Pharmacology & Therapeutics 136 (2012) 283–294

Contents lists available at SciVerse ScienceDirect

Pharmacology & Therapeutics

journal homepage: www.elsevier.com/locate/pharmthera

Associate editor: B. Teicher

Alternative transcription and alternative splicing in cancer

Sharmistha Pal, Ravi Gupta, Ramana V. Davuluri *

Center for Systems and Computational Biology, The Wistar Institute, Philadelphia, PA, USA
Molecular and Cellular Oncogenesis Program, The Wistar Institute, Philadelphia, PA, USA

ARTICLE INFO

Keywords:
Alternative transcription
Alternative splicing
Transcript variants
Protein isoforms
Cancer diagnostics
Cancer therapeutics

ABSTRACT

In recent years, the notion of "one gene makes one protein that functions in one signaling pathway" in mammalian cells has been shown to be overly simplistic. Recent genome-wide studies suggest that at least half of the human genes, including many therapeutic target genes, produce multiple protein isoforms through alternative splicing and alternative usage of transcription initiation and/or termination. For example, alternative splicing of the vascular endothelial growth factor gene (*VEGFA*) produces multiple protein isoforms, which display either pro-angiogenic or anti-angiogenic activities. Similarly, for the majority of human genes, the inclusion or exclusion of exonic sequences enhances the generation of transcript variants and/or protein isoforms that can vary in structure and functional properties. Many of the isoforms produced in this manner are tightly regulated during normal development but are misregulated in cancer cells. Altered expression of transcript variants and protein isoforms for numerous genes is linked with disease and its prognosis, and cancer cells manipulate regulatory mechanisms to express specific isoforms that confer drug resistance and survival advantages. Emerging insights indicate that modulating the expression of transcript and protein isoforms of a gene may hold the key to impeding tumor growth and act as a model for efficient targeting of disease-associated genes at the isoform level. This review highlights the role and regulation of alternative transcription and splicing mechanisms in generating the transcriptome, and the misuse and diagnostic/prognostic potential of alternative transcription and splicing in cancer.

Contents

1. Introduction . . . 284
2. The roads to transcriptome and proteome diversity . . . 284
3. Regulation of alternative events . . . 284
4. Transcript variants and protein isoforms in cancer . . . 286
5. Why is the cancer transcriptome different? . . . 287
6. Transcript variants and protein isoforms as cancer diagnostics and therapeutic targets . . . 289
7. Role of transcript variants and protein isoforms in drug resistance . . . 290
8. Concluding remarks . . . 291
Conflict of interest statement . . . 291
Acknowledgments . . . 292
References . . . 292

Abbreviations: AFE, Alternative first exon; ALE, Alternative last exon; AR, Androgen receptor; EMT, Epithelial mesenchymal transition; mRNA, Messenger RNA; NcRNA, Non-coding RNA; Pol II, RNA polymerase II; Pre-mRNA, Premature RNA; TF, Transcription factor; TKI, Tyrosine kinase inhibitor; UTR, Untranslated region; TSS, Transcription start site; NGS, Next-generation sequencing.

* Corresponding author at: Center for Systems and Computational Biology, The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA. Tel.: 215 495 6903; fax: 215 495 6848.
E-mail address: rdavuluri@wistar.org (R.V. Davuluri).

0163-7258/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.pharmthera.2012.08.005

1. Introduction

Transcriptome and proteome complexity has evolved not through acquisition of new genes but from alternative transcription and alternative splicing mechanisms that generate a variety of transcripts from a single gene. Transcriptional events that involve the use of alternative promoters and/or transcriptional termination sites result in multiple pre-mRNAs that can further undergo alternative splicing to generate a plethora of transcript variants corresponding to a single gene. The resulting transcript variants either translate to alternative proteins, known as protein isoforms, which may vary in structure and functional properties, or remain as non-coding transcripts. For example, *SRA1* produces both NcRNA and protein isoforms that function as co-activators for TFs (Ulveling et al., 2011), and the transcript variants of *BDNF*, which differ in their first exon (5′UTR), translate the same preproBDNF protein (Timmusk et al., 1993; Pruunsild et al., 2007). Therefore, a gene can yield transcript variants that differ either in their regulatory UTRs or/and protein coding regions; thereby expanding the complexity of mammalian transcriptomes and proteomes.

Currently, the human genome contains annotations for 28,526 genes that express a total of 120,145 transcripts, of which 80,932 are protein coding and 39,213 are non-coding transcripts (Ensembl 65). Based on the current annotations, while 65% of the genes produce multiple transcripts, with an average of six transcript variants per gene, the remaining 35% are single transcript genes. While the alternative transcription and splicing mechanisms have long been known for some genes, e.g. *BDNF*, *MYC*, and *FGFR* (Johnson et al., 1991; Miki et al., 1992; Timmusk et al., 1993; Hann, 1995), their prevalence in almost all multi-exon genes has been recently realized with the increasing application of high-throughput experimental methods, such as exon arrays and NGS methods. In fact, recent survey of various mammalian tissues, developmental stages and cell-lines suggests that at least 90% of the protein-coding genes use alternative transcriptional and alternative splicing events (Barrera et al., 2008; Pan et al., 2008; Wang et al., 2008; Kalsotra & Cooper, 2011; Pal et al., 2011; Sun et al., 2011). Therefore, the mammalian transcriptome and proteome is far more diverse than expected from "one gene → one mRNA → one protein" paradigm.

Henceforth, for simplicity, the transcript variants and the protein isoforms produced from a single gene locus will be referred to as isoforms. Emerging evidence indicate that the alternative isoforms are regulated by a combination of genetic and epigenetic regulatory mechanisms and their balanced expression is tightly linked to one another. Further, there is now growing evidence linking aberrant use of alternative isoforms and cancer formation, several oncogenes and tumor-suppressor genes (e.g. *LEF1*, *TP63*, *TP73*, *HNF4A*, *RASSF1* and *BCL2L1*) are known to have multiple promoters and alternative transcript variants, and moreover, it is known that the aberrant use of one isoform over another in some of these genes is directly linked to cancerous cell growth (Davuluri et al., 2008; Rajan et al., 2009). In this review, we summarize the mechanisms of isoform production, regulation of isoform expression, the role of specific isoforms in oncogenesis and drug resistance, and discuss the potential of isoforms as biomarkers and therapeutic targets. For recent reviews on the technological developments in genome-wide identification of alternative transcription and alternative splicing, see Refs. Martin and Wang (2011) and Lenhard et al. (2012).

2. The roads to transcriptome and proteome diversity

The gene regulatory mechanisms that produce multiple isoforms from a gene can be broadly classified into two groups (1) alternative transcriptional mechanisms and (2) alternative splicing mechanisms (see Fig. 1). Pol II transcribes a gene to produce pre-mRNA, which undergoes splicing to form a mature RNA that is either translated to protein or functions as NcRNA. When alternative transcriptional mechanisms are at play, isoforms arise due to the production of multiple pre-mRNAs, whereas alternative splicing produces isoforms through modulation of the splicing pattern of a single pre-mRNA. Though transcription of pre-mRNA and its splicing to mature RNA are two distinct events, they are coupled and interdependent, with splicing being co-transcriptional (Listerman et al., 2006; Kornblihtt, 2007). Alternative transcription events include alternative promoter and transcription termination usage, alternative splicing covers exon skipping, mutually exclusive exons (special case of exon skipping), alternative 5′ and 3′ splice site usage, intron retention, and alternative polyadenylation. In literature, the term "alternative splicing" has been used to describe both alternative transcription and splicing events, but in this review we highlight the mechanistic differences. Alternative promoters are defined as regions of the same gene that recruit the Pol II machinery and produce pre-mRNAs with distinct non-overlapping first exons called alternative first exons (AFEs). For some genes, overlapping first exons are observed due to multiple TSSs within a given promoter/first exon, but these events are not alternative promoter usage because the regulatory promoter (−1000 bp to +500 bp of TSS) remains the same. Similarly, when the pre-mRNAs end with distinct non-overlapping last exons (alternative last exon-ALE), they arise from alternative transcription termination events. In contrast, the variations in the pattern of intron removal, exon joining, and the addition of a poly A tail on a single pre-mRNA result in alternatively spliced mature RNAs. These various alternative events have been identified in different cells and tissues by application of NGS based methods in recent genome-wide studies (Sultan et al., 2008; Wang et al., 2008; Pal et al., 2011). It is estimated that there are 263,772 exons in the human genome. Approximately 22% of these exons participate in alternative splicing phenomena, while ~21% and 14% of these exons represent AFE and ALE, respectively, highlighting the importance of alternative promoter and transcription termination in producing transcriptome and proteome diversity (Table 1). Consistent with this global analysis, approximately 81% of multi-transcript genes that are expressed during postnatal cerebellum development used alternative transcriptional mechanisms (Pal et al., 2011), while 37–60% of genes used alternative promoters in different cells and tissues (Carninci et al., 2006; Kimura et al., 2006; Pal et al., 2011; Sun et al., 2011). Furthermore, many novel gene promoters both within gene and in inter-genic regions were identified (Carninci et al., 2006; Gupta et al., 2010; Illingworth et al., 2010; Pal et al., 2011; Sun et al., 2011), and many genes possess previously unknown alternative transcription termination and polyadenylation sites (Tian et al., 2005, 2007). Hence, though most studies have indicated alternative splicing as the major driving force for transcriptome and proteome diversity, recent findings indicate that alternative transcription mechanisms are equally, if not more, important in generating this diversity. Understanding this difference is critical in order to design the experimental and computational approaches to study the isoform-level gene regulatory networks and infer the dynamic gene interactions across different physiological and disease conditions (Fig. 1C).

3. Regulation of alternative events

Most isoforms are expressed in a tissue/cell-specific (spatial) and development time/stage-specific (temporal) manner, suggesting that the alternative events responsible for isoform expression are tightly regulated. For example, it was found that 53% of the alternative promoter-driven transcript variants were differentially expressed during mouse cerebellar development (more than 2 fold change across four developmental stages) and ~17% of the alternative promoters were differentially bound by Pol II among the five adult tissues surveyed (Pal et al., 2011; Sun et al., 2011). Recent studies have shown the involvement of both genetic and epigenetic