Seminars in Cancer Biology 22 (2012) 417–427

Contents lists available at SciVerse ScienceDirect

Seminars in Cancer Biology

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

Review

Complex changes in alternative pre-mRNA splicing play a central role in the epithelial-to-mesenchymal transition (EMT)

Claude C. Warzecha[a], Russ P. Carstens[a,b,*]

[a] Department of Medicine, University of Pennsylvania, Perelman School of Medicine, United States
[b] Department of Genetics, University of Pennsylvania, Perelman School of Medicine, United States

ARTICLE INFO

Keywords:
Alternative splicing
ESRP
EMT
Epithelial cells
Mesenchymal cells
Metastasis

ABSTRACT

The epithelial-to-mesenchymal transition (EMT) is an important developmental process that is also implicated in disease pathophysiology, such as cancer progression and metastasis. A wealth of literature in recent years has identified important transcriptional regulators and large-scale changes in gene expression programs that drive the phenotypic changes that occur during the EMT. However, in the past couple of years it has become apparent that extensive changes in alternative splicing also play a profound role in shaping the changes in cell behavior that characterize the EMT. While long known splicing switches in FGFR2 and p120-catenin provided hints of a larger program of EMT-associated alternative splicing, the recent identification of the epithelial splicing regulatory proteins 1 and 2 (ESRP1 and ESRP2) began to reveal this genome-wide post-transcriptional network. Several studies have now demonstrated the truly vast extent of this alternative splicing program. The global switches in splicing associated with the EMT add an important additional layer of post-transcriptional control that works in harmony with transcriptional and epigenetic regulation to effect complex changes in cell shape, polarity, and behavior that mediate transitions between epithelial and mesenchymal cell states. Future challenges include the need to investigate the functional consequences of these splicing switches at both the individual gene as well as systems level.

1. Introduction

The organ structures of all eumetazoans arise from the interactions of two basic tissue types that have fundamental differences in their physiology and behavior: the epithelium and the mesenchyme. The epithelium is composed of epithelial cells tightly bound through several types of interactions including adherens junctions, desmosomes, and tight junctions. Typically epithelial cells are polarized, with an apical surface that faces a lumen and a basolateral surface that rests on a thin extracellular matrix called the basal lamina. These properties enable the epithelium to form a semi-permeable layer that separates compartments and function in protection, directional absorption, and secretion. The mesenchyme consists of cells which have a largely undefined shape due to their relatively dynamic cytoskeleton. Mesenchymal cells, unlike epithelial cells, do not display tight cell–cell contacts, and instead interact primarily with the extracelluar matrix, enabling them to migrate in three-dimensional space.

The epithelial and mesenchymal cell phenotypes are not static as both cell types have the ability to undergo major cellular transformations that affect the morphology and behavior of the cell. These transformations are respectively known as the epithelial-to-mesenchymal transition (EMT) and the mesenchymal-to-epithelial transition (MET) and are essential for normal vertebrate development [1]. For example, during gastrulation, cells of the invaginating epithelium undergo an EMT and migrate inward to become the primary mesenchyme. Mesenchymal cells can subsequently undergo the reverse process of MET to form epithelial structures, a process that has been well documented in formation of the renal tubular epithelium during renal organogenesis [2]. While the EMT has essential roles during developmental, there is also evidence that this process can be hijacked in certain pathophysiologic conditions. For example, there is strong evidence that the EMT is one mechanism by which carcinomas spread beyond the primary tumor site and acquire invasive properties that can promote metastasis [3].

The molecular mechanisms that induce and drive the cellular conversions of EMT and MET have long been the focus of intense investigation. These studies have primarily focused on transcriptional regulation to account for changes in total transcript and

Abbreviations: AS, alternative splicing; ESRP, epithelial splicing regulatory protein; RBP, RNA binding protein; RRM, RNA recognition motif; SRN, splicing regulatory network.

* Corresponding author at: University of Pennsylvania, Perelman School of Medicine, Room 413 Hill Pavilion, 380 S. University Ave., Philadelphia, PA 19104, United States. Tel.: +1 215 573 1838; fax: +1 215 898 0189.
E-mail address: russcars@upenn.edu (R.P. Carstens).

1044-579X/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.semcancer.2012.04.003

protein levels in epithelial or mesenchymal cells. For example, much work has revealed important roles for the Snail, Twist, and Zeb family of transcription factors and the signaling pathways that activate them [4,5]. However, a new avenue of investigation into the genetic program that impacts the EMT is beginning to emerge. Recent findings from our lab and several other groups has revealed that vast, coordinated changes in alternative splicing (AS) occur during the EMT and profoundly alter cellular phenotypes and behaviors. Notably, the genes that are regulated by AS during the EMT are generally not regulated transcriptionally. Thus, AS adds a post-transcriptional layer of gene regulation that functions in concert with transcriptional alterations. In this review we will briefly cover the key aspects of AS followed by a description of the recent studies performed to uncover an EMT associated splicing network and the factors that regulate it, including the recently discovered epithelial-specific splicing regulators ESRP1 and ESRP2. We will also discuss several genes showing AS changes during the EMT and speculate on the functional impact. We will conclude by placing the recent analyses in the context of the EMT and discuss some future directions for the field.

2. Alternative splicing regulation

Recent studies using high-throughput sequencing have shown that nearly all multi-exon genes within the human genome produce multiple mature mRNAs [6,7]. Therefore, AS represents a critical mechanism for vastly expanding the protein-coding potential from a limited genome. Small changes in peptide sequence can alter a protein in many ways including localization, interactions with other proteins, enzymatic activity, and post-translational modifications as just a few examples [8]. Exemplifying its importance in regulating gene expression, many AS events are tightly regulated in a cell type or tissue specific manner, at different developmental stages, or in response to extra-cellular stimuli and activation of specific signaling pathways [9,10].

The different types of commonly observed AS events are illustrated in Fig. 1. An exon that is included or excluded from the mRNA is called a cassette exon. In addition to simple cassette exons that are singly spliced or skipped, often multiple adjacent cassette exons are spliced or skipped in tandem or spliced in a mutually exclusive manner. Exons can also be shortened or lengthened by the presence of alternative 3′ or 5′ splice sites on either end of the exon. AS can also lead to alternative polyadenylation (APA) through use of alternative 3′ splice sites (APA3) or 5′ splice sites (APA5) that are present in upstream alternative 3′ terminal exons. It is noteworthy that these events leading to use of alternative polyA sites alter the composition of the 3′ untranslated region (UTR) and thus, also subject the mRNAs to differential regulation by microRNAs and other mRNA stability or translation factors.

Splicing is achieved by the spliceosome, a macromolecular machine composed of five small ribonucleoproteins (snRNPs) and hundreds of additional proteins [11]. The precision of the reaction is accomplished through a coordinated series of complex RNA–protein interactions that can occur post- or co-transcriptionally. Exons are identified by the basal splicing apparatus through the recognition of conserved consensus sequences that are present at the upstream intron/exon boundary (the 3′ splice site) and the downstream exon/intron boundary (the 5′ splice site) (Fig. 2). The 3′ splice site consists of an invariant AG at the end of the intron and an upstream polyprimidine tract (PPT). In addition, the 3′ splice site is associated with a branchpoint sequence (BPS) that is usually located 20–40 nt upstream. The 5′ splice site is a 9 nt sequence of which the invariant GU is present at the start of the intron. Splice sites that are weak matches to the consensus sequences are less efficiently recognized by spliceosomal components and thus these sequence elements are one determinant of whether an exon is spliced or skipped.

In addition to the core splice site sequences, splicing is also influenced by *cis*-elements located within the regulated exons or flanking introns that can act to either enhance or suppress exon recognition and are therefore defined as exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs), or intronic splicing silencers (ISSs). One function of the *cis*-elements is to act as binding sites for RNA binding proteins (RBPs) that bind these sequences and function either to facilitate or inhibit splicing at nearby splice sites (Fig. 2). The most well characterized splicing regulatory proteins are the generally ubiquitously expressed heterogeneous nuclear ribonucleoproteins (hnRNP) and the serine–arginine rich or SR proteins [12,13]. Members of both families bind to degenerate sequence motifs within pre-mRNA transcripts to facilitate constitutive splicing as well as AS through interactions with specific *cis*-elements and components of the basal splicing machinery. The SR proteins have predominantly been shown to bind and mediate the functions of ESEs, whereas the hnRNP proteins have been shown to bind ESSs, ISSs, and ISEs. Additional families of splicing factors have been characterized including several with more cell type-specific expression patterns. Among these more specialized regulators are the neuronal NOVA proteins, the neural and muscle enriched Fox proteins, and the CELF and Muscleblind family of proteins that display both spatial and temporal expression patterns in muscle and nervous tissues [10]. The regulation of an AS event is largely thought to be under combinatorial control, whereby multiple RBPs bound to the pre-mRNA influence the splicing outcome and that this balance may be tipped by the presence or absence of one or more splicing factors such as those with cell-type or condition-specific expression [14].

Several studies have been undertaken to comprehensively profile AS events regulated by cell-type or developmental-stage specific splicing factors [15,16]. Analysis of the regulated targets characterized in these studies found that many of the co-regulated splicing events occur within transcripts that encode proteins that function in biologically coherent pathways that are relevant to the tissue or cell type in which the splicing factor is expressed. For example, Nova regulates the splicing of gene transcripts that are enriched for those encoding proteins that function in processes relevant to neurons, such as synaptic transmission [15]. This work thereby established principles in which a splicing regulatory network (SRN) regulated by a specific splicing factor coordinates the expression of many protein isoforms and adds a layer of gene expression regulation independent of transcription. Such coordinated modules of post-transcriptional regulation by RNA binding proteins have been described as post-transcriptional operons [17]. An important implication of these observations is that the identification of SRNs will further define protein interaction networks and biological pathways that can impact cell morphology and function.

3. Evidence for an epithelial splicing regulatory network

Prior to the advent of high-throughput technologies, AS studies were performed on an individual gene basis and the identification of novel splice variants arose through individual cloning of mRNAs and ESTs. Nevertheless, earlier literature described several AS events in which specific splice variants showed complete or preferential expression in epithelial cells versus mesenchymal cells and/or showed changes during the EMT. These examples include the genes encoding the Fibroblast growth factor receptor 2 (FGFR2), CD44, p120-catenin (CTNND1), and Mena (ENAH) proteins. Each of these proteins has been well characterized and this section provides an overview of what is known about their isoform-specific functions. Of particular interest is that all four genes encode