Embracing the complexity of matricellular proteins: the functional and clinical significance of splice variation.

Katrina Viloria and Natasha J Hill

Diabetes and Cardiovascular Research Group, Kingston University, Kingston-upon-Thames, UK

Address for correspondence: Natasha Hill, School of Life Sciences, Kingston University, Penrhyn Road, Kingston-upon-Thames, UK, Tel +44 (0)20 8417 2519, Email: n.hill@kingston.ac.uk

Abstract

Matricellular proteins influence wide-ranging fundamental cellular processes including cell adhesion, migration, growth and differentiation. They achieve this through both interactions with cell surface receptors and regulation of the matrix environment. Many matricellular proteins are also associated with diverse clinical disorders including cancer and diabetes. Alternative splicing is a precisely regulated process that can produce multiple isoforms with variable functions from a single gene. To date, the expression of alternate transcripts for the matricellular family has been reported for only a handful of genes. Here we analyse the evidence for alternative splicing across the matricellular family including the SPARC, thrombospondin, tenascin and CCN families. We find that matricellular proteins have double the average number of splice variants per gene, and discuss the types of domain affected by splicing in matricellular proteins. We also review the clinical significance of alternative splicing for three specific matricellular proteins that have been relatively well characterised: osteopontin, tenascin-C and periostin. Embracing the complexity of matricellular splice variants will be important for understanding the sometimes contradictory function of these powerful regulatory proteins, and for their effective clinical application as biomarkers and therapeutic targets.

Keywords – matricellular; alternative splice variation; osteopontin; periostin; tenascin-C

Abbreviations – extracellular matrix (ECM), discoidin domain receptors (DDRs), thrombospondin-1 (TSP-1), transforming growth factor-β (TGF-β), CYR61, CTGF, NOV protein family (CCN), secreted protein acidic and rich in cysteine (SPARC), osteopontin (OPN), matrix metalloproteinase (MMP), TGFβ induced protein (TGFβIp), esophageal carcinoma (EAC), transcript support level (TSL), coding sequence (CDS), untranslated region (UTR), tenascin-C (TNC), fibronectin type III-like (FNIII), fibrinogen-like globe (FBG), fasciclin-like 1 (FAS-1), periostin (POSTN).

Introduction

The extracellular matrix (ECM) is a protein scaffold that provides structural support to tissues, but is also increasingly recognised for the role it plays in providing context to the cell's physical and chemical environment [1]. The matrix activates specific cellular signalling pathways through binding of integrins and discoidin domain receptors (DDRs), for example. The synergy between integrin and growth factor signalling pathways further regulates cellular responses [2]. In addition, the matrix can sequester cytokines and growth factors to achieve high local concentrations protected from degradation, and regulate the activation and release of growth factors [3]. The physical properties of the matrix such as rigidity and tension control mechanosensory pathways essential for correct growth, migration and differentiation [2]. The matrix is also capable of remodelling to mediate repair and to regulate cell growth, migration and differentiation [4]. The distinct matrix composition of individual tissues is an important variable determining tissue-specific characteristics, but as yet remains relatively undefined.

Matricellular proteins are an important component of the extracellular matrix. Examples of matricellular proteins include the SPARC family, thrombospondins, tenascins, osteopontin, the CCN family and periostin. Although distinguished from structural matrix proteins such as collagen they nevertheless play an essential regulatory role, and are typically induced during tissue remodelling and repair as well as in many disease states [5]. Matricellular proteins can influence matrix formation. For example, SPARC binds to structural matrix components such as collagen and fibronectin, and SPARC null mice have defective ECM that results in cataracts and allows increased tumour growth [6-8]. Collagen fibril formation during the foreign body response is also affected in SPARC null mice, as well as in mice lacking TSP-1 or TSP-2 [9].

Matricellular proteins also modulate cellular responses to extracellular signals [10]. Many matricellular proteins interact with both growth factors and integrins to regulate cell growth, adhesion and motility. For example, TSP-1 is a physiologically relevant activator of TGFβ: it binds to latent TGFβ via specific sequences in the thrombospondin repeat region, and causes a conformational change that activates

TGFβ [11,12]. TSP-1 can also transactivate EGFR and is involved in the regulation of VEGF, FGF, and PDGF signalling [12]. Tenascin-X was also recently identified as an activator of TGFβ [13]. SPARC similarly regulates signalling by a number of growth factors, both directly and indirectly, including TGFβ, VEGF, PDGF, FGF and HGF [14,15]. TGFβ in turn regulates the expression of a number of matricellular proteins, including SPARC, CCN1, TSP-1, TSP-2 and periostin [16-19].

In terms of integrin binding, most matricelllar proteins bind to multiple integrin receptors, for example, SPARC, CCN1, OPN, periostin and fibulin-5 [5]. Matricellular proteins can therefore directly regulate cell-matrix interactions with subsequent effects on cell adhesion and migration. The capacity of matricellular proteins to modulate integrin signalling also allows indirect regulation of growth factor signalling through the intersection of these pathways. Although we still have much to learn about the effects of matricellular proteins in regulating cell responses to structural matrix proteins, some interesting illustrations are emerging. For example, TSP-1 binding to fibronectin regulates the interaction of TSP-1 with integrin α3β1 [20]. Along similar lines, studies identifying the location of SPARC binding on fibrillar collagen suggest that SPARC may influence the interaction between collagen and integrins, and potentially also collagen interactions with DDR2 proteins [21].

Matricellular proteins such as SPARC and TSP-1 were originally defined by their de-adhesive properties [22], although more recently added members such as osteopontin and TSP2 are now known to promote cell adhesion [10]. Although in many cases the effects of matricellular proteins on cell adhesion relate to integrin binding, the effects of TSP-2 on cell adhesion are thought to be mediated by MMP regulation, though altered matrix and growth factor interactions may also play a role [10]. Many matricellular proteins in fact regulate MMPs, for example the SPOCK family of proteins exhibit a complex regulation of MMP activity [23] and SPARC can upregulate expression of multiple MMPs in cancer cells [24]. Through MMP regulation matricellular proteins can further influence not just cell adhesion and migration but also cell growth, since MMPs can activate and release growth factors such as VEGF and TGFβ [25]. Interestingly, many matricellular proteins are also substrates of MMPs and other proteases,

and cleavage of matricellular proteins can expose new active domains and binding sites, or disrupt existing motifs [26,27]. As an illustration, thrombin cleavage of osteopontin reveals a cryptic non-RGD epitope for integrin binding [28], and cleavage of SPARC produces fragments with increased affinity for collagen [21]. Furthermore, cleavage of Hevin/SPARCL1 by MMP3 generates a "SPARC-like fragment" associated with neovascularisation in glioma and that may compensate for loss of SPARC expression in some tissues [29,30]. Matricellular proteins also regulate protease activity more widely than just MMPs, for example, CCN1 inhibits ADAMTS4 activity [19]. This highly complex web of interactions between different matricellular proteins and other regulatory matrix components reinforces the inter-connectedness underpinning homeostatic regulation of the matrix environment.

Through both interactions with cell surface receptors and regulation of the matrix environment, matricellular proteins are therefore able to influence wide-ranging fundamental cellular processes including cell adhesion, migration, growth and differentiation. There is a correspondingly large body of literature demonstrating the involvement of matricellular proteins in many and varied human disorders such as cancer, diabetes, cardiovascular disease and ocular disorders. For example, Osteopontin is over-expressed in the stroma of a variety of cancers and increases malignancy and metastasis both *in vivo* and *vitro* [31]. High levels of SPARC expression in the tumour-stroma are similarly associated with disease progression in many cancers, including pancreatic cancer [4]. However, in other types of cancer, such as ovarian and prostate cancers, SPARC expression is downregulated [32]. This complexity is also reflected functionally, since although SPARC is overexpressed by the pancreatic tumour stroma, SPARC has been shown to inhibit pancreatic cancer cell growth *in vitro* [33]. Similar contradictions are observed with CCN2: while poor prognosis is associated with low CCN2 expression in breast cancer [34], poor prognosis is associated with increased expression in pancreatic cancer [35] and chondrosarcomas [36].

Increased levels of SPARC are also associated with type 2 diabetes, obesity and insulin resistance, both in serum and in adipose tissue [37]. SPARC expression is regulated by metabolic parameters such as

insulin and glucose in both adipose tissue and in pancreatic stellate cells [15,37]. Increased expression of SPARC is likely to have negative effects in the pancreas, since SPARC inhibits β cell responses to growth factors and inhibits β cell survival [15]. However, it has also been suggested that ectopic SPARC expression in β cells can actually promote insulin secretion at high levels of glucose [38], perhaps consistent with the regulation of insulin release by integrin signalling [39]. Therefore while the majority of evidence suggests a pathogenic role for SPARC in diabetes, it remains possible that under some circumstances SPARC may have beneficial effects in β cells. Similarly, TSP-1 expression in islet endothelial cells is important for islet function [40], and metabolic parameters such as high glucose also regulate TSP-1 expression [41]. As for SPARC, elevated adipose expression of osteopontin also occurs in obesity, and is thought to predispose to insulin resistance [42]. Interestingly, osteopontin expression is also increased in type 1 diabetes patients [43]. Furthermore, the regulation of TGF-β and other growth factors by SPARC, osteopontin, TSP-1, and members of the CCN family have all been linked with diabetic complications such as nephropathy and retinopathy [44-47].

The contradictory effects of matricellular proteins can therefore be seen at both the functional and clinical level, as illustrated above for the tissue-specific effects of proteins such as SPARC. The specific molecular mechanisms underlying this context dependence are almost entirely unknown, yet this must be addressed if we are to harness the potential therapeutic value of these powerful regulatory proteins. A number of matricellular proteins are currently involved in clinical trials [5,48], and it will therefore be important that we better understand mechanisms behind their controversial effects.

The complexity in both form and function of matricellular proteins is likely to explain many of the controversies and tissue-specific effects. Specific matricellular proteins are frequently observed at a range of different molecular weights, as revealed for example by western blotting. The multiple isoforms of these proteins are likely to result from a number of factors, including proteolytic cleavage as discussed above. Matricellular proteins are also subject to a range of post-translational modifications, including glycosylation, phosphorylation and transglutaminase-mediated crosslinking.

Where multiple isoforms with different forms or degrees of post-translational modification are present, these isoforms are likely to have distinct functions. For example, transglutaminase cross-linking of osteopontin [49] and tissue-specific glycosylation of SPARC both result in differential collagen affinity [21].

Alternative splicing also is also likely to contribute to matricellular protein diversity. However, the expression of alternate transcripts for the matricellular family has been reported for only a handful of genes [5]. In this review we will examine for the first time the evidence for alternative splicing across the matricellular family, highlighting specific areas for further study. We also illustrate the clinical significance of alternative splicing for three specific matricellular proteins for which splice variants have been relatively well-characterised: osteopontin, tenascin-C and periostin.

Overview of splice variation in matricellular proteins

Alternative splicing is a precisely regulated process that can produce multiple isoforms with variable functions from a single gene. Alternative splicing is tissue dependent and is regulated by tissue specific *cis*- and *trans*- splicing activators that influence spliceosome assembly [50]. We performed an analysis of the splice variants identified in the ENSEMBL database for 22 matricellular proteins, including the SPARC, thrombospondin, tenascin and CCN families, as well as osteopontin and periostin. As shown in Table 1, the number of splice variants listed ranges from 1 to 26, with a mean of 8 variants in the set of matricellular genes analysed (sd=5.3, n=22). Less than half of these are predicted by ENSEMBL to encode a potentially functional protein that would be not be degraded by cellular quality control mechanisms (100/276 variants for all genes analysed). Many non-coding transcripts have retained introns, and such transcripts are typically restricted to the nucleus and thought to be the result of splicing errors, although functionally relevant cases of transcripts with retained introns have been described [51]. However, 17 protein coding variants are identified for *SPOCK3*, and 10 for *SPARCL1/Hevin*. Other members of the SPARC family with relatively high numbers of variants listed are *SPARC* and *SPOCK1*, with 5 predicted protein-coding variants each. Other matricellular proteins of

interest in this regard are tenascin C, tenascin XB and osteopontin, having 8, 6 and 5 predicted protein variants, respectively.

Recent estimates of the prevalence of alternatively spliced transcripts based on the ENCODE project suggest that, of the 24,000 genes in the human genome, 92–97% express alternatively spliced variants, with an average of 4 isoforms per gene and on average 2.5 protein coding variants per gene [52,53]. Matricellular proteins therefore have on average around double the average number of splice variants per gene, and double the number of protein coding variants per gene. As discussed below, there is growing evidence for the importance of splice variants of matricellular genes in human health and disease. However, there is clearly still much work to be done in this area. Of the matricellular genes analysed here, only 72/100 protein coding variants listed in ENSEMBL contain the complete coding sequence (CDS), and this is a particular problem for some members of the SPARC family. For example, *SPARC* (1/5 complete CDS), *SPARCL1* (3/10 complete CDS) and *SPOCK1* (2/5 complete CDS). Furthermore, less than 40% of the protein-coding variants have a high support level, classified by ENSEMBL as transcript support level 1 (TSL1). TSL1 indicates that the transcript is supported by at least one non-suspect mRNA [54]. Transcripts with support lower than TSL1 are typically based only on either one or more EST sequences, or on a suspect mRNA sequence, or are not supported by either an EST or mRNA sequence (TSL5). Further work will therefore be required to identify and validate predicted variants of these matricellular proteins, in particular the SPARC family.

One critical question will be to determine the abundance of these matricellular splice variants. There is evidence that many splice variants identified by high throughput methods are at low frequency, and that these low frequency variants are typically not evolutionarily conserved and therefore of questionable functional significance [55]. Furthermore, it has been estimated that around 80% of transcripts in human tissue represent the major isoform, suggesting relatively minor roles for alternative splice variants in most cases [51]. However, the number of variants identified for matricellular genes suggests that many matricellular genes may be amongst the exceptions. Splicing is

also altered in disease states such as cancer, and we discuss the evidence for the clinical importance of splice variants of three matricellular genes below. It will also be important to determine cell specific expression patterns of matricellular splice variant transcripts and their relative abundance in different tissues, and to gain insight into the factors that regulate their expression. While RNA seq data from the Illumina Body Map project, for example, can be accessed through ENSEMBL and provides tissue specific information on intron-spanning sequencing reads, data for alternative transcript expression abundance in different cell/tissue types is currently not available. Splice variant databases are also not currently linked to high throughput proteomics data such as mass spectrometry. However, for coding variants where detailed structural and domain functional data is available for the primary isoform it is possible to use bioinformatics tools to make predictions about the function of alternatively spliced proteins, which can then be tested experimentally.

While many of the matricellular splice variants collated in ENSEMBL predict a change in protein sequence, there is also striking splice variation in exons encoding 5'- and 3'-UTR regions of these genes. The potential importance of variation in the 5'- and 3'-UTRs should also not be underestimated, as our understanding grows of how the secondary structure of these UTR can affect transcript stability, for example by miRNA-mediated degradation [56], and also regulate the translational machinery [57]. In support of the importance of UTR regions in matricellular protein expression, TSP-1 in particular has a long 3'UTR that has been analysed in detail, and is known to contain AU- rich elements that regulate mRNA stability [17]. Indeed, the regulation of TSP-1 expression by glucose occurs at the level of translation and is mediated by miR-467 binding to the 3'-UTR [58,59]. SPARC expression is also regulated by non-coding RNA, and in particular is regulated by miR029a/b in nasopharangeal cancer [60]. The potential importance of UTR regions should therefore not be overlooked when considering variants affecting UTR rather than coding regions.

Osteopontin splice variants – evidence for association with specific cancer types.

Osteopontin is a member of the SIBLING family of glycoproteins and is expressed from the *SPP1* gene on human chromosome 4. As show in Figure 1, a total of 6 splice variants were identified across the two databases, and in Table 2 we have integrated the IDs for these splice variants in ENSEMBL, Genbank and the common names used in the literature. Three splice variants of OPN have been known for some time, widely referred to as OPN-a, OPN-b and OPN-c in the literature, and have been relatively well characterised, while OPN-4 and OPN-5 were only very recently identified [61]. Variant 201 listed in ENSEMBL has not been previously described and appears to affect exon 7 outside the splicing region, although this variant is listed as TSL-2 and remains to be confirmed. Interestingly, the recently described OPN-5 is now the longest transcript and therefore the reference sequence at NCBI, though this transcript is not listed in ENSEMBL. This unusual variant has an alternate transcriptional start site, and is discussed further below.

The osteopontin (OPN) protein is relatively unstructured and lacks predicted domains identified by bioinformatics searches [62,63]. However, the regions involved in ligand binding have been characterized, as indicated in Figure 1A. There are three integrin-binding motifs in the central region of the protein, plus a C-terminal region responsible for CD44 binding and containing an EF-hand calcium binding motif [64]. However, the N-terminal region adjacent to the signal peptide that is affected by splicing lacks any known ligand binding domains [65], although this region does include O-linked glycosylation and phosphorylation sites, and may also contain sites of interaction with the extracellular matrix [26]. According to Anborgh *et al*, the exon labeled 'exon 5' in Figure 1 contains glutamine residues that undergo cross-linking by transglutaminase [31]. This exon is absent in both OPN-c and OPN-4 variants, suggesting that these variants may be unable to form polymeric structures (Table 2 and Figure 1). Cross-linking enhances the ability of osteopontin to bind collagen [49], suggesting that OPN-c and OPN-4 may have reduced collagen binding. Furthermore, the alternatively spliced exons 5 and 6 contain putative phosphorylation sites (exon numbering as indicated in Figure 1), suggesting that the various splice variants identified will contain distinct phosphorylation patterns [66]. The binding of

osteopontin to integrins is known to be regulated by phosphorylation[67], and it has been suggested that OPN-a and OPN-b, but not OPN-c, can activate integrin receptors [68]. This would imply that phosphorylation of exon 5 determines integrin binding, though this would need experimental confirmation.

There is also further evidence for functional and clinical diversity of these variants. While RNA seq data in ENSEMBL suggests that OPN-a is expressed in all tissues tested (ENSEMBL accessed March 2016), multiple splice variants of osteopontin are found in cancer cells. For example, in a study by He *et al*, OPN-c was detected in breast cancer but not normal tissue, and OPN-c over-expression was found to promote breast cancer cell invasion *in vitro*. In contrast, OPN-a promoted cell adhesion [65]. A follow up study with increased number of patient samples found a more complex picture, but demonstrated strong association between nuclear OPN-c and mortality in patients with early onset breast cancer, while expression of OPN-a/b in the cytoplasm was associated with poor prognosis [69]. In pancreatic cancer the picture is subtly different. While OPN-a was found to be highly expressed, OPN-c was associated with metastatic disease and OPN-b expression showed a trend towards association with poor survival rates [70]. Similarly, in gastric cancer and glioma while OPN-a, OPN-b and OPN-c were all expressed, only OPN-b and OPN-c were associated with the severity of disease symptoms, and in particular OPN-c was again found to be associated with metastasis [71-73].

In a study by Tilli *et al*, a mechanistic analysis helped to shed additional light on the effects of individual isoforms. Stable expression of OPN-a, OPN-b and OPN-c variants in prostate cancer cells revealed that both OPN-b and OPN-c promote xenograft growth, as well as increasing proliferation, migration and invasion. Expression of OPN-b and OPN-c also increased expression of MMP-2, MMP-9 and VEGF [74]. In a similar study by Lin *et al,* stable expression of OPN-b in esophageal carcinome (EAC) cells increased migration and enhanced invasion compared to cells expressing OPN-c. However, OPN-c expression dramatically reduced cell adhesion, and the authors suggest that OPN isoforms have distinct effects on cell behaviour but can work together to promote EAC progression (Lin, 2015). This is consistent with

the clinical evidence described above suggesting that both OPN-b and OPN-c are overexpressed in some cancers, for example pancreatic [70,75], gastric [71], lung [75], and prostate [74], while OPN-c may be most strongly associated with other cancers, such as breast cancer [65,75,76] and ovarian cancer [77].

These and other studies suggest the potential value of OPN isoforms as biomarkers for specific cancers. This idea was tested by Hartung and Weber in a study analysing blood serum from patients with a range of cancers using an OPN isoform specific RT-PCR assay. Increased expression of both OPN-b and OPN-c was observed in the serum of pancreatic cancer patients while OPN-c was elevated in the serum of lung as well as in breast cancers [75]. OPN-b was also elevated in the serum of lung cancer patients, and when a cutoff of 2 SD over normal was used, could detect over 40% lung cancers.

Diabetes is a risk factor for pancreatic cancer, and osteopontin expression is also increased in diabetes patients [78,79]. Recently, Sarosiek *et al* examined the expression of specific osteopontin splice variants in patients with pancreatic lesions and with diabetes or obesity. In patients with pancreatic lesions, OPN-b was found in 48% of sera, OPN-c in 34%, and both variants in 5% of sera; neither variant was detected in normal sera. However, in patients with pancreatic lesions who also had diabetes and/or obesity, OPN-b was completely absent and the frequency of OPN-c detection was significantly increased. Logistic regression modelling showed that the odds of having diabetes and/or obesity in a patient with OPN-c were seven times higher than patients without OPN-c. This study demonstrates an association between diabetes/obesity and OPN-c expression, and suggests the possibility that osteopontin splice variants may be involved in the link between pancreatic cancer and diabetes/obesity [79].

OPN-5 was recently described by Lin *et al* [61] and is now the longest transcript identified. As mentioned above, this variant has a unique additional exon (exon 4) containing an alternative translation start site. Interestingly, since the signal peptide is encoded by exon 2, the OPN-5 protein created using the downstream translation start site therefore lacks the signal peptide. The Phobius

software for predicting signal peptides clearly identifies AA 1-16 of the other osteopontin variants as a signal peptide, but does not identify a signal peptide sequence in OPN-5 [80], results that are confirmed using the SignalP prediction tool. It is therefore likely that OPN-5 encodes an intracellular isoform. Translation of the OPN-5 transcript using ExPASy indicates that the upstream start codon utilized by all other OPN isoforms creates an upstream open reading frame (uORF) in the OPN-5 transcript, due to the presence of an in-frame stop codon, as illustrated in Figure 1D. Only the downstream translational start site can therefore create a functional protein. Interestingly, uORF are important regulators of translation, particularly in response to nutritional and metabolic stress [81]. The OPN-5 uORF may therefore serve to regulate the translational efficiency of the OPN-5 protein under certain metabolic conditions.

Although matricellular proteins typically contain a signal peptide that targets the protein for secretion, and osteopontin is generally regarded as a secreted protein, osteopontin has been detected intracellularly. Nuclear localisation of OPN has been described [82] and an intracellar isoform of osteopontin (OPNi) has also been characterised. This intracellular form is a truncated protein that lacks the N-terminus, including the signal peptide, and is instead localised to the cytoplasm. Interestingly, this protein was not thought to be a splice variant, but instead is thought to arise from the use of an alternate downstream translation start site [83], although other possibilities have also been discussed [28,84]. OPNi has distinct functions to secreted OPN, and in particular plays an important role in innate immune cells, for example dendritic cells and NK cells [83,84], where it is involved in signalling transduction pathways downstream of innate immune receptors [28]. It will therefore be interesting to test whether innate immune cells express the OPN-5 variant.

The second splice variant recently described by Lin *et al*, OPN-4, is the shortest transcript and lacks all the alternatively spliced exons 4, 5 and 6 [61]. This study found that all 5 characterised isoforms of OPN, including OPN-4 and OPN-5, are co-overexpressed in primary oesophageal cancer [61]. However, no other studies have yet examined the function of OPN-4 and OPN-5 or their potential role in disease.

While the predicted molecular weight of the proteins encoded by the 5 identified osteopontin splice variants ranges from 30.8 kDa (OPN-4) to 37.2 kDa (OPN-5; predicted molecular weights calculated using Expasy), the primary observed band in western blotting is typically around 70 kDa, and multiple bands ranging from 41 to 75 kDa are frequently observed [31]. However, osteopontin undergoes extensive post-translational modification, with serine/threonine phosphorylation, glycosylation, tyrosine sulfation and cross-linking, and osteopontin is also a substrate for thrombin and matrix metalloproteinases. The relationship between expression of particular splice variants and observed protein isoforms of different molecular weights is therefore complex and challenging to unravel.

In summary, osteopontin mediates cell adhesion and regulates signalling by interacting with a range of binding partners such as CD44 and integrins. There is evidence that alternative splicing creates isoforms with distinct post-translational modifications that are likely to affect the capacity to interact with integrins and collagen. Expression of specific isoforms has been shown to be associated with diseases such as cancer and diabetes, though more recently described variants remain as yet relatively uncharacterised. Understanding the role of these isoforms and their expression has great potential in their use as biomarkers and therapeutic targets.

Tenascin-C splice variants as therapeutic targets

Tenascin-C (TNC) is a large matricellular glycoprotein that forms a 1080–1500 kDa hexamer, and is encoded by the *TNC* gene on human chromosome 9. Like many other matricellular proteins TNC expression is normally limited in the adult but is upregulated during tissue remodelling, for example during wound repair, inflammation and cancer. Indeed, increased expression of tenascin-C by tumour-stroma cells is associated with a wide range of cancers [85]. Tenascin-C modulates cell adhesion, migration, proliferation and survival [86], and there are ongoing clinical trials in which TNC is targeted for the treatment of various disorders including heart disease, inflammatory disease and cancer [86,87].

There are 20 human TNC splice variants identified in the ENSEMBL and NCBI Gene databases (see Table 3 for further detail), and the exon structures of the subset of variants present in ENSEMBL are shown in Figure 2B as an illustration. As shown in Figure 2A, tenascin-C contains a C-terminal fibrinogen-like globe (FBG) domain and a region towards the N-terminus containing EGF-like repeats. Tenascin-C is a low affinity ligand for EGF receptor and binding is mediated by the EGF-like repeats [88]. The tenascin-C protein also contains a central domain consisting of fibronectin type III-like (FNIII) repeats that binds various integrin receptors and growth factors [85,89]. Alternative splicing exclusively affects this central FNIII domain and determines the number of FNIII-like repeat units.

There is extensive evidence for functional diversity in alternative splice variants of tenascin-C. This topic has been recently reviewed in detail [90] so only a brief overview is given here. TNC variants are typically divided into 'long' and 'short' variants, where long TNC variants can contain up to 17 FNIII like repeats. The small TNC variant, TNC-S, (220 kDa) binds to fibronectin, while the large variant, TNC-L, (320 kDa) does not [91]. In contrast, it is uniquely the large TNC variants that bind with high affinity to annexin II [92,93]. Furthermore, while short variants promote focal adhesion formation and cell adhesion, long variants typically inhibit the formation of focal adhesions and promote cell migration [94]. These observations have been important for rationalising previous observations that tenascin-C could either promote or inhibit cell adhesion, and are an excellent illustration of how the identification of alternative splice variants can resolve existing functional controversies.

Interestingly, examination of the splice variants in the Gene database allowed us to identify a previously undescribed variant (XM_011518630) that contains a novel alternatively spliced exon containing a stop codon (exon 18 in Table 3 and Figure 2). This variant contains the alternatively spliced exons 11-17, and therefore would be predicted to encode a protein with a long FNIII-repeat domain, but would lack completely the fibrinogen-like globe domain. The FBG domain is important for proteoglycan and integrin binding [95,96]. It has also been suggested that the tenascin-C FBG domain can activate TLR4,

although it is not known whether this is a direct effect [97]. It will therefore be of interest to further investigate the expression and function of this novel TNC variant.

Physiologically, there is strong evidence that the expression of alternative TNC splice variants is developmentally regulated and tissue-specific [90]. Furthermore, RNA seq data from the Illumina Body Map project suggests that the long TNC-001 variant is detected in lung, colon and lymph tissues, while the short TNC-012 variant is detected in breast, ovary and prostate tissue (accessed via ENSEMBL, March 2016). There have also been a large number of studies suggesting that in particular the long TNC isoforms are associated with many different types of cancer, though this may depend on tumour type. A study by Borsi *et al* demonstrated that expression of long variants that include the alternatively spliced region was associated with high stromal cellularity [98]. More recent studies using antibodies specific for the alternatively spliced regions have shown that TNC variants containing these regions are absent in normal tissue but expressed at the tumour invasion front and are associated with metastasis. For example, a TNC antibody specific to exon 14 detected a high molecular weight isoform in basement membranes and stromal fibroblasts of breast cancer as well as around the vasculature. This isoform was strongly expressed in the stroma of highly invasive breast tissues but was largely undetected in benign and normal tissue [99]. Similarly, antibodies with high affinity for the spliced region detected strong expression of large TNC variants in high grade astrocytomas, particularly near vascular structures and proliferating cells [100]. There is currently widespread interest in using the tumour specific expression of TNC variants to target anti-cancer therapies, and growing interest in developing variant-specific TNC antagonists [90].

Functional and clinical significance of periostin splice variants

Periostin is encoded by the *POSTN* gene on human chromosome 13. Similar to osteopontin and tenascin-C, periostin binds a number of integrins and regulates cell adhesion and migration [101,102]. Periostin is also a marker of tumour progression in many types of human cancer, including pancreatic,

ovarian, and breast carcinomas, and is thought to promote a tumour-supportive microenvironment [101,103,104].

Nine *POSTN* splice variants are identified in ENSEMBL and Gene databases, as detailed in Table 4 (the naming of isoforms used by Morra *et al* is used throughout the text for simplicity) [18,105]. The Periostin protein includes four fasciclin-like 1 (FAS-1) repeats that mediate integrin binding, as shown in Figure 3. Alternatively splicing exclusively affects the C-terminal region that has no known conserved functional domains and appears to be intrinsically disordered. The functional effect of alternative splicing is therefore difficult to predict. Hoersch *et al* investigated diversity and isoforms of periostin from an evolutionary perspective, and hypothesized that the C' domain, while lacking in known functional domains, may assume a structure of multiple consecutive beta strands when binding to sequential beta strand elements of fibronectin domains [106]. This region does, however, contain predicted heparin binding domains and also an anomalous nuclear localisation signal [106,107]. The periostin C-terminal region is also thought to regulate cell–matrix interactions through binding of additional ECM proteins such as collagen, fibronectin and tenascin-C [108]. Interestingly, TGFβ induced protein (TGFβIp) is a paralog of periostin that has the same domain structure except for the absence of this C' region [109-111]. Similarly, the *Drosophila* protein midline fasciclin (MFAS) is also classified as a periostin homologue and also lacks this C' region, although is alternatively spliced at its unique N' domain [112]. Furthermore, the capacity of periostin to regulate invasiveness and lung metastasis also maps to the C-terminus of the protein [113,114], supporting the functional significance of alternative splice variants.

Periostin splice variant expression is regulated during development in multiple tissues, with additional isoforms expressed during foetal development in both lung and kidney [105]. Tissue-specific expression has also been observed, with isoform 6 detected in renal but not lung tissue [105]. Multiple isoforms were also identified in normal adult kidney (isoforms 3, 5 and 8). Interestingly, isoform 8 was detected more frequently in renal cell carcinoma than in matched normal tissue [18]. Of the

alternatively spliced exons 17-21, this isoform contains exon 20 alone. Since only the C-terminal exons 14 to 23 were examined this isoform remains to be fully characterised, and it remains to be shown whether this isoform produces a functional protein. Multiple periostin isoforms were also identified in thyroid carcinoma, but expression was not different to normal tissue [115]. In support of highly tissue specific expression patterns, the RNA seq data in ENSEMBL suggests that while isoform 1 is expressed in breast, ovary, testes and heart tissue, isoform 3 is expressed in adipose, colon, kidney, lymph and prostate, isoform 5 is expressed in the thyroid, while the novel isoform POSTN-003 is expressed in adrenal tissues (ENSEMBL accessed March 2016).

Although periostin is frequently overexpressed in different cancers, periostin mRNA is downregulated in bladder cancer. Kim *et al* showed that normal bladder tissues expressed four different periostin splice variants, corresponding to isoforms 1, 2, 3 and 4 [114]. The normal bladder therefore expresses a distinct splice variant prolife to the normal kidney. Importantly, all bladder cancer tissues examined had lost expression of the full length mRNA (isoform 1), and 48% of bladder cancer tissues had lost expression of alternatively spliced versions of periostin. Instead, Isoform 4 was the most commonly expressed periostin variant in bladder cancer tissues. Functional analysis showed that this variant lacks the capacity to inhibit cell invasion and metastasis that were observed for both isoforms 1 and 3. This suggests that a shift from full length to alternatively splice variants of periostin may be involved in the metastasis of bladder cancer [114]. Isoform 4 contains only exons 19 and 20 of the alternatively spliced exons, and like Isoform 8 that is associated with renal cell carcinoma, this isoform may also lack important functional domains.

However, further studies are needed to more fully characterise periostin splice variants. Of the 9 potential splice variants identified in the databases, only four have been fully sequenced and studied in some detail [18]. Analysis of the predicted proteins encoded by all variants is currently lacking, and further work is needed to investigate, for example, whether the predicted protein coding periostin

isoforms indeed produce functional proteins, and how binding to collagen and other matricellular components is affected in these isoforms.

Perspectives and conclusions

There are interesting parallels between the regions affected by alternative splicing in periostin and osteopontin. Both are regions lacking in conserved functional domains [65,106]. Furthermore, both spliced domains are thought to contain heparin-binding sites that mediate matrix binding [18,26,106,107]. This suggests that alternative splicing in these matricellular proteins produces variants with altered matrix binding properties; although this will require experimental testing to substantiate. Interestingly, the correspondence between alternatively spliced exons and unstructured proteins or regions, rather than in a globular domain, is consistent with current thinking in terms of the tolerance of protein structures to accommodate variation. It has been observed that splicing is coupled with protein disorder, as such proteins are naturally less prone to mis-folding and degradation [52]. However, this is not exclusively the case, and around 28% alternatively spliced variants have split protein domains [52]. For tenascin-C, although the complete structure of this large protein is not known, the structure must allow for variable FNIII repeat numbers. Protein repeats can form either open or closed structures, and open structures tend to be more tolerant of variable copy number [116], suggesting that the tenascin-C FNIII repeats may form an open structure.

We have identified that the protein encoded by the recently described OPN-5 isoform lacks a signal peptide and is therefore likely to be located intracellularly. Intracellular osteopontin is important in innate immune cells, yet was previously thought not to occur as a result of alternative splicing (80). It will therefore be important to test the expression of OPN-5 in innate immune cells. Similarly, the study of splice variants of other matricellular proteins may also identify novel intracellular variants, and therefore provide an alternative explanation of the intracellular staining pattern observed for many matricellular proteins [117]. Endocytosis from the extracellular environment has been observed for some matricellular proteins and is currently thought to be responsible for the presence of these

particular matricellular proteins within the cell. For example, there is evidence that SPARC is endocytosed from the extracellular environment and subsequently present in both cytoplasm and associated with the nuclear matrix [118]. A similar mechanism of internalisation into the cytoplasm via an endosomal pathway has also been described for CTGF (CCN2) [119,120]. However, it remains possible that intracellular splice variants are also expressed for these or other matricellular proteins.

For the matricellular proteins osteopontin, tenascin-C and periostin, there is clear evidence for tissue-specific expression of multiple splice variants, and growing interest in their clinical significance as both prognostic markers and therapeutic targets. In particular, identification of tenascin-C splice variants with opposing functions has helped to resolve controversy over the role of this protein in cell adhesion. The identification of splice variants for the majority of matricellular proteins in ENSEMBL and Gene databases suggests that further research into potential matricellular splice variants is warranted. However, dissecting the diversity of splice variants and their functional significance is challenging. Often, multiple matricellular isoforms are detected at the protein level using traditional techniques such as western blotting, but it can be difficult to decipher splice variants from isoforms resulting from post-translational modifications and enzymatic cleavage. Furthermore, splice variants themselves are subject to additional post-translational modifications, and may not even be detected by existing antibodies. Proteomics techniques such as mass spectrometry analysed in the context of tissue-specific alternative transcript databases will be required to determine the degree to which splice variation at the transcript level is reflected at the protein level. In addition, the effect of variants affecting 5'- and 3'-UTR regions are likely to be reflected in changes in protein abundance rather than protein sequence. However, despite these challenges, we suggest that embracing the complexity of matricellular protein splice variation will be an important step towards more fully understanding their function and for effective clinical application of these powerful regulatory proteins.

Figure and Table legends

Figure 1: Alternative splicing of the human osteopontin gene (*SPP-1*) and associated protein domains. **(A)** The osteopontin (OPN) protein is relatively unstructured and lacks conserved domains. Instead, the regions involved in ligand binding are indicated. Splicing affects a region in the N-terminal region of the protein following the signal peptide (SP; AA 31-72 in the OPN-a sequence), shown in orange. This region does not contain any known ligand binding domains but does include O-linked glycosylation and phosphorylation sites [26]. There are also three integrin-binding motifs in the central region of the protein, plus a C-terminal region responsible for CD44 binding and containing an EF-hand calcium binding motif [64]. The site of Thrombin cleavage (AA 169 in the OPN-a sequence) is shown, and OPN is also cleaved by MMPs at AA 166, 201 and 210. Heparin binding sites have been identified at AA 151-160 and 276-283 [26,64]. Diagram is not to scale. **(B)** The ENSEMBL database identifies 5 protein coding alternative transcripts for the *SPP-1* gene, resulting from the alternative splicing of 2 exons (blue box). **(C)** The NCBI Gene database also identifies 5 alternative protein coding transcripts for *SPP-1*, but identifies an additional variant not in ENSEMBL ('OPN-5') that includes a novel exon upstream of the alternatively spliced exons. This exon contains an alternative translation start site. Little is as yet known about the structure or function of this variant, but if the downstream translation start site is used the protein would be expected to be present intracellularly since it lacks the signal peptide for secretion. The common OPN variant names are indicated in (B) and (C), and Table 3 provides an overview of the different IDs. All databases accessed in November 2015. **(D)** The OPN-5 transcript contains an upstream open reading frame (uORF) that may serve to regulate expression of the downstream protein coding sequence. The presence of the uORF also implies that the OPN-5 transcript can only produce a protein lacking a signal peptide, as discussed in the text. OPN-a is shown for comparison (transcripts not to scale).

Figure 2: Alternative splicing of the human tenascin-C gene (*TNC*) and associated protein domains. **(A)** At the protein level, each tenascin-C monomer consists of a signal peptide targeting the

protein for secretion (SP), an N' terminal Tenascin Assembly (TA) domain responsible for hexamer formation and a run of 14.5 Epidermal Growth Factor-Like (EGF-L) repeats. Towards the C-terminus are between 8-17 Fibronectin Type III-Like (FNIII) repeats, and the number of repeats is determined by alternative splicing; plus a C-terminal Fibrinogen-Like Globe (FBG) domain. The FNIII repeats shown in white are expressed in all variants, while the orange boxes indicate alternatively spliced FNIII repeats. Diagram is not to scale. **(B)** The ENSEMBL database identifies 6 alternative protein coding transcripts for the *TNC* gene, resulting from the alternative splicing of 9 exons towards the 3' end of the coding sequence (blue box). The original publication by Nies *et al* describing the human TNC cDNA sequence and domain structure was used to identify the nucleotide sequence corresponding to each named FNIII repeat and identify the corresponding exon in the ENSEMBL database [121]. Additional splice variants identified in the NCBI Gene database are given in Table 3. All databases accessed in November 2015.

Figure 3: Alternative splicing of the human periostin gene (*POSTN*) and associated protein domains. **(A).** The periostin protein consists of a signal peptide (SP) that targets the protein for secretion, a cysteine -rich EMI domain, and four fasciclin-like 1 (FAS-1) repeats. The second and fourth FAS1 domains contain integrin binding domains (black diamonds). The alternatively spliced C' region (shown in orange) has no known functional domains and appears to be intrinsically disordered. The functional effect of alternative splicing is therefore difficult to predict. However, there are predicted heparin binding domains in the C' tail (blue circle), and also an anomalous nuclear localisation signal (green oval) [106,107]. Diagram is not to scale. **(B).** The ENSEMBL database identifies 6 alternative protein coding transcripts for *POSTN*, resulting from the alternative splicing of 5 exons towards the 3' end of the coding sequence (blue box). Additional splice variants identified in the NCBI Gene database are given in Table 4. All databases accessed in November 2015.

Table 1: Analysis of splice variants of matricellular genes in the ENSEMBL database. The number of human variants for the main matricellular proteins listed in the ENSEMBL database (accessed August 2015). The final column indicates the number of variants for which there is a high transcript support

level (TSL). TSL1 is the highest support level, defined as when transcripts are supported by at least one non-suspect mRNA. CDS = coding sequence. Other matricellular proteins not listed here include: R-spondins, short fibulins, small leucine rich proteoglycans (SLRPs), autotaxin, pigment epithelium defived factor (PEDF) and plasminogen activator inhibitor-1 (PAI-s).

Table 2: Osteopontin variants encoded by the *SPP-1* gene.

The splice variant identifiers used in the literature, in ENSEMBL and in Genbank are cross-referenced and the alternatively spliced exons in each variant given. The number of predicted amino acids (AA) and the transcript support level (TSL) from the ENSEMBL database are also indicated. Matricellular proteins undergo extensive post-translational modification and also proteolytic cleavage, so the observed molecular weight of each variant may not correspond to that predicted by the amino acid sequence. *Number predicted amino acids not given as coding sequence (CDS) is incomplete at the 3' end in ENSEMBL. All databases accessed in November 2015.

Table 3: Tenascin-C splice variants encoded by the *TNC* gene.

See the legend for Table 2 for further detail.

Table 4: Periostin splice variants encoded by the *POSTN* gene.

See the legend for Table 2 for further detail. The common names used in Morra *et al* [18] and Kim *et al [114]* are also given.

References

[1]. Cox TR, Erler JT. Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. Dis Model Mech 2011 Mar;4(2):165-178.

[2]. Kim SH, Turnbull J, Guimond S. Extracellular matrix and cell signalling: the dynamic cooperation of integrin, proteoglycan and growth factor receptor. J Endocrinol 2011 May;209(2):139-151.

[3]. Schultz GS, Wysocki A. Interactions between extracellular matrix and growth factors in wound healing. Wound repair and regeneration 2009;17(2):153-162.

[4]. Arnold SA, Brekken RA. SPARC: a matricellular regulator of tumorigenesis. Journal of cell communication and signaling 2009;3(3-4):255-273.

[5]. Murphy-Ullrich JE, Sage EH. Revisiting the matricellular concept. Matrix Biology 2014;37:1-14.

[6]. Gilmour DT, Lyon GJ, Carlton MB, Sanes JR, Cunningham JM, Anderson JR, et al. Mice deficient for the secreted glycoprotein SPARC/osteonectin/BM40 develop normally but show severe age-onset cataract formation and disruption of the lens. EMBO J 1998 Apr 1;17(7):1860-1870.

[7]. Brekken RA, Puolakkainen P, Graves DC, Workman G, Lubkin SR, Sage EH. Enhanced growth of tumors in SPARC null mice is associated with changes in the ECM. J Clin Invest 2003 Feb;111(4):487-495.

[8]. Yan Q, Clark JI, Wight TN, Sage EH. Alterations in the lens capsule contribute to cataractogenesis in SPARC-null mice. J Cell Sci 2002 Jul 1;115(Pt 13):2747-2756.

[9]. Morris AH, Kyriakides TR. Matricellular proteins and biomaterials. Matrix Biology 2014;37:183-191.

[10]. Bornstein P, Sage EH. Matricellular proteins: extracellular modulators of cell function. Curr Opin Cell Biol 2002;14(5):608-616.

[11]. Schultz-Cherry S, Chen H, Mosher DF, Misenheimer TM, Krutzsch HC, Roberts DD, et al. Regulation of transforming growth factor-beta activation by discrete sequences of thrombospondin 1. J Biol Chem 1995 Mar 31;270(13):7304-7310.

[12]. Sweetwyne MT, Murphy-Ullrich JE. Thrombospondin1 in tissue repair and fibrosis: TGF-β-dependent and independent mechanisms. Matrix Biology 2012;31(3):178-186.

[13]. Alcaraz LB, Exposito JY, Chuvin N, Pommier RM, Cluzel C, Martel S, et al. Tenascin-X promotes epithelial-to-mesenchymal transition by activating latent TGF-beta. J Cell Biol 2014 May 12;205(3):409-428.

[14]. Rivera, L.B., Bradshaw, A.D., Brekken, R.A. The regulatory function of SPARC in vascular biology. Cellular and Molecular Life Sciences 2011;68:3165-3173.

[15]. Ryall, C., Viloria, K., Lhaf, F., Walker, A., King, A., Jones, P., Mackintosh, D., McNeice, R., Kocher, H., Flodstrom-Tullberg, M., Edling, C. and Hill, N. Novel Role for Matricellular Proteins in the Regulation of Islet β Cell Survival. Journal of Biological Chemistry 2014;289(44):30614-30624.

[16]. Wrana J, Overall C, Sodek J. Regulation of the expression of a secreted acidic protein rich in cysteine (SPARC) in human fibroblasts by transforming growth factor beta. Comparison of transcriptional and post-transcriptional control with fibronectin and type I collagen. Eur J Biochem 1991;197(2):519-528.

[17]. Stenina-Adognravi O. Invoking the power of thrombospondins: Regulation of thrombospondins expression. Matrix Biology 2014;37:69-82.

[18]. Morra L, Moch H. Periostin expression and epithelial-mesenchymal transition in cancer: a review and an update. Virchows Archiv 2011;459(5):465-475.

[19]. Chijiiwa M, Mochizuki S, Kimura T, Abe H, Tanaka Y, Fujii Y, et al. CCN1 (Cyr61) Is Overexpressed in Human Osteoarthritic Cartilage and Inhibits ADAMTS-4 (Aggrecanase 1) Activity. Arthritis & Rheumatology 2015;67(6):1557-1567.

[20]. Rodrigues RG, Guo N, Zhou L, Sipes JM, Williams SB, Templeton NS, et al. Conformational regulation of the fibronectin binding and alpha 3beta 1 integrin-mediated adhesive activities of thrombospondin-1. J Biol Chem 2001 Jul 27;276(30):27913-27922.

[21]. Bradshaw AD. The role of SPARC in extracellular matrix assembly. Journal of cell communication and signaling 2009;3(3-4):239-246.

[22]. Bornstein P. Thrombospondins as matricellular modulators of cell function. J Clin Invest 2001 Apr;107(8):929-934.

[23]. Nakada M, Miyamori H, Yamashita J, Sato H. Testican 2 abrogates inhibition of membrane-type matrix metalloproteinases by other Testican family proteins. Cancer Res 2003;63:3364-3369.

[24]. Bradshaw A. Diverse biological functions of the SPARC family of proteins. International Journal of Biochemsitry and Cell Biology 2012;44:480-488.

[25]. Mott JD, Werb Z. Regulation of matrix biology by matrix metalloproteinases. Curr Opin Cell Biol 2004;16(5):558-564.

[26]. Scatena M, Liaw L, Giachelli CM. Osteopontin: a multifunctional molecule regulating chronic inflammation and vascular disease. Arterioscler Thromb Vasc Biol 2007 Nov;27(11):2302-2309.

[27]. Lund SA, Giachelli CM, Scatena M. The role of osteopontin in inflammatory processes. Journal of cell communication and signaling 2009;3(3-4):311-322.

[28]. Inoue M, Shinohara ML. Intracellular osteopontin (iOPN) and immunity. Immunol Res 2011;49(1-3):160-172.

[29]. Weaver M, Workman G, Schultz CR, Lemke C, Rempel SA, Sage EH. Proteolysis of matricellular protein Hevin by matrix metalloproteinase-3 produces a SPARC-like fragment (SLF) associated with neovasculature in a murine glioma model. Journal of Cellular Biochemistry 2011;112:3093-3102.

[30]. Barker TH, Framson P, Puolakkainen PA, Reed M, Funk SE, Sage EH. Matricellular homologs in the foreign body response: hevin suppresses inflammation, but hevin and SPARC together diminish angiogenesis. The American journal of pathology 2005;166(3):923-933.

[31]. Anborgh PH, Mutrie JC, Tuck AB, Chambers AF. Pre-and post-translational regulation of osteopontin in cancer. Journal of cell communication and signaling 2011;5(2):111-122.

[32]. Nagaraju GP, Dontula R, El-Rayes BF, Lakka SS. Molecular mechanisms underlying the divergent roles of SPARC in human carcinogenesis. Carcinogenesis 2014 May;35(5):967-973.

[33]. Sato N, Fukushima N, Maehara N, Matsubayashi H, Koopman J, Su GH, et al. SPARC/osteonectin is a frequent target for aberrant methylation in pancreatic adenocarcinoma and mediator of tumor-stromal interactions. Oncogene 2003;22(32):5021-5030.

[34]. Jiang WG, Watkins G, Fodstad O, Douglas-Jones A, Mokbel K, Mansel RE. Differential expression of the CCN family members Cyr61, CTGF and Nov in human breast cancer. Endocr Relat Cancer 2004 Dec;11(4):781-791.

[35]. Hartel M, di Mola FF, Gardini A, Zimmermann A, Di Sebastiano P, Guweidhi A, et al. Desmoplastic reaction influences pancreatic cancer growth behavior. World J Surg 2004;28(8):818-825.

[36]. Shakunaga T, Ozaki T, Ohara N, Asaumi K, Nishida K, Kawai A, et al. Expression of connective tissue growth factor in cartilaginous tumors. Cancer 2000;89(7):1466-1473.

[37]. Kos K, Wong S, Tan B, Gummesson A, Jernas M, Franck N, et al. Regulation of the fibrosis and angiogenesis promoter SPARC/osteonectin in human adipose tissue by weight change, leptin, insulin, and glucose. Diabetes 2009 Aug;58(8):1780-1788.

[38]. Harries LW, McCulloch LJ, Holley JE, Rawling TJ, Welters HJ, Kos K. A role for SPARC in the moderation of human insulin secretion. PloS one 2013;8(6):e68253.

[39]. Kaido T, Yebra M, Cirulli V, Rhodes C, Diaferia G, Montgomery AM. Impact of defined matrix interactions on insulin production by cultured human beta-cells: effect on insulin content, secretion, and gene transcription. Diabetes 2006 Oct;55(10):2723-2729.

[40]. Olerud J, Mokhtari D, Johansson M, Christoffersson G, Lawler J, Welsh N, et al. Thrombospondin-1: an islet endothelial cell signal of importance for beta-cell function. Diabetes 2011 Jul;60(7):1946-1954.

[41]. Stenina-Adognravi O. Invoking the power of thrombospondins: Regulation of thrombospondins expression. Matrix Biology 2014;37:69-82.

[42]. Kahles F, Findeisen HM, Bruemmer D. Osteopontin: a novel regulator at the cross roads of inflammation, obesity and diabetes. Molecular metabolism 2014;3(4):384-393.

[43]. Fierabracci A, Biro PA, Yiangou Y, Mennuni C, Luzzago A, Ludvigsson J, et al. Osteopontin is an autoantigen of the somatostatin cells in human islets: identification by screening random peptide libraries with sera of patients with insulin-dependent diabetes mellitus. Vaccine 1999;18(3):342-354.

[44]. Kos K, Wilding JP. SPARC: a key player in the pathologies associated with obesity and diabetes. Nature reviews endocrinology 2010;6(4):225-235.

[45]. Nicholas SB, Liu J, Kim J, Ren Y, Collins AR, Nguyen L, et al. Critical role for osteopontin in diabetic nephropathy. Kidney Int 2010;77(7):588-600.

[46]. Kong P, Cavalera M, Frangogiannis NG. The role of thrombospondin (TSP)-1 in obesity and diabetes. Adipocyte 2014;3(1):81-84.

[47]. Twigg S, Cooper M. The time has come to target connective tissue growth factor in diabetic complications. Diabetologia 2004;47(6):965-968.

[48]. Sawyer AJ, Kyriakides TR. Matricellular proteins in drug delivery: Therapeutic targets, active agents, and therapeutic localization. Adv Drug Deliv Rev 2016.

[49]. Kaartinen MT, Pirhonen A, Linnala-Kankkunen A, Maenpaa PH. Cross-linking of osteopontin by tissue transglutaminase increases its collagen binding properties. J Biol Chem 1999 Jan 15;274(3):1729-1735.

[50]. Stein S, Lu ZX, Bahrami-Samani E, Park JW, Xing Y. Discover hidden splicing variations by mapping personal transcriptomes to personal genomes. Nucleic Acids Res 2015 Dec 15;43(22):10612-10622.

[51]. Gonzàlez-Porta M, Frankish A, Rung J, Harrow J, Brazma A. Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. Genome Biol 2013;14(7):R70.

[52]. Hegyi H, Kalmar L, Horvath T, Tompa P. Verification of alternative splicing variants based on domain integrity, truncation length and intrinsic protein disorder. Nucleic Acids Res 2011 Mar;39(4):1208-1219.

[53]. Tress ML, Martelli PL, Frankish A, Reeves GA, Wesselink JJ, Yeats C, et al. The implications of alternative splicing in the ENCODE protein complement. Proc Natl Acad Sci U S A 2007 Mar 27;104(13):5495-5500.

[54]. Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR. The ensembl regulatory build. Genome Biol 2015 Mar 24;16:56-015-0621-5.

[55]. Pickrell JK, Pai AA, Gilad Y, Pritchard JK. Noisy splicing drives mRNA isoform diversity in human cells. PLoS Genet 2010;6(12):e1001236.

[56]. Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R. Control of translation and mRNA degradation by miRNAs and siRNAs. Genes Dev 2006 Mar 1;20(5):515-524.

[57]. Hughes TA. Regulation of gene expression by alternative untranslated regions. Trends in Genetics 2006;22(3):119-122.

[58]. Bhattacharyya S, Marinic TE, Krukovets I, Hoppe G, Stenina OI. Cell type-specific post-transcriptional regulation of production of the potent antiangiogenic and proatherogenic protein thrombospondin-1 by high glucose. J Biol Chem 2008 Feb 29;283(9):5699-5707.

[59]. Bhattacharyya S, Sul K, Krukovets I, Nestor C, Li J, Adognravi OS. Novel tissue-specific mechanism of regulation of angiogenesis and cancer growth in response to hyperglycemia. J Am Heart Assoc 2012 Dec;1(6):e005967.

[60]. Qiu F, Sun R, Deng N, Guo T, Cao Y, Yu Y, et al. Mir-29a/b enhances cell migration and invasion in nasopharyngeal carcinoma progression by regulating sparc and col3a1 gene expression. PloS one 2015;10(3):e0120969.

[61]. Lin J, Myers AL, Wang Z, Nancarrow DJ, Ferrer-Torres D, Handlogten A, et al. Osteopontin (OPN/SPP1) isoforms collectively enhance tumor cell invasion and dissemination in esophageal adenocarcinoma. Oncotarget 2015 Sep 8;6(26):22239-22257.

[62]. Fisher L, Torchia D, Fohr B, Young M, Fedarko N. Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin. Biochem Biophys Res Commun 2001;280(2):460-465.

[63]. Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res 2015 Jan;43(Database issue):D213-21.

[64]. O'Regan A, Berman JS. Osteopontin: a key cytokine in cell-mediated and granulomatous inflammation. Int J Exp Pathol 2000;81(6):373-390.

[65]. He B, Mirza M, Weber G. An osteopontin splice variant induces anchorage independence in human breast cancer cells. Oncogene 2006;25(15):2192-2202.

[66]. Gimba E, Tilli T. Human osteopontin splicing isoforms: known roles, potential clinical applications and activated signaling pathways. Cancer Lett 2013;331(1):11-17.

[67]. Weber GF, Zawaideh S, Hikita S, Kumar VA, Cantor H, Ashkar S. Phosphorylation-dependent interaction of osteopontin with its receptors regulates macrophage migration and activation. J Leukoc Biol 2002 Oct;72(4):752-761.

[68]. Chae S, Jun H, Lee EG, Yang S, Lee DC, Jung JK, et al. Osteopontin splice variants differentially modulate the migratory activity of hepatocellular carcinoma cell lines. Int J Oncol 2009;35(6):1409-1416.

[69]. Zduniak K, Ziolkowski P, Ahlin C, Agrawal A, Agrawal S, Blomqvist C, et al. Nuclear osteopontin-c is a prognostic breast cancer marker. Br J Cancer 2015;112(4):729-738.

[70]. Siddiqui AA, Jones E, Andrade D, Shah A, Kowalski TE, Loren DE, et al. Osteopontin splice variant as a potential marker for metastatic disease in pancreatic adenocarcinoma. J Gastroenterol Hepatol 2014;29(6):1321-1327.

[71]. Tang X, Li J, Yu B, Su L, Yu Y, Yan M, et al. Osteopontin splice variants differentially exert clinicopathological features and biological functions in gastric cancer. International journal of biological sciences 2013;9(1):55.

[72]. Yan W, Qian C, Zhao P, Zhang J, Shi L, Qian J, et al. Expression pattern of osteopontin splice variants and its functions on cell apoptosis and invasion in glioma cells. Neuro Oncol 2010 Aug;12(8):765-775.

[73]. Saitoh Y, Kuratsu J, Takeshima H, Yamamoto S, Ushio Y. Expression of osteopontin in human glioma. Its correlation with the malignancy. Lab Invest 1995 Jan;72(1):55-63.

[74]. Tilli TM, Mello KD, Ferreira LB, Matos AR, Accioly MTS, Faria PA, et al. Both osteopontin-c and osteopontin-b splicing isoforms exert pro-tumorigenic roles in prostate cancer cells. Prostate 2012;72(15):1688-1699.

[75]. Hartung F, Weber GF. RNA blood levels of osteopontin splice variants are cancer markers. Springerplus 2013 Dec;2(1):110-1801-2-110. Epub 2013 Mar 14.

[76]. Mirza M, Shaughnessy E, Hurley JK, Vanpatten KA, Pestano GA, He B, et al. Osteopontin-c is a selective marker of breast cancer. International journal of cancer 2008;122(4):889-897.

[77]. Tilli TM, Franco VF, Robbs BK, Wanderley JL, da Silva FR, de Mello KD, et al. Osteopontin-c splicing isoform contributes to ovarian cancer progression. Mol Cancer Res 2011 Mar;9(3):280-293.

[78]. Huxley R, Ansary-Moghaddam A, De González AB, Barzi F, Woodward M. Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies. Br J Cancer 2005;92(11):2076-2083.

[79]. Sarosiek K, Jones E, Chipitsyna G, Al-Zoubi M, Kang C, Saxena S, et al. Osteopontin (OPN) isoforms, diabetes, obesity, and cancer; what is one got to do with the other? A new role for OPN. Journal of Gastrointestinal Surgery 2015;19(4):639-650.

[80]. Phobius.sbc.su.se,. Phobius. 2016.

[81]. Araujo PR, Yoon K, Ko D, Smith AD, Qiao M, Suresh U, et al. Before it gets started: regulating translation at the 5′ UTR. Comp Funct Genomics 2012;2012.

[82]. Junaid A, Moon MC, Harding GE, Zahradka P. Osteopontin localizes to the nucleus of 293 cells and associates with polo-like kinase-1. Am J Physiol Cell Physiol 2007 Feb;292(2):C919-26.

[83]. Shinohara ML, Kim HJ, Kim JH, Garcia VA, Cantor H. Alternative translation of osteopontin generates intracellular and secreted isoforms that mediate distinct biological activities in dendritic cells. Proc Natl Acad Sci U S A 2008 May 20;105(20):7235-7239.

[84]. Leavenworth JW, Verbinnen B, Wang Q, Shen E, Cantor H. Intracellular osteopontin regulates homeostasis and function of natural killer cells. Proc Natl Acad Sci U S A 2015 Jan 13;112(2):494-499.

[85]. Yoshida T, Akatsuka T, Imanaka-Yoshida K. Tenascin-C and integrins in cancer. Cell adhesion & migration 2015;9(1-2):96-104.

[86]. Midwood KS, Orend G. The role of tenascin-C in tissue injury and tumorigenesis. Journal of cell communication and signaling 2009;3(3-4):287-310.

[87]. Spenlé C, Saupe F, Midwood K, Burckel H, Noel G, Orend G. Tenascin-C: Exploitation and collateral damage in cancer management. Cell adhesion & migration 2015;9(1-2):141-153.

[88]. Swindle CS, Tran KT, Johnson TD, Banerjee P, Mayes AM, Griffith L, et al. Epidermal growth factor (EGF)-like repeats of human tenascin-C as ligands for EGF receptor. J Cell Biol 2001 Jul 23;154(2):459-468.

[89]. Chi-Rosso G, Gotwals PJ, Yang J, Ling L, Jiang K, Chao B, et al. Fibronectin type III repeats mediate RGD-independent adhesion and signaling through activated β1 integrins. J Biol Chem 1997;272(50):31447-31452.

[90]. Giblin SP, Midwood KS. Tenascin-C: form versus function. Cell adhesion & migration 2015;9(1-2):48-82.

[91]. A. Ghert M, Qi W, P. Erickson H, A. Block J, P. Scully S. Tenascin-C splice variant adhesive/anti-adhesive effects on chondrosarcoma cell attachment to fibronectin. Cell Struct Funct 2001;26(3):179-187.

[92]. Chung CY, Erickson HP. Cell surface annexin II is a high affinity receptor for the alternatively spliced segment of tenascin-C. J Cell Biol 1994 Jul;126(2):539-548.

[93]. Chung CY, Murphy-Ullrich JE, Erickson HP. Mitogenesis, cell migration, and loss of focal adhesions induced by tenascin-C interacting with its cell surface receptor, annexin II. Mol Biol Cell 1996 Jun;7(6):883-892.

[94]. Murphy-Ullrich JE, Lightner VA, Aukhil I, Yan YZ, Erickson HP, Hook M. Focal adhesion integrity is downregulated by the alternatively spliced domain of human tenascin. J Cell Biol 1991 Nov;115(4):1127-1136.

[95]. Milev P, Fischer D, Häring M, Schulthess T, Margolis RK, Chiquet-Ehrismann R, et al. The fibrinogen-like globe of tenascin-C mediates its interactions with neurocan and phosphacan/protein-tyrosine phosphatase-ζ/β. J Biol Chem 1997;272(24):15501-15509.

[96]. Yokoyama K, Erickson HP, Ikeda Y, Takada Y. Identification of amino acid sequences in fibrinogen gamma -chain and tenascin C C-terminal domains critical for binding to integrin alpha vbeta 3. J Biol Chem 2000 Jun 2;275(22):16891-16898.

[97]. Zuliani-Alvarez L, Midwood KS. Fibrinogen-Related Proteins in Tissue Repair: How a Unique Domain with a Common Structure Controls Diverse Aspects of Wound Healing. Advances in wound care 2015;4(5):273-285.

[98]. Borsi L, Carnemolla B, Nicolò G, Spina B, Tanara G, Zardi L. Expression of different tenascin isoforms in normal, hyperplastic and neoplastic human breast tissues. International journal of cancer 1992;52(5):688-692.

[99]. Adams M, Jones JL, Walker RA, Pringle JH, Bell SC. Changes in tenascin-C isoform expression in invasive and preinvasive breast disease. Cancer Res 2002 Jun 1;62(11):3289-3297.

[100]. Carnemolla B, Castellani P, Ponassi M, Borsi L, Urbini S, Nicolo G, et al. Identification of a glioblastoma-associated tenascin-C isoform by a high affinity recombinant antibody. The American journal of pathology 1999;154(5):1345-1352.

[101]. Gillan L, Matei D, Fishman DA, Gerbin CS, Karlan BY, Chang DD. Periostin secreted by epithelial ovarian carcinoma is a ligand for alpha(V)beta(3) and alpha(V)beta(5) integrins and promotes cell motility. Cancer Res 2002 Sep 15;62(18):5358-5364.

[102]. Baril P, Gangeswaran R, Mahon P, Caulee K, Kocher H, Harada T, et al. Periostin promotes invasiveness and resistance of pancreatic cancer cells to hypoxia-induced cell death: role of the β4 integrin and the PI3k pathway. Oncogene 2007;26(14):2082-2094.

[103]. Erkan M, Kleeff J, Gorbachevski A, Reiser C, Mitkus T, Esposito I, et al. Periostin creates a tumor-supportive microenvironment in the pancreas by sustaining fibrogenic stellate cell activity. Gastroenterology 2007;132(4):1447-1464.

[104]. Shao R, Bao S, Bai X, Blanchette C, Anderson RM, Dang T, et al. Acquired expression of periostin by human breast cancers promotes tumor angiogenesis through up-regulation of vascular endothelial growth factor receptor 2 expression. Mol Cell Biol 2004 May;24(9):3992-4003.

[105]. Morra L, Rechsteiner M, Casagrande S, von Teichman A, Schraml P, Moch H, et al. Characterization of periostin isoform pattern in non-small cell lung cancer. Lung Cancer 2012;76(2):183-190.

[106]. Hoersch S, Andrade-Navarro MA. Periostin shows increased evolutionary plasticity in its alternatively spliced region. BMC Evol Biol 2010 Jan 28;10:30-2148-10-30.

[107]. Norris RA, Moreno-Rodriguez R, Hoffman S, Markwald RR. The many facets of the matricelluar protein periostin during cardiac development, remodeling, and pathophysiology. Journal of cell communication and signaling 2009;3(3-4):275-286.

[108]. Norris RA, Damon B, Mironov V, Kasyanov V, Ramamurthi A, Moreno-Rodriguez R, et al. Periostin regulates collagen fibrillogenesis and the biomechanical properties of connective tissues. J Cell Biochem 2007;101(3):695-711.

[109]. SKONIER J, NEUBAUER M, MADISEN L, BENNETT K, PLOWMAN GD, Purchio A. cDNA cloning and sequence analysis of βig-h3, a novel gene induced in a human adenocarcinoma cell line after treatment with transforming growth factor-β. DNA Cell Biol 1992;11(7):511-522.

[110]. Ma C, Rong Y, Radiloff DR, Datto MB, Centeno B, Bao S, et al. Extracellular matrix protein betaig-h3/TGFBI promotes metastasis of colon cancer by enhancing cell extravasation. Genes Dev 2008 Feb 1;22(3):308-321.

[111]. Horiuchi K, Amizuka N, Takeshita S, Takamatsu H, Katsuura M, Ozawa H, et al. Identification and characterization of a novel protein, periostin, with restricted expression to periosteum and periodontal ligament and increased expression by transforming growth factor β. Journal of Bone and Mineral Research 1999;14(7):1239-1249.

[112]. Mosher DF, Johansson MW, Gillis ME, Annis DS. Periostin and TGF-β-induced protein: Two peas in a pod? Crit Rev Biochem Mol Biol 2015;50(5):427-439.

[113]. Kim CJ, Yoshioka N, Tambe Y, Kushima R, Okada Y, Inoue H. Periostin is down-regulated in high grade human bladder cancers and suppresses in vitro cell invasiveness and in vivo metastasis of cancer cells. International journal of cancer 2005;117(1):51-58.

[114]. Kim CJ, Isono T, Tambe Y, Chano T, Okabe H, Okada Y, et al. Role of alternative splicing of periostin in human bladder carcinogenesis. Int J Oncol 2008;32(1):161-169.

[115]. Bai Y, Nakamura M, Zhou G, Li Y, Liu Z, Ozaki T, et al. Novel isoforms of periostin expressed in the human thyroid. Japanese clinical medicine 2010;1:13.

[116]. Andrade MA, Petosa C, O'Donoghue SI, MuÈller CW, Bork P. Comparison of ARM and HEAT protein repeats. J Mol Biol 2001;309(1):1-18.

[117]. Perbal B. Alternative splicing of CCN mRNAs…. it has been upon us. Journal of cell communication and signaling 2009;3(2):153-157.

[118]. Gooden M, Vernon R, Bassuk J, Sage E. Cell cycle-dependent nuclear location of the matricellular protein SPARC: association with the nuclear matrix. J Cell Biochem 1999;74(2):152-167.

[119]. Wahab NA, Brinkman H, Mason RM. Uptake and intracellular transport of the connective tissue growth factor: a potential mode of action. Biochem J 2001 Oct 1;359(Pt 1):89-97.

[120]. Kubota S, Hattori T, Shimo T, Nakanishi T, Takigawa M. Novel intracellular effects of human connective tissue growth factor expressed in Cos-7 cells. FEBS Lett 2000;474(1):58-62.

[121]. Nies DE, Hemesath TJ, Kim JH, Gulcher JR, Stefansson K. The complete cDNA sequence of human hexabrachion (Tenascin). A multidomain protein containing unique epidermal growth factor repeats. J Biol Chem 1991 Feb 15;266(5):2818-2823.

Figure 1

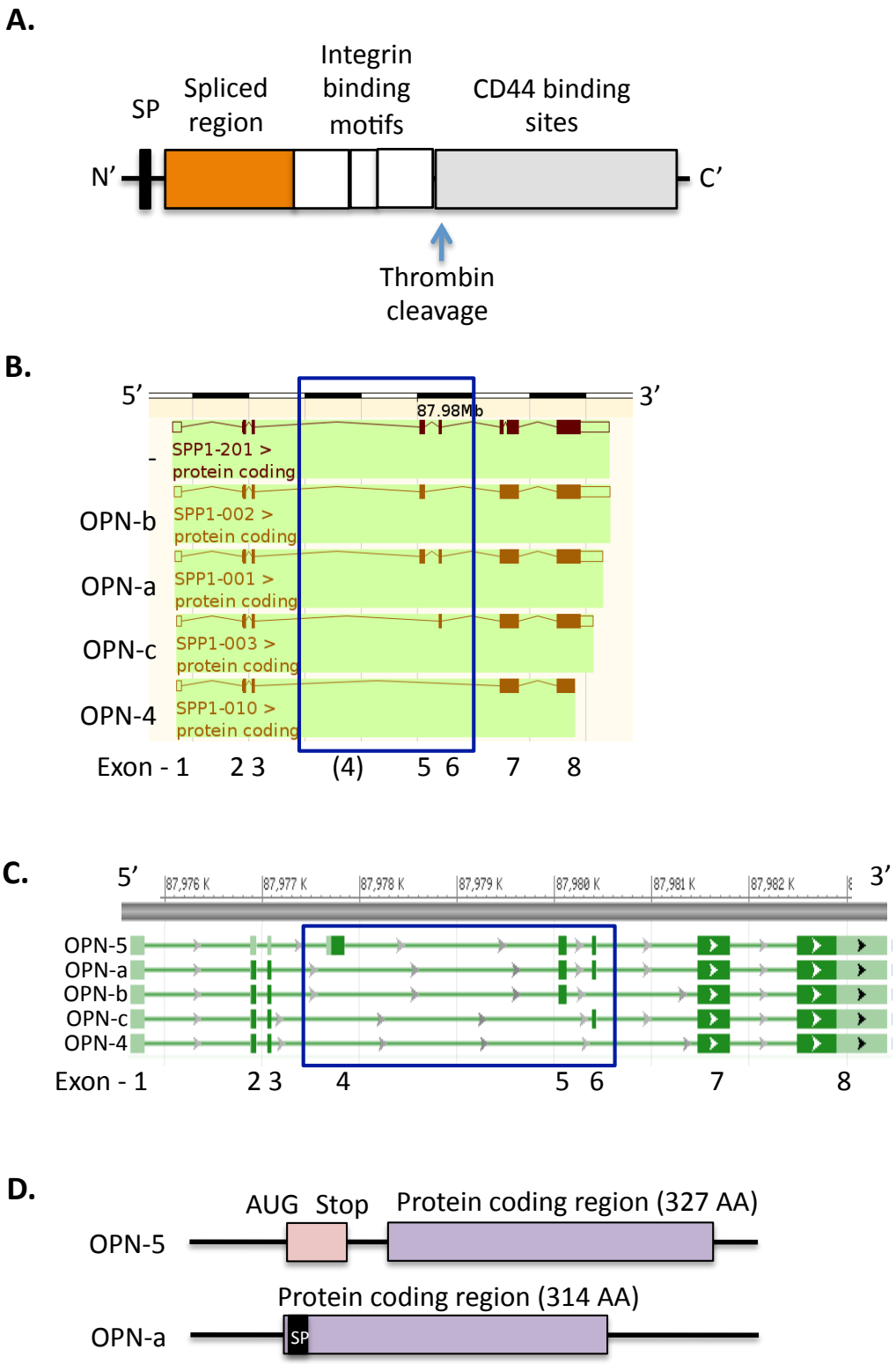
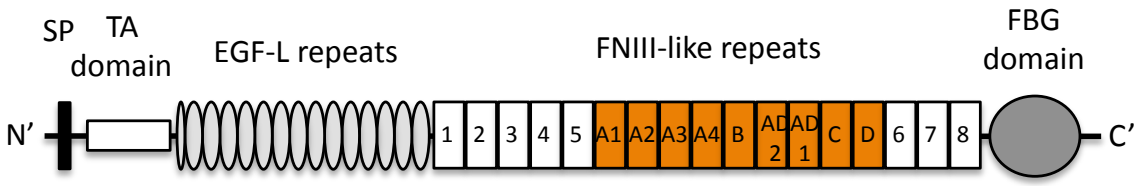

A.
SP
Spliced region
Integrin binding motifs
CD44 binding sites
N'
C'
Thrombin cleavage
B.
5'
3'
87.98Mb
SPP1-201 > protein coding
SPP1-002 > protein coding
SPP1-001 > protein coding
SPP1-003 > protein coding
SPP1-010 > protein coding
OPN-b
OPN-a
OPN-c
OPN-4
Exon - 1 2 3 (4) 5 6 7 8
C.
5'
3'
87,976 K
87,977 K
87,978 K
87,979 K
87,980 K
87,981 K
87,982 K
OPN-5
OPN-a
OPN-b
OPN-c
OPN-4
Exon - 1 2 3 4 5 6 7 8
D.
AUG
Stop
Protein coding region (327 AA)
OPN-5
Protein coding region (314 AA)
OPN-a
SP

Figure 2

A.


SP
TA domain
EGF-L repeats
FNIII-like repeats
FBG domain
N'
1
2
3
4
5
A1
A2
A3
A4
B
AD2
AD1
C
D
6
7
8
C'

B.

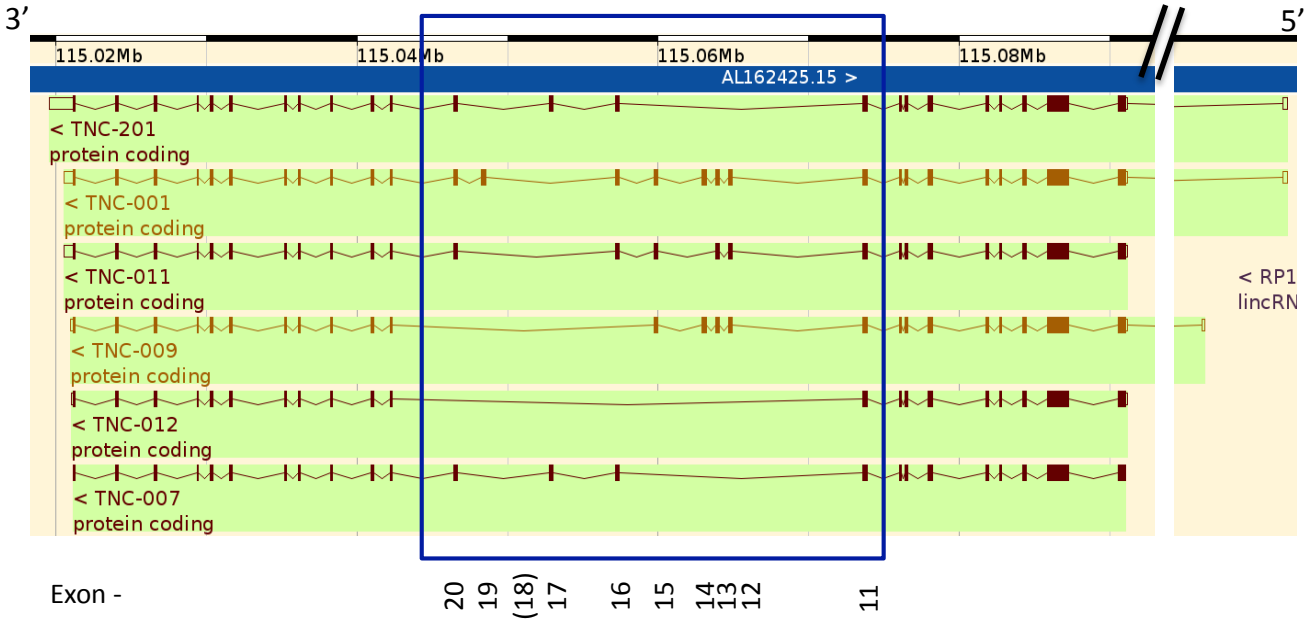
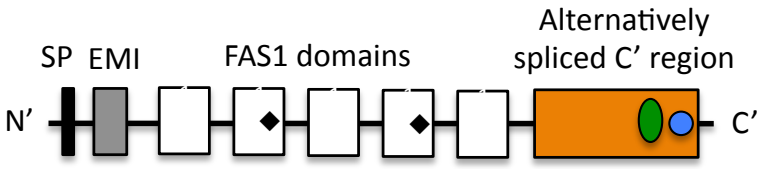

3'
5'
115.02Mb
115.04Mb
115.06Mb
115.08Mb
AL162425.15 >
< TNC-201
protein coding
< TNC-001
protein coding
< TNC-011
protein coding
< TNC-009
protein coding
< TNC-012
protein coding
< TNC-007
protein coding
< RP1
lincRN
Exon -
20
19
(18)
17
16
15
14
13
12
11

Figure 3

A.


SP
EMI
FAS1 domains
Alternatively
spliced C' region
N'
C'

B.

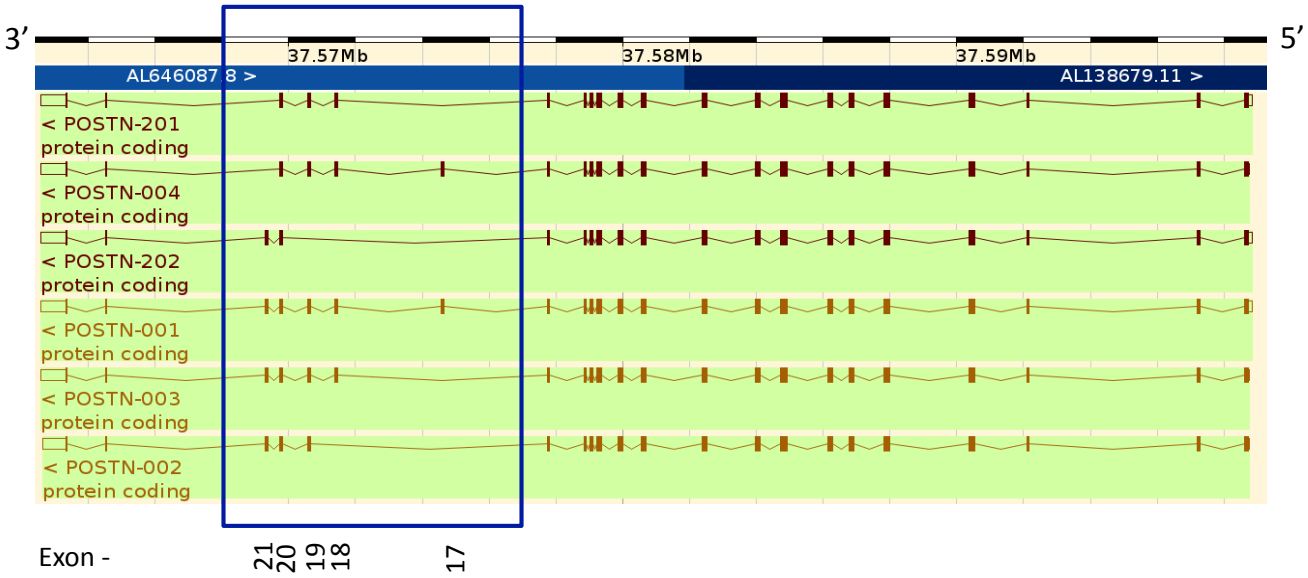

3'
5'
37.57Mb
37.58Mb
37.59Mb
AL646087.8 >
AL138679.11 >
< POSTN-201
protein coding
< POSTN-004
protein coding
< POSTN-202
protein coding
< POSTN-001
protein coding
< POSTN-003
protein coding
< POSTN-002
protein coding
Exon -
21
20
19
18
17

Table 1

Matricellular Family	Gene name	Description	Other identifiers	Total number ENSEMBL human variants	Number predicted protein coding variants	Number complete CDS listed	Number complete CDS with high confidence (TSL-1)
SPARC	*SPARC*	secreted protein, acidic, cysteine-rich (osteonectin)	*ON*	9	5	1	1
	SPARCL1	SPARC-like 1 (hevin)	*MAST9*	10	10	3	1
	SPOCK1	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1	*SPOCK, testican-1, TIC1*	10	5	2	2
	SPOCK2	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2	*KIAA0275, testican-2*	9	4	4	2
	SPOCK3	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 3	*testican-3*	26	17	13	3
	SMOC1	SPARC related modular calcium binding 1	-	5	2	2	2
	SMOC2	SPARC related modular calcium binding 2	*SMAP2*	6	4	3	2
	FSTL1	follistatin-like 1	*FRP, FSL1*	10	4	2	1
Thrombospondin	*THBS1*	thrombospondin 1	*THBS, THBS-1, TSP, TSP-1, TSP1*	8	2	1	1
	THBS2	thrombospondin 2	*TSP2*	6	3	2	2
	THBS3	thrombospondin 3	-	11	4	4	3
	THBS4	thrombospondin 4	-	7	2	2	1
	COMP	cartilage oligomeric matrix protein	*EDM1, EPD1, MED, PSACH, THBS5*	4	3	3	2
Tenascin	*TNC*	tenascin C	*DFNA56, HXB, MGC167029, TN*	13	8	6	3
	TNR	tenascin R	-	3	3	2	0
	TNN	tenascin N	*TN-W*	3	3	3	0
	TNXB	tenascin XB	*TNXB1, TNXB2, TNXBS, XB, XBS*	9	6	5	2
Osteopontin	*SPP1*	secreted phosphoprotein 1	*BNSP, BSPI, ETA-1, OPN*	11	5	4	3
CCN	*CYR61*	cysteine-rich, angiogenic inducer, 61	*CCN1, GIG1, IGFBP10*	3	2	2	1
	CTGF	connective tissue growth factor	*CCN2, IGFBP8*	1	1	1	1
	NOV	nephroblastoma overexpressed	*CCN3, IGFBP9*	2	1	1	1
Periostin	*POSTN*	periostin, osteoblast specific factor	*OSF-2, periostin, PN*	10	6	6	5
			Total	176	100	72	39
			mean variants/gene	8.0	4.5	3.3	1.8

Table 2

Common name	ENSEMBL variant	AA	TSL	Genbank ID	Exons present in alternatively spliced region
OPN-a	001	314	1	NM_001040058	5, 6
OPN-b	002	300	1	NM_000582	5
OPN-c	003	287	1	NM_001040060	6
OPN-4	010	*	2	NM_001251829	
OPN-5	-	-	-	NM_001251830	4, 5, 6
-	201	292	2	-	5, 6

Table 3

ENSEMBL variant	AA	TSL	Genbank ID	Exons present in alternatively spliced region
TNC-001	2201	1	NM_002160	11-16
TNC-201	1838	5	XM_006717098	11, 16, 17, 20
TNC-011	2019	5	XM_006717098	11, 12, 13, 15, 16, 20
TNC-009	1928	1	XM_006717100	11-15
TNC-007	1838	1	XM_006717098	11, 16, 17, 20
TNC-012	1564	5	XM_005251975	11
-			XM_005251974	11, 20
-			XM_011518629	11,16,20
-			XM_011518626	11, 16, 17, 19, 20
-			XM_011518624	11, 12, 15, 16, 17, 19, 20
-			XM_006717101	11, 12
-			XM_011518628	11-13
-			XM_006717097	11, 12, 13, 15, 16, 19, 20
-			XM_011518623	11, 12, 13, 15, 16, 17, 19, 20
-			XM_005251973	11-14
-			XM_011518625	11-16
-			XM_005251972	11-16, 20
-			XM_011518622	11-17, 20
-			XM_006717096	11-17, 19, 20
-			XM_011518630	11-19

Table 4

Common name (Morra 2011)	Common name (Kim, 2008)	ENSEMBL variant	AA	TSL	Genbank ID	Exons present in alternatively spliced region
“isoform 1”	“WT”	POSTN-001	836	1	NM_006475	17-21
“isoform 3	“variant II”	POSTN-201	781	1	NM_001136935	18, 19, 20
		POSTN-003	809	1	NM_001286665	18-21
“isoform 7”		POSTN-202	749	1	NM_001286666	20, 21
“isoform 2”	“variant III”	POSTN-002	779	1	NM_001135934	19, 20, 21
“isoform 5”		POSTN-004	808	5	XM_005266231	17-20
“isoform 4”	“variant I”	-			NM_001135936	19, 20
“isoform 8”		-			NM_001286667	20
“isoform 6”		-			XM_005266232	17, 19, 20, 21