood MJ (2013) Clinical trials using antisense oligonucleotides in duchenne muscular dystrophy. *Hum Gene Ther* **24(5)**:479-88.

Krett NL, Pillay S, Moalli PA, Greipp PR, Rosen ST (1995) A variant glucocorticoid receptor messenger RNA is expressed in multiple myeloma patients. *Cancer Res* **55(13)**:2727-9.

Lambert CA, Garbacki N, 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 IM, 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, Leung YK (2013) Differential expression of estrogen receptor beta isoforms in prostate cancer through interplay between transcriptional and translational regulation. *Mol Cell Endocrinol* **376(1-2)**:125-35. doi: 10.1016/j.mce.2013.06.023.

Lee Y, Rio DC (2015) Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* **84**:291-323. doi: 10.1146/annurev-biochem-060614-034316. Review.

Letunic I, Copley RR, Bork P (2002) Common exon duplication in animals and its role in alternative splicing. *Hum Mol Genet* **11(13)**:1561-7.

Li Y, He B, Gao J, Liu QS, Liu R, Qu G, Shi J, Hu L, Jiang G (2018) Methylmercury exposure alters NA splicing in human neuroblastoma SK-N-SH cells: Implications from proteomic and post-transcriptional responses. *Env Pollution* **238**:213-221.

Lim KR, Maruyama R, Yokota T (2017) Eteplirsen in the treatment of Duchenne muscular dystrophy. *Drug Des Devel Ther* **11**:533-545. doi: 10.2147/DDDT.S97635. Review.

Lin YS, Yasuda K, Assem M, Cline C, Barber J, Li CW, Kholodovych V, Ai N, Chen JD, Welsh WJ, Ekins S, Schuetz EG (2009) The major human pregnane X receptor (PXR) splice variant, PXR.2, exhibits significantly diminished ligand-activated transcriptional regulation. *Drug Metab Dispos* **37(6)**:1295-304. doi:10.1124/dmd.108.025213

Lundell K, Thulin P, Hamsten A, 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, Kino T (2017) Single-Nucleotide Variations of the Human Nuclear Hormone Receptor Genes in 60,000 Individuals. *J Endocr Soc* **2(1)**:77-90. doi: 10.1210/js.2017-00406.

Magomedova L, Tiefenbach J, Zilberman E, Le Billan F, Voisin V, Saikali M, Boivin V, Robitaille M, Gueroussov S, Irimia M, Ray D, Patel R et al. (2019) ARGLU1 is a transcriptional coactivator and splicing regulator important for stress hormone signaling and development. *Nucleic Acids Res*. **47(6)**:2856-2870. doi: 10.1093/nar/gkz010.

Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schütz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P, Evans RM (1995) The nuclear receptor superfamily: the second decade. *Cell* **83(6):**835-9. Review.

Martinez-Montiel N, Rosas-Murrieta NH, Anaya Ruiz M, Monjaraz-Guzman E, Martinez-Contreras R (2018) Alternative Splicing as a Target for Cancer Treatment. *Int J Mol Sci* **19(2)**. pii: E545. doi: 10.3390/ijms19020545. Review.

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. doi: 10.1038/msb.2012.67.

McClorey G, Moulton HM, Iversen PL, Fletcher S, Wilton SD (2006) Antisense oligonucleotide-induced exon skipping restores dystrophin expression in vitro in a canine model of DMD. *Gene Ther* **13(19)**:1373.

McClorey G, Fall AM, Moulton HM, Iversen PL, Rasko JE, Ryan M, Fletcher S and Wilton SD (2006) Induced dystrophin exon skipping in human muscle explants. *Neuromuscul Disord* **16(9-10)**: 583-590.

McGuire J, Okamoto K, Whitelaw ML, Tanaka H, 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(45)**:41841-9.

Melangath G, Sen T, Kumar R, Bawa P, Srinivasan S, 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(12)**:e0188159.

Merkin J, Russell C, Chen P, Burge CB (2012) Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338(6114)**:1593-9. doi: 10.1126/science.1228186.

Modrek B, Lee C (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased rate of exon creation/ loss. *Nat Genet* **34**:177-180.

Mourich, DV, Iversen, PL (2009) Splicing in the immune system: potential targets for therapeutic intervention by antisense mediated alternative splicing. *Current Opinion in Molecular Therapeutics* **11(2):** 124-132.

Mourich DV, Oda SK, Schnell FJ, Crumley SL, Hauck LL, Moentenich CA, Marshall NB, Hinrichs DJ, 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(2)**:114-26.

Mueller GC, Herranen AM, Jervell KF (1958) Studies on the mechanism of action of estrogens. *Recent Prog Horm Res* **14**:95-129.

Nakai A, Seino S, Sakurai A, Szilak I, Bell GI, DeGroot LJ (1988) Characterization of a thyroid hormone receptor expressed in human kidney and other tissues. *Proc Natl Acad Sci U S A* **85(8)**:2781-5.

Okonechnikov K, Golosova O, Fursov M; UGENE team (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28(8)**:1166-7. doi: 10.1093/bioinformatics/bts091.

Pan Q, Shai O, Lee LJ, Frey BJ, 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 Therapeutics* **24(3)**: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, RARalpha1DeltaB and RARalpha1DeltaBC, define a new class of nuclear receptor isoforms. *Nucleic Acids Res* **29(24)**:4901-8.

Popadin KY, Gutierrez-Arcelus M, Lappalainen T, Buil A, Steinberg J, Nikolaev SI, Lukowski SW, Bazykin GA, Seplyarskiy VB, Ioannidis P, Zdobnov EM, Dermitzakis ET, Antonarakis SE (2014) Gene age predicts the strength of purifying selection acting on gene expression variation in humans. *Am J Hum Genet.* **95(6)**:660-74. doi: 10.1016/j.ajhg.2014.11.003.

Poola I, Speirs V (2001) Expression of alternatively spliced estrogen receptor alpha mRNAs is increased in breast cancer tissues. *J Steroid Biochem Mol Biol* **78(5)**:459-69.

Prakash C, Zuniga B, Song CS, Jiang S, Cropper J, Park S, Chatterjee B (2015) Nuclear Receptors in Drug Metabolism, Drug Response and Drug Interactions. *Nucl Receptor Res* **2**. pii: 101178.

Qu S, Zhong Y, Shang R, Zhang X, Song W, Kjems J, Li H (2017) The emerging landscape of circular RNA in life processes. *RNA Biol* **14(8)**:992-999. doi: 10.1080/15476286.2016.1220473. Review.

Qu W, Cingolani P, Zeeberg BR, Ruden DM (2017) A bioinformatics-based alternative mRNA splicing code that may explain some disease mutations in conserved in animals. *Front Genet* **8**:38. doi: 10.3389/fgene.2017.00038.

Rastinejad F, Huang P, Chandra V, Khorasanizadeh S (2013) Understanding nuclear receptor form and function using structural biology. *J Mol Endocrinol* **51(3)**:T1-T21. Review.

Reitzel AM, Pang K, Ryan JF, Mullikin JC, Martindale MQ, Baxevanis AD, 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(1)**:3. doi: 10.1186/2041-9139-2-3.

Richer JK, Lange CA, Wierman AM, Brooks KM, Tung L, Takimoto GS, Horwitz KB (1998) Progesterone receptor variants found in breast cells repress transcription by wild-type receptors. *Breast Cancer Res Treat* **48(3)**:231-41.

Rochel N, Wurtz JM, Mitschler A, Klaholz B, Moras D (2000) The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand. *Mol Cell* **5(1)**:173-9.

Sablin EP, Woods A, Krylova IN, Hwang P, Ingraham HA, Fletterick RJ (2008) The structure of the corepressor Dax-1 bound to its target nuclear receptor LRH-1. *Proc Natl Acad Sci* **105:**18390-18395.

Savkur RS, Burris TP (2004) The coactivator LXXLL nuclear receptor recognition motif. *J Pept Res* **63(3)**:207-12. Review.

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.

Schimmer BBP, White PC (2010) Minireview: Steroidigenic Ractor 1: Its roles in differentiation, development, and disease. *Mol Endocrinol* **24(7):**1322-1337.

Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479(7371)**:74-9. doi: 10.1038/nature10442.

Sladek FM (2011) What are nuclear receptor ligands? *Mol Cell Endocrinol* **334(1-2):**3-13.

Smith CW, Valcarcel 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, Graur D. (2002) Alu containing exons are alternatively spliced. *Genome Res* **12**:1060-1067.

Sorek R, Shamir R, Ast G (2004) How prevalent is functional alternative splicing in the human genome? *Trends Genet* **20(2)**:68-71. Review.

Sprung CN, Li J, Hovan D, McKay MJ, Forrester HB (2011) Alternative transcript initiation and splicing as a response to DNA damage. *PLoS One* **6(10)**: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, Aiyar SE, Li R (2007) Cofactor of BRCA1 modulates androgen-dependent transcription and alternative splicing. *J Steroid Biochem Mol Biol*. **107(3-5)**:131-9.

Suetsugi M, Su L, Karlsberg K, Yuan YC, Chen S (2003) Flavone and isoflavone phytoestrogens are agonists of estrogen-related receptors. *Mol Cancer Res* **1(13)**:981-91.

Syed YY (2016) Eteplirsen: First Global Approval. *Drugs* **76(17)**:1699-1704. Review.

Takacs M, Petoukhov MV, Atkinson RA, Roblin P, Ogi FX, Demeler B, Potier N, Chebaro Y, Dejaegere A, Svergun DI, et al. (2013) The asymmetric binding of PGC-1α to the ERRα and ERRγ nuclear receptor homodimers involves a similar recognition mechanism. *PLoS ONE* **8(7):**e67810.

Taladriz-Sender A, Campbell E, Burley GA (2019) Splice-switching small molecules: A new therapeutic approach to modulate gene expression. *Methods* **167**:134-142.

Tan JJ, Ong SA, Chen KS (2011) Rasd1 interacts with Ear2 (NR2F6) to regulate renin transcription. *BMC Mol Biol* **12:**4.

Thierry-Mieg D, 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, Zanger UM (2015) The truncated splice variant of peroxisome proliferator-activated receptor alpha, PPARα-tr, autonomously regulates proliferative and pro-inflammatory genes. *BMC Cancer* **15**:488. doi: 10.1186/s12885-015-1500-x.

Ule J, Blencowe BJ (2019) Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol Cell* **76(2)**:329-345. Review.

Unutmaz D (2009) RORC2: the master of human Th17 cell programming. *Eur J Immunol* **39(6)**:1452-5. doi: 10.1002/eji.200939540.

van der Vaart M, Schaaf MJ (2009) Naturally occurring C-terminal splice variants of nuclear receptors. *Nucl Recept Signal* **7**:e007. doi:10.1621/nrs.07007. Review.

van Roon-Mom WM, 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. doi: 10.1038/ncomms3480.

Villaseñor-Altamirano AB, Watson JD, Prokopec SD, Yao CQ, Boutros PC, Pohjanvirta R, Valdés-Flores J, Elizondo G (2019) 2,3,7,8-Tetrachlorodibenzo-p-dioxin modifies alternative splicing in mouse liver. *PLoS One* **14(8):**e0219747.

Vuong LM, Chellappa K, Dhahbi JM, Deans JR, Fang B, Bolotin E, Titova NV, Hoverter NP, Spindler SR, Waterman ML, Sladek FM (2015) Differential Effects of Hepatocyte Nuclear Factor 4α Isoforms on Tumor Growth and T-Cell Factor 4/AP-1 Interactions in Human Colorectal Cancer Cells. *Mol Cell Biol* **35(20)**:3471-90. doi: 10.1128/MCB.00030-15.

Wallace BD, Redinbo MR (2013) Xenobiotic-sensing nuclear receptors involved in drug metabolism: a structural perspective. *Drug Metab Rev* **45(1)**:79-100. doi: 10.3109/03602532.2012.740049. Review.

Wan Y, Larson DR (2018) Splicing heterogeneity: separating signal from noise. *Genome Biol* **19(1)**:86.

Wang W, Chen J, Luo L, Li Y, Liu J, 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(2)**:227-239.

Weatheritt RJ, Sterne-Weiler T, Blencowe BJ (2016) The ribosome-engaged landscape of alternative splicing. *Nat Struct Mol Biol* **23(12)**:1117–1123. 10.1038/nsmb.3317

Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, Zheng H, Goity A, van Bakel H, et al. (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158(6)**:1431-1443. doi:10.1016/j.cell.2014.08.009.

Wilton S, Fall AM, Harding PL, McClorey G, Coleman C, Fletcher S (2007) Antisense oligonucleotide-induced exon skipping across the human dystrophin gene transcript. *Mol Ther* **7**:1288-1296.

Xing Y, Lee CJ (2004) Negative selection pressure against premature protein truncation is reduced by alternative splicing and diploidy. *Trends Genet* **20(10)**:472-5.

Xing Y, Lee C. (2005) Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. *Proc Natl Acad Sci U S A* **102(38)**:13526-31.

Yang P, Wei X, Zhang J, Yi B, Zhang GX, Yin L, Yang XF, Sun J (2016) Anti-thrombotic effects of Nur77 and Nor1 are mediated through upregulating thrombomodulin expression in endothelial cells. *Arterioscler Thromb Vasc Biol* **36(2):**361-369.

Yang X, Coulombe-Huntington J, Kang S, Sheynkman GM, Hao T, Richardson A, Sun S, Yang F, Shen YA, Murray RR, Spirohn K, Begg BE et al. (2016) Widespread Expansion of Protein Interaction

Capabilities by Alternative Splicing. *Cell* **164(4)**:805-17. doi:10.1016/j.cell.2016.01.02Yang Z, 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, Hoffman E (2009) Efficacy of systemic morpholino exon-skipping in Duchenne dystrophy dogs. *Ann Neurol* **65(6)**:667-76. doi: 10.1002/ana.21627.

Zaharieva E, Chipman JK, 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, 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(3)**:1436-41.

Zennaro MC, Souque A, Viengchareun S, Poisson E, Lombès M (2001) A new human MR splice variant is a ligand-independent transactivator modulating corticosteroid action. *Mol Endocrinol* **15(9)**:1586-98.

Zhang X, Peng Q, Li L, Li X (2016) Recognition of alternatively spliced cassette exons based on a hybrid model. *Biochem Biophys Res Commun*. **471(3)**:368-72. doi: 10.1016/j.bbrc.2016.02.022.

Zhao Y, Zhang K, Giesy JP, Hu J (2015) Families of nuclear receptors in vertebrate models: characteristic and comparative toxicological perspective. *Sci Rep* **5**:8554. doi: 10.1038/srep08554.

Zhi X, Zhou XE, He Y, Zechner C, Suino-Powell KM, Kliewer SA, Melcher K, Mangelsdorf DJ, Xu HE (2014) Structural insights into gene expression by the orphan nuclear receptor SHP. *Proc Natl Acad Sci* **111**:839-844.

Zhou R, Chun RF, Lisse TS, Garcia AJ, Xu J, Adams JS, Hewison M (2015) Vitamin D and alternative splicing of RNA. *J Steroid Biochem Mol Biol.* **148**:310-7. doi: 10.1016/j.jsbmb.2014.09.025. Review.

Zhu H, Huang Y, Su H, Ma Y, Tao Y, Liao DJ, Liu Y, Feng Z (2018) Identification of a novel human estrogen receptor-α splice variant able to enhance malignant biological behaviors of breast cancer cells. *Oncol Lett* **15(4)**:5339-5344. doi: 10.3892/ol.2018.7970.

Zhu LY, Zhu YR, Dai DJ, Wang X, Jin HC (2018) Epigenetic regulation of alternative splicing. *Am J Cancer Res* **8(12)**:2346-2358. Review.

FIGURE LEGENDS

Figure 1.) Stereo Meta-Analysis of Human Nuclear Receptor Transcript Variants and SNPs. The total number of mRNA transcript variants and single nucleotide polymorphisms (SNPs) for each of the 48 human nuclear receptors, and the processed NR pseudogene FXRβ (NR1H5), were catalogued using the NIH Aceview and Ensembl databases (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 CYP 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), SHP-1 (3), Dax-1 (4) and RORβ (4) display the fewest number of transcript variants. Each NR genes is associated with ~6,985 SNPs, compared to only 848 SNPs per gene for the related CYP 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), SHP-1 (407), 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 (9,373). Interestingly, the xenobiotic sensor CAR (NR1I3), 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.

Figure 2.) Cassette Exon Visualization Pictographs (CEViP) 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 transcript. The

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

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, 2 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 (2/3 complete) an arginine (Arg or 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 3-base codon at each node of the transcript. We utilized this CEViP methodology to analyze the organization of cassette exons in all 48 human NR genes and the complete analysis is shown in **Supplemental Figure 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, post-splicing, are easily visualized. Here we highlight the 4 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, that provide the proper frame for joining two bordering exons in its absence; and 4) Multi-exonic 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.

Figure 3.) Cassette Exon Structure of the Prototype Nuclear Receptor NR2A1. (**A**) The crystal structure of human NR2A1 (HNF4α; PDB: 4IQR), shown here as a cylindrical cartoon diagram shows the structural basis for both the DNA-binding domain (DBD) and the ligand-binding domain (LBD) of a class

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

I NR bound to a DR1 DNA binding element from the HNF4α promoter. The DBD is comprised of 3 alpha helices (α1, α2, and the c-terminal extension helix (CTE)) partially encoded by exons 2 (gray cylinder) and 3 (red cylinder). The LBD is comprised of 10 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 positioning of putative cassette exons in HNF4α are highlighted. Exons 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 co-regulatory 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 **Figure 2**; individual exons are colored using the same scheme used in **Figures 3A** and **3B**. The positioning of cassette exons with respect to the DBD and LBD are highlighted, and the 7 natural splice variants (ΔExon5, ΔExon7, ΔExon7,8, ΔExon7-9, ΔExon8, ΔExon8,9, and ΔExon9) encoded within HNF4α's reference gene assembly are shown.

Figure 4.) CEViP Modeling of the Vitamin D Receptor. (**A**) A schematic representation of the human vitamin D receptor gene (VDR; NR1I1) is shown using the CEViP modeling method, which has been augmented to highlight key secondary structures in the NR gene family, including the DNA-binding domain (DBD; shown in orange) and the ligand-binding domain (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 are 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, due to 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) a 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 Figure 1**. (**B**) The canonical LBD of the VDR is comprised 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 **Figure 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 ($1,25(OH)_2D_3$) binding pocket in the LDB 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 catalogued 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 Figure 1**).

Figure 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 to all other NR families. Interestingly, these NR genes lack a canonical DNA-binding domain (DBD), but encode a classical ligand-binding domain (LBD) segmented across 2 standard, non-cassette 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 Family 4 NRs (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 Figure 1**). Phylogenetic analysis of NR gene coding

sequences (see **Supplemental Figure 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 Figure 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 4-5 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.

Figure 6.) Alternative Splicing of Metabolic Genes Refine Adaptive Physiological 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 CYP 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 co-evolved in concert with related NR and CYP genes, as exemplified by the interplay among PXR, CYP3A4, and P-glycoprotein (also known as multidrug resistance protein 1 (MDR1)), a well characterized ABC-transporter, encoded by the ABCB1 gene (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 physiological responses to the environment through splice variant selection that promotes a return to cellular homeostasis after both acute and chronic endo-xenobiotic exposures.

TABLES

TABLE 1. Summary of Human Nuclear Receptor Variants.

NR Family (# genes)	**Transcript Size** (Bases)	**Exons**	**Amino Acids** (Residues)	**Molecular Weight** (kDa)	**Transcript Variants**	**SNPs**	**SNPs per Variant**[†]
NR0 (2)	1011-2021	1-2	54-470	6.3-51.7	7	871	124
NR1 (20)	1256-10998	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)	2169-13762	5-11	344-984	38.7-107.1	211	111,634	529
NR4 (3)	2465-4996	7-8	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[*]	1011-10998	1-15	54-984	6.3-114.2	1004	342,359	341[1]

[*] The analysis includes all 48 human NR genes plus the processed human pseudogene FXRβ (NR1H5).
[†]Derived as Total SNPs/Total Splice Variants; Equivalent to 0.003 splice variants/SNPs. The mean average for Total SNPs/Total Splice Variants for all 48 primary human NR genes is 473 SNPs/variant or 0.009 variants/SNP (not shown). The median for the 48 human NR genes is 191 SNPs/variant or 0.005 variants/SNP.

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

Table 2. Representative Nuclear Receptor Association with Pathology.

NR Gene	Group	Disease Associations	SNP/GWAS[1]
DAX-1	NR0B1	X-Linked Adrenal Hyperplasia; Azoospermia	Flucloxacillin-Induced Liver Injury X-Linked Adrenal Hypoplasia Autism and Schizophrenia
SHP	NR0B2	Hepatocellular Carcinoma, Lipidystrophy, and Obesity	Body Mass Index (BMI), Glucose/Insulin Resistance
THRα	NR1A1	Resistance to Thyroid Hormone	-
RARβ	NR1B2	Neural Tube Defects	-
RARγ	NR1B3	Anthracycline Cardiotoxicity	4 Khz Hearing Threshold
PPARα	NR1C1	-	Insulin Resistance
PPARβ/δ	NR1C2	-	Altered Blood Glucose
PPARγ	NR1C3	-	Type II Diabetes (T2D)
REV-erbα	NR1D1	Disordered Circadian Clock	Altered White Blood Cell Count
REV-erbβ	NR1D2	Chagas Cardiomyopathy	Severe Acne
RORα	NR1F1	-	Alcoholism, Alzheimer Disease, ALS, Asthma, Diabetic Nephropathies
RORβ	NR1F2	Circadian Rhythm, Retinal Degeneration	Stroke, Tobacco Use, Type II Diabetes, Macular Degeneration
LXRβ	NR1H2	Gall Bladder Cancer Risk	Elevated vascular endothelial growth factor (VEGF) levels
LXRα	NR1H3	Coronary Artery Disease, T2D	Supranuclear Palsy; Metabolic Syndrome
VDR	NR1I1	Vitamin D-Resistant Rickets	Inflammatory Bowel Disease; Microbiome
CAR	NR1L3	Zellweger Syndrome	Blood Metabolite Levels
HNF4α	NR2A1	Diabetes Mellitus, Adenocarcinoma, Nephropathies	Rheumatoid Arthritis, T2D
RXRα	NR2B1	Adenocarcinoma, Attention-Deficit/Hyperactivity Disorder (ADHD), Bipolar Disorder, Autism	Macular Degeneration, ALS, Hypertension
RXRβ	NR2B2	Alcoholism, Cancer (NSCLC)	Wegener's Granulomatosis
RXRγ	NR2B3	Share with NR2B1/NR2B2	Age of Onset Menarche
TR2	NR2C1	Prostate Neoplasms	Alopecia
TR4	NR2C2	Alzheimer Disease, Vitamin A Deficiency	-
TLX	NR2E1	Human Lymphoid Leukemia	Increased Reaction Time
PNR	NR2E3	Retinitis Pigmentosa 37	Body Height, HVA:MHPG Ratio
ERR-β	NR3B2	Deafness	ADHD Symptoms
GR	NR3C1	Rheumatoid Arthritis, Asthma	BMI
MR	NR3C2	Hypertension, Pseudohypoaldosteronism	Breast Neoplasms, Nicotine Dependence, Akathisia, Schizophrenia
AR	NR3C4	Androgen Insensitivity	Elevated LDL Cholesterol, Alopecia
NURR1	NR4A2	Parkinson's Disease, ADHD	Age Onset Menarche
LRH-1	NR5A2	Pancreatic Cancer	Crohn's Disease, Ulcerative Colitis, Pancreatic Neoplasms
GCNF	NR6A1	Prostate Cancer Marker	Macular Degeneration, Narcolepsy

[1] Data obtained in early 2019 from www.GWASCentral.org and GWAS Catalog www.ebi.ac.uk/gwas/search?query

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

TABLE 3. Nuclear Receptor Variants Associated with Human Disease.

Isoform	Variant	Related Disease	Reference
NR1A1	TRα2	Tissue specific processing in kidney and liver TRα2 expression associated with breast cancer survival.	Nakai et al. 1988 Jerzak et al. 2015
NR1B1	RARα1ΔBC RARA fusion	Functional interaction with RARα1 PKRAR1a:RARA fusion in promyleocytic leukemia.	Parrado et al. 2001 Catalano et al. 2007
NR1C1	PPARα	Truncated variants exert antitumor effects.	Thomas et al. 2015
NR1C2	PPARδ-trunc	3'-truncations (ΔEx4-9) are dominant negatives.	Lundell et al. 2007
NR1C3	PPARγ Ex3'	Interfere with PPARγ1; chemotherapeutic resistance.	Kim et al. 2007
NR1F3	ROR-γt	Drive Th17 cell differentiation.	Unutmaz D. 2009
NR1I1	VDR Ex1f	Under-expression 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	HNF4α2	Colon cancer tumor suppressor; HNF4α8 is not	Vuong et al. 2015
NR2B2	RXRβ2	4 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	ERα ΔEx3, ΔEx3,7, ΔEx6, ΔEx2,4,5; ΔEx7; ΔEx5,7	16 variants associated with Breast Cancer, Endometrial Adenocarcinoma; Dominant negative ΔEx3, ΔEx2,4,5, ΔEx3,7, ΔEx6 enriched in cancer cells; ΔEx7 induction regulated by estradiol; ΔEx5,7 induced by tamoxifen, does not bind DNA promoter element.	Poola and Speirs 2001 Horvath et al. 2002
NR3A2	ERβ2, β5	Differential expression in Prostate Cancer; Use of 5'-most exons (0K, 0N) drive ERβ2, β5 expression.	Lee et al. 2013
NR3C1	GR var	Promotes glucocorticoid resistance in multiple myeloma.	Krett et al. 1995
NR3C3	PRΔ2,Δ4, PRΔ5,6, PRΔ6; delta6/2 ASV	Found in normal and breast cancer tissue; PRΔ5,6, PRΔ6 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 co-expressed with AR-V7, share similar 3' cryptic exon; predictive of drug resistance.	Dehm et al. 2008 Antonarakis et al. 2014 Kohli et al. 2017

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

Table 4. Human Nuclear Receptor Variant Localization and Expression.

Isoform	Variant Name[1]	Localization/Features[2]	Tissue
NR0B1	DAX1-var a,b	Cytoplasmic; ΔLBD	Carcinoid Lung, Adrenals
NR0B2	SHP-var a, b	Mitochondrial; ΔLBD	Liver, Kidney
NR1B2	RARβ (4) variant	Cytoplasmic; subnuclear compartments	Mammary epithelial cells
NR1B3	RARγ-var n -var o	ΔLBD ΔDBD	Esophageal Tumor Embryonic Stem Cells
NR1C1	PPARα-var e	Nucleus; ΔDBD	Choriocarcinoma
NR1C2	PPARβ/δ-var g -var m	Cytoplasmic; ΔLBD Nucleus	Prostate/Ovarian Cancer Uterus
NR1C3	PPARγ-var j -var c	Nuclear Cytoplasmic	Liver Lung, Colon Macrophage
NR1F1	RORα-var a, b & d	Cytoplasmic; No membrane associations	Testis, brain
NR1I2	PXR-var c -var f	Cytoplasmic; ΔLBD Nucleus; ΔDBD	Liver, Spleen, Colon, Stomach, Tumor Tissue
NR1I3	CAR-var h -var w -var bb, cc	Cytoplasmic; ΔLBD/ ΔDBD Nucleus; ΔDBD Mitochondrial; ΔDBD	Liver
NR2B2	RXRβ-var e -var f	Nucleus; ΔDBD Cytoplasmic; ΔLBD/ ΔDBD	B-Cell Testis, Skin
NR2B3	RXRγ-var c	Cytoplasmic; ΔLBD (unspliced)	Eye
NR2F2	COUP-TFII-var g	Endoplasmic reticulum	Kidney, Lung, Uterus
NR2F6	EAR2-var b -var c -var d	Nuclear Cytoplasmic; ΔLBD Mitochondrial	Brain Lung, Kidney, Brain Eye, Kidney
NR3C3	PR-var d PR-var g	Nucleus Cytoplasmic ; ΔLBD	Breast Cancer Adipose, Aorta
NR3B1	ERRα-var c	Cytoplasmic	Lung, Colon, Kidney
NR4A2	NURR1-var i	Cytoplasmic; ΔLBD	Lung

[1]Human nuclear receptor splice variants (var) associated with alternative subcellular trafficking or tissue distribution currently listed in the NIH Aceview Database: https://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/

[2] Sequence-based predictions of alternative subcellular trafficking for NR variants; ΔLBD refers to loss of a complete LBD coding sequence in the transcript variant; ΔDBD refers to loss of a complete DBD coding sequence in the variant transcript.

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

FIGURES

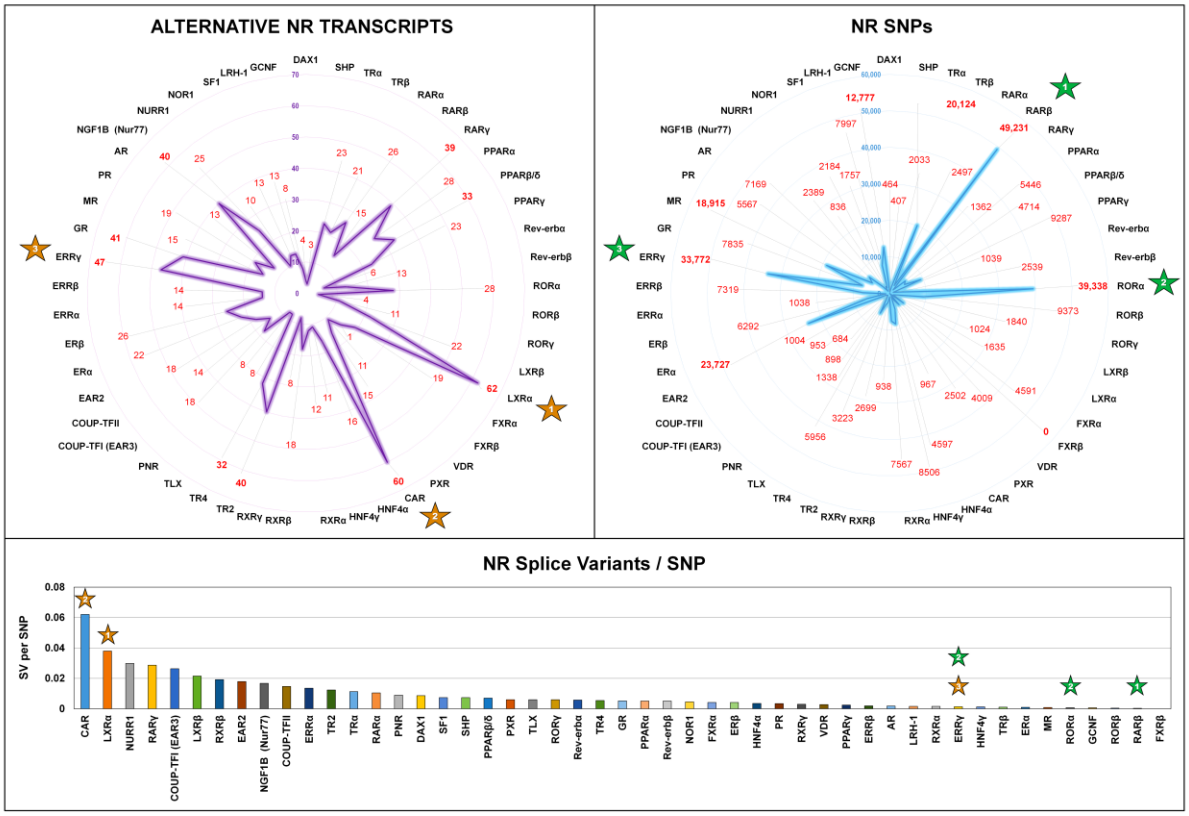

Figure 1

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

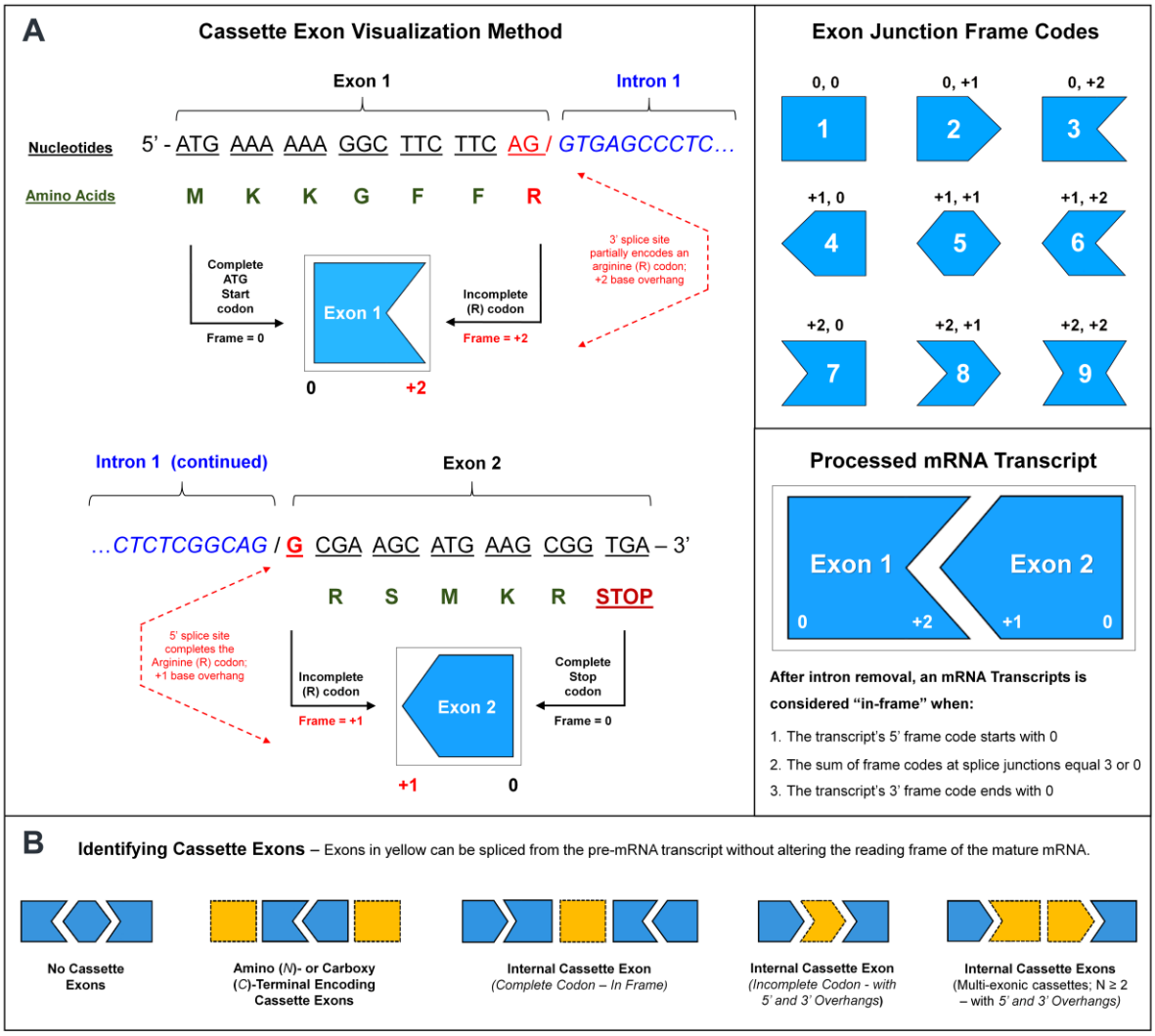

Figure 2

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

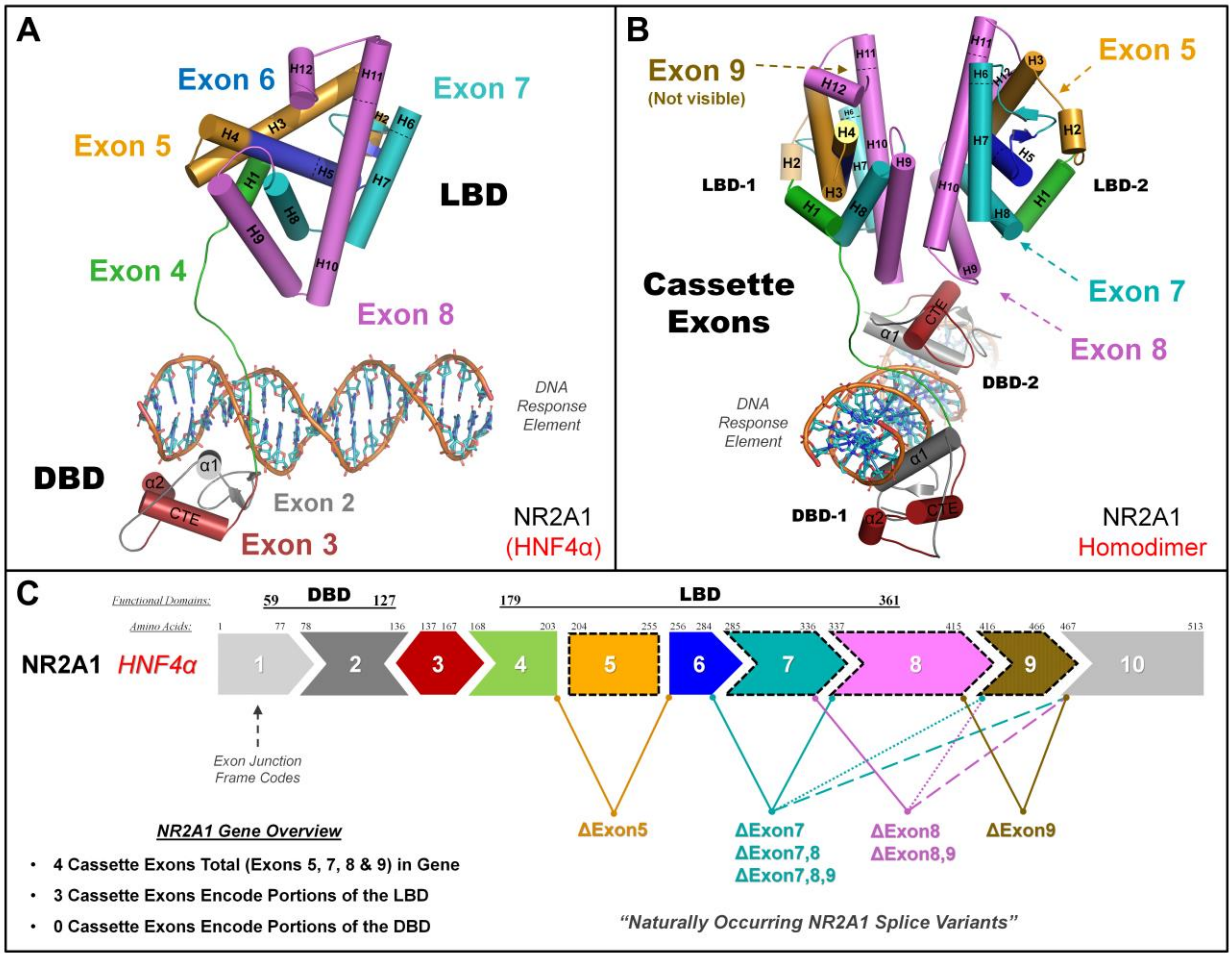

Figure 3

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

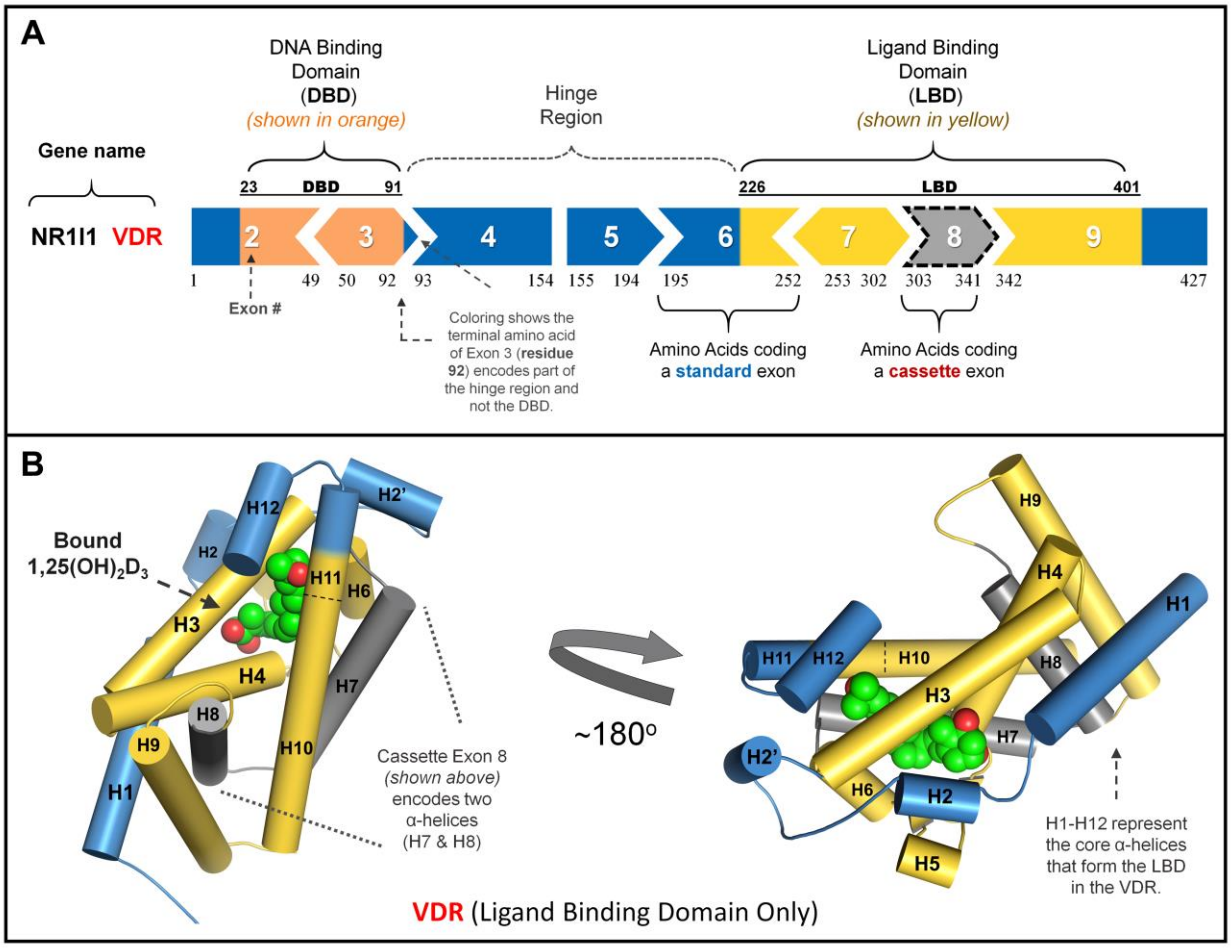

Figure 4

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

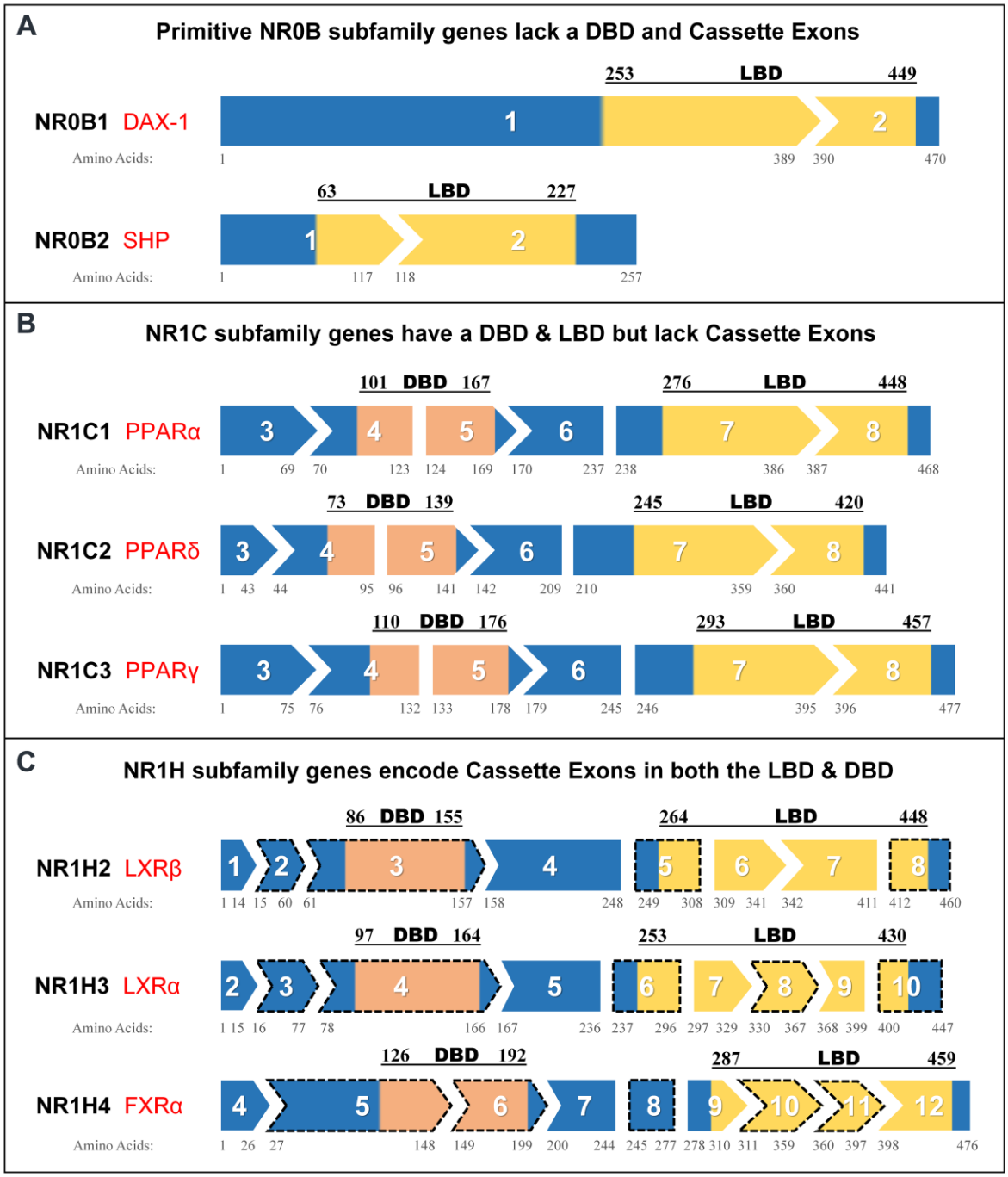

Figure 5

DMD Fast Forward. Published on January 24, 2020 as DOI: 10.1124/dmd.119.089102
This article has not been copyedited and formatted. The final version may differ from this version.

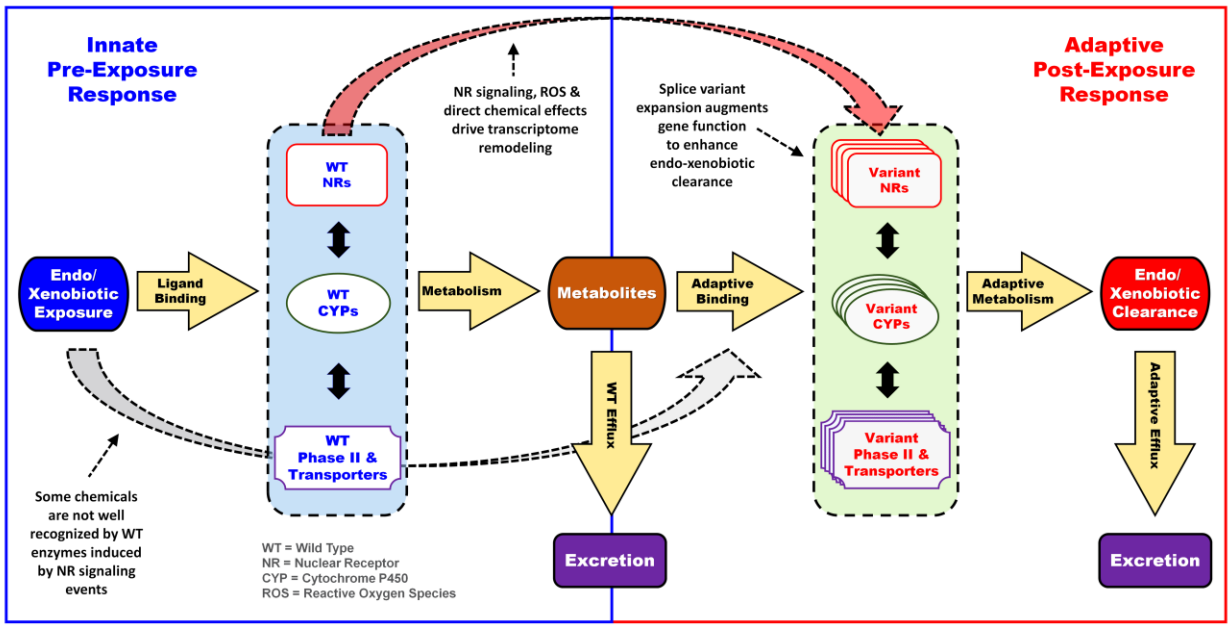

Figure 6